



構造活性相関部会・ニュースレター <1 October, 2012>

SAR News No.23

「目次」

//// Perspective/Retrospective ////

SciNetS : セマンティックウェブ技術を活用した創薬のための情報基盤

豊田 哲郎 … 2

//// Cutting Edge ////

ChEMBL 及び創薬関連 Open Database の動向

池田 和由 … 6

医薬品有害事象データベースを用いたデータマイニング

五島 誠、奥野 恭史… 12

//// Activities ////

< 報告 >

構造活性フォーラム 2012 開催報告「GPCR 研究の最前線」

石黒 正路 … 19

Hansch-藤田法 50 周年記念シンポジウム開催報告

中川 好秋 … 20

< 会告 >

第 40 回 構造活性相関シンポジウム

… 21

//// Perspective/Retrospective ////

SciNetS : セマンティックウェブ技術を活用した創薬のための情報基盤

理化学研究所 豊田哲郎

1. はじめに

ライフサイエンスの研究活動を行う上で、遺伝子情報やタンパク質構造など大規模で多種多様なデータを体系的に蓄積した情報基盤は不可欠なものになってきており、最先端技術で計測した新しいデータを真に深く理解するためには、既存の知識体系と照らし合わせる必要がある。従来は、発見された個別の知識は論文に記載される形で公開されていたが、近年にみられるようにデータ解析手法がハイスループット化したことで、論文には載せきれない1セットの知識体系(データベース)の形式で公開されるようになった。このため、多種多様なライフサイエンスデータを体系的に整理して提供する情報基盤が求められていた。そこで理研では、ライフサイエンスの多様なデータを格納し、同時にその関係性を体系的に整理することができる研究基盤として「SciNetS (サイネス、Scientists' Networking System)」を開発してきた。SciNetS は、理研が生み出すデータの公開窓口の機能を果たすだけでなく、それらの成果同士を結びつけることで、個別研究だけでは発見し得なかった、より高度な科学的知見を積極的に探索することを目指している。このために、各種データを繋ぎ合わせて相関解析をあらゆる角度から自動的に行う機能も盛り込まれている。

実際に SciNetS 上では、哺乳類や植物、タンパク質などを中心としたデータが 192 のデータベースプロジェクトについて公開されており、未公開データを扱うものまで含めると約 400 のデータベースプロジェクトが構築されている。また、扱われるデータはすべてセマンティックウェブ[1]に準拠しており、コンピュータがデータ間の関係性の意味を解釈し、高度な知識処理を行うのに適した形で提供されている(図1)。

2. セマンティックウェブに準拠した統合データベースの構築

SciNetS では多数のデータベースプロジェクトから公開されたデータテーブル群を、プロジェクト横断的に、標準語彙としてのオントロジーで階層的に束ねることが容易にできる。そのため、様々なコンセプトの統合データベース(DB)を比較的簡単に構築できる。これまでも哺乳類、植物、タンパク質を中心とする統合データベースを公開している。一例として理研哺乳類統合DBでは、理研内の FANTOM[2]、the ENU mutagenesis program[3]、RIKEN Cerebellar Development Transcriptome Database[4]、バイオリソース DB[5-7]など 9DB と、関連する 19 の理研外 DB、計 90 万件のデータをセマンティックウェブ[1]に準拠した形で公開している[8]。これら全てのデータは、ゲノム科学にも対応した上位オントロジーである YAMATO-GXO (Yet Another More Advanced Top-level Ontology-Genetics Ontology) [9,10]を標準語彙として用いることで、体系化されている。このような試みは、ライフサイエンス系の大規模 DB としては世界初の試みである。公開されたデータはすべてセマンティックウェブの標準形式(RDF、OWL、Turtle、N-Triples、RDFa など)で提供されているだけでなく、生命科学者が利用しやすいテーブル形式でも提供されている。

さらに SciNetS では、データベースプロジェクトごとに詳細なアクセス制限が付けられている。そのため、公開/非公開混合の DB 統合にも適しており、理研哺乳類統合 DB では、これらのデータベースプロジェクト機能を使うことで、約半年という短時間に統合 DB を構築することができた。

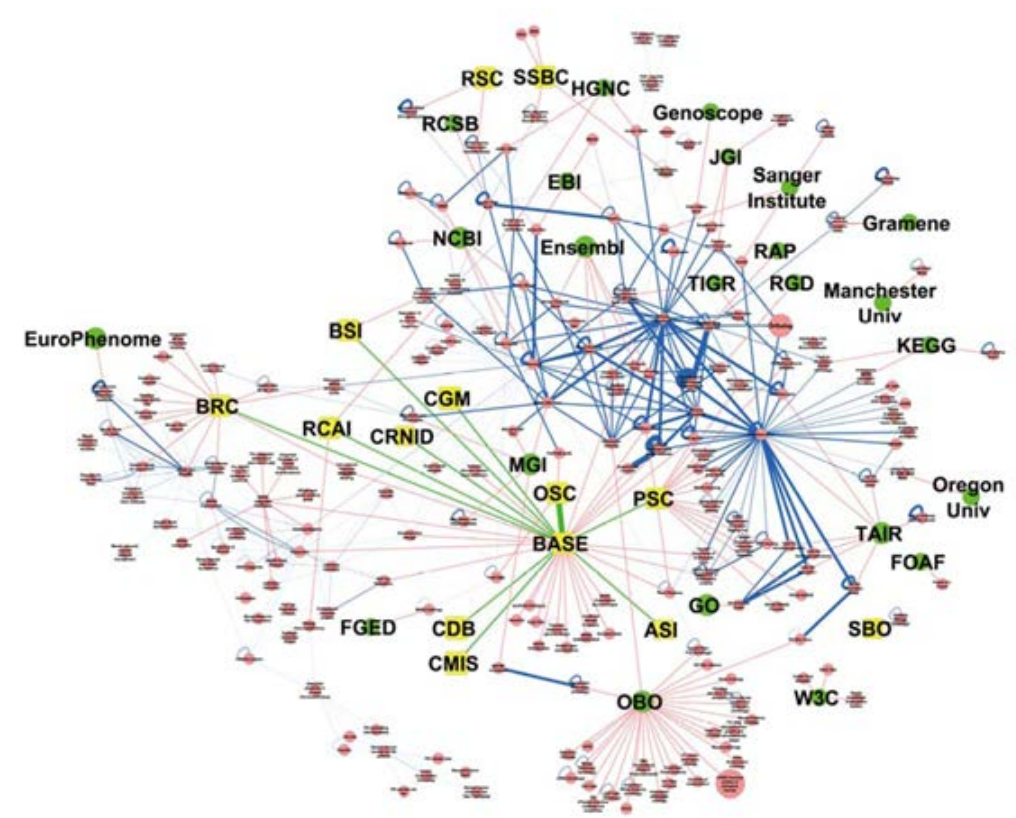


図1. SciNetS上でデータ統合されているデータベースプロジェクトネットワーク
 桃色のノードは個々の「データベースプロジェクト」、黄色と緑のノードは「現実の組織」（黄色：理研内センター、緑：理研外）を表している。青線はデータ間のリンク数をその数に応じた太さで示し、桃線はデータとそれを生み出した組織の関係を示している。

3. SciNetSを支える推論的相関検索エンジン、GRASE

SciNetSでは、セマンティックウェブに準拠したデータ構造をしているため、単なる文字列検索だけではなく、セマンティックリンクでつながるデータを追った推論的相関検索に対応している（図2）。例えば、“diabetes”で検索すると、OMIMにおける“diabetes”の説明と共に、関連する遺伝子の一覧、さらにその遺伝子とオルソログなどの関係でつながった遺伝子（直接 diabetes の文字を含まないものも含む）のリストを得ることができる。また、検索結果の表示においては、データごとにキーワードとの関連性を統計値で順位付けする「GRASE（General and Rapid Association Study Engine）」という検索エンジンを独自に開発している[11]。GRASEはSciNetSのみならず、ポジショナルクローニングによる候補遺伝子のランキングを行うウェブアプリケーション PosMed（Positional Medline）においても使われている[12,13]。PosMedでは図3に示す様に、染色体領域とキーワードを入力することで、キーワードに関連の強い領域内の遺伝子を順位付けし、遺伝子同定を補助する。実際にマウスの the ENU mutagenesis program においても活用され、65 以上の変異マウスの遺伝子同定に貢献してきた[14]。

また推論的相関検索エンジン GRASE は、データベースサーチエンジンとしての高い技術が認められ、2011 年に米国特許が成立した。

The screenshot displays the SciNetS web interface for the RIKEN ENU Mouse Lines database. The main content area shows detailed information for the **Gdf5<Rgs451> (MGI Alleles and Phenotypes: MGI Allele)**. Key features highlighted by annotations include:

- アレル名** (Allele name): The specific allele identifier.
- SciNetSで定義されたIDとURI** (ID and URI defined in SciNetS): The unique identifier and URI for the allele.
- MGIの同じアレルデータへのリンク** (Link to the same MGI allele data): A link to the corresponding MGI allele data.
- MGIのGdf5遺伝子へのリンク** (Link to the MGI Gdf5 gene): A link to the MGI gene entry for Gdf5.
- CDT-DBでのGdf5発現データ** (Gdf5 expression data in CDT-DB): A link to the Gdf5 expression data in the CDT-DB.
- ENU誘発変異アレルへのリンク** (Link to ENU-induced mutant allele): A link to the ENU-induced mutant allele.
- 画像はMouse overで拡大** (Image can be enlarged by mouse over): A note indicating that the image of the mouse can be enlarged by hovering over it.
- 詳細な理研フェノタイプ情報** (Detailed RIKEN phenotype information): A link to the detailed RIKEN phenotype information.
- オリジナルDBへのリンク集** (Collection of links to original databases): A collection of links to the original databases.

図2. SciNetSでの各データの表示例

セマンティックリンクでつながったデータをユーザが一望できる画面。左のコンテンツタブの中には、セマンティックリンクでつながりのある SciNetS プロジェクトの一覧が表示される。また Links タブの中には、セマンティックリンクでつながる各データのオリジナルデータベースへのリンクを集めている。右上の新着情報では、セマンティックリンクでつながるデータの更新情報が確認できる。

4. 統合的な相関解析の自動化と創薬への応用

SciNetS ではセマンティックウェブ形式で多数のデータベースが繋がって収納されている (図1)。このため、任意のデータベースの部分集合を投入すると、それらとリンク関係で繋がっている他のデータベースの中から相関関係が強いものを広範囲にわたって自動的に探索してくれる (この機能は現時点では非公開)。よくあるケースでは、ある生物種でマイクロアレイ実験から抽出した遺伝子群を投入すると、それと繋がっている Gene Ontology (遺伝子の機能オントロジー) の中からその遺伝子群に統計的に有意に多く含まれているものを探索するケースが多い。SciNetS では、Gene Ontology に限らず、様々な理研内外由来のデータセットと繋がっているため、これらの関係を使って Gene Ontology 以外の相関も自動解析し、新しい発見を導くことができる。この機能は、新しいデータセットを加えれば加えるほど強力になっていくため、創薬ターゲットの発見や機能解釈に応用することを期待している。

(A) 問い合わせの入力

search gene condition genomic interval species mouse

Select Interval with OmicBrowse chromosome 1 position from 90M to 140M clear version NCBIm37

keyword diabetes OR insulin gene name search clear recent 4 years 4 show 20

(B) 推論検索に用いるドキュメント指定

All Hits Total Hits: 114 (0.981 sec) Simple Mode

☒ mouse mutant ☒ HsPPI ☒ human gene record ☒ MEDLINE (sentence)

☒ mouse gene record ☒ OMIM ☒ REACTOME ☒ rat gene record

Associate the keyword with entities co-cited within the same sentences

Further associate the entities with entities co-cited within the same sentences

(C) ランキングされた候補遺伝子結果

1. **Insig2, insulin induced gene 2**

Insig2 26 docs Position: Mm.1:12320934-123224363 Link to: MGI

21 hits

diabetes OR insulin

2. **Il10, interleukin 10**

Il10 20845 docs P value: 1.06E-3417 Position: Mm.1:132916422-132921547 Link to: MGI

8899 docs P value: 1.06E-3417

Tnf 94466 docs Position: Mm.17:35336326-35338952 Link to: MGI

3871 hits P value: 2.96E-4095

diabetes OR insulin

3. **Serpinb2, serine (or cysteine) peptidase inhibitor, clade B, member 2**

Serpinb2 828 docs P value: 3.39E-1149 Position: Mm.1:109408000-109422169 Link to: MGI

712 docs P value: 3.39E-1149

Serpine1 5879 docs Position: Mm.5:137537378-137548129 Link to: MGI

1222 hits P value: 2.04E-1293

diabetes OR insulin

4. **Adora1, adenosine A1 receptor**

Adora1 2755 docs P value: 5.18E-1062 Position: Mm.1:136095800-136132004 Link to: MGI

1310 docs P value: 5.18E-1062

Adenosine 118618 docs Link to: Original

1279 hits P value: 1.09E-1353

diabetes OR insulin

(D) 候補遺伝子リストとダウンロード

download annotation

download rank list

set as target

add comment

draw

check all

1. ☐ Insig2
2. ☐ Il10
3. ☐ Serpinb2
4. ☐ Adora1
5. ☐ Ptpnc
6. ☐ Adipor1
7. ☐ Cxcr4
8. ☐ Mapkapk2
9. ☐ Capn10
10. ☐ Bcl2
11. ☐ Ramp1
12. ☐ Pigr
13. ☐ Lcl
14. ☐ Dbl
15. ☐ Pcdcl
16. ☐ Il19
17. ☐ Myog
18. ☐ Gpc1
19. ☐ Tnfrsf2
20. ☐ Ctse
21. ☐ Sox13
22. ☐ Ugt1a1
23. ☐ Sclt
24. ☐ Pam
25. ☐ Nr5a2

図3. PosMedの検索結果画面

マウスの1番染色体90M bpから140M bpの領域において、“diabetes OR insulin”というキーワードに関連の強い遺伝子のランキングを行った結果。遺伝子のランキングにおいては、推論的相関検索エンジン「GRASE」が使われている。

参考文献

- [1] Berners-Lee, T., et al.: Scientific American, 284(5): 34-43, 2001.
- [2] FANTOM Consortium, Suzuki, H., et al.: Nature Genetics, 41(5): 553-562, 2009.
- [3] Masuya, H., et al.: Mammalian Genome, 15(5): 404-411, 2004.
- [4] Sato, A., et al.: Neural Networks, 21(8): 1056-1069, 2008.
- [5] Yoshiki, A., et al.: Experimental Animals, 58(2): 85-96, 2009.
- [6] Yokoyama, K. K., et al.: Experimental Animals, 59(2): 115-124, 2010.
- [7] Nakamura, Y.: Experimental Animals, 59(1): 1-7, 2010.
- [8] Masuya, H., et al.: Nucleic Acids Research, 39(Database issue): D861-870, 2011.
- [9] 榎屋啓志: 人工知能学会誌, 25(4): 485-492, 2010.
- [10] Masuya, H. and Mizoguchi, R.: Interdisciplinary Ontology, 2: 35-44, 2009.
- [11] Kobayashi, N. and Toyoda, T.: Bioinformatics, 24(7): 1002-1010, 2008.
- [12] Yoshida, Y., et al.: Nucleic Acids Research, 37(Web Server issue): W147-152, 2009.
- [13] Makita, Y., et al.: Plant and Cell Physiology, 50(7): 1249-1259, 2009.
- [14] Masuya, H., et al.: Journal of Bioinformatics and Computational Biology, 5(6): 1173-1191, 2007.

//// Cutting Edge ////

ChEMBL 及び創薬関連 Open Database の動向

EMBL-EBI 池田和由

1. はじめに

近年、創薬研究にとって有用な大規模データの公開が進んでいる。例えば、米国 NIH Molecular Libraries Screening Center Networks (MLSCN) の成果が公開され、以前は大手製薬企業の研究者しか入手できなかった HTS データが中小製薬企業やアカデミックの研究者にも利用できるようになった。また、有料でしか手に入らなかった大量の SAR 情報も、オープンデータ化が進んでいる。こういった動きの中で、欧米を中心に熱帯病創薬への Public-Private Partnership が進展し、GSK 社や Novartis 社などがマラリアのスクリーニングデータをパブリックドメインへ無償で提供している。一方、医薬品の安全性や副作用予測に重要な毒性データに関しては、日本のトキシコゲノミクスや米国 DrugMatrix などの大型プロジェクトの成果が公開されている。このような飛躍的に増加するオープンデータを集積し、創薬に有用な情報ソースとなっているのが、PubChem や ChEMBL データベースである。本稿では、筆者が開発に携わった ChEMBL の開発状況を中心に、その他の関連データベース (PubChem、BindingDB、Binding MOAD 及び PDBbind) と、それらの一次情報を活用した二次的創薬データベース (canSAR 及び SARfari) について、以下に紹介する。

2. ChEMBL データベース

ChEMBL (ケンブル : <https://www.ebi.ac.uk/chembl>) は、医薬品及び開発化合物の SAR 情報のデータベースである[1]。前身は、企業 (Inpharmatica 社、その後 Galapagos NV 社) が開発した有償データベース (StARLite) であり、2008 年にウェルカム財団の支援によりパブリックドメインへ移管された。現在は、英国ケンブリッジ郊外にある EBI (European Bioinformatics Institute) が開発・管理をしている。その特長は、化学及び生物学の専門家によって収集・整理された質の高い SAR 情報が無償で提供されている点である。データベースの更新も定期的 (約 3 ヶ月に一度) に行われている。利用方法は、ウェブからの検索だけでなく、全データをダウンロードしてインハウスで使うことも可能であるため、アカデミアだけではなく企業ユーザーにも利用されている。特に、企業では自社のデータと ChEMBL のデータを融合させ、研究の効率化を図っているところもある。

データベースの概要

2012 年 7 月にリリースされた Version 14 には、121 万個の化合物とそれらを用いた 64 万件のアッセイ実験による 1 千万件以上の活性 (アフィニティ) 情報が登録されている。ChEMBL の情報源は、Medicinal Chemistry の論文であるが、対象となっている主な雑誌は、Journal of Medicinal Chemistry、Bioorganic & Medicinal Chemistry Letters、Antimicrobial Agents & Chemotherapy、Journal of Natural Products である。その中でも J.M.C. は 1980 年からデータ抽出を行っており、アカデミアや製薬企業における過去 30 年以上の SAR 研究の結果がカバーされている。これらの雑誌に SAR の論文が掲載されると、i) 化合物の情報、ii) ターゲットの情報、iii) アッセイ実験と生物活性値の情報を抽出し、その後キュレーターによってチェックがなされ、最終的にそれらの情報が Oracle データベースへ登録される。

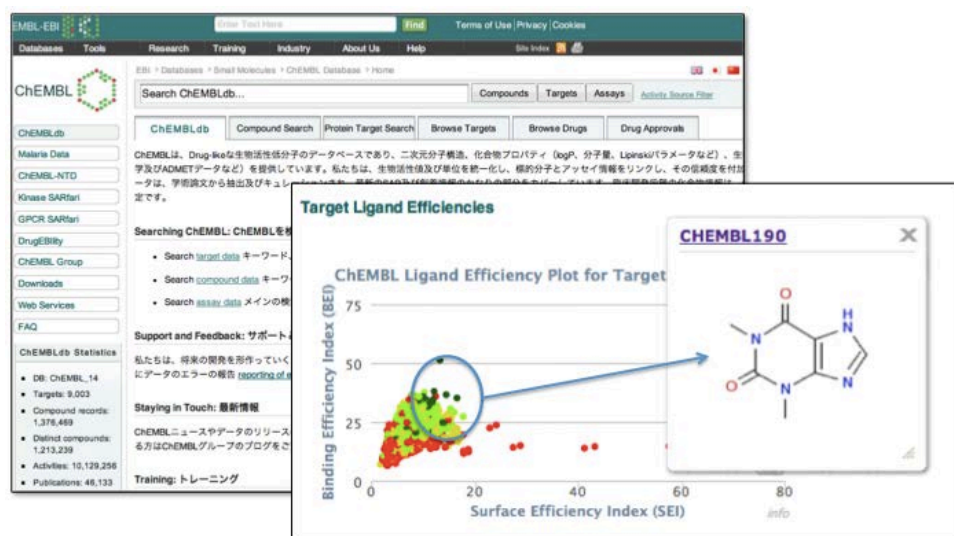


図1:ChEMBLトップページ日本語版（左上）とリガンド効率プロット（右下）

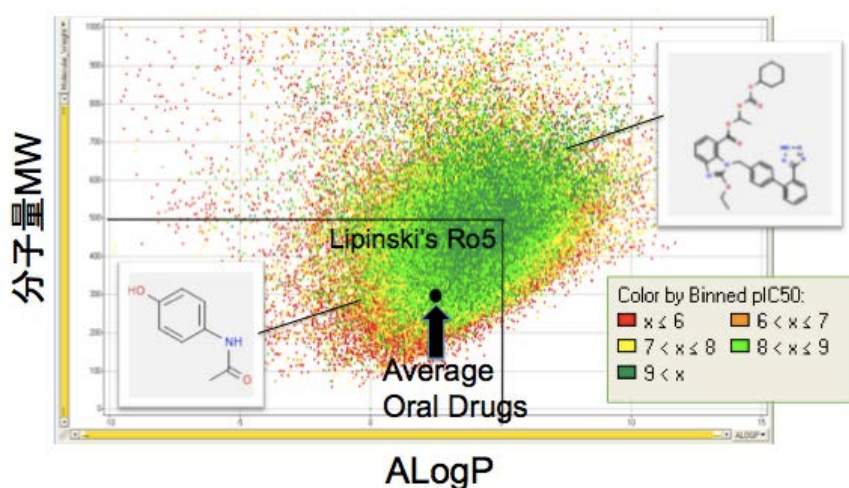


図2:ChEMBLの化合物のケミカルスペースとアフィニティの関係

データの詳細

i) 化合物（分子）データには、平面構造、プロパティ、医薬品（低分子・抗体）情報が含まれる。化合物データベースでは、化合物構造の重複をチェックして、同じ構造には同一の ID（認識番号）をつけ管理することが一般的である。ChEMBL では、化合物の登録に Standard InChI[2]を用いており、この InChI を使うことで、全ての化合物構造に固有の ID を割り振っている。また、ChEMBL には、約 2500 個（抗体 43 個を含む）の医薬品分子も登録されている。最新の承認薬に関しても、FDA Orange Book から随時情報をアップデートしている。さらに、化合物の臨床試験データ及び特許情報も新たに追加され、ウェブ上でこれらの詳細を見ることができる。

化合物プロパティの情報には、分子量、PSA、ALogP、Lipinski の Rule of Five (Ro5) などがあり、これらは科学ソフトウェア (Pipeline Pilot) を用いて計算している。最近、FBDD やリー

ドの最適化に重要な指標として注目をされているのがリガンド効率 LE (Ligand Efficiency) であるが、ChEMBL では、BEI (Binding Efficiency Index) と SEI (Surface Efficiency Index) という 2 つの異なる LE 値 (活性値をそれぞれ分子量及び PSA で割ったもの) を導入している[3]。図 1 の右下は、Adenosine A_{2A} 受容体の結合アッセイに関する LE プロットの例であり、各点が化合物、点の色はアフィニティの強さ (赤<黄<緑) を表している。この図では、リード最適化において有利な物性をもつと考えられる化合物 (分子量と PSA が比較的小さいにもかかわらず高活性であるもの) が右上に存在し、すぐにその構造を確認することができる。また、図 2 は、ChEMBL に存在する化合物のケミカルスペースを可視化したものである。この図では各点が ChEMBL の化合物であり、それらを分子量と ALogP の二次元空間に射影して作成している。この図から、ChEMBL の化合物は空間に広く分布し、よって様々な物性の化合物が存在しており、そのうち高活性の化合物 (緑色の点) の中心が、低分子経口薬の平均値よりも少し大きいことがわかる。

ii) ターゲットは、約 9 千種類登録されており、ほとんどはタンパク質である (核酸、Cell-Line や Whole Organism も少ないがある)。ターゲットの生物種の内訳は、ヒト、ラット、マウスで半分以上を占めている。ChEMBL では、生物の専門家がターゲットをタンパク質の機能やリガンドの種類によって細かく階層的に分類している。活性値が多く登録されているターゲットのクラスは上位から順に、GPCR (17%)、キナーゼ (16%)、プロテアーゼ (6%)、核内受容体 (4%)、イオンチャネル (3%) となっている。GPCR は古くから知られている有名な創薬ターゲットである。一方、キナーゼは最近 10 年のデータ増加が顕著なターゲットである。現在、ChEMBL では、創薬に特に重要なこの 2 つのタンパク質ファミリーを対象とした SARfari と呼ばれるデータベースを別に開発しており、その詳細については後で述べる。

iii) アッセイには、Binding Assay、Functional Assay 及び ADMET Assay の 3 種類がある。Binding Assay は、化合物とターゲットの結合を直接測定したアッセイである。Binding Assay によって同定される活性値のタイプは様々であり、IC₅₀、K_i、%Inhibition、K_d が多い。ChEMBL が独自に標準化した活性値である Standard Value は、単位を nM に統一して比較できるように工夫している。Functional Assay は、GPCR などの Cell-Based から Disease-derived Cell Line、Whole Organism のアッセイまで幅広く、数としても全体の半分以上を占めている。アッセイの実験条件は、Assay Description の項目に詳細が記載されている。ADMET Assay は、全体の 17%含まれているが、その検索は Assay Description に対するキーワード検索で行うことができる。例えば、“Bioavailability”で検索すると 5929 件、“Hepatotoxicity”では 292 件、“Brain-to-Plasma”では 577 件の検索結果を得ることができる。これらの情報は、化合物の薬物動態及び安全性にとって有用であり、製薬会社にとっても魅力的なデータであろう。

ChEMBL と PubChem のデータ共有

現在、ChEMBL と PubChem ではデータ共有をしており、それぞれのデータベースのアップデート時に、お互いのフォーマットに従うデータの取り込みを行なっている。ChEMBL では、PubChem Bioassay から化合物とタンパク質の相互作用に関わる活性値 (IC₅₀、K_i など) が存在するデータのみを取り込んでいる。現在、ChEMBL の中に収録されている生物活性の数で見ると、実は PubChem 由来のデータ (58%) の方が文献由来のデータ (37%) よりも多い (図 3 左)。一方で、アッセイの数では、ほとんどが文献由来のデータである (図 3 右)。これは、ChEMBL が個別の論文で報告された活性値を収録したものであるのに対して、PubChem には大規模なアッセイ実験から発生した大量の実験値が納められているためである。また、PubChem と ChEMBL の化合物のオーバーラップを調査したところ、興味深いことに両方で同一の化合物は

2%以下であった。これも、スクリーニング用の化合物ライブラリーと文献化合物の構造が違うためであろう。

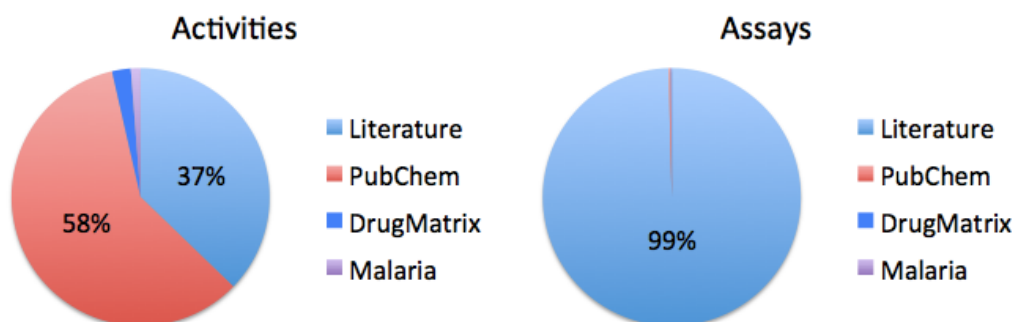


図3:ChEMBLの生物活性 (Activities) とアッセイ (Assays) の情報源の内訳

3. その他の創薬関連 Open Database の紹介

PubChem

PubChem は、米国 NCBI が管理する化学分子とアッセイのデータベースである[4]。PubChem には、化合物検索の PubChem Compound とアッセイ検索の PubChem Bioassay がある。PubChem Bioassay が、化合物を用いたアッセイの実験結果が収録されたデータベースであり、その特長は、HTS を使った大規模な化合物ライブラリーによる大量のスクリーニング結果が収録されている点である。主な情報源は、米国 NIH Chemical Genomics Center を始めとするスクリーニングセンターである。現在、5 千件を超えるアッセイのデポジット（文献由来を除く）があり、3 千万個以上もの分子情報を含んでいる。このうち、活性が確認された化合物は 45 万個ほどで全体の 2%以下である。現在、PubChem には ChEMBL 由来のアッセイも多く含まれている。PubChem では、ダウンロード機能以外にもアッセイや化合物の解析ツールなどが提供されている。

BindingDB, Binding MOAD & PDBbind

BindingDB は、カリフォルニア大学のグループが開発している低分子とタンパク質のアフィニティ情報のデータベースである[5]。タンパク質と低分子リガンドの結合情報が 91 万件以上登録されている。活性値としては、 K_i 、 IC_{50} 、 K_d 、 EC_{50} などの活性値が登録されており、PubChem や PDSP K_i Database[6]からのデータもインポートされている。また、PDB の立体構造上の結合情報をベースとした代表的なデータベースとして、Binding MOAD[7]や PDBbind[8]が知られている。これらのデータベースの利点は、リガンドとタンパク質とのアフィニティ情報がその複合体構造とリンクされており、物理的な相互作用に裏付けされている点である。

canSAR

canSAR は、英国 ICR (Institute of Cancer Research) が開発した癌研究に特化した創薬情報データベースである[9]。情報源としては、ChEMBL、BindingDB などの文献由来の SAR データ以外に、Cancer Cell Lineを用いたスクリーニングデータ (NCI-60)、遺伝子発現データ及び RNAi 実験の情報なども取り込んでおり、癌に関連するターゲットと化合物の複雑な生物化学情報を高度にインテグレーションしている。canSAR では、ChEMBL チームが開発した Druggability 予測ツール[10]を用いることで、癌の治療薬のターゲットとなりえるタンパク質の探索と、それが低分子薬の標的として妥当かどうかの評価も調べることができる。

Kinase/GPCR SARfari

SARfari は、ChEMBL チームが開発しているキナーゼと GPCR ファミリーに特化した SAR データベースである[11]。どちらも活性化化合物と活性値データは、ChEMBL から定期的にアップデートしている。Kinase SARfari では、プロテインキナーゼドメインに結合する化合物が 5 万個以上と、それらの化合物が関与する 53 万件以上の活性情報が登録されている。さらに、臨床試験の途中または中断した候補薬 (Clinical Candidates) の情報も独自に収集しており、現在 258 個の候補が登録されている。これらの情報は、将来の副作用予測の観点からも有用な情報であろう。Binding Site Search では、キナーゼドメインの配列アライメントから、リガンドと相互作用している部位のみを抽出し、その物理化学的な性質を考慮したリガンド結合部位の類似性を検索することができる。例えば、リガンド結合部位が酷似しているキナーゼの活性化化合物情報を集めてフォーカストライブラリーの設計に役立てることができる。また、類似性密度 (Neighbourhood Density) を計算することで、そのターゲットの選択性を考察することも可能である。3D Structure では、キナーゼの立体構造を比較することができるので、SBDD への活用も可能である (図 4 右下)。

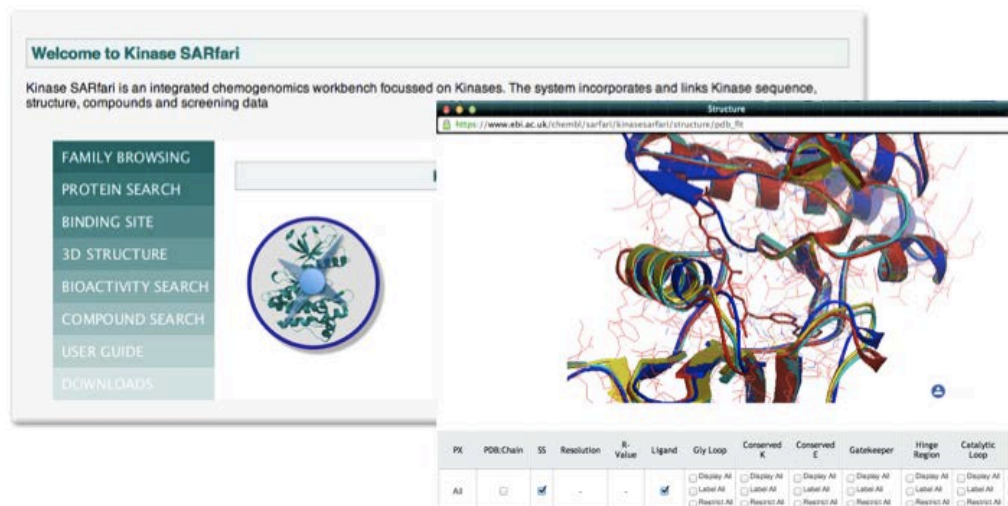


図4:Kinase SARfariのトップページと3D Structure Viewer

4. おわりに

昨年、MIABE (Minimum Information About a Bioactive Entity) と呼ばれる生物活性データベースにおけるデータ共有のため指針が提案され、そこには今回紹介したデータベースの開発チームの多くが参加した[12]。今後、EU-OPENSREEN を始めとする大型プロジェクト[13]の進展によって、創薬関連のオープンデータは益々増大することが予想されるが、それと同時に PubChem と ChEMBL のデータ共有に代表されるようなデータベース間の連帯も進んでいくと筆者は考えている。

謝辞

本稿の執筆に関して、貴重なアドバイスを頂いた株式会社ファルマデザインの古谷利夫様、井上直子様、そして ChEMBL チームの Anne Hersey と Anna Gaulton に感謝申し上げます。

参考文献

- [1] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D1100-7. Epub 2011 Sep 23.
- [2] IUPAC Standard InChI: <http://www.iupac.org/home/publications/e-resources/inchi.html>
- [3] Abad-Zapatero C, Perisic O, Wass J, Bento AP, Overington J, Al-Lazikani B, Johnson ME. *Drug Discov Today.* 2010 Oct;15(19-20):804-11. Epub 2010 Aug 19.
- [4] Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Zhou Z, Han L, Karapetyan K, Dracheva S, Shoemaker BA, Bolton E, Gindulyte A, Bryant SH. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D400-12. Epub 2011 Dec 2.
- [5] Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. *Nucleic Acids Res.* 2007 Jan;35(Database issue):D198-201. Epub 2006 Dec 1.
- [6] Roth BL, Kroeze WK, Patel S, Lopez E. *The Neuroscientist.* 2000 6:252-262.
- [7] Benson ML, Smith RD, Khazanov NA, Dimcheff B, Beaver J, Dresslar P, Nerothin J, Carlson HA. *Nucleic Acids Res.* 2008 Jan;36(Database issue):D674-8. Epub 2007 Nov 30.
- [8] Wang R, Fang X, Lu Y, Yang CY, Wang S. *J Med Chem.* 2005 Jun 16;48(12):4111-9.
- [9] Halling-Brown MD, Bulusu KC, Patel M, Tym JE, Al-Lazikani B. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D947-56. Epub 2011 Oct 19.
- [10] DrugEBility: <https://www.ebi.ac.uk/chembl/drugability/structure>
- [11] Bellis LJ, Akhtar R, Al-Lazikani B, Atkinson F, Bento AP, Chambers J, Davies M, Gaulton A, Hersey A, Ikeda K, Kruger FA, Light Y, McGlinchey S, Santos R, Stauch B, Overington JP. *Biochem Soc Trans.* 2011 Oct;39(5):1365-70.
- [12] Orchard S et al. *Nat Rev Drug Discov.* 2011 Aug 31;10(9):661-9. Review.
- [13] EU-OPENSREEN: <http://www.eu-openscreen.de/>

///// Cutting Edge /////

医薬品有害事象データベースを用いたデータマイニング

(株)京都コンステラ・テクノロジーズ 五島 誠
 京都大学 大学院薬学研究科 奥野恭史

1. はじめに

薬物治療の重大な問題の一つに副作用が挙げられる。副作用は、患者の QOL を損ねるだけでなく、2 次的医療費の発生や予期せぬ副作用による死亡事件など、国内外でたびたび社会問題となってきた。米国 FDA の調査では、2003 年度の 30 億件の処方のうち、200 万件超の副作用が報告され、これによる 2 次的な医療費支出は 8.4 兆円と試算されている。更には、10 万人以上が副作用により死亡している（死亡要因の第 5 位が副作用）との報告がなされている。

新薬は、定められた臨床試験をパスした上で市場に登場するのであるが、それでも市販後に、予期せぬ副作用などが起こり得る。このことは、臨床試験と市販後の環境との違いで説明されている。臨床試験では、通常、対象患者数は数百人程度で用法用量は画一的である。他の医薬品の併用は制限されており、投与期間は相対的に短い。妊産婦に投与することではなく、専門的な医療施設で注意深いモニタリングのもと実施される。一方、市販後では、数十万人以上の規模で患者に投与される。用法用量も併用薬もさまざま、時には高齢者、妊産婦などに投与することもあり、専門医以外の医師も使用する。このように、市販前の臨床試験においては想定される全ての副作用を見出すことは事実上不可能であり、市販後の日常診療下での処方を通じて更なるエビデンスを収集することが必要となっている。このような背景から、世界各国の規制当局では、製薬会社に市販後における有害事象発生状況等の調査をすること（市販後調査制度）を義務付けるなど、医薬品安全性監視体制が整備されている。

本稿では、医薬品安全性監視体制の一貫として、我が国より先駆けて取り組まれてきた米国 FDA の有害事象自発報告システム Adverse Event Reporting System (AERS) について説明のうえ、本データの医療現場利用を目指して我々が開発した Web ツールシステム CzeekV と、それによる医薬品有害事象情報のデータマイニング例について紹介する。

2. 副作用と監視ツール

ファーマコビジランス

新薬の承認前試験は一般的には厳選、限定されたものであり、販売承認後にはより多くの人々、多様な人々に使われることとなる。薬の投与を受ける患者は臨床試験の対象患者に比べると、より病的である可能性もあり、多くの併用薬剤を投与されている可能性もある。薬剤投与対象患者には性別や人種別・年齢層別の違いも存在するうえ、妊婦の可能性もある。それら全てのケースを市販前に試験することは不可能である。だからこそ、新薬には未知のリスクが多く含まれていることを認識し、市販後の情報を活用することで、リスクヘッジを行うことが重要である[1,2]。

ファーマコビジランス”Pharmacovigilance”（医薬品安全性監視）とは、「医薬品の有害な作用または医薬品に関連するその他の問題の検出・評価・理解・予防に関する科学と活動」と WHO によって定義され、各国の規制当局では、市販後の薬剤情報収集と評価を科学的かつ適切に実施し、副作用被害を最小限にすることを目的とするファーマコビジランス体制が構築されている。

シグナル

ファーマコビジランスでは、シグナルを検出し、更なる評価によってそれを副作用のシグナルとして特定することが最も重要である。それにより、これまで特定されていない、あるいは認識すらされていなかった副作用が明確になり、リスクマネジメントも可能になる。

WHO はファーマコビジランスにおけるシグナルを、「有害事象とある薬剤の間に因果関係があるかもしれないという報告情報。ただし、その因果関係が未知のもの、あるいは過去に不完全にしか記録されていないもの。シグナルを形成するには、事象の重篤性及び情報の質に応じて、通常複数の症例報告が必要である」と定義している。また、Hauben M らの報告には「単一あるいは複数の情報源（観察や実験）から得られた情報であり、それらは、介入と事象もしくは関連した事象の組み合わせ、あるいは有害もしくは有用な事象の間に示唆された、新たな潜在的な因果関連や、すでに知られていた関係での新たな側面を示すものであり、検証をするに足りる十分な可能性がある」とある[3,4]。

ここで重要なことは、シグナルを特定した後の安全性対策であり、理想的には、新たに特定された副作用に対する薬物の作用原因を確定し、有害作用を臨床的に特徴づけ、副作用頻度などのリスクを定量化し、リスクの防止あるいは最小化のための規制措置をとり、これらの情報を医療従事者や患者に迅速に伝達することである。

ファーマコビジランスデータベース

シグナルの検出には、日々の医療現場で発生する有害事象を出来る限り網羅的に収集することが必要となる。これに対し、現在、欧米を中心とする各国では、医療機関や製薬会社からの有害事象の自発報告システムを整備しており、膨大なボリュームとなるデータベースを構築、運営している（表1）。ほとんどのデータベースは、規制当局などの関係者のみの利用に制限されているが、一部のデータベースは入手可能となっている。公的に利用可能な米国 FDA の自発報告システムについては、以下に詳細説明する。

〈表1：世界のファーマコビジランスデータベース〉

国	名称	作成年	情報源の国	個別症例の数
オーストラリア	ブルーカードシステム	1971	同国の自発報告	約 20 万件
カナダ	カナダビジランス	1965	同国の自発報告	約 20 万件
	CAEFISS (予防接種有害事象)	1987	同国の自発報告	N/A
ヨーロッパ連合	EudraVigilance	2001	EU と、 EU 外の重篤な未知の副作用	N/A
フランス	ANPV	1985	同国の自発報告	約 30 万件
日本	副作用情報マネジメントシステム	2003	同国の自発報告(医療機関・製薬企業) 海外(製薬)	約 40 万件
オランダ	Lareb	1985	同国の自発報告	N/A
スウェーデン	SWEDIS	1965	同国の自発報告	約 10 万件 (2007)
英国	イエローカード (Sentinel)	1963	同国の自発報告 Eu と EU 外外国症例	N/A
米国	AERS	1969	米国及び世界中	約 450 万件
	VAERS (Vaccine)	1990	海外は未知重篤のみ製薬企業から FDA へ提出	約 20 万件
WHO	Vigibase	1968	WHO 薬剤モニタリングプログラムへの参加国からの自発報告	約 400 万件

FDA 有害事象自発報告システム

FDA は、1969 年より、有害事象の自発報告システム Adverse Event Reporting System (AERS) の稼働を開始した。このシステムの特徴は、製薬メーカー、医療専門家のみならず、患者や患者の家族も、自由に報告できる点にある。収集される情報は、現場主導のエンドユーザーからの生の情報であり、非常に貴重な情報源であると認識されている。1997 年に現在の報告システムとして再整備され、医療専門家・患者・製薬企業など様々な関係者の自発的な報告による膨大なレポートデータ（2012 年 6 月時点で 450 万件超）が集約されている。情報源は米国内が中心ではあるが、外国で発現した有害事象報告についても製薬会社から FDA へ提出が求められており、日本を含む世界中のデータが集められている。データは一定期間公開されており様々な利用が可能である。患者名・医療機関名・所在地は公開されておらず、個人が特定されることはない。しかし、同じ症例に対して異なる人物・機関から報告される重複データがあったり、ユーザーの専門性に依存した報告データのバラツキなどが見られたりなど、自発報告システムならではの問題点が指摘されている。

AERS の具体的なデータ構成は（表 2）に示す 7 つのサブセットから成る。このようなデータの集積により、例えば、抗インフルエンザウイルス剤タミフルにより異常行動が起こり易いなどといった情報はもちろんのこと、タミフルにより異常行動が起こり易い年齢層、性別などといったより詳しい情報の収集も可能となる。

〈表 2 : AERS のデータ構成〉

AERS ID	情報の内容
DEMO	症例情報（患者の性別、年齢など）
DRUG	使用医薬品情報（どの医薬品を使用したか）
REAC	有害事象情報（どのような有害事象が起こったか）
INDI	適応疾患情報（疾患名など）
RPSR	報告者情報（報告者に関する情報など）
OUTC	転帰情報（患者がどうなったか）
THER	治療期間情報（治療期間など）

シグナル検出のためのデータマイニング手法

シグナル検出の情報源となるファーマコビジランスデータベースは、数百万件にもおよぶ膨大なデータの塊であり、これら膨大なデータソースから意味のある「医薬品と有害事象との関係性」を抽出するには、適当な統計学的手法やデータマイニング手法の適用が必須である。このために開発され、現在、世界の代表的な規制当局で用いられている 4 つの統計学的指標（シグナル検出指標）を紹介する。これらのシグナル検出指標は、テキストマイニングにおける共起性の算出と原理的に同じものであり、AERS データの最小単位である報告文書に医薬品タームと有害事象タームが共起する度合いを統計学的にスコア化したものと言える。

具体的には、登録されている全報告を、まず、特定の有害事象が起こっているか否かに基づいて 2 群に大別する。続いて、2 群個々について、特定の医薬品が投与されているか否かに基づいて 2 群に大別する。このように作成した 2×2 分割表に基づいて以下のシグナル検出指標をそれぞれ算出し、解析の対象とした医薬品と有害事象との間の因果関係が推定される。

Proportional Reporting Ratio (PRR) :

医薬品ごとに特定の有害事象の報告割合を算出し、それらの比率で表したものである。英国医薬庁 MHRA・欧州医薬品庁 EMEA で採用されている。

Reporting Odds Ratio (ROR) :

注目すべき医薬品と副作用の共起回数におけるオッズ比である。PRR と同様のプロセスを経て算出されるが、算出方法が異なる。オランダ薬剤監視センター Lareb で採用されている。

Information Component (IC) :

注目する医薬品に関する報告がなされる確率、および注目する副作用に関する報告がなされる確率から、注目する医薬品と副作用の組が報告される期待確率を推定し、この推定された期待確率と実際の報告数から得られる実測確率の比率を対数スコア化することで、シグナルを検出する。この対数スコアを **Information Component** と称する。ここで期待確率と実測確率は、ベイズ流アプローチの一つである **Bayesian Confidence Propagation Neural Network (BCPNN)** により推定される。世界保健機関 WHO で採用されている。

Empirical Bayes Geometric Mean (EBGM) :

実際の報告数と期待値の比率で表現した相対リスクをシグナル検出指標としたものである。ここで、実際の報告数をポアソン分布の平均とする相対リスクを定義し、この相対リスク値の事後分布を 2 つのガンマ分布の混合分布から推定する。この推定された相対リスク値の事後分布から算出される幾何平均値が **Empirical Bayes Geometric Mean** である。FDA で採用されている。

以上の 4 つの指標は、理論的背景は異なるものの、いずれも、特定の有害事象が特定の医薬品に由来する可能性の指標となる。詳細は紙面の都合上割愛するが、それぞれの指標を用いる規制当局により統計学的有意性に基づく閾値が設定されており、その閾値を満たせば、“シグナルが検出された”と判断している。またこれらのシグナル検出に対する統計的アプローチは膨大なデータベースから効率的にシグナルを検出する一手段に過ぎず、データマイニング手法固有の感度と特異度の問題は、他分野で応用されているデータマイニング手法と同様に発生する。我々もシグナルマネジメントを含めて多面的なアプローチの開発を進めており、本稿ではその全体概要と我々が構築したシステム (**CzeekV**) を用いた手法を示す。

更に注意すべき点は、このような大規模データベースを用いた計算によるシグナル検出は、シグナル特定の出発点に過ぎないことである。従って、シグナル計算によって検出されたその有害事象に関しては、実際の報告レポートに記載されている詳細情報（用法用量、併用薬、患者背景や有害事象の経緯など）の精査や、更なる疫学調査を行うなどの詳細な評価を行った上で、副作用のシグナルとして特定されなければならない。

CzeekV

上述の通り、米国 AERS のデータは現場主導の生の臨床情報として非常に有用な情報であるが、登録データの重複や誤入力などの自発報告システム特有の問題点を有していた。また、FDA から我々が直接入手できる情報は、総数 450 万件もの関連情報が列挙された CSV ファイルのみであり、検索ツールなどのサポートもされておらず、現場の医療従事者のそのままの利用は事実上不可能な状況であった。そこで我々はこれらの背景を鑑み、重複データの削除・入力ミスの修正・薬剤整理・医療従事者による翻訳など独自のキュレーション作業を経て、システムとして活用可能なデータベースを構築し、2011 年 8 月に、キーワード検索を実現する WEB アプリケーション『**CzeekV**』をリリースした。**CzeekV** (2012 年 9 月現在) では、医薬品 7,684 件、有害事象 22,540 件が登録されており、医薬品名や有害事象名からの日本語高速検索機能により、容易に AERS の報告レポートの検索が可能である。更に、登録されている全ての医薬品と有害事象の膨大な組合せ (11,607,205 ペア) について、世界主要規制当局で使用されているシグナル値 (医薬品と有害事象の統計的な関連性の強さを示す統計指標 (上述)) の計算を予め行った計算結果を付与することにより、統計的な客観性に基づく医薬品と有害事象の関係性評価も実現している。現在、**CzeekV** は、<https://www.czeek.com> よりアクセス可能であり、無償の検索ページと製薬企業・薬学/疫学研究者・医療従事者向けの有償の詳細情報ページのサービスを行っている。

以下に詳細機能を紹介する。

〈レポート内容〉

一つの報告レポートには、副作用の発生報告日・国以外に、「患者情報：年齢・性別・体重（個人は特定されない）」、「使用薬物情報：薬剤名（併用薬も記載あり）・薬剤タイプ・適応症・投与経路・用法用量・治療期間・副作用への役割（第一被疑薬など）等」、「副作用情報：投与後の副作用内容・薬剤停止後の変化・薬剤再使用後の変化・転帰情報等」が記載されている。

副作用情報		
投与後の副作用	血圧の低下、循環虚脱、転倒、フラッシング、意識喪失、頻脈、嘔吐症状	
薬剤停止後副作用の変化	実施しない	
薬剤再使用後副作用の変化	実施しない	
患者の転帰	入院が必要または入院期間の延長	
死亡日時		
レポート関連情報		
報告タイプ:速報	レポート状態:初報告	製薬会社における月製番号:
製薬会社への報告日:1997/10/29	FDAへの報告日:1997/11/10	報告を送付した日:
製薬会社の報告ID:JP97106238A	報告製薬会社の名称:ELI LILLY AND COMPANY	報告書コピー要請番号:1000002242-5
電子的報告:	製薬会社が知っている:	自発報告の開示の可否:

〈図1：レポート内容（一部抜粋）〉

これらの項目毎で検索できるようになっており、利用者が探索したい項目で絞込み検索ができるようになっている。

〈検索機能〉

薬剤一般名から関連する副作用を容易に調べる機能があり、全レポート内の報告件数や詳細な情報とともにデータとして表示する。副作用との関連性だけでなく、適応症との関連性についても検索可能である。検索ボックスは3つ用意されており、薬剤・副作用・適応症それぞれから検索できる。条件設定として国別・性別・年齢別による絞込みも可能である。

特筆すべき点は、現場主導の報告レポートを検索対象としているため、添付文書には記載されていない副作用項目も見出すことができる点である。検索結果で得られる薬剤－副作用の関連性の強さについては、詳細情報（下記）にて示されるシグナルスコア値や該当するレポート内容を吟味して判断することになる。

なお、当該ページの検索は無償で使用可能である。



〈図2：検索結果画面〉

〈詳細情報〉

詳細情報では、転帰情報・商品リスト・主成分が同じ他の製剤を情報として記載している。また、副作用項目・件数を表示して、4種類のシグナル計算手法により算出した各スコア値を副作用毎に表記している。各規制当局で用いられているシグナル閾値以上のスコア値を赤字で表示している。各スコア値でソーティングも可能であるため、シグナルマネジメントに有効活用できる。

全件数上位

副作用

適応症

絞り込み:

副作用名	件数	シグナル 総合ランキング	FDAスコア (米国食品医薬品局)	WHOスコア (世界保健機関)	EMAスコア (欧州医薬品庁)	Larebスコア (蘭国薬剤監視センター)		
リンパ球刺激試験陽性	61	1位	15.79	3.56	19.95	15.71		
間質性肺疾患	441	2位	10.63	3.36	11.66	10.69		
肝機能異常	501	3位	10.26	3.31	11.2	10.33		
皮膚粘膜眼症候群	22	4位	23.8	3.21	36.91	24.82		
薬疹	138	5位	9.21	3.07	10.71	9.12		
上気道の炎症	32	6位	13.08	3.06	18.22	13.02		
ニューモシスティス肺炎	110	7位	8.08	2.87	9.56	7.98		
播種性血管内凝固症候群	252	8位	7.5	2.84	8.41	7.47		

ページ

1 / 104

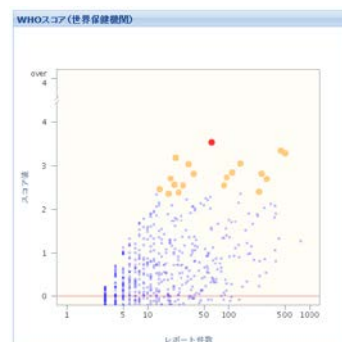
1 - 30 of 310

〈図3：詳細情報画面のスコア値〉

〈シグナル分布表示〉

スコア値とレポート件数に合わせてプロットした分布図である。シグナル計算によって得られるスコア値は統計的な有意性を示す指標であり、必ずしもレポート件数と比例するものではない。この分布図を用いることで、関連する副作用項目の全体の中での位置づけを、スコア値とレポート件数の両方の項目により相対的に評価をすることができる。

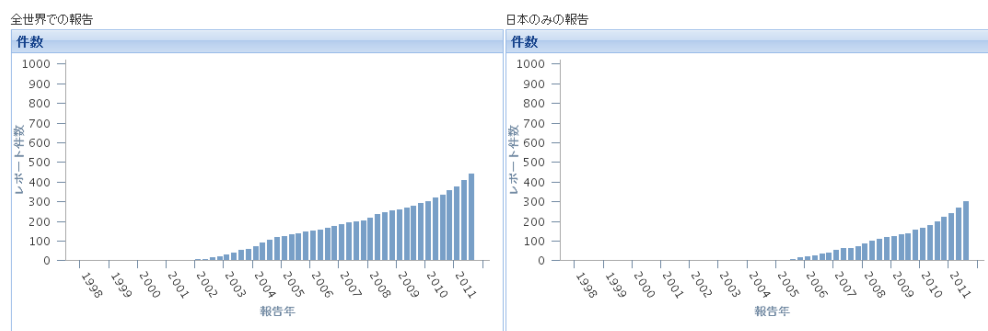
シグナル分布図には分布表示を行う際に設定した薬剤に関する全副作用が青色で（スコア値の上位 20 件は黄色で）表示されている。それに対してユーザーによって検索された副作用項目が赤色で表示されているため、副作用項目間での比較が容易である。



〈図 4：シグナル分布図〉

〈年代別ヒストグラム〉

ファーマコビジランスでは、新薬が市場に出た時点から有害事象発生率の経過をモニタリングすることが重要である。CzeekV が情報源とする AERS データは四半期ごとに更新されるため、年代別ヒストグラムでは、副作用の報告件数ならびにシグナルスコア値の四半期ごとの時系列変化をプロットしている。シグナルスコア閾値を越える時期や副作用報告件数の増加率などを確認することができる。また、全世界の報告と日本の報告を両方表示しているので比較検討しやすくなっている。



〈図5：レポート件数のヒストグラム〉

〈統計解析による同種同効薬比較〉

統計解析機能は、任意に指定した複数の薬剤について、副作用項目やそのシグナルスコア値を比較することができる。例えば同種同効薬のみを検索画面で抽出し登録することで、各薬剤と各副作用の関連性を一つのテーブルとして比較することが可能となる。

閾値以上の値をもつ項目は赤色で表示されているので、どの薬剤/副作用項目が関連性の高いものであるかを視覚的に判断できる。処方時の薬剤選択補助ツールやファーマコビジランスの研究ツールとして様々な活用が可能である。



〈図 6：統計解析/同種同効薬比較〉

3. CzeekV を用いた研究事例

我々が論文報告を行った CzeekV を用いたデータ解析に関する研究を 2 例紹介する。

白金系抗がん剤の代表的な有害事象との関連性に関する比較研究

2004～2009 年までの AERS データ 1,644,220 件を対象に、複数の白金系抗がん剤シスプラチン、カルボプラチン、オキサリプラチンについて代表的な有害事象との関連性の比較解析を行った。有害事象のシグナル検出には、WHO や FDA などの代表的な規制当局が用いる 4 つの統計指標 (PRR, ROR, IC, EBGm) を用いた。結果として 2004～2009 年までの有害事象報告 1,644,220 件によると、シスプラチン、カルボプラチン、オキサリプラチンいずれも、悪心・嘔吐・急性腎不全・好中球減少症・血小板減少症・末梢神経障害との関連性が確認され、特に、悪心とシスプラチン、急性腎不全とシスプラチン、血小板減少症とカルボプラチンと、末梢神経障害とオキサリプラチンとの間に、他の白金系抗がん剤よりも強い相関が確認された[5]。

白金系抗がん剤と過敏性反応の関連性に関する研究

2004～2009 年までの AERS データ 1,644,220 件を対象に、4 つのシグナル計算 (PRR, ROR, IC, EBGm) を行い、シスプラチン、カルボプラチン、オキサリプラチンと過敏症との関連性における統計的評価を行った。更に、オキサリプラチン誘発性の過敏性反応に対するデキサメタゾンおよびジフェンヒドラミンの相互作用効果を Excess2 という統計指標により評価した。結果として 2004～2009 年までの AERS 有害事象報告によると、シスプラチンよりも、カルボプラチンやオキサリプラチンの投与で過敏性反応の発症例が多いことが明らかになった。また、オキサリプラチン誘発性の過敏性反応について、デキサメタゾンの併用は、穏やかな (mild) 過敏性反応の発症を抑えたが、重大/致死的な過敏性反応には影響を及ぼさず、ジフェンヒドラミン併用による影響は認められなかった[6]。

4. おわりに

CzeekV はファーマコビジランスを担う新ツールとしてリリースされてまだ 1 年である。まさに新しい分野を切り開いている最中であるが、CzeekV を通じて現場の医師・薬剤師との討論を行う中で、医療現場の生の情報を集積したデータベースの有用性やそのデータマイニングの必要性を痛切に感じるばかりである。我々も本研究開発を通じて、ファーマコビジランスという分野を初めて学ぶようになったが、欧米諸国に比べ、日本におけるファーマコビジランスの意識は現状においてお世辞にも高いとは言えない。我々の研究開発をきっかけに、今後更なる医療現場の高度化と医薬品の適切使用による患者の安全性向上が促進され、世界有数のファーマコビジランス体制が確立されることを期待する。

参考文献

- [1]Edwards IR, Aronson JK : Lancet 356, 1255-1259, 2000.
- [2]栄田敏之, 角山香織, 奥野恭史 : 人工知能学会誌 26, 126-130, 2011.
- [3]Hauben M, Bate A : Drug Discov. Today 14, 343-357, 2009.
- [4]藤田利治 : 薬剤疫学 14, 27-36, 2009.
- [5]Sakaeda T, Kadoyama K, Okuno Y : Int. J. Med. Sci. 8, 487-491, 2011.
- [6]Sakaeda T, Kadoyama K, Yabuuchi H, Nijima S, Seki K, Shiraishi Y, Okuno Y : Int. J. Med. Sci. 8, 332-338, 2011.

///// Activities /////

構造活性フォーラム 2012 開催報告
「GPCR 研究の最前線」構造活性フォーラム 2012
実行委員長 石黒正路

味覚や臭覚において重要な役割を持つ G タンパク質結合型受容体 (GPCR) は、最大の受容体ファミリーを形成し、医薬品の標的として広く注目されています。さらに、機能解析に向けた結晶構造解析により、受容体の構造変化とリガンドの認識様式等が解明されつつあるなど、より合理的なリガンド (医薬品など) のデザインにも応用される研究が進んでいます。本フォーラムでは、GPCR 研究分野における第一線の研究者を招いて講演していただき、GPCR 研究の最前線を体系的に理解し、分子設計・医薬品化学における応用について討論しました。

当日は、梅雨の季節だけあって小雨が降っていたが、会場には 75 名にも及ぶ多数の参加者が集まりました。今回は、製薬業界にとって関心が高い分野である GPCR の最前線をフォーラムのテーマにしたためか、大学関係者に比べ、製薬メーカーを中心とした企業の研究者の方が多くいらっしゃいました。参加者としては、薬理、モデリング、合成を専門とする分野の研究者が中心だったため、各講演に対する質疑応答では、活発な議論が繰り広げられ、本フォーラムの関心度の高さが窺えました。

最後になりましたが、貴重なお時間を割き、最後の総合討論にまでお付き合い下さいましたシンポジストの先生方、座長の先生方に御礼申し上げます。

なお、次回の構造活性フォーラム 2013 は、理化学研究所・本間光貴先生が実行委員長を務められ、例年どおり 6 月頃に開催される予定です。

《午前の部》

多様な GPCR の機能と構造

新潟薬科大学 石黒正路

GPCR をターゲットにした「構造に指南された創薬」を目指すための三つの戦略

京都大学 小林拓也

《午後の部》

GPCR モデリングとインシリコスクリーニング

産業技術総合研究所 広川貴次

味覚受容体における味物質認識機構

東京大学 三坂 巧

オーファン GPCR 研究に基づく創薬ターゲット探索

武田薬品工業 森 正明

//// Activities ////

Hansch-藤田法 50 周年記念シンポジウム開催報告

シンポジウム実行委員会代表
中川好秋

Hansch-藤田法 50 周年記念シンポジウムを 8 月 25 日 (土) に京都大学医学部の芝蘭会館にて、第 29 回藤田カンファレンスと時期を合わせる形で開催した。まず Corwin Hansch 先生に黙祷を捧げたのち、藤田カンファレンス幹事長岡島伸之氏の挨拶を頂戴した。続いて、藤田稔夫先生に感謝状と記念品 (目録) をお渡しするという簡単なセレモニーを行った。講演会の午前中は、Hansch-藤田法の誕生と 50 年の歴史およびその物理化学的背景についてのセッションを行った。午後の最初の 2 題は、農薬と医薬分野からそれぞれの応用研究の紹介、休憩をはさんで Hansch-藤田法の展望についてのお話を頂戴した。講演者の一人として、(故) Hansch 教授の研究室を継承されているアメリカポモナ大学の Selassie 教授にも来日頂いた。閉会にあたり、元日本農薬学会会長の梅津憲治様に講評をお願いした。プログラムと要旨は基本的には英語で作成し、演題も英語であったが、ここでは、Selassie 教授の講演タイトルを除いて日本語で紹介させて頂いた。

「構造活性相関における複雑さをときほどく」

藤田稔夫 (京都大学名誉教授)

「超熱力学的 QSAR アプローチにおけるパラメータに関する物理化学的考察」

加納健司 (京都大学大学院農学研究科)

「農薬科学における分子設計と作用機構研究への QSAR の応用」

中川好秋 (京都大学大学院農学研究科)

「ADME に着目した医薬研究への QSAR の応用」

赤松美紀 (京都大学大学院農学研究科)

「QSAR のさらなる発展に向けて：ハンシュー 藤田からの贈り物」

中馬 寛 (徳島大学ヘルスバイオサイエンス研究部)

「Mechanistic QSAR: Alive and Well!!」

Cynthia Selassie (Pomona College)

本シンポジウムは、2011 年 5 月に Hansch 先生が他界され、その追悼のシンポジウムということで計画したが、Hansch-藤田法の論文が Nature 誌に掲載されてから、今年がちょうど 50 年目にあたるということで、50 周年記念シンポジウムとして開催させて頂いた。藤田先生は QSAR の創始者であることから、構造活性相関部会として開催してはどうかという話もあったが、藤田カンファレンス (1981 年に藤田門下生の勉強会として発足) の主催とし、部会には協賛を頂く形で開催させて頂いた。Hansch-藤田法の誕生に関しては、SAR News No. 21 において、藤田先生が Obituary Tribute として「Corwin Hansch 先生を偲ぶ - QSAR の誕生まで -」という論文を紹介されておられますので、そちらをご覧ください。

本シンポジウムには 141 名 (一般 112 名、学生 29 名)、懇親会にも 100 名を超える参加があり大盛況であった。本シンポジウムにご参加頂きました皆様には、この場を借りて御礼申し上げます。講演要旨にモノクロで掲載した Hansch 先生と藤田先生のツーショットのカラー写真を載せさせて頂いた。



//// Activities ////

〈会告〉

第40回構造活性相関シンポジウム

会 期：2012年11月29日（木）～11月30日（金）

会 場：岡崎市図書館交流プラザ・りぶら（愛知県岡崎市康生通西4-71）

主 催：日本薬学会構造活性相関部会

共 催：日本化学会、日本農芸化学会、日本分析化学会、日本農薬学会

第1日目（11月29日）

10:30 - 10:35 開会

10:35 - 11:15 一般講演（会場：りぶらホール）

座長：赤松美紀

KO01 LERE-QSAR 解析による bilinear model の新しい解釈：トリアジン誘導体とジヒドロ葉酸還元酵素との複合体形成

○山内香子, 相原薫, 坂本修平, 杉本拓弥, 吉田達貞, 中馬寛（徳島大院薬）

KO02 LERE-QSAR 解析の酵素触媒反応への適用：トリプシンによる置換馬尿酸フェニルの加水分解反応のミカエリス・メンテン定数および反応速度定数の定量的解析

○馬島彬, 坂本修平, 杉本拓弥, 吉田達貞, 中馬寛（徳島大院薬）

11:25 - 12:10 招待講演（会場：りぶらホール）

座長：清水 良

KI01 ウイルス由来感染機構を有する非ウイルス性DDSキャリアー：

バイオナノカプセルの開発

黒田俊一（名古屋大学大学院生命農学研究科）

13:30 - 14:30 一般講演（会場：りぶらホール）

座長：大田雅照

KO03 ドッキングおよび構造活性相関に基づく脱皮ホルモン受容体結合様式の予測

○蒲池沙織, 瀧本征佑, 赤松美紀, 宮川恒, 多田俊治, 中川好秋（大阪府大院理, 京大院農）

KO04 ネオニコチノイド系殺虫剤の定量的構造活性相関解析

○西脇寿, 栗山光博, 長岡ひかる, 加藤聡, 赤松美紀, 山内聡, 首藤義博（愛媛大農, 京大院農）

KO05 非経験的フラグメント分子軌道法を活用した高活性 Protein Kinase CK2 阻害剤の設計

○呉竜英, 侯増輝, 木下誉富, 武井義則, 安江美里, 三須良介, 鈴木大和, 中村真也, 大野浩章, 村田克美, 北浦和夫, 平澤明, 大石真也, 藤井信孝, 仲西功（近畿大薬, 京大薬, 大阪府大院理）

14:40 - 16:40 ポスターセッション（会場：3階会議室 301-303）

14:40 - 15:40 奇数番発表

15:40 - 16:40 偶数番発表

16:50 - 17:50 特別講演 (会場：りぶらホール)

座長：高橋由雅

KS01 創薬における安全性評価：毒作用発現とその分子毒性学的考究
堀井郁夫 (ファイザー・昭和大学薬学部)

18:30 - 懇親会 (会場：岡崎ニューグランドホテル)

ポスターセッション [11月29日 14:40 - 16:40] (会場：3階会議室 301-303)

- KP01** 抗炎症フラボノイドの NO 産生抑制作用に関する構造活性相関
大河原充好, 根本崇志, 則武隆佑, 大根谷章浩, 高宮知子, 北中進, ○飯島洋 (日本大薬)
- KP02** IT 創薬技術を用いた Endomorphin-1 の非ペプチド化研究
○酒匂佑介, 神田泰彦, 服部一成, 福西快文, 中村春木 (塩野義製薬, 産総研, 阪大蛋白研)
- KP03** ATP・基質結合部位阻害を目的とした新規 Pim1 阻害剤探索
○幸瞳, 津金沢恵子, Lorien Parker, 田仲昭子, 横山茂之, 本間光貴 (理研 SSBC)
- KP04** 非経験的分子軌道法計算を用いたリレンザおよびその誘導体とインフルエンザ・ノイラミニダーゼとの結合自由エネルギー変化の QSAR 解析 - Fragment Based LERE-QSAR
○的場弘, 比多岡清司, 河野明大, 芝田雄登, 吉田達貞, 中馬寛 (徳島大院薬)
- KP05** β -セクレターゼ阻害剤の Lead Evolution と Optimization: 分子科学計算による薬物-タンパク質間相互作用からの定量的考察
○相原薫, 山内香子, 吉田達貞, 中馬寛 (徳島大院薬)
- KP06** リガンド-タンパク質の複合体形成に及ぼす溶媒効果の検討: LERE-QSAR 解析における水和自由エネルギー項の定量的評価
○比多岡清司, 的場弘, 河野明大, 芝田雄登, 原田政隆, 坂本修平, 吉田達貞, 中馬寛 (徳島大院薬)
- KP07** 新規な in silico FBDD 手法
Hugues-Olivier Bertrand, Hongwei Huang, Jurgen Koska, Jiabo Li, Tien Luu, Jon Sutter, Adrian Stevens, ○高岡雄司, Deqiang Zhang (Accelrys Inc.)
- KP08** FMO+PBSA による DJ-1 リガンド結合エネルギーのインシリコ解析
○重光保博 (長崎工技セ)
- KP09** CYP2B6 の基質認識に対する一塩基多型の影響のドッキングシミュレーションによる評価
○小林佳奈, 小田彰史, 平塚真弘, 山乙教之, 広野修一, 高橋央宜 (東北薬大, 金沢大院薬, 阪大蛋白研, 東北大院薬, 北里大薬)
- KP10** 有機フッ化合物のバクテリアに対する毒性と構造との関係
○三島淳, 森田馨, 石原良美, 高野二郎 (東海大院理)
- KP11** 市販(Q)SAR ソフトによる予測結果の信頼性評価に関する検討
○赤堀有美 (CERI)
- KP12** 異性化アスパラギン酸を含むペプチド基質とヒト PIMT との相互作用様式の推定
○小田彰史, 野地郁彦, 小林佳奈, 高橋央宜, 福吉修一, 中垣良一 (金沢大院薬, 東北薬大, 阪大蛋白研)

- KP13** カテゴリーアプローチによる化学物質の生物濃縮性の評価
○池永裕, 櫻谷祐企, 山田隼 (NITE)
- KP14** オープンアクセス可能な広範な化合物に対応する高等動物への毒性予測システムの検討
○寺本泉瑠, 宇根琢, 岡本晃典, 川下理日人, 高木達也 (阪大院薬, 阪大薬, 阪大微生物病研)
- KP15** CYP2D6.1 (野生体型) / CYP2D6.17 (変異体型) 阻害に関するイン・シリコ予測研究
○半田耕一, 中込泉, 山乙教之, 合田浩明, 広野修一 (北里大薬, 富山化学工業)
- KP16** Sucrase 及び Isomaltase への LAB 化合物の結合様式に関する理論的研究
○小関準, 中込泉, 足立伊佐雄, 加藤敦, 広野修一 (北里大薬, 富山大病院薬)
- KP17** イン・シリコ創薬技術に基づいたヒトセリンラセマーゼ阻害剤の創製研究
○合田浩明, 和田亮吾, 李杰, 帯田孝之, 水口峰之, 森寿, 豊岡尚樹, 広野修一 (北里大薬, 富山大院理工, 富山大院薬, 富山大院医)
- KP18** α -glucosidase 阻害剤 α -1-C-Butyl-LAB の Maltase への結合様式に関する理論的研究
○中込泉, 小関準, 足立伊左雄, 山乙教之, 合田浩明, 加藤敦, 広野修一 (北里大薬, 富山大病院薬)
- KP19** 酵素リゾチームに対するリガンド分子の結合サイト: 分子力学計算と等温滴定カロリーメトリー
○松浦誠, 田浦俊明 (愛知県立大院情報)
- KP20** NTG スーパーグラフを利用したリード骨格の探索
○大西貴之, 高橋由雅 (豊橋技科大院工)
- KP21** 分子進化計算を用いた創薬候補化学構造設計システムの開発
○守國拓史, 高橋由雅 (豊橋技科大院工)
- KP22** GPCRSARfari データに基づくリガンド/GPCR 相互作用の予測
○菅谷昇義 (ファルマデザイン)
- KP23** バイオインフォマティクス観点からのターゲットタンパク質に対するリガンド結合部位の探索方法の開発
○吉山晃太郎, 松岡美里, 加納和彦, 小松克一郎, 梅山秀明, 岩舘満雄 (中央大理工, 国立感染症研, 北里大理)
- KP24** ホモロジーモデリング法の複数のコンセンサス法による構造選択の評価と二次構造相対配置最適化
○佐藤亘, 宮下貴弘, 梅山秀明, 岩舘満雄 (中央大理工)
- KP25** 精錬過程の自動化による大規模データベースからの基本活性構造抽出
○高田直人, 大森紀人, 岡田孝 (関学大院理工)
- KP26** 基本活性構造を用いた副作用のシグナル検出
○大森紀人, 吉岡祐一, 加藤祐樹, 岡田孝 (関学大院理工)
- KP27** 副作用に関わる基本活性構造の頻出部分グラフマイニング法による抽出
○中尾寛郎, 猪口明博, 大森紀人, 阿部武丸, 岡田孝, 鷲尾隆 (関学大院理工, 阪大産研)
- KP28** 遺伝的アルゴリズムを用いた分子ライブラリデザインシステムの開発
○森本孝朗, 加藤博明 (豊橋技科大院工)
- KP29** 近傍フラグメント表現に基づくタンパク質配列特徴解析システムの開発
○永松晃一, 加藤博明 (豊橋技科大院工)

- KP30** 予備的化合物選択のためのリガンドベースバーチャルスクリーニング手法の開発
(その2) : フィンガープリントと記述子を組み合わせた化合物選択手法の開発と評価
○郡司久恵, 関達徳, 山崎広之, 西端芳彦 (北里大薬)
- KP31** ヒト主要 P450 分子種によるモデル化合物 Tebufenozide の代謝物同定および In Silico
予測
○矢川勝太, 城谷直紀, 十川萌, 生城真一, 榊利之, 原田俊幸, 宮川恒, 平井伸博, 赤松
美紀 (京大院農, 富山県立大工)
- KP32** Biological Evaluation on New Inhibitors of Indoleamine 2, 3-Dioxygenase
○李昕, 田中実, 滝川修, 横山祐作 (東邦大薬, 長寿研)
-

第2日目 (11月30日)

9:30 - 10:50 一般講演 (会場: りぶらホール)

座長 : 久保寺英夫

- KO06** A Combined QM/MM (ONIOM) and QSAR Approach to the Study of Complex Formation of
Matrix Metalloproteinase-9 with a Series of Biphenylsulfonamides -LERE QSAR Analysis-
○吉田達貞, 比多岡清司, 馬島彬, 杉本拓弥, 的場弘, 坂本修平, 野々下航, 林敬久,
中馬寛 (徳島大院薬)
- KO07** ジアゼピン化合物の軸不斉エネルギー障壁の高精度・高速モンテカルロシミュレーション
による推算
○森上賢治, 射手園佳子, 西本昌弘, 大田雅照 (中外製薬)

座長 : 山下富義

- KO08** 反応前駆体を考慮に入れた毒性予測 QSAR 式の開発
○古濱彩子, 青木康展, 白石寛明 (国立環境研)
- KO09** ToxBay: ベイジアンネットによる反復投与毒性評価システム
○岡田孝, 大森紀人, 堀川裕志, 山川真透, 森幸雄, 櫻谷祐企, 山田隆志, 阿部武丸,
山添康, 林真 (関学大理工, NITE, 東北大薬, 安評センター)

11:10 - 12:10 一般講演 (会場: りぶらホール)

座長 : 本間光貴

- KO10** 分子進化アルゴリズムを利用した薬物の構造設計
○河合健太郎, 永田尚也, 高橋由雅 (科研製薬, 豊橋技科大院工)
- KO11** 情報学的スクリーニングによる自己免疫疾患遺伝子とその FAMS を用いた複合体モデ
リングおよびインシリコスクリーニング
○岩舘満雄, 田口善弘, 梅山秀明 (中央大理工)
- KO12** 蛋白質-リガンド複合体データベースからのフラグメント同定と化合物 3 次元データベ
ース検索に基づくリード探索法の開発
○山乙教之, 広野修一 (北里大薬)

13:30 - 14:50 一般講演（会場：りぶらホール）

座長：飯島 洋

KO13 ヒト酸性キチナーゼを標的とした新規喘息治療薬開発を目指したイン・シリコ創薬研究
○若杉昌輝, 合田浩明, 廣瀬友靖, 菅原章公, 山本剛, 塩見和朗, 砂塚敏明, 大村智,
広野修一（北里大薬, 北里生命科学研）

KO14 タンパク質立体構造データベースに対する高速類似部分構造検索法の開発
○寺師玄記, 渋谷哲朗, 竹田-志鷹真由子（北里大薬, 東大医科研）

座長：粕谷 敦

KO15 Short chain dehydrogenase/reductase ファミリタンパク質の構造進化
○前田美紀（農業生物資源研）

KO16 機械学習を用いた induced fit 部位の予測と kinase DFG loop などへの応用
○高谷大輔, 佐藤朋広, 幸瞳, 田仲昭子, 本間光貴, 横山茂之（理研(横浜)）

15:00 - 15:45 招待講演（会場：りぶらホール）

座長：藤原英明

KI02 複合糖質の構造生物学：創薬標的としての糖鎖
加藤晃一（自然科学研究機構岡崎統合バイオサイエンスセンター・
名古屋市立大学大学院薬学研究科）

15:45 - 15:50 閉会

参加登録予約申込：11月9日（金）締切

詳細は、ホームページ上の参加登録予約申込要領をご覧ください。

参加登録費：[一般]予約 8,000 円、当日 9,000 円 [学生]予約 2,000 円、当日 3,000 円

※要旨集前送希望の場合は、郵送料 1,000 円を別途申し受けます。

※費用振込み後、参加取り消しによる返金には応じられません。

懇親会：11月29日（木）18:30 から（会場：岡崎ニューグランドホテル）

[一般]予約 7,000 円、当日 8,000 円 [学生]予約 3,000 円、当日 4,000 円

お問合せ・申込先：

〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

豊橋技術科学大学 情報・知能工学系 分子生命情報学研究室

第40回構造活性相関シンポジウム実行委員会 加藤博明

TEL: 0532-44-6879 FAX: 0532-44-6873

E-mail: sar40@mbi.cs.tut.ac.jp

http://sar.mbi.cs.tut.ac.jp/sar40/

//// Activities ////

構造活性相関部会の沿革と趣旨

1970年代の前半、医農薬を含む生理活性物質の活性発現の分子機構、立体構造・電子構造の計算や活性データ処理に対するコンピュータの活用など、関連分野のめざましい発展にともなって、構造活性相関と分子設計に対する新しい方法論が世界的に台頭してきた。このような情勢に呼応するとともに、研究者の交流と情報交換、研究発表と方法論の普及の場を提供することを目的に設立されたのが本部会の前身の構造活性相関懇話会である。1975年5月京都において第1回の「懇話会」(シンポジウム)が旗揚げされ、1980年からは年1回の「構造活性相関シンポジウム」が関係諸学会の共催の下で定期的開催されるようになった。

1993年より同シンポジウムは日本薬学会医薬化学部会の主催の下、関係学会の共催を得て行なわれることとなった。構造活性相関懇話会は1995年にその名称を同研究会に改め、シンポジウム開催の実務担当グループとしての役割を果たすこととなった。2002年4月からは、日本薬学会の傘下組織の構造活性相関部会として再出発し、関連諸学会と密接な連携を保ちつつ、生理活性物質の構造活性相関に関する学術・研究の振興と推進に向けて活動している。現在それぞれ年1回のシンポジウムとフォーラムを開催するとともに、部会誌のSAR Newsを年2回発行し、関係領域の最新の情勢に関する啓蒙と広報活動を行っている。

本部会の沿革と趣旨および最新の動向などの詳細に関してはホームページを参照頂きたい。

(http://bukai.pharm.or.jp/bukai_kozo/index.html)

編集後記

日本薬学会構造活性相関部会誌SAR News第23号をお届けいたします。今回は、創薬および関連分野のデータベースをテーマとしました。Perspective/Retropectiveでは、豊田哲郎先生(理化学研究所)に、統合データベースSciNetSの構築と今後の創薬への応用について解説していただきました。また、Cutting Edgeでは、池田和由先生(EMBL-EBI)に、オープンデータの化合物SAR情報データベースChEMBLおよび関連データベースについてご紹介いただき、また、五島誠先生(京都コンステラ・テクノロジーズ)と奥野恭史先生(京都大学)には、医薬品有害事象データベース(CzeekV)と、それを利用したデータマイニング例について、ご紹介いただきました。今後も、医薬品の探索・開発研究の様々な面において、データベース、特に公開データベースの活用が、ますます重要になっていくと考えられます。データベース自体に加え、データベース間の連携と統合、さらに活用例についての各先生のご解説は、現状を理解し、また将来への展望を考える上で、たいへん有用な情報ではないかと思われます。先生方には、大変お忙しい中ご執筆いただき、心よりお礼申し上げます。このSAR Newsが今後とも構造活性相関研究の先端情報と展望を会員の皆様にご提供できることを、編集委員一同願っております。

(編集委員会)

SAR News No.23 平成24年10月1日

発行：日本薬学会 構造活性相関部会長 高橋 由雅

SAR News 編集委員会

(委員長) 粕谷 敦

福島 千晶

飯島 洋

竹田一志 鷹 真由子

久保寺 英夫

* 本誌の全ての記事、図表等の無断複写・転載を禁じます。