

《第 32 回情報処理技術検討交換会》

## PubChem の有用性① 緒論, 毒性, 新規性

### 日本製薬情報協議会関西支部勉強会での試みと 毒性・新規性調査への適用の可能性

石田 修\*<sup>1</sup>, 中嶋尚志\*<sup>2</sup>, 中村規子\*<sup>3</sup>, 本田直樹\*<sup>4</sup>,  
羽田幸代\*<sup>5</sup>, 西野順子\*<sup>6</sup>, 中尾泰美\*<sup>7</sup>, 寺浦一正\*<sup>8</sup>,  
斉藤 郁\*<sup>9</sup>, 鄭 兆雄\*<sup>10</sup>, 市岡剛宏\*<sup>11</sup>

[抄録] PubChem は Substance, Compound, BioAssay の 3 データベースから成り立っており, 公開以来収録ソース数・収載化合物数の増加が著しい。日本製薬情報協議会関西支部では PubChem の有用性について, 毒性, 新規性, 薬理作用, 活性情報の各方面より検証した。毒性面からの検証の結果 PubChem 独自の毒性情報は掲載されていなかったため, さらなる検証は行わなかった。化合物の新規性調査によく利用される CAS<sup>®</sup> Registry ファイルに未収載の化合物が PubChem に収録されていることが見いだされたが, そのうちの多くは化合物収録のタイムラグが原因で PubChem のみに収録されているものであった。新規性調査には Registry と併せて PubChem を利用したほうがよいと思われた。

[キーワード] PubChem, 毒性調査, ChemIDplus, 新規性調査, 化合物構造

#### 1. はじめに

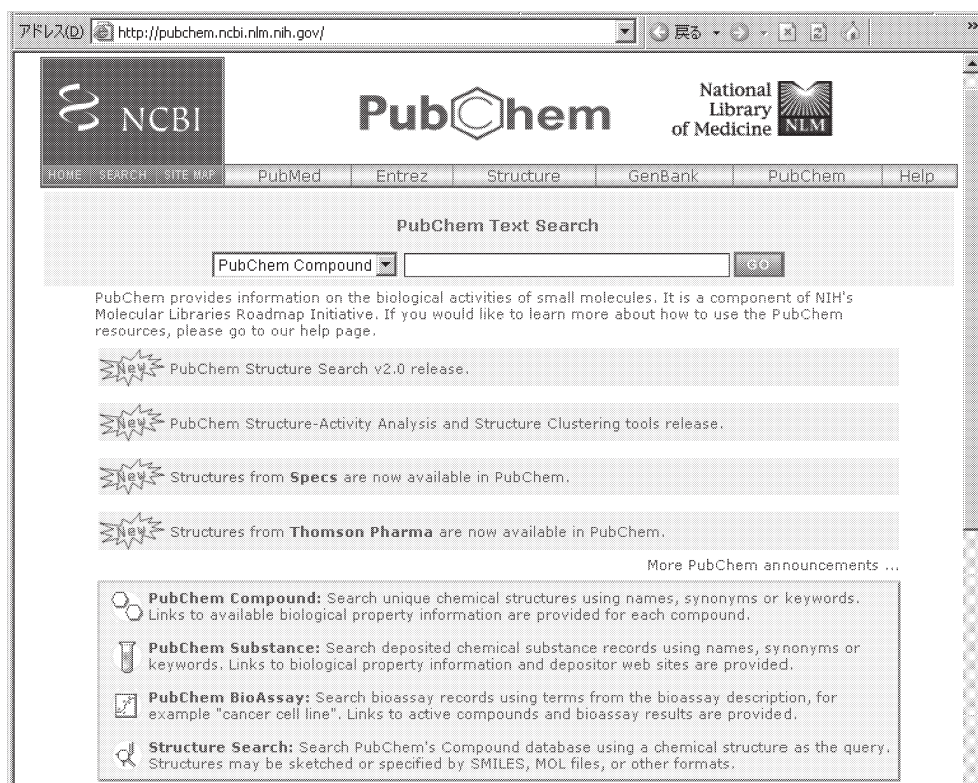
PubChem<sup>1)</sup> (図 1) は米国国立医学図書館 (National Library of Medicine, NLM) が 2004 年に一般公開した, 無料でアクセスできる化合物のライブラリーデータベースである。化合物ごとに付与されたリンクから化合物の物性情報や生物学的情報などのソースにアクセスすることができる。

PubChem は, PubChem Compound, PubChem Substance および PubChem BioAssay の 3 つのデータベースで構成され, 各々がリンク付けされ密接に関連している。PubChem Substance に収録された情報のうち, 化学的情報が確認されたユニークな構造の化合物については PubChem Compound にも収録され, PubChem Substance に収録された化合物の生物活性アッセイ方法は PubChem BioAssay に収録されることである<sup>2)</sup>。情報源については, 2007 年 2 月中

旬時点で PubChem Substance は 54 ソース, PubChem BioAssay は 22 ソースであった (表 1)<sup>3)</sup>。日本製薬情報協議会 (Pharmaceutical Information Association of Japan, 以下 PIAJ と略す) 関西支部の勉強会で PubChem の検討を取り上げた 2006 年 2 月の時点では, PubChem Substance の情報源は 34 ソース, PubChem BioAssay については 9 ソースであったことから, 約 1 年の間に, PubChem Substance については 1.5 倍以上に, PubChem BioAssay においては 2.5 倍近くに情報源が増えたことになる。

また, 2007 年 1 月末時点で PubChem Substance に 1,550 万以上の物質が, PubChem Compound には 1,016 万以上の物質が収録され, PubChem BioAssay には 395 アッセイが収録されていた。2006 年 3 月頃には, PubChem Substance に約 1,032 万物質, PubChem Compound に約 520 万物質, PubChem BioAssay に 195 アッセイが収録されていたことを考えると, 短期間のうちに急激に収録化合物数およびアッセイ数をのぼしていることがわかる (図 2)。PubChem か

(著者所属等は p. 162 に掲載)

図1 PubChem フロントページ<sup>1)</sup>

らのアナウンス一覧<sup>4)</sup>を見れば、ほぼ1カ月もしくはさらに短い期間でデータベースが刻々と更新されていることが窺える。

PIAJ 関西支部では、PubChem がどのようなデータベースであるのか、どのような使い方ができるか関心を持ち、勉強会のテーマとして取り上げた。勉強会を経る過程で PubChem の利用方法についての検討事項を挙げていき、最終的に以下の4つの事項について、関西支部内でチームに分かれて種々検討を行った。

- (1)毒性チーム：PubChem で得られる毒性データと商用データベースとの比較検討
  - (2)新規性・構造チーム：新規性調査の観点での PubChem の検討
  - (3)薬理チーム：PubMed での文献検索を目的とする PubChem の利用検討
  - (4)活性チーム：PubChem 中の創薬候補標的分子に対する情報や活性データの検討
- 本投稿では(1)の毒性チーム、(2)の新規性・構造

チームでの検証結果について記載し、(3)の薬理チーム、(4)の活性チームの検証結果については本号「PubChem の有用性②」および「PubChem の有用性③」にてそれぞれ紹介する。

## 2. 毒性チームでの検討

### 2.1. 毒性調査における適用の可能性

我々製薬企業は、患者さん達へより安全で有効性の高い医薬品を提供することをミッションとしている。したがって、医薬品の安全性に加えて毒性情報に関しても強い関心を持っている。創薬において真の意味で必要な毒性情報とは、臓器・組織・細胞レベルで薬物が示す目的外の生理作用である。皮膚刺激性などの外的な毒性、致死量など生物の生死に関わる情報は、生活や労働環境に関わる毒性情報として比較的広く公開されている。一方、臓器・組織・細胞レベルといったマクロレベルでの毒性情報は、非公開データが多く入手が困難である。

表1 PubChem Substance の情報源

Data Source Category	Data Source Name
• Biological Properties	DiscoveryGate, ChemBank, DTP/NCI, NIAID, ChemBlock, MTDP, Cambridge-Soft Corporation, LipidMAPS, BindingDB, ChEBI, PDSP, xPharm, Diabetic Complications Screening, NINDS Approved Drug Screening Program, Lead-Scope, SGCoxCompounds, SGCStoCompounds, PANACHE, Structural Genomics Consortium
• Chemical Reactions	UPCMLD, CMLD-BU
• Imaging Agents	MOLI, MICAD
• Journal Publishers	Thomson Pharma, Prous Science Drugs of the Future, Nature Chemical Biology, Web of Science
• Metabolic Pathways	KEGG, BioCyc, BIND, UM-BBD
• Molecular Libraries Screening Center Network	San Diego Center for Chemical Genomics, NCGC, NMMLSC, Emory University Molecular Libraries Screening Center, CC_PMLSC, PCMD
• NIH Substance Repository	MLSMR
• Physical Properties	NIST, ChemExper Chemical Directory, NIST Chemistry WebBook, NMRShift-DB
• Protein 3 D Structures	MMDB, SMID
• Substance Vendors	ZINC, ChemBridge, ASINEX, Specs, Total TOSLab Building-Blocks
• Theoretical Properties	ChemDB
• Toxicology	ChemIDplus, EPA DSSTox
• なし	SPRESI, KUMGM

注) PubChem Source Information<sup>3)</sup> を基に作成

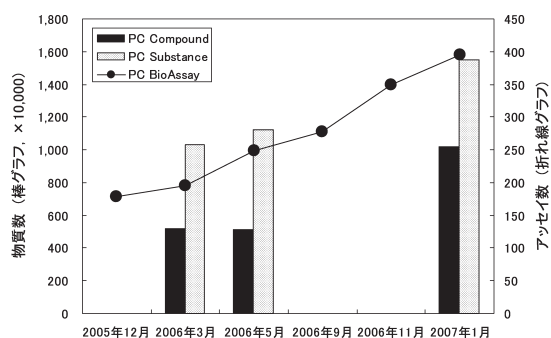


図2 PubChem 収録物質数, アッセイ数の推移

構造活性相関データは、特定の薬理活性と構造式の間を示すデータであり、その研究テーマのみに有用である。それに対して、化合物の構造と毒性の強さの関係を示す構造毒性相関データは、研究テーマにかかわらず役立つことから、より幅広い有用性を示す。この汎用性が創薬研究におけ

る毒性ファクトデータの重要性を高めているが、信頼性の高い情報は先に述べた理由で入手が困難である。このような背景から PubChem Compound 検索結果画面に「NLM Toxicology」のリンクが存在することを知り、従来入手が困難であったマクロレベルでの毒性情報が存在するのではないかと期待し検証を行った。

## 2.2. 検証結果

検証した結果、「NLM Toxicology」からのリンク先は、TOXNET 上のデータベースの1つである ChemIDplus の検索結果画面であり、PubChem 独自の毒性データは存在しないことが判明した。TOXNET に含まれる各種毒性データベースに関してさらなる検証を行うことも考えてみた。しかし、勉強会での目的は PubChem の理解を深めることであるという原点に立ち返り、「毒性チーム」は TOXNET に対するさらなる検証

The image shows a series of screenshots from the ChemIDplus Lite website. The top screenshot displays the search results for 'Acetone' (RN: 67-64-1) with its chemical structure and a table of reported doses and effects. The middle screenshot shows the 'Basic Information' section with a 'File Locator' and 'Internet Locator' containing various database links. The bottom-left screenshot is an 'International Chemical Safety Card' for Acetone, detailing its hazards and safety measures. The bottom-right screenshot is the 'NIOSH Pocket Guide to Chemical Hazards' for Acetone, providing exposure limits and physical properties.

**ChemIDplus Lite Record - Acetone (RN: 67-64-1)**

Organism	Test Type	Route	Reported Dose (Normalized Dose)	Effect	Source
rod	LD50	intrapertoneal	50mg/kg(5000mg/kg)	BEHAVIORAL/COMA GASTROINTESTINAL ALTERATION IN GASTRIC SECRETION	Acute toxic Experimentelle Pathologie und Pharmakologie, Vol. 59, Pg. 216, 1954.
rod	LD50	oral	50mg/kg(5000mg/kg)	BEHAVIORAL/ATAXIA MUSCULOSKELETAL JOINTS	Acute toxic Experimentelle Pathologie und Pharmakologie, Vol. 59, Pg. 216, 1954.
rod	LD50	subcutaneous	50mg/kg(5000mg/kg)		Acute toxic Experimentelle Pathologie und Pharmakologie, Vol. 59, Pg. 216, 1954.
guinea pig	LD50	ip			Toxicology and Applied Pharmacology, Vol. 17, Pg. 329, 1971.

**International Chemical Safety Cards - ACETONE** (ICSC: 0087)

2-Propanone  
Dimethyl ketone  
Methyl ketone  
C<sub>3</sub>H<sub>6</sub>O / CH<sub>3</sub>COCH<sub>3</sub>  
Molecular mass: 58.1

TYPES OF HAZARD/EXPOSURE	ACUTE HAZARDS/SYMPTOMS	PREVENTION	FIRST AID/FIRE FIGHTING
FIRE	Highly flammable.	NO open flames, NO sparks, and NO smoking.	Powder, alcohol-resistant foam, water in large amounts, carbon dioxide.

**NIOSH Pocket Guide to Chemical Hazards - Acetone**

RTCS # AL3150000  
CAS # 67-64-1  
UPDATE: November 2006  
MW: 58.08  
MF: C<sub>3</sub>H<sub>6</sub>O

**Exposure Limits**

NIOSH REL: TWA 250 ppm (580 mg/m <sup>3</sup> )	
OSHA PEL: TWA 1000 ppm (2400 mg/m <sup>3</sup> )	
IDLH 2500 ppm [10%LEL] See: 67-64-1	Conversion 1 ppm = 2.39 mg/m <sup>3</sup>

**Physical Description**

MW: 58.1	BP: 133°F	FRZ: -140°F	Sol
----------	-----------	-------------	-----

図3 RTECS データへのアクセス方法 (ChemIDplus<sup>9</sup>) にて“acetone”を検索した結果とリンク先ページ

は行わず、「薬理チーム」へ合流することで発展的に解消することとした。

しかし、上記の検証過程で知り得た ChemID-

plus に関する知見を以下に簡単に紹介したい。

[RTECS データへのアクセス方法]

調べた限りでは以下の3つのリンクからのアク

セスが可能であるが、レコード数と1レコードあたりの情報量に違いが見られた (図3)。

① Basic Information 中の [Toxicity] からのリンク

- ・レコード数: 147,778レコード (2007/02/12現在)
- ・情報量: RTECS の <TOXICITY DATA> セクションの収録データのみが参照可能。

(データ自体も最新のものではなく、古いデータであることが ChemIDplus の HELP 画面に記載されている。)

② Internet Locator 中の [NIOSH ICSC] からのリンク

- ・レコード数: 1,071レコード (2007/02/12現在)
- ・情報量: RTECS のすべてのセクションの収録データが参照可能と思われる。

(①からのリンク先に比べると新しいデータが収録されている。)

③ Internet Locator 中の [NIOSH Pocket Guide] からのリンク

- ・レコード数: 635レコード (2007/02/12現在)
- ・情報量: RTECS のすべてのセクションの収録データが参照可能と思われる。

(①からのリンク先に比べると新しいデータが収録されている。)

[ChemIDplus からのリンク先]

毒性関連のデータベースを中心に 80 以上のリンクが設定されており、大きく分けて以下の3つのグループで構成されている。

① File Locator: NLM 関連のデータベースへのリンク

② Internet Locator: Web 上で公開されている生物医学関連データベースへのリンク

③ Superlist Locator: Web 上の規制情報へのリンク

2.3. 毒性調査における適用の可能性: まとめ

以上、簡単に ChemIDplus に関して紹介したが、ChemIDplus と PubChem はどちらも NLM が提供する構造型検索可能な化合物辞書データベースという点で姉妹関係にある。両者を比較した場合、前者は、NLM の提供する各種データベースに索引された化合物のみを収録対象としている

ため、レコード数は約 38 万程度と化合物データベースとしては比較的小規模である。一方、後者は生理活性を有する低分子化合物を広く収録対象としているため、レコード数も 1,500 万レコード以上と大規模である点が ChemIDplus との大きな相違点となっている。

毒性情報に関する PubChem 独自の情報を求めた当初の期待は見事に裏切られる結果となったが、今後 BioAssay Database に相当する詳細な毒性情報が提供されることを願いたい。

### 3. 新規性・構造チームでの検討

#### 3.1. 新規性調査における適用の可能性

製薬企業の研究部門では、特許取得にあたり化合物の新規性調査を行う。その際よく利用されるのが、Chemical Abstracts Service (CAS<sup>®</sup>) の Registry ファイルである。新規性調査を行う上で PubChem はどのような位置付けとなるのか、製薬企業にとって PubChem は新規性調査に欠かせないデータベースなのかという観点に立ち、Registry ファイルとの比較を4方向から行った。以降、特に明記していない場合の Registry ファイルは STN<sup>®</sup> で利用した。

##### 3.1.1. 開発化合物収録状況の比較

製薬企業でよく利用されている Prous Science 社の Integrity<sup>®</sup> を用い、開発段階の高い化合物 (フェーズ 2 から上市) と低い化合物 (前臨床以前) を企業名と分子量で絞った。該当する化合物をそれぞれ 24 件、50 件抽出し、その収録状況を比較した。その結果、それぞれの収録率は Registry ではともに 100% であるのに対し、PubChem では、96%、24% であった。

##### 3.1.2. 部分構造検索におけるヒット化合物の比較

①低分子セレンウム縮合環系化合物を選択し、PubChem および Registry にて部分構造検索を行った。その結果、PubChem および Registry でのヒット件数はそれぞれ 32 化合物、199 化合物で、後者は前者をすべて包含していた。PubChem ヒット化合物での CAS 登録番号<sup>®</sup> 付与率は 94% であった。

##### ②医薬品「タミフル<sup>®</sup>」(リン酸オセルタミビ

ル) の基本骨格をもとに PubChem および Registry にて部分構造検索を行った。その結果, PubChem および Registry のヒット件数はそれぞれ 58 化合物, 229 化合物であった。PubChem には Registry と異なる 9 つの立体異性体の記載はあったが, 置換基が異なる化合物は見られなかった。

### 3.1.3. PubChem 収録のうち BioAssay で活性のある化合物の Registry での収録状況

PubChem BioAssay の “Discovery of Novel Allosteric Agonists of the M4 Muscarinic Receptor” (AID=412) で活性のある 72 化合物のうち, 58 化合物について SciFinder<sup>®</sup> で同一化合物があるか確認した。その結果, 58 化合物のうち 13 化合物 (約 22%) が PubChem にのみ収録されていた。なお, SciFinder<sup>®</sup> に収録されていた化合物はいずれも 「No Reference」との表示があった。また, 他のデータベース (Beilstein, MDL<sup>®</sup> Patent Chemistry Database, Integrity<sup>®</sup>) にも収録はなかった。検討数日後に, PubChem にのみ収録されていた 13 化合物については SciFinder<sup>®</sup> に収録されており, 一部の化合物には出典も記載されていることを確認した。

### 3.1.4. PubChem 低分子レポジトリ化合物の Registry での収録状況

National Institutes of Health (NIH) の Molecular Libraries Small Molecule Repository (MLSMR) に収録された化合物の中から, 低分子化合物, 高分子化合物を各 15 化合物ずつ選択し, Registry での収録状況を見た。その結果, 全 30 化合物中, 高分子化合物 1 件が PubChem のみの収録であった。

## 3.2. 新規性調査における適用の可能性： まとめ

以上の結果から, すでに開発化合物としてデータベースに掲載されている化合物の調査については, Registry でほぼ十分であるが, 新規性調査には Registry と併せて PubChem を利用したほうがよいと思われた。また, PubChem は研究分

野で新たな種を探すツールとしても有用である可能性が示唆された。

## 4. おわりに

PubChem の膨大な化合物情報は, 現在のところ無償で提供されている。冒頭でも触れたが, PubChem は 2004 年 9 月のリリース以来, 約 2 年の間に急速に収録化合物数および収録アッセイ数をのぼしており, インターフェースも含めて今後も目まぐるしく変化していくものと思われる。PIAJ 関西支部勉強会で検討した, 後の 2 報の結果もあわせて, 情報担当者としては今後も PubChem の動向に注目していくべきと思われる。

なお, 本研究の成果は製薬企業の情報調査の実務者の集まりである日本製薬情報協議会 (PIAJ) で得たものである。PIAJ では, 毎年テーマを決めて研究しており, 関心のある企業情報担当者のさらなる参加を期待したい<sup>6)</sup>。

執筆にあたり, (株)サンメディア・松下茂氏および PIAJ 会長の大正製薬(株)・小河邦雄氏に多大なご協力・ご助言を賜りました。この場をお借りし, 深謝いたします。

## 参 照 U R L

- 1) PubChem. (online), available from <<http://pubchem.ncbi.nlm.nih.gov/>>, (accessed 2007-02-27).
- 2) NLM. “PubChem Help”. PubChem (online), available from <<http://pubchem.ncbi.nlm.nih.gov/help.html>>, (accessed 2007-01-30).
- 3) NLM. “PubChem Source Information”. PubChem (online), available from <<http://pubchem.ncbi.nlm.nih.gov/sources/>>, (accessed 2007-02-19).
- 4) NLM. “What’s new in PubChem”. PubChem (online), available from <<http://pubchem.ncbi.nlm.nih.gov/pcnews.html>>, (accessed 2007-01-30).
- 5) ChemIDplus. (online), available from <<http://chem.sis.nlm.nih.gov/chemidplus/chemidlite.jsp>>, (accessed 2007-02-13).
- 6) 日本製薬情報協議会 関連サイト <<http://piaj.sub.jp/ring/>>, (accessed 2007-02-28).

(原稿受付：2007.2.28)