

《第 32 回情報処理技術検討交換会》

## PubChem の有用性② 薬理作用

### Pharmacological Action の付与状況とその応用

網本 淳子<sup>\*12</sup>, 石田 修<sup>\*1</sup>, 松尾 政芳<sup>\*13</sup>, 中江 貴彦<sup>\*14</sup>, 中 嶋 尚 志<sup>\*2</sup>,  
羽田 幸代<sup>\*5</sup>, 斉藤 郁<sup>\*9</sup>, 鄭 兆 雄<sup>\*10</sup>, 市岡 剛 宏<sup>\*11</sup>

[抄録] PubChem は様々な Source から多数の化合物を収録し、豊富な検索機能を備えている。しかも無償で使用できることからその有効活用方法を模索した。今回、PubChem に MeSH の Pharmacological Action が付与された記録があることに着目し、その付与状況を MeSH と比較することで検証した。化合物としては MeSH の約 91% が PubChem に収録されていること、付与された薬理作用は MeSH とほぼ 100% 同等であることが確認できた。その結果、構造と活性情報の得られる化合物データベースとしての利用以外にも、PubMed を検索する際の化合物辞書としての利用、例えば部分構造や物性条件と薬理作用を組み合わせた化合物群に関する PubMed の文献を容易に得られるなど、MeSH を用いた検索では従来難しかった事例にも対応可能なことが見出された。

[キーワード] PubChem, NCBI, MeSH, PubMed, 薬理作用, 情報検索

#### 1. はじめに

PubChem (<http://pubchem.ncbi.nlm.nih.gov/>) の概要については本号「PubChem の有用性①」で述べられているので本稿では割愛するが、筆者らは PubChem 中の医薬品や生理活性物質の記録の多くに NLM のシソーラスである MeSH が付与されていることに着目し、特に Pharmacological Action 項目の付与状況を検証することによって、その応用の可能性を検討したのでここにその結果を報告する。なお本稿は、日本製薬情報協議会<sup>1)</sup> 関西支部勉強会における検討結果を第 32 回情報処理技術検討交換会 (2006 年 11 月 30 日) にて発表し、その後補足を加えたものである。そのため最新の状況とは異なる部分もあるかもしれないがご容赦願いたい。

##### 1.1. PubChem Compound と PubChem Substance

PubChem を構成する 3 つのデータベースのう

ち、PubChem Compound (以下 PC Compound) と PubChem Substance (以下 PC Substance) は共に物質単位で、PC Substance についてはさらに Source ごとに記録が構成されており、相互にリンクしている。図 1 に両者の関係を模式図で示す。例えば Compound A が 2 つの Source を持つ場合、PC Substance には Source ごとに 2 件の記録が存在するが、PC Compound にはこれらを統合した 1 記録が作成される。ただしすべての物質がこのような関係にあるわけで

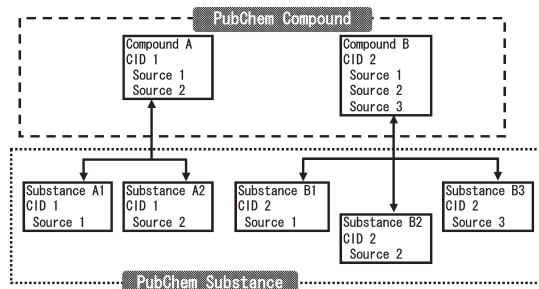


図 1 PubChem Compound と PubChem Substance の関連

(著者所属等は p. 162 に掲載)

The image shows two side-by-side screenshots of the PubChem Compound Summary page for CID 3308. The left screenshot (Page 1 of 2) shows the main summary with a chemical structure and various tabs. The right screenshot (Page 2 of 2) shows detailed properties, descriptors, and category information. Numbered callouts 1 through 7 point to specific sections: 1. Compound Summary, 2. MeSH, 3. Synonyms, 4. Properties, 5. Descriptors, 6. Category, and 7. Exports.

図2 PubChem Compound のレコード構成  
 注) <http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=3308> に追記

はなく、核酸配列や高分子化合物など、PC Substance のみで対応する PC Compound のレコードが存在しない化合物もある。次に、PC Compound のレコード構成を図2に示す。

## 2. PubChem の応用の可能性

PubChem は、本号「PubChem の有用性①」で報告されているように多岐にわたる Source から多数の物質を収録している。また表1のとおり構造検索 (部分構造検索, 類似構造検索も可能) を含む多彩な検索項目があり<sup>2)</sup>, しかも無償で利用できるというのは大きなメリットと考えられ、製薬企業、特に創薬の現場においては化合物の構造と Pharmacological Action との相関性を示すデータを必要とされるケースが多いため、その利用価値を見出すことを目的に今回の検討を行った。

### 2.1. 生理活性物質データベースとしての利用

化合物の構造と Pharmacological Action の両者を収録・検索できるデータベースは種々存在するが、古くから知られている伝統的化合物から開

発候補品として創製されるごく新しいものまでを広くカバーするものは限られている。化合物のデータベースとしては最大級の CAS<sup>®</sup> の REGISTRY ファイルでは、柔軟な条件による構造検索には対応できても、何らかの Pharmacological Action もしくは具体的な Action を有するという条件を指定しての検索には現状では対応できない。PubChem における Pharmacological Action の付与状況を検証することにより、生理活性物質データベースとしての有用性とその位置付けを把握することを目的とした。

### 2.2. PubMed の検索ツールとしての利用

医学文献データベースである MEDLINE は NCBI より PubMed として無償で提供されていることや、階層構造を持つシソーラスの MeSH を用いた索引により包括的で適合性の高い検索が可能なることから、医学・薬学の研究現場や医療従事者に広く普及している。MeSH 自体も NCBI より MeSH Database<sup>3)</sup> として検索・参照が可能であり、その結果を元に PubMed の文献集合が容易に得られるように設計されている。しかし、

表1 PubChem Compound および PubChem Substance の検索項目

	検索フィールド		別名
全体	All		[ALL]
	Filter		[Filter]
収録源	SourceName		[SRC, SRCNAM, SRCNAME]
	SourceID		[SRID, SRCID]
	SourceReleaseDate	※ S	[SRD, SRDAT, RLSDAT]
	SourceCategory		[SRCC, SRCCAT, SRCCATG]
識別番号	Uid		[UID]
	CompoundID		[CID]
	ComponentCID	※ S	[CCID]
	NeutralizedCID	※ S	[NCID]
	StandardizedCID	※ S	[SCID]
	SubstanceID		[SID]
名称	Synonym		[SYNO]
	CompleteSynonym		[CSYN, CSYNO]
	IUPACName		[UPAC, IUPAC]
	InChI		[inchi]
MeSH	MeSHTerm		[MSHT, MESHT]
	MeSHTreeNode		[MSHN, MESHTN]
	MeSHDescription		[MHD]
	PharmAction		[PHMA, PHARMA]
BioAssay	BioAssayID		[BAID, AID]
	ActiveAid		[AA]
	ActiveAidCount		[AC, ACNT]
	InactiveAid		[IA]
	InactiveAidCount		[IC, ICNT]
	ActiveAidRatio		[AAR]
	TotalAidCount		[TAC]
	AssaySourceName		[ASRC, ASRCNAM, ASRCNAME]
CHEMISTRY	AtomChiralCount		[ACC, ACCNT]
	AtomChiralDefCount		[ACDC, ACDCNT]
	AtomChiralUndefCount		[ACUC, ACUCNT]
	BondChiralCount		[BCC, BCCNT]
	BondChiralDefCount		[BCDC, BCDCNT]
	BondChiralUndefCount		[BCUC, BCUCNT]
	Complexity		[CPLX]
	CovalentUnitCount		[CUC, CUCNT]
	Element		[ELMT, EL]
	ExactMass		[EMAS, EXMASS]
	HeavyAtomCount		[HAC, HACNT]
	HydrogenBondAcceptorCount		[HBAC, HBACNT]
	HydrogenBondDonorCount		[HBDC, HBDCNT]
	IsotopeAtomCount		[IAC, IACNT]
	MolecularWeight		[MW, MWT, MOLWT]
	MonoisotopicMass		[MMAS, MIMASS]
	RotatableBondCount		[RBC, RBCNT]
	TautomerCount		[TC, TCNT, TTMC]
	TotalFormalCharge		[TFC, CHG, CHRG]
	TPSA		[TPSA]
XLogP		[XLGP, LOGP]	
その他	DepositDate	※ S	[DDAT, DEPDAT]
	ModifiedDate	※ S	[ModifyDate]
	CreateDate	※ C	[CreateDate]
	Comment	※ S	[CMT]

注) ※ C: PubChem Compound のみ, ※ S: PubChem Substance のみ (PubChem Help<sup>2)</sup> を基に作成)

表2 Pharmacological Action を持つ MeSH 語のリスト (抜粋)

RecordUI	MeSH	Pharmacological Action
D 000001	Calcimycin	Anti-Bacterial Agents   Ionophores
D 000002	Temefos	Insecticides
D 000040	Abscisic Acid	Plant Growth Regulators
D 000070	Acebutolol	Adrenergic beta-Antagonists   Anti-Arrhythmic Agents   Antihypertensive Agents
D 000072	Acedapsone	Anti-Infective Agents   Antimalarials   Leprosstatic Agents
D 000074	Acenocoumarol	Anticoagulants
D 000075	Acepromazine	Antipsychotic Agents   Dopamine Antagonists

特定の Pharmacological Action が付与された化合物グループの文献集合を容易に得ることはできても、部分構造や分子量など化合物の化学的特性に基づく条件を加味することや、あらかじめグループ化された化合物群から特定の化合物を除外したグループに対する文献集合を得ることは容易ではなかった。具体例を挙げると、「特定の部分構造を持つ serine proteinase inhibitors の文献」や「taxol 以外の抗癌剤と taxol の効果を比較した文献」などを MeSH と PubMed で検索することは困難であった。しかし MeSH の代わりに PubChem を利用する、つまり PubChem で作成した化合物グループに関連する PubMed の文献集合を得ることで、前述のような事例への対応も可能になると予想された。ただし、そのためには MeSH 中で Pharmacological Action が付与されている化合物の PubChem 収録率および Pharmacological Action の付与率が充分であることが必要とされるため、今回これらについての検証を行った。

### 3. 方法

#### 3.1. Pharmacological Action が付与された MeSH

NLM のダウンロードサイト<sup>4)</sup>より Pharmacologic Actions of a given substance-2006 (XML) を入手し、これを基に表2のような 6,314 語のリストを作成した。なお、この中には MEDLINE のレコード中で MeSH フィールドに索引される MeSH Heading のほか、化合物

表3 MeSH Term と Substance Name の例

1: Angiotensin II Type 1 Receptor Blockers [Pharmacological Action]

1-Sarcosine-8-Isoleucine Angiotensin II (*MeSH Term*)

candesartan (*Substance Name*)

candesartan cilexetil (*Substance Name*)

eprosartan (*Substance Name*)

irbesartan (*Substance Name*)

Losartan (*MeSH Term*)

olmesartan medoxomil (*Substance Name*)

Saralasin (*MeSH Term*)

telmisartan (*Substance Name*)

フィールドに索引される Supplementary Concept Records (MeSH データベース中では (*Substance Name*) と表示) を含み、以後単に MeSH 語と記載した場合でもこれらの両者を指す。Pharmacological Action として Angiotensin II Type 1 Receptor Blockers が付与された MeSH 語<sup>5)</sup> の例を表3に示す。

#### 3.2. Pharmacological Action が付与された PubChem 化合物

同一化合物の重複がない PC Compound において、すべての Pharmacological Action 用語を INDEX で参照しながら、これらの OR 演算で検索される 11,345 化合物 (2006.7.18 現在) を得た。これらについて、化合物識別番号である CID・検索結果の表示形式で Brief を選択した時に表示される代表的名称・Synonyms の3項目か

表4 Pharmacological Action が付与された PubChem Compound のリスト (抜粋)

CID	Brief 名称	Synonyms
40486	Calcimycin	Calcimycin, Ionophore A 23187, Antibiotic A 23187, Spectrum_001976, CBiol_001742, KBioGR_000296, KBioSS_000296, KBioSS_002542, DivK 1 c_000618, KBio 1_000618...
5392	temephos	temephos, Temefos, Etidronsaeure, Tetrafenphos, Abathion, Biothion, Difenphos, Abaphos, Bithion, Nimitex...
3035210	(+)-ABC	(+)-ABC, U 71184, U-71184, 101222-80-4, 1H-Indole-2-carboxamide, N-(2-((4,5,8,8a-tetrahydro-7-methyl-4-oxocyclopropa(c)pyrrolo (3,2-e)indol-2(1H)-yl)carbonyl)-1H-indol-5-yl)-
100787	Abc 907	Abc 907, Abc-907, 3-Chloro-6-hydrazinopyridazine, 3-Chloro-6-hydrazino-pyridazine, 6-Chloro-3-hydrazino-pyridazine, NSC367616, ZINC01225177, 3(2H)-Pyridazinone, 6-chloro-, hydrazone, 1/C4H5ClN4/c5-3-1-2-4(7-6)9-8-3/h 1-2H, 6H2, (H, 7,9, 17284-97-8)
65141	Trizivir	Trizivir, Epzicom, Ziagen, Abacavir sulfate, ABC sulfate, 1592U89 sulfate, (component of) Trizivir, Abacavir sulfate [USAN], HSDB7154, DRG-0257...

らなるリスト (表4) を作成した。ただし PubChem の一覧表示において Synonyms 形式を選択した際、多数の同義語を持つ化合物でも先頭から 10 個程度しか表示されないため、作成したリストもすべての Synonym を含むものではない。また、この時点では PubChem のリストに Pharmacological Action を追加することはできなかった。

### 3.3. MeSH と PubChem の化合物比較

両者の収録化合物を比較するにあたり、MeSH の化合物が PubChem に収録されているか否かを調査した。まず第1段階として MeSH と PC Compound の代表的名称を Microsoft® Excel の VLOOKUP 関数を用いて比較し、この照合に失敗したものについては第2段階として MeSH の名称と PubChem の Synonyms を照合した。その結果、MeSH と PubChem の Brief 名称が一致したものが 3,019 件、MeSH と PubChem Synonym が一致したものが 1,135 件であり、この段階で一致を確認できないものが 2,158 件存在した。ただし前述のように PubChem のリストに含めた Synonym は先頭から 10 個に制限されているため、PubChem のレコード中で 11 個目以降に記載された Synonym と MeSH が一致する可能性は十分にあり得る。そこで第3段階として表2のリストを基に検索式を作成し、実際

表5 検証に使用した検索式の例

"Acepromazine" [MeSHTerm]<sub>(A)</sub>  
and "Acepromazine" [CSYN]<sub>(B)</sub>  
and "Antipsychotic Agents" [PharmAction]<sub>(C)</sub>  
and "Dopamine Antagonists" [PharmAction]<sub>(C)</sub>

に PubChem を検索してヒットが得られるかどうかで PubChem 収録の有無を判断した。作成した検索式の例を表5に示す。この例では、下線(A)と(B)の部分で「"Acepromazine" が Complete synonym フィールド (接尾辞 CSYN) に存在する物質には、MeSH Term フィールドに "Acepromazine" が付与されている」という論理式になり、この条件でヒットが得られれば、その MeSH 語は PubChem に収録されているとみなすことができる。なお次の段階として、付与されている Pharmacological Action の同一性についての検証も予定していたので、下線(C)のようにその MeSH 語に付与されている Pharmacological Action が PubChem に存在するか否かを確認する検索語も追加して検索を実施した。第1章で述べたように PC Compound ではなく PC Substance のみに収録されている物質も存在するため、ヒットが得られなかった場合は PC Substance でも検索を行い、さらにそれでもヒットしない場合は順次「(A) and (B)」「(A)のみ」

表 6 MeSH 6,314 件の PubChem 収録状況

化合物の照合方法	結果
(1) MeSH と PubChem Brief 名称が一致	3,019 件
(2) MeSH と PubChem Synonym が一致	1,135 件
(3 a) 検索式(A) * (B) * (C) PC Compound	891 件
(3 b) 検索式(A) * (B) * (C) PC Substance	477 件
(3 c) 検索式(A) * (B) PC Substance	6 件
(3 d) 検索式(A) PC Substance	218 件
(4) (1)~(3)で確認できないもの	568 件

(A)MeSHTerm, (B)CompleteSynonym, (C)PharmAction

表 7 MeSH と PubChem での化合物の対応

MeSH	CID	IUPAC (PubChem)
Flurbiprofen	3394	2-(3-fluoro-4-phenyl-phenyl)propanoic acid
〃	41984	sodium 2-(3-fluoro-4-phenyl-phenyl)propanoate
〃	5282495	sodium 2-(3-fluoro-4-phenyl-phenyl)propanoate dihydrate
〃	92337	(2 R)-2-(3-fluoro-4-phenyl-phenyl)propanoic acid
〃	72099	(2 S)-2-(3-fluoro-4-phenyl-phenyl)propanoic acid

CID: PubChem Compound ID

と検索条件を緩和し、ヒットが得られない原因がどの検索語に起因するものかがわかるようにした。これらの結果を表6に示す。(3 d)に示す218件は、MeSH 語単独ではヒットするにもかかわらず同一の語がSynonymsに存在することが機械的に判断できなかったものである。ここでMeSHとPubChemにおける化合物の扱いの違いに注意する必要がある。PubChemでは塩・水和物・立体異性体などは別々の化合物として登録されているのに対し、MeSHではこれらは区別されず同一のMeSH語が付与される。その例を表7に示す。つまり実際にはFlurbiprofen Sodiumに関する文献であっても、PubMed (MEDLINE)に索引されるのはFlurbiprofenということを見ると、ここでは両者は対応関係にあるとみなすべきである。この観点より(3 d)の218件について、MeSH Databaseの該当レコードを参照し、そのEntry Term (付与の対象となる化合物)とPubChemのSynonymsを目視で比較した。その結果、218件とも単複数形の差・塩の有無による差・呼称の略称(ex. Immunoglobulin GとIgG)などの差で機械的照合の際に

は名称が一致しなかったが、索引されるMeSH語というレベルでは同一とみなされるものであった。これらの結果を総合すると6,314件のうちPubChemで収録が確認できないものは568件(約9.0%)であった。

### 3.4. Pharmacological Actionの同一性比較

前項で述べたように表6の(3 a)~(3 d)では、MeSHから作成した検索式を用いて実際にPubChemを検索し、付与されたPharmacological Actionが一致するかどうかを検証した。Pharmacological ActionのデータがMeSHとPubChemで異なるもの、つまりPharmacological Actionに関する検索語を除外すればヒットするものは(3 a)~(3 c)の1,374件中(3 c)の6件のみであり、その原因は複数のPharmacological Actionの一部しかPubChemに入力されていなかったり、表現が若干異なったりしていたもので、件数的にも誤差範囲のエラーとみなすことができると考えた。

#### 3.4.1. プログラムを使用した検証

前項の検証は、日本製薬情報協議会関西支部会のメンバーで分担し、1件ずつ手動で検索を実行

した。次に別の方法として、PubChemの化合物識別番号(CID)のデータを用い、プログラムを使用してPharmacological Actionのデータ取得を試みた。PubChemを含むEntrezではEntrez Utilities Web Serviceと呼ばれるApplication Programming Interface(API)を提供しており、それに従ってスクリプトを作成すればよいが、今回は独自のプログラムを作成したのでその方法について言及する。例えばフェノール(CID=996)の情報は以下のURLで参照できる。

<http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=996>

つまり、996と書かれている箇所を書き換えれば任意のCIDの化合物情報を参照できる。また、上記URLの内容はHTMLと呼ばれるマークアップ言語で作成された文書であり、一般の利用者はInternet Explorerなどのブラウザを通してきれいに整形された文書として閲覧しているが、ここを機械的にパース(構文解析)してPharmacological Action:以降の各文字列(CID=996の例ではAnti-Infective Agents, LocalDisinfectants, Sclerosing Solutions)を取得するプログラムを記述すればよい。表6に(1)で示した3,019件は、MeSHとPubChemで代表的名称を照合した際に各MeSH語に対応するPubChem化合物の識別番号(CID)を取得済みであったので、このプログラムを適用して得られたPharmacological ActionとMeSHのそれとを比較した結果、両者が同一であったものは2,986件であった。ただし例外があって、例えばCID=100094の化合物の場合は2つのMeSH語がリンクされている関係で、最初に表示される画面上ではPharmacological Action:は隠れている。プログラムによる照合ではこうした例外的な事象すべてに機械的に対応することはほぼ不可能なので、エラーとなった33件については手作業による検索と確認を行った結果、PubChemのレコードに記載されているPharmacological Actionとダウンロードサイトより入手したMeSHのデータとの一致を確認できたものが32件であった。残りの1件(CID: 5386591, Ajoene)は、ダウ

ンロードサイトより得たデータには5個のPharmacological Actionが記載されていたが、PubChemのレコードにはそのうちの4個のみが記載されていた。表6の(2)については化合物の照合時に対応するCIDを取得していなかったため、この方法による確認作業は行っていない。

ここで注意すべき点として、プログラムを使用して情報を一括取得する場合には、決してサーバに必要以上の負荷をかけないことに細心の注意を払う必要がある。プログラムを用いる自動的な情報取得は、人が操作するのと違って非常に短い間隔で連続動作を行うので、サーバに大変な負荷を与えてしまう危険性がある。組織内に設置されたインハウスのデータベースならばともかく、こうした公共のデータベースを使う以上はネットワーク負荷について留意するのが最小限のマナーである。今回の場合は、1クエリごとに30秒の間隔をとった。

#### 3.4.2. PubChem FTP

PubChemはNCBIが提供する化合物データベースである。Entrezに統合され、PubMedその他NCBIの様々なサービスの1つとして共通に利用できるが、データベースの収載内容、即ちPC Compound, PC Substance, PC BioAssayの3つはFTPでも提供され、以下のURLから取得が可能である。

<ftp://ftp.ncbi.nih.gov/pubchem/Compound>

<ftp://ftp.ncbi.nih.gov/pubchem/Substance>

<ftp://ftp.ncbi.nih.gov/pubchem/Bioassay>

FTPというのは、インターネットの黎明期から存在するネットワーク上でファイルを管理し供給するための手法である。WWWのほうが後発の技術であるためInternet ExplorerなどのブラウザでもFTPからのファイル取得は可能である。ただし、HTTP(WWW)とはプロトコルが異なるため、組織のFirewallで許可されていない場合があったり、許可されていてもブラウザのプロキシの設定を然るべく行わないと閲覧はできないので注意が必要である。実際、今回の検討メンバーの中でも所属企業のPC端末からではアクセスできないケースも見られた。これらのファイルはXMLや化学構造情報を交換するための

SDF 形式で提供されているが、それぞれのファイルは非常に大きいためパースに非常に大きなリソースが必要である。このためか CID と SID (それぞれ PC Compound, PC Substance の識別番号) の関連や、CID と Synonym の関連だけを Compound ディレクトリの下に Extras ディレクトリで提供しているが、ここにはほかにも CID-MeSH, MeSH-Pharm というファイルが格納されており (表 8)、これらが筆者らが今回興味を持っている Pharmacological Action に関連があると考え、その内容を調査した。PubChem の出力やダウンロード機能では、特定条件で検索される集合を対象に MeSH や Pharmacological Action を一括して抽出することができなかったため、前項までに記述したような検証を行ってきたが、これらの FTP 上のデータが CID と MeSH の対応および MeSH とその Pharmacological Action の対応であることがこの段

階になって判明した。同じ NLM 上のデータであるので当然といえば当然かもしれないが、表 6(2) の 1,135 件について、ダウンロードサイトから得られた Pharmacological Action と FTP 上の MeSH-Pharm を照合させたところ、1 件を除いて一致した。一致しなかった 1 件 (MeSH: tetramethylpyrazine) は、ダウンロードサイトには 8 個の Pharmacological Action が記載されていたのに対し、MeSH-Pharm および実際の PubChem 上ではこのうち 6 個しか記載がなかった。これらの結果を総合すると、確認対象とした表 6(1)~(3c) までの 5,528 件中、5,520 件 (約 99.9%) で Pharmacological Action の一致が確認できた。

#### 4. PubMed 検索ツールとしての応用

前章までの結果において、PubChem のレコードには MeSH とほぼ同等の Pharmacological

表 8 Pharmacological Action に関連情報が取められた FTP とその内容

```
$ lftp ftp://ftp.ncbi.nih.gov/pubchem/Compound/Extras/
lftp ftp.ncbi.nih.gov:/pubchem/Compound/Extras> ls
-r--r--r-- 1 ftp      anonymous 2361912 Jan 28 08:28 CID-MeSH
-r--r--r-- 1 ftp      anonymous 44643831 Jan 28 08:31 CID-Parent. gz
-r--r--r-- 1 ftp      anonymous 83399765 Jan 28 08:36 CID-SID. gz
-r--r--r-- 1 ftp      anonymous 59843277 Jan 28 08:31 CID-Synonym. gz
-r--r--r-- 1 ftp      anonymous 281877 Jan 28 08:28 MeSH-Pharm
-r--r--r-- 1 ftp      anonymous 3327 Nov 5 01:30 README-Extras
```

```
lftp ftp.ncbi.nih.gov:/pubchem/Compound/Extras> more CID-MeSH
```

```
1 Acetylcarnitine
4 monoisopropanolamine
6 Dinitrochlorobenzene
```

(以下略)

```
lftp ftp.ncbi.nih.gov:/pubchem/Compound/Extras> more MeSH-Pharm
```

(中略)

```
Acemetacin      Anti-Inflammatory Agents, Non-Steroidal
Acenocoumarol   Anticoagulants
Acephate        Insecticides
Acepromazine    Antipsychotic Agents      Dopamine Antagonists
Acetaminophen   Analgesics, Non-Narcotic
```

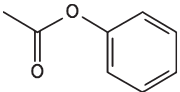
(以下略)

注) lftp という UNIX 上の FTP クライアントソフトを使って操作している



Action が付与されていることが確認できたので、その応用例として、PubChem を利用した PubMed の検索事例 (表 9) を紹介する。まず PC Compound において、Pharmacological Action として serine proteinase inhibitors が付与されたものを検索すると 37 化合物が検索された。これを #2 に示す部分構造検索の結果で絞り込むと 9 化合物となった。ここで Display 形式より PubChem via MeSH を選択すると (図 3), これらの 9 化合物に関する 1,057 件の PubMed 文献リストが表示される。PubMed via MeSH を選択する前に必要に応じて化合物リストのチェックボッ

表 9 PubChem を利用した PubMed の検索事例 (2007.2.22 実施)

Search Queries	Result
# 1 Search "serine proteinase inhibitors" [PharmAction]	37
# 2 Select 141971 document(s)	141,971
 substructure search	
# 3 Search #1 AND #2	9
# 4 PubMed via MeSH for PubChem Compound (Search #1 AND #2)	1,057

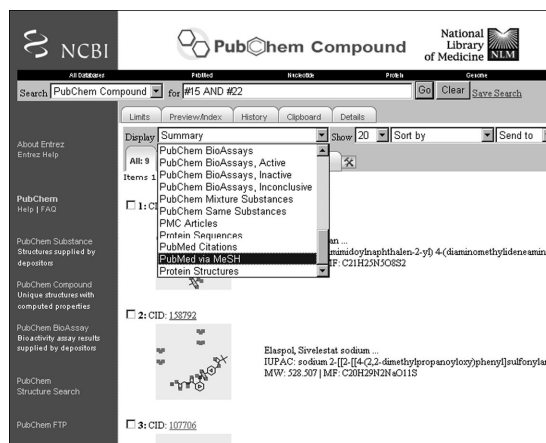


図 3 PubChem の化合物集合を対象にした PubMed の文献集合作成

クスに印を入れておけば、選択した化合物のみに関連する文献を表示することもできる。従来の MeSH を用いた MEDLINE の検索では、構造的観点や薬理的にあらかじめ分類された MeSH の階層構造を反映させることしかできなかったが、PubChem は表 1 に示す検索項目や構造検索機能を有しているため、MeSH よりも検索対象とする化合物の選定が自在に行える利便性がある。ただし、MeSH に収録されているが PubChem には対応する化合物が存在しないものは PubMed の検索対象外となることや、PC Compound と PC Substance のどちらを使用するかによって PubMed の検索対象となる化合物が変わってくるにも注意を払う必要がある。後者に関しては、PC Compound には構造の特定できない物質や高分子・抗体類はあまり収録されていないので、こういった物質を対象から除外したい場合などにはうまく利用できるケースもあると考えられる。

## 5. おわりに

MeSH とその Pharmacological Action, PubChem の化合物 ID と MeSH を対応させるデータ (それぞれ MeSH-Pharm, CID-MeSH) が PubChem の FTP サイトに存在することがわかった。よって PubChem に付与されている Pharmacological Action は MeSH のものと同一であることが想像できたが、実際にいくつかの方法でこれらを検証し確認したことは有意義であったと考えている。今回の検証により、PubChem を構造と活性情報の得られる化合物データベースとして利用する際には、その収録範囲は「MeSH で Pharmacological Action が付与されている物質」とほぼ同等と位置付けることができる。また、Pharmacological Action から PubMed を検索する際には、一部の化合物は対象外となるものの、PubChem を化合物辞書的に利用することによって、MeSH を用いた検索では従来難しかった事例にも対応可能なことを見出した。

## 参 照 URL

- 1) 日本製薬情報協会 関連サイト <<http://piaj>.

- sub.jp/ring/), (参照 2007-02-28).
- 2) PubChem Help <<http://pubchem.ncbi.nlm.nih.gov/help.html>>, (accessed 2007-02-09).
  - 3) MeSH Database <<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=mesh>>, (accessed 2007-02-09).
  - 4) Medical Subject Headings-Files Available to Download <<http://www.nlm.nih.gov/mesh/filelist.html>>, (accessed 2006-07-11).
  - 5) MeSH Database 検索結果 <[http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=mesh&list\\_uids=82047228&dopt=Full](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=mesh&list_uids=82047228&dopt=Full)>, (accessed 2007-02-16).

(原稿受け: 2007.2.28)