

生理活性脂質データベース [LIPID_{BANK} for Web] の完成から システムティックバイオアクティブデータベース (SBLD) の構築へ

大島美恵子

1. はじめに

脂質とは、水に溶けにくく、有機溶媒に溶けやすい物質と定義される化学物質で、生体では、分解されてエネルギーになるほかに、細胞膜の主要成分であり、また生体におけるさまざまな生理活性を担う物質である。筆者は国立国際医療センター研究所に在職中、プロジェクトリーダーとして脂質情報のデータベース構築に関与した。この生理活性脂質データベース [LIPID_{BANK} for Web] は、科学技術振興事業団のデータベース化支援事業の助成を受けて、インターネットで脂質化合物に関する情報を発信するシステムとして構築されたもので、化学構造の異なる多くの脂質を、ひとつのデータベースとしてまとめたところに特徴がある。また、世界で初めて構築された脂質データベースであり、日本が独自に開発し発信するデータベースである。その構造は、27種類の脂質の名称・構造・物理化学的性質・スペクトル・生理活性・分布などを収録したファクトデータと、それらの根拠となる文献情報データと画像データからなっている。このデータベース [LIPID_{BANK} for Web] について説明し、また一般の方が利用しやすいように、糖脂質データを例にとって簡単に利用法をまとめたい。さらに、今後このデータベースを発展させるための、新たな総合的脂質データベースの構築計画を提案したい。

2. 生理活性脂質データベース [LIPID_{BANK} for Web] とは何か

生理活性脂質データベース [LIPID_{BANK} for Web] は、脂質に関するデータを化合物ごとに整理してデータベース化したもので、インターネット上で広く公開することを目的とし、特別なソフトや機器がなくても利用可能なシステムとして開発したものである。[LIPID_{BANK} for Web] で扱うデータには、文字・数値情報以外に化学構造式の描画データ及びクロマトグラム等のグラフィックデータが含まれる。インターネット上でポピュラーな画像を扱えるデータ提供システムとして、WWW (World Wide Web) を基本に、検索サービスを構築し、また、一般に広く公開するシステムとして、ユーザ管理や課金処理を行わず、誰でも自由に情報を取得できるようになっている。過去に文部省の科学研究費を受けて構築された MS-DOS 版 脂質情報データベース [LIPID_{BANK}] のプロトタイプシステムを基にしており、これを科学技術振興事業団の助成により、新たにインターネットで脂質化合物に関する情報を発信するシステムとして再構築したものである。

平成8年度後期に科学技術振興事業団が生理活性脂質データベースプロジェクトへの支援（平成11年10月までの3年間、年間約3500万円の予算）を決定したことから、平成8年12月から、科学技術振興事業団と国立国際医療センター研究所との共同研究として本格的にデータベース構築を開始した。同時に日本脂質生化学研究会会員を主体とした「生理活性脂質データベース構築検討会」が発足した。各脂質項目毎に「データベース構築責任者」を置き、この責任者のもとで、データ提供者が文献を基にした脂質情報を市販のデータベースソフト（ファイルメーカー pro、Microsoft Access など）で入力したものを事務局に送り、これを事務局で書式変換してデータベースサーバへ収録した。データ提供者総数は、現在130名に及ぶ。

3. データ集積について

平成8年末からの3年間で6000件を目標にデータ集積を開始し、平成9年度1500件、平成10年度2000件、平成11年度800件を集積し、平成11年9月までにさらに1000件以上を追加集積し、ほぼ目標数に達した。本事業の代表者としての筆者の定年退官をひかえ、平成11年7月末にサーバを日本脂質生化学研究会事務局が存在する東京大学医学部に移転し、平成11年10月から本格的に一般公開を開始した。データベース公開後も科学技術振興事業団からの継続予算が計上されており、平成13年度現在、さらにデータ修正およびデータ集積が行われている。

現在、[LIPID_{BANK} for Web]の本体は、東京大学医学部構内に置かれたサーバ上で稼働しており、インターネットへの接続は、東京大学構内のLAN回線を経由して行われている。

(URL : <http://lipid.bio.m.u-tokyo.ac.jp>)

データを集積した脂質項目は、以下の27種類である。()内は平成13年6月における集積数である。27種類の化合物は16種類にグループ分けして収録されている。DERIVED LIPIDSのグループにはFATTY ACID、LONG CHAIN ALCOHOL、LONG CHAIN ALDEHYDE、LONG CHAIN BASE and CERAMIDEが[§]、FAT SOLUBLE VITAMINには、CAROTENOID、COENZYME Q、VITAMIN A、VITAMIN D、VITAMIN E、VITAMIN Kが[§]、GLYCOLIPIDのグループにはglycoSPHINGOLIPID、glycoGLYCEROLIPID and OTHERSが[§]、PHOSPHOLIPIDにはGLYCEROPHOSPHOLIPID、PAF、SPHINGOPHOSPHOLIPIDがそれぞれグループ分けされている。

1. ACYLGLYCEROL アシルグリセロール (574)
2. BILE ACID (CHOLANOID) 胆汁酸 (コラノイド) (550)
3. DERIVED LIPIDS FATTY ACID 脂肪酸 (695)
4. DERIVED LIPIDS LONG CHAIN ALCOHOL 長鎖アルコール (57)
5. DERIVED LIPIDS LONG CHAIN ALDEHYDE 長鎖アルデヒド (108)
6. DERIVED LIPIDS LONG CHAIN BASE and CERAMIDE 長鎖塩基とセラ

ミド(138)

7. ETHER TYPE LIPID エーテル型脂質 (489)
8. FAT SOLUBLE VITAMIN CAROTENOID カロテノイド (80)
9. FAT SOLUBLE VITAMIN COENZYME Q COENZYME Q (30)
10. FAT SOLUBLE VITAMIN VITAMIN A ビタミンA (10)
11. FAT SOLUBLE VITAMIN VITAMIN D ビタミンD (699)
12. FAT SOLUBLE VITAMIN VITAMIN E ビタミンE (76)
13. FAT SOLUBLE VITAMIN VITAMIN K ビタミンK (31)
14. GLYCOLIPID glycoSPHINGOLIPID スフィンゴ糖脂質 (583)
15. GLYCOLIPID glycoGLYCEROLIPID and OTHERS グリセロ糖脂質他 (115)
16. HOPANOID ホパノイド (1)
17. ISOPRENOID イソプレノイド (111)
18. LIPOAMINO ACID リポアミノ酸 (5)
19. LIPOPOLYSACCHARIDE リポ多糖 (35)
20. LIPOPROTEIN リポ蛋白 (12)
21. MYCOLIC ACID ミコール酸 (203)
22. PHOSPHOLIPID GLYCEROPHOSPHOLIPID グリセロ磷脂質 (167)
23. PHOSPHOLIPID PAF 血小板活性化因子 (123)
24. PHOSPHOLIPID SPHINGOPHOSPHOLIPID スフィンゴ磷脂質 (26)
25. PROSTANOID プロスタノイド (221)
26. STEROID ステロイド (479)
27. WAX ワックス (14)

4. システム構成の概要

システム構成概要は以下のとおりである。

[LIPID_{BANK} for Web]は、インターネットに接続するUNIXワークステーション (WS)サーバで稼働しており、利用者は各自のWWW Browser (In-

ternet Explorer, Netscape 等)を通してアクセスする。

サーバ：Sun Enterprise 450ワークグループサーバ、ディスク4.2Gbyte、メモリー512Mbyte

ソフトウェア：OS Solaris Ver.2.6、Webサーバ Sun Web Server Ver.1.0

DBMS Oracle Ver.8.0.3、DBMS 接続ツール Zolar Ver.4.0.4

5. データベースの操作法

ここでは27種類の脂質化合物データの中で、スフィンゴ糖脂質データの検索を中心にその操作法を解説する。スフィンゴ糖脂質は、筆者が40年以上にわたって研究対象としてきた物質であり、ヒトの赤血球の膜や細胞の膜を構成する重要な成分である。

5-1. LIPID_{BANK} for Web へのアクセス

ブラウザアプリケーションを用いて <http://lipid.bio.m.u-tokyo.ac.jp> にログインすると、LIPID_{BANK} for Web のホームページが表示される。ここには画面下部にデータベース構築協力者の氏名がリストアップされており、どのような人がデータベースを構築しているのか知ることができる。

5-2. 検索化合物の種類の選択

画面中央の[Next page →]のボタンを押すと脂質メニュー画面に進む。日本語表示に設定されている場合は、ギリシャ文字等の特殊文字を正確に表示するため、ブラウザの文字コードを「欧米 (西ヨーロッパ言語)」に設定する。(日本語表示のままでは特殊文字は文字化けして表示される)

脂質メニュー画面 (LIPID MENU) で、27種類の脂質化合物の中から検索したい化合物を選びクリックする。糖脂質 (GLYCOLIPID) は脂質部分の違いにより、スフィンゴ糖脂質 (glycoSPHINGOLIPID) とグリセロ糖脂質 (glycoGLYCEROLIPID and OTHERS) の二つに大別されている。ここでは、

スフィンゴ糖脂質 (glycoSPHINGOLIPID) をクリックし、検索画面へ移る。

5-3. スフィンゴ糖脂質の検索条件の設定

検索画面は、左右に2分割され、左画面に表示されている検索メニューにキーワードや数値を入力し[Search]ボタンを押すと検索が開始される。左ウインドウの検索メニューは、キーワード検索 (KEY WORD)、分類検索 (CLASSIFICATION)、脂質部分 (アグリコン) の種類による検索 (SPECIES OF AGLYCON) 他があるが、これは脂質項目によって異なる。

メニューにキーワードや文字、数値などを入力した後、[Search]ボタンを押すと検索が開始され、しばらく待つと検索結果が右ウインドウに表示される。画面上部には検索条件を満たす化合物の数が表示され、20件ずつ化合物名が表示される。慣用名がある場合は、左から慣用名が太字で、次いで系統名が表示される。それぞれの化合物名の左側に[VIEW]ボタンが配置され、データの詳細を見たい化合物の左側にあるこのボタンを押し、詳細画面に移る。詳細画面は別ウインドウとして表示される。

検索メニューウインドウで、検索条件をまったく入力せずに[Search]ボタンを押すと、収録されている全データの件数と化合物名のリストが表示され。この操作で特定の化合物のデータベースへの収録件数を知ることができる。2001年9月現在583件のスフィンゴ糖脂質データが収録されている。

検索条件を設定した後、キーワード検索項目の下部にある[☐ CHECK the box to search ALL LIPID CLASSES.]の☐にカーソルを合わせてここをチェックすると、入力された検索条件で27化合物全てを検索対象として検索を開始する。脂質メニュー画面で選択した種類の化合物のみならず、他の種類の化合物を含めて設定した検索条件を満たす化合物を検索したい時に利用する。

検索条件が設定されているときは、ヒットした化合物名のリストの上に、検索されたデータ数と共に、どのような条件を満たす化合物であるのかがわかるように、設定した検索条件が表示されている。検索条件を満たす化合物がない場合は、「0 RECORDS FOUND」と表示されるので、再度左ウインドウで検索条件を入力後「Search」ボタンを押す。前回の検索条件をクリアして新たな検索条件を設定するには、[Clear]ボタンを押す。

複数の検索条件を設定した場合は、それぞれの検索条件が“AND”条件で検索される。画面の「？」ボタンをクリックすると、検索条件設定のためのヘルプ（英文）が別ウインドウで表示されるので参考にされたい。

検索メニューは化合物毎に異なり、糖脂質では、糖脂質の名称や含まれる臓器、組織名、生物活性などのKey Wordsを入力する「KEY WORD」検索、中性、酸性、塩基性、両性などの糖脂質の種類から検索する「CLASSIFICATION」検索、アシル基やステロールなどのアグリコンの種類から検索する「SPECIES OF AGLYCON」検索、結合している単糖の数やシアル酸の数、フコースの数などから検索する「NUMBER」検索の他、IUPAC/IUBMBによる名称からの検索、糖鎖構造からの検索、脂肪酸の種類、長鎖塩基、グリセロールの構造などからの検索もできる。また、「NUMBER」検索では、ある特定の数値に合致するもののみならず、数値の範囲を指定して検索ができる。

5-4. データの詳細表示

検索リストから、各化合物の詳細データを見るには、化合物名の左にある[VIEW]ボタンを押す。別ウインドウが開き、データ番号、データ提供者名に続いて、化合物の系統名、慣用名、略名および化学構造式が表示される。糖脂質の糖鎖およびセラミドは3文字記号で表されている。構造式はChemDrawで描かれており、[Download]ボタンを押すと構造データをダウンロードすることが出来る。

次いで生物活性 (BIOLOGICAL ACTIVITY)、物理化学的性質 (PHYSICAL AND CHEMICAL PROPERTIES)、UV、IR、NMR、MASSなどのスペクトルデータ (SPECTRAL DATA)、TLCや液体クロマトグラム、ガスクロマトグラムなどのクロマトデータ (CHROMATOGRAM DATA)、所在・分布 (SOURCE)、化学合成法 (CHEMICAL SYNTHESIS)、代謝データ (METABOLISM)、遺伝情報 (GENETIC INFORMATION)、備考 (NOTE) および文献リストが表示される。

各項目内に書かれたスペクトル番号やクロマト番号をクリックすると該当するスペクトル、クロマトグラム表示される。また、文献番号をクリックすると該当する文献にジャンプするようになっている。

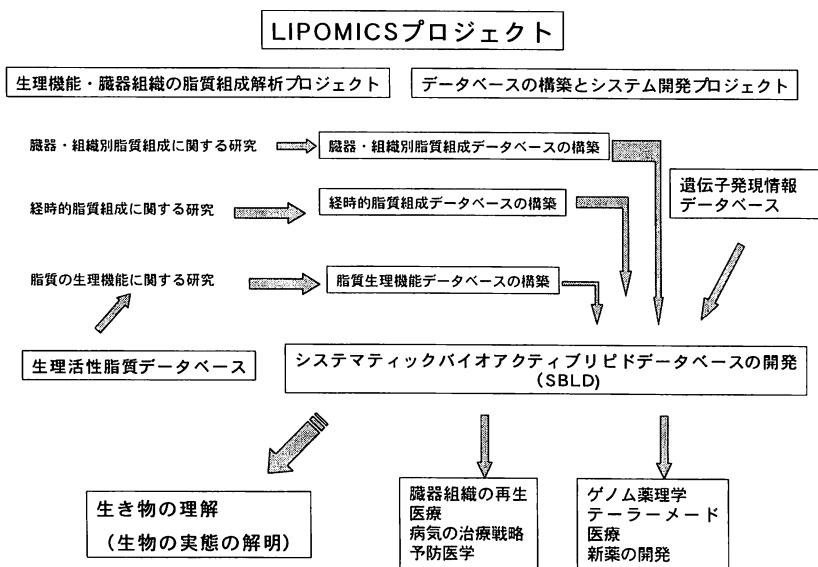
以上、操作方法を簡単にまとめたが、ぜひ実際にアクセスして検索してみたい。

6. 生理活性脂質データベース [LIPID_{BANK} for Web] の運営について

以上のように、生理活性脂質データベース [LIPID_{BANK} for Web] は、3年間でほぼ原型が完成したが、今後このデータベースが常に正しい情報を提供するためには、データの修正作業が必須である。そのため、データベースを日本脂質生化学研究会のもとにおき、脂質生化学を専門とする会員の責任において修正を継続している。現在抱えている問題にはつぎのようなものがある。

1. 生理活性脂質データベースの開発は、日本において世界に先駆けて行われたプロジェクトである。データベースに収録すべき脂質の数はまだ多く、一方脂質化学の進歩により、情報として加えるべき項目は増加している。従って3年間という短い期間でデータを収録するのではなく、データを継続的に集積し、内容を充実させる必要がある。また、雑誌投稿の際に、新規脂質をデータベースに報告してもらうシステムを作ることが必要である。
2. 検索項目と検索方法については新たな検討と修正が必要である。検索項目と検索方法の修正費用については、科学技術振興事業団 (JST) のデータベース化支援の継続予算では充分ではなく、新規の研究費に期待している。
3. 他の生物科学関連の構造データベース (たんぱく質、酵素、核酸、糖質他) との関係を強め、世界のデータベースとリンクしてゆく必要がある。本プロジェクトの外部評価は大変よく、これについては平成13年8月3日付および、9月28日付の科学新聞を参照されたい。

7. 新規プロジェクト「LIPOMICS」の提案から システマティック バイオアクティブリピドデータベース (SBLD) の構築へ



LIPOMICS プロジェクトの概要

ゲノム遺伝子の塩基配列の全解明を目前にし、ゲノム研究後に進めるべき課題（ポストゲノムプロジェクト）として、遺伝子の直接の産物であるタンパク質の構造や機能を研究するプロテオミクスの研究が世界各国で開始されている。

一方、生体成分としての脂質は、基本的な生体構成要素であると共に、多彩な生理機能を有することが解明されつつある。しかし脂質は、そのカテゴリーの中に多種多様な化合物を包含することからデータベース化が難しく、その電子情報化は遺伝子やタンパク質の情報電子化と比べ著しく遅れていた。日本においてこのたび世界に先駆けて「生理活性脂質データベース」が構築されたことから、さらにこれを有効に利用するための包括プロジェクト、LIPOMICSプロジェクトをポストゲノムプロジェクトとして提案したい。このLIPOMICS

プロジェクトは、新たに脂質化合物の遺伝子発現と脂質化合物の生体における機能発現に関するデータベースを構築して、すでに我々が開発した「生理活性脂質データベース」とともに、ゲノムデータやタンパク質データと相互にリンクした新しいデータベースとして、生体情報ネットワークを構築しようとするものである。本プロジェクトは、バイオインフォマティクス（生物情報学）の一分野でもある。すなわち、生物をまるごと理解するために、生体を構成する脂質化合物の一つ一つが、細胞のどの部分に存在し、それが時間と共にどのように変化するのか、さらには細胞の集合体である臓器、組織における分子レベルの変化と遺伝子の発現とを関連づけたデータベース Systematic Bioactive Lipid Database（SBLD）を構築しようという提案である。

具体的には、以下のような研究をまとめてゆく必要がある。

1. 生体成分である化合物ごとに、受精から発生、成長、老化、死に至るまでの動的情報をコンピューター化して、生体成分の化合物の動的情報データベースを構築する。
2. 脂質化合物のもつ生物学的活性情報をコンピューター化して、化合物を基礎とする生理機能データベースを構築する。
3. ゲノム解析プロジェクトのもたらす遺伝子発現情報と、上に述べた化合物の動的情報データベースおよび生理機能データベースの情報を対応づけるコンピュータシステム[Systematic Bioactive Lipid Database（SBLD）]を構築する。
4. このSBLDは、SNP（一塩基変異型）データベースとリンクすることにより、オーダーメイド医療における基礎データベースとして、また再生医療の基礎データベースとして機能することができる。

このような総合的脂質データベース（SBLD）を構築し、これを既存のゲノムデータベースやタンパク質データベースとさらに相互リンクすることにより、生体全体を網羅した生体情報ネットワークを構築することができる。バイオインフォマティクス（生物情報学）は、ここにおいて初めて「生物のまるごと情報」を観ることができるようになる。

8. おわりに

生命科学分野では、間もなくヒトの全染色体(ゲノム)遺伝子の塩基配列のすべてが決定されるという時代を迎え、いわゆるポストゲノムプロジェクトについての議論が喧しい。

特に医療分野では、難病を克服するためのゲノム創薬や、個人の体質にあった投薬と治療を可能にするオーダーメイド医療が可能になるとの期待から、研究者のみならず、新しい産業の創出をもくろむ企業が、さまざまなプロジェクト計画をたてている。一方、決定された膨大なゲノム情報を管理するには、コンピューターによる情報処理が必須であり、その意味で、バイオインフォマテックス(生命情報科学)の進歩は、ヒトゲノム解明後の生命科学研究の進歩に大きな影響を及ぼす。

一方、ポストゲノムプロジェクトとして、現在政府はじめ企業の多くの人々が期待しているのは、ゲノムがつくるまるごとのたんぱく質(プロテオーム)である。しかしながらそれらの人々の発想には、ゲノムがつくるたんぱく質が生体内で機能をもつには、さらに多くの生体物質の存在とその情報が必須であるという認識が欠けている。すなわちゲノム情報をもとに作られるたんぱく質(プロテオーム)が生体内で酵素活性などの機能を発現するには、さまざまなたんぱく質以外の分子が関与するという認識である。例えば、糖鎖や脂質分子によるたんぱく質の修飾や脂質による機能発現のための環境作りなどである。これらのたんぱく質以外の物質も、ゲノム情報をもとに生体内で合成されているのだが、その過程は、プロテオームのように単純ではなくて、いくつかのゲノム情報を必要とする複雑な合成過程である。

現在、政府はじめ企業が研究費をつぎ込み、活発化しているバイオインフォマテックス(生命情報科学)分野は、遺伝子の直接の産物であるタンパク質(プロテオーム)の構造や機能を研究するプロテオミクスの研究である。しかし、生物をシステムとして理解する上で必要なのは、遺伝子やたんぱく質情報の上に、機能発現のための糖質や脂質の情報を加えて、生物体全体を網羅した生体情報ネットワークを構築することである。その視点から、グライコーム(糖

質)、リピオーム (脂質) という言葉も産まれてきている。「21世紀は脳の世紀」といわれているが、その主成分である脂質成分 (リピオーム) の情報をデータベース化し、ゲノムやたんぱく質の情報と密接にリンクする総合的生体情報ネットワークを構築されて初めて、「情報器官」としての脳を理解することができる。バイオインフォマテックス (生命情報科学) のなかで、このように生物をシステムとして理解するシステムバイオロジー分野の研究の発展が望まれている。

ここで提案したシステマティックバイオアクティブリピドデータベース (SBLD) の構築は、世界中でまだどの国でも行われておらず、世界に先駆けで日本がこれから構築できる分野である。世界各国でポストゲノムプロジェクト競争が開始されている中で、日本が世界の国々と肩を並べて研究を進めるには、他国でまだ手がつけられていない新規分野の開拓が必要である。脂質分野では糖脂質やプロスタグランジン研究など、また糖鎖構造解析など、歴史的にも世界をリードしてきた実績のあるわが国において、システマティックバイオアクティブリピドデータベース (SBLD) の構築を進める意義は大きい。

文 献

Kiyohiro Watanabe, Yasugi Etsuko and Mieko Oshima, How to Search the Glycolipid data in "LIPID_{BANK} for Web", Newly Developed Lipid Database in Japan, Trends In Glycoscience and Glycotechnology, Vol. 12, p175-184 (2000).