資料１

Regulatory Application of QSARs

Weida Tong, Ph.D

Director, Division of Bioinformatics and Biostatistics
National Center for Toxicological Research (NCTR), FDA

資料１

化学物質の規制におけるQSARsの適用

Weida Tong, Ph.D

Director, Division of Bioinformatics and Biostatistics
National Center for Toxicological Research (NCTR), FDA

Outline

- Who we are and what we do?
- How to develop a good QSAR model?
- Our lessons-learned for predicting endocrine disruptors
  - How the models were developed and why they were developed in such a way?
  - Regulatory evaluation of our model by EPA
- The difference between in-house tools and commercial tools
  - Introduce Decision Forest and Mold2
- What's the best practice for regulatory application with QSARs

アウトライン

- 私達は何者で何をしているか？
- 良いQSARモデルの開発方法は?
- 内分泌かく乱物質の予測で私達が学んだこと
  ーどのようにモデルを開発し、なぜそのように開発したか？
  ー EPAによる私達のモデルの規制評価
- 組織内開発ツールと市販ツールの違い
  ーDecision Forest とMold2の紹介
- 化学物質の規制のためのQSARsのベストプラクティス（成功例）

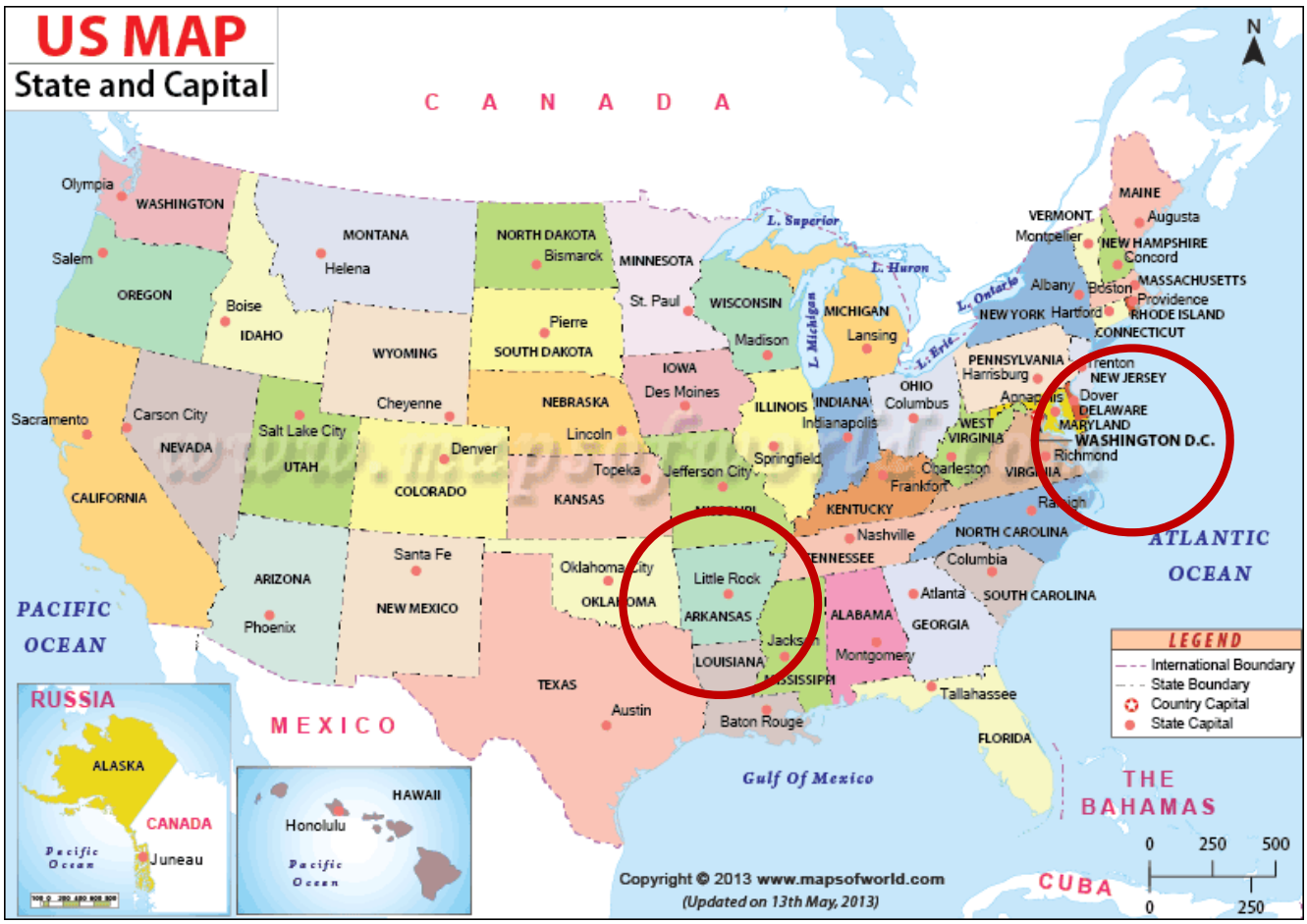
Food and Drug Administration (FDA)

1. Center for Drug Evaluation and Research (CDER)
2. Center for Biologics Evaluation and Research (CBER)
3. Center for Food Safety and Nutrition (CFSAN)
4. Center for Device and Radiological Health (CDRH)
5. Center for Veterinary Medicine (CVM)
6. Center for Tobacco Products
7. Office of Regulatory Affairs (ORA)
8. **National Center for Toxicological Research (NCTR)**

米国･食品医薬品庁 (FDA)

1. 医薬品評価研究センター (CDER)
2. 生物学的製剤評価研究センター (CBER)
3. 食品安全･応用栄養センター (CFSAN)
4. 医療機器･放射線保健センター (CDRH)
5. 動物薬センター(CVM)
6. たばこ製品センター
7. 規制事務局(ORA)
8. **国立毒性研究センター(NCTR)**


US MAP
State and Capital
C A N A D A
N
Olympia
WASHINGTON
Salem
OREGON
Boise
IDAHO
MONTANA
Helena
WYOMING
Cheyenne
NORTH DAKOTA
Bismarck
Pierre
SOUTH DAKOTA
MINNESOTA
St. Paul
NEBRASKA
Lincoln
Denver
COLORADO
IOWA
Des Moines
Sacramento
Carson City
NEVADA
Salt Lake City
UTAH
CALIFORNIA
Topeka
KANSAS
Jefferson City
Springfield
ILLINOIS
Santa Fe
NEW MEXICO
ARIZONA
Phoenix
Oklahoma City
OKLAHOMA
Little Rock
ARKANSAS
Jackson
LOUISIANA
Baton Rouge
TEXAS
Austin
Nashville
TENNESSEE
KENTUCKY
Frankfort
Charleston
WEST VIRGINIA
Richmond
VIRGINIA
Raleigh
NORTH CAROLINA
Columbia
SOUTH CAROLINA
Atlanta
GEORGIA
Montgomery
ALABAMA
Tallahassee
FLORIDA
Gulf Of Mexico
THE BAHAMAS
CUBA
MEXICO
PACIFIC OCEAN
ATLANTIC OCEAN
L. Superior
L. Michigan
L. Huron
L. Erie
L. Ontario
MICHIGAN
Lansing
Madison
INDIANA
Indianapolis
OHIO
Columbus
PENNSYLVANIA
Harrisburg
Annapolis
Trenton
NEW JERSEY
Dover
DELAWARE
MARYLAND
WASHINGTON D.C.
NEW YORK
Albany
Hartford
VERMONT
Montpelier
MAINE
Augusta
NEW HAMPSHIRE
Concord
MASSACHUSETTS
Boston
Providence
RHODE ISLAND
CONNECTICUT
LEGEND
International Boundary
State Boundary
Country Capital
State Capital
0 250 500
0 250
RUSSIA
ALASKA
CANADA
Pacific Ocean
Juneau
HAWAII
Honolulu
Pacific Ocean
Copyright © 2013 www.mapsofworld.com
(Updated on 13th May, 2013)


US MAP
State and Capital
C A N A D A
N
Olympia
WASHINGTON
Salem
OREGON
Boise
IDAHO
MONTANA
Helena
WYOMING
Cheyenne
NORTH DAKOTA
Bismarck
Pierre
SOUTH DAKOTA
MINNESOTA
St. Paul
NEBRASKA
Lincoln
Denver
COLORADO
IOWA
Des Moines
Sacramento
Carson City
NEVADA
Salt Lake City
UTAH
CALIFORNIA
Topeka
KANSAS
Jefferson City
Springfield
ILLINOIS
Santa Fe
NEW MEXICO
ARIZONA
Phoenix
Oklahoma City
OKLAHOMA
Little Rock
ARKANSAS
Jackson
LOUISIANA
Baton Rouge
TEXAS
Austin
Nashville
TENNESSEE
KENTUCKY
Frankfort
Charleston
WEST VIRGINIA
Richmond
VIRGINIA
Raleigh
NORTH CAROLINA
Columbia
SOUTH CAROLINA
Atlanta
GEORGIA
Montgomery
ALABAMA
Tallahassee
FLORIDA
Gulf Of Mexico
THE BAHAMAS
CUBA
MEXICO
PACIFIC OCEAN
ATLANTIC OCEAN
L. Superior
L. Michigan
L. Huron
L. Erie
L. Ontario
MICHIGAN
Lansing
Madison
INDIANA
Indianapolis
OHIO
Columbus
PENNSYLVANIA
Harrisburg
Annapolis
Trenton
NEW JERSEY
Dover
DELAWARE
MARYLAND
WASHINGTON D.C.
NEW YORK
Albany
Hartford
VERMONT
Montpelier
MAINE
Augusta
NEW HAMPSHIRE
Concord
MASSACHUSETTS
Boston
Providence
RHODE ISLAND
CONNECTICUT
LEGEND
International Boundary
State Boundary
Country Capital
State Capital
0 250 500
0 250
RUSSIA
ALASKA
CANADA
Pacific Ocean
Juneau
HAWAII
Honolulu
Pacific Ocean
Copyright © 2013 www.mapsofworld.com
(Updated on 13th May, 2013)

NCTR Office of Research

1. Division of Biochemical Toxicology
2. Division of Neurotoxicology
3. Division of Microbiology
4. Division of Genetic and Molecular Toxicology
5. Division of Systems Biology
6. Division of Bioinformatics and Biostatistics

NCTR 研究部門

1. 生物化学毒性部
2. 神経毒性部
3. 微生物部
4. 遺伝子・分子毒性部
5. システム生物部
6. 生物情報・生物統計部

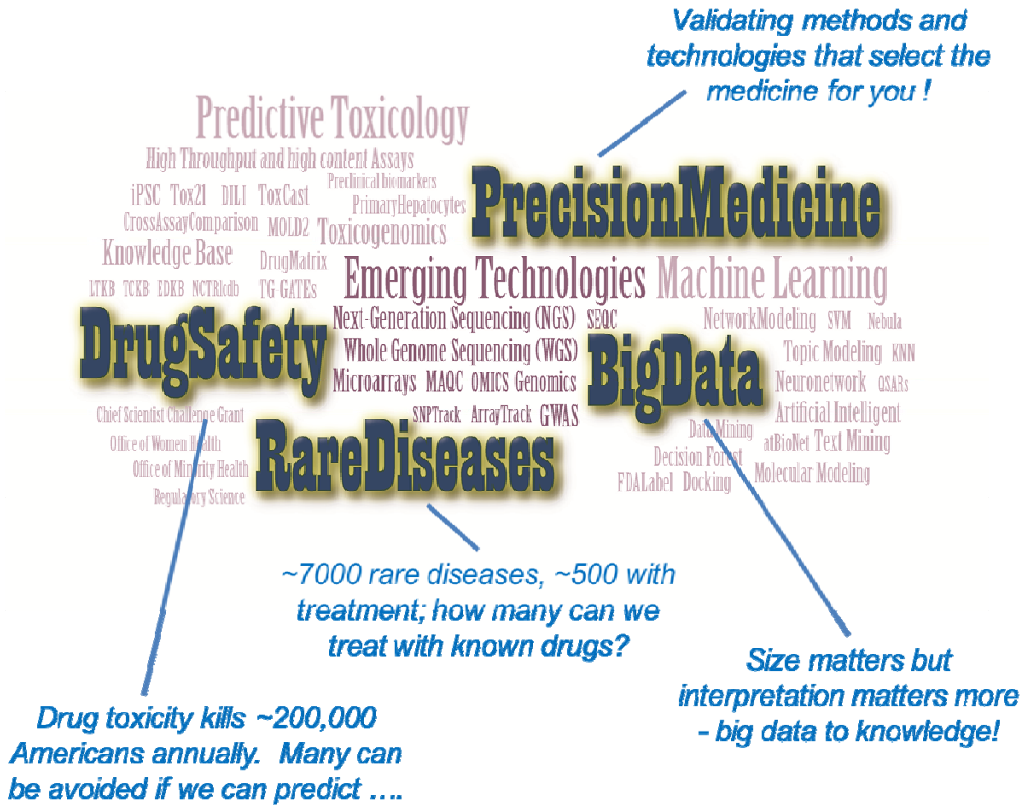
Division Overview

- Established on May 20th, 2012
- Three branches
  - Bioinformatics branch is centered around research
  - Biostatistics branch focuses on research and service
  - Scientific Computing branch is service oriented
- Current staffs
  - ~50 FTEs (including postdoc fellows)
- 40% in Research and 60% in Support/Service

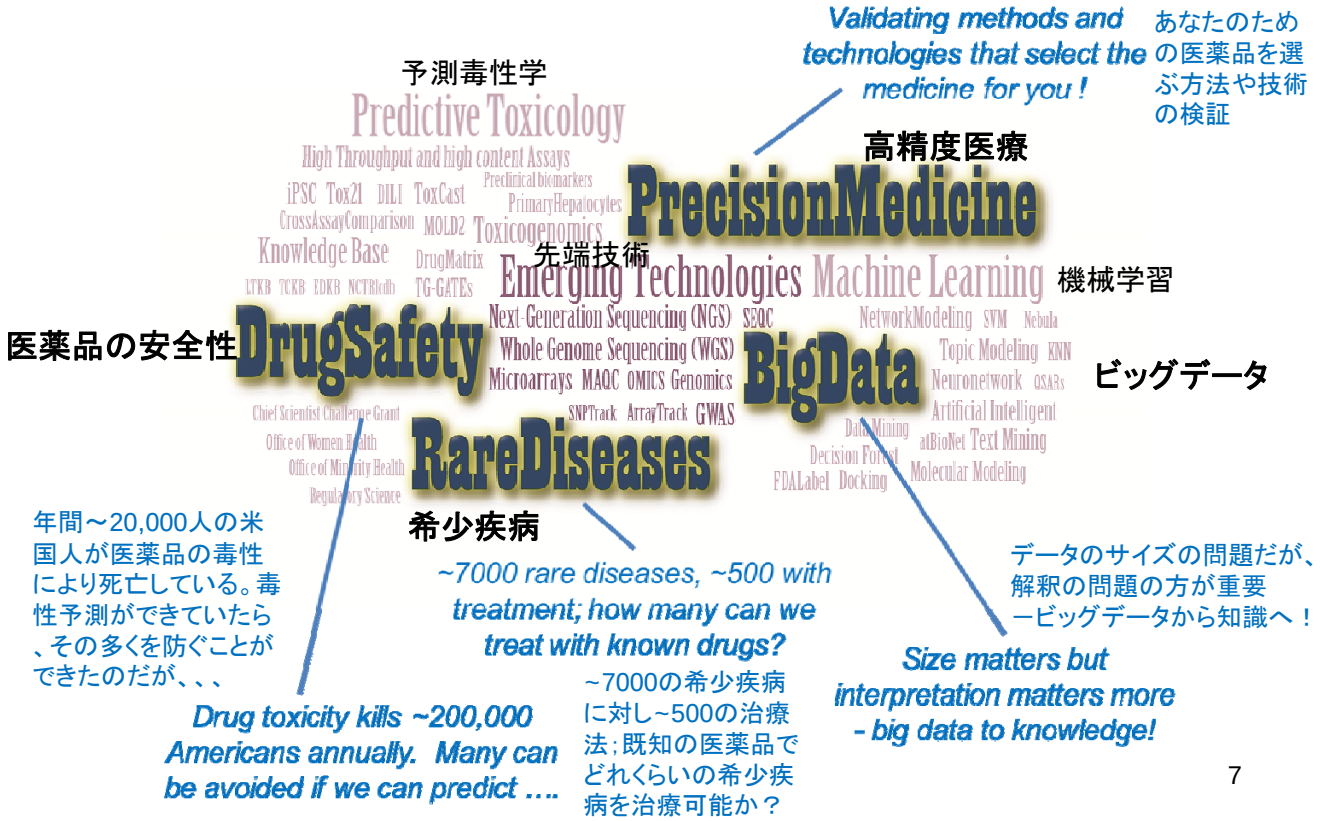
生物情報・生物統計部の概要

- 2012年5月20日設立
- 3つの部門
  - 生物情報部門は研究中心
  - 生物統計部門は研究とサービスに携わる
  - 科学計算部門はサービス中心
- 現在のスタッフ
  - ~50 フルタイム当量 (ポスドクフェロー含む)
- スタッフの40%は研究、60%はサポート/サービスに従事

Division Research Priorities



部内での優先度が高い研究



Foretelling toxicity:

FDA researchers work to predict risk of liver injury from drugs

By Cassandra Willyard

In December 2014, the US Food and Drug Administration (FDA) approved a new drug cocktail, from the Chicago-based pharmaceutical company AbbVie, to treat hepatitis C infection. Less than a year later, the agency warned that the cocktail, Viekira Pak, and another, newer AbbVie hepatitis C therapy could cause serious liver injury in individuals with advanced liver disease. The agency noted that it had received reports of at least 26 cases of liver injuries that might have been caused by the drugs. Of these, ten patients experienced liver failure so severe that they either needed a transplant or died.

The news came as a shock to many people, and AbbVie's share prices tumbled. However, Weida Tong, a researcher at the FDA's National Center for Toxicological Research (NCTR) in Jefferson, Arkansas, could have predicted this outcome. He and his colleagues had recently developed an algorithm to assess a drug's potential for causing liver injury. Tong's team had not assessed these particular drugs before they were approved, but after the agency issued its warning, the researchers entered the data for Viekira Pak into their algorithm and found that it predicted the drug cocktail might have toxic effects on the liver.

When a drug receives FDA approval, the presumption is that it is safe. However, liver injury can be hard to predict, and animal studies do not always identify compounds that might harm human livers. Even human safety studies can miss the signs, in part because the potential for injury can depend on an individual's genetic makeup. "In the area of liver safety, I don't believe there's been any progress whatsoever in the last 30 years," says Paul Watkins, a toxicologist and director of the Institute for Drug Safety Sciences, a joint venture between The Hamner Institutes for Health Sciences and the University of North Carolina at Chapel Hill. Tong and his four-member team hope to change that by developing models that can predict which medicines might cause trouble, before drugmakers embark on costly clinical trials and dangerous drugs reach the public.

Researchers have devised many ways of assessing whether a drug will harm the liver. Watkins and his colleagues have constructed an *in silico* liver called DILIsym to model liver injury. Other researchers are creating three-dimensional mini-livers or seeding liver tissue onto plastic chips to identify toxic drugs, and some groups have bioengineered mice to carry human liver tissue. Tong is taking a less sensational approach by devising mathematical models to predict the risk of liver injury, but he is doing it from within the walls of the world's largest national drug regulatory agency.

Model student

Tong, a bioinformatics buff, began to work on drug-induced liver injury, or DILI, eight years ago. Although there was a wealth of information on the topic, he noticed that the data were scattered. So he became a collector, combing the literature for information that might be useful for building predictive models. As part of this effort, Tong knew that he would first need to develop a scheme for classifying existing drugs according to their potential for causing liver injury. So he and his colleagues turned to the drugs' full labeling information, which is found in the US National Library of Medicine's DailyMed database. These labels are dozens of pages long and contain more than a dozen sections, but the researchers homed in on just three: boxed warning, warnings and precautions, and adverse reactions. The team searched the labels for key words that might indicate liver harm, such as 'hepatitis' or 'fatty liver'. This methodology enabled them to sort nearly 300 FDA-approved drugs into three DILI categories: of 'most concern', of 'less concern' and of 'no concern' (*Drug Discov. Today* **16**, 697–703, 2011). "Even though FDA drug labels are not almighty perfect to address

8

予測毒性学:
FDAの研究者が医薬品による肝障害リスクを予測する研究を実施

NEWS FEATURE

Foretelling toxicity:

FDA researchers work to predict risk of liver injury from drugs

By Cassandra Willyard

© 2016 Nature America, Inc. All rights reserved.

In December 2014, the US Food and Drug Administration (FDA) approved a new drug cocktail, from the Chicago-based pharmaceutical company AbbVie, to treat hepatitis C infection. Less than a year later, the agency warned that the cocktail, Viekira Pak, and another, newer AbbVie hepatitis C therapy could cause serious liver injury in individuals with advanced liver disease. The agency noted that it had received reports of at least 26 cases of liver injuries that might have been caused by the drugs. Of these, ten patients experienced liver failure so severe that they either needed a transplant or died.

The news came as a shock to many people, and AbbVie's share prices tumbled. However, Weida Tong, a researcher at the FDA's National Center for Toxicological Research (NCTR) in Jefferson, Arkansas, could have predicted this outcome. He and his colleagues had recently developed an algorithm to assess a drug's potential for causing liver injury. Tong's team had not assessed these particular drugs before they were approved, but after the agency issued its warning, the researchers entered the data for Viekira Pak into their algorithm and found that it predicted the drug cocktail might have toxic effects on the liver.

When a drug receives FDA approval, the presumption is that it is safe. However, liver injury can be hard to predict, and animal studies do not always identify compounds that might harm human livers. Even human safety studies can miss the signs, in part because the potential for injury can depend on an individual's genetic makeup. "In the area of liver safety, I don't believe there's been any progress whatsoever in the last 30 years," says Paul Watkins, a toxicologist and director of the Institute for Drug Safety Sciences, a joint venture between The Hamner Institutes for Health Sciences and the University of North Carolina at Chapel Hill. Tong and his four-member team hope to change that by developing models that can predict which medicines might cause trouble, before drugmakers embark on costly clinical trials and dangerous drugs reach the public.

Researchers have devised many ways of assessing whether a drug will harm the liver. Watkins and his colleagues have constructed an *in silico* liver called DILIsym to model liver injury. Other researchers are creating three-dimensional mini-livers or seeding liver tissue onto plastic chips to identify toxic drugs, and some groups have bioengineered mice to carry human liver tissue. Tong is taking a less sensational approach by devising mathematical models to predict the risk of liver injury, but he is doing it from within the walls of the world's largest national drug regulatory agency.

Model student

Tong, a bioinformatics buff, began to work on drug-induced liver injury, or DILI, eight years ago. Although there was a wealth of information on the topic, he noticed that the data were scattered. So he became a collector, combing the literature for information that might be useful for building predictive models. As part of this effort, Tong knew that he would first need to develop a scheme for classifying existing drugs according to their potential for causing liver injury. So he and his colleagues turned to the drugs' full labeling information, which is found in the US National Library of Medicine's DailyMed database. These labels are dozens of pages long and contain more than a dozen sections, but the researchers homed in on just three: boxed warning, warnings and precautions, and adverse reactions. The team searched the labels for key words that might indicate liver harm, such as 'hepatitis' or 'fatty liver'. This methodology enabled them to sort nearly 300 FDA-approved drugs into three DILI categories: of 'most concern', of 'less concern' and of 'no concern' (*Drug Discov. Today* **16**, 697–703, 2011). "Even though FDA drug labels are not almighty perfect to address

450 VOLUME 22 | NUMBER 5 | MAY 2016 NATURE MEDICINE

8

Toxicological Knowledge Base Development

"Knowledgebase" development involves (1) data collection and (2) predictive modeling (e.g., QSARs):

- Endocrine Disruptor Knowledge Base (EDKB): ~8000 chemicals
- Liver Toxicity Knowledge Base (LTKB): ~2000 drugs
- Liver Cancer Knowledge Base (NCTRlcdb): ~1000 chemicals
- Tobacco Constituents Knowledge Base (TCKB): ~9000 tobacco constituents

毒性学知識ベースの開発

"知識ベース"の開発は(1)データ収集と(2)予測モデル(例:QSARs)が必要:

- 内分泌かく乱知識ベース (EDKB): ~8000 化学物質
- 肝毒性知識ベース (LTKB): ~2000 医薬品
- 肝腫瘍知識ベース (NCTRlcdb): ~1000 化学物質
- たばこ成分知識ベース (TCKB): ~9000 たばこ成分

Outline

- Who we are and what we do?
- How to develop a good QSAR model?
- Our lessons-learned for predicting endocrine disruptors
  - How the models were developed and why they were developed in such a way?
  - Regulatory evaluation of our model by EPA
- The difference between in-house tools and commercial tools
  - Introduce Decision Forest and Mold2
- What's the best practice for regulatory application with QSARs

アウトライン

- 私達は何者で何をしているか？
- 良いQSARモデルの開発方法は?
- 内分泌かく乱物質の予測で私達が学んだこと
  ーどのようにモデルを開発し、なぜそのように開発したか？
  ー EPAによる私達のモデルの規制評価
- 組織内開発ツールと市販ツールの違い
  ーDecision Forest とMold2の紹介
- 化学物質の規制のためのQSARsのベストプラクティス（成功例）

How to Achieve a Robust QSAR Model – Rule of Thumb

- Data size: need a large dataset
  - Usually we stuck with what we have, but the larger the dataset is, the more robust a QSAR model will be
- Biological activity data: Quality does matter
  - The high quality of activity data helps; garbage-in and garbage-out
  - Cover a broad range of activity space; more than 3 order of magnitude for "Quantitative" model
- Chemical Structure: Diversity is a good thing! Cover as many chemical classes as possible
- Modeling approach: Advocating the democratic (consensus) approaches

頑健なQSARモデルの管理方法 —経験より

- データサイズ: 大規模なデータセットが必要
  - たいてい私達は何を持っているかにとらわれてしまうが、データセットが大きいほどQSARモデルは頑健になる。
- 生物学的活性データ:質が重要
  - 高品質の活性データは解析を有効にする;ガラクタを入れればガラクタが出てくる
  - 広い範囲の活性スペースをカバーすること; "定量"モデルには3桁以上必要
- 化学構造: 多様性は良いこと！できるだけ多くのケミカルクラスをカバーするとよい
- モデリングアプローチ:民主的(合意)アプローチを提唱すること

How We Know We Had a Good Model?

1. **Cross-validation** to assess whether a robust model can be developed based on this dataset
2. **Permutation test** to ensure the observation is not due to chance
3. **Validation sets** to assess the performance of the model derived from the training set
4. **Additional validation sets** from the literature to further validate the model
5. **Applicability domain** assessment to identify the drug categories for which the model will perform better

良いモデルであることをどのように確認するか？

1. **クロスバリデーション** このデータセットに基づき頑健なモデルを開発できたかを評価する方法
2. **順列テスト** 結果が偶発的ではないことを確認する方法
3. **バリデーションセット** トレーニングセットからつくられたモデルを評価するデータセット
4. **追加バリデーションセット** モデルの妥当性をさらに評価するための文献から得たデータセット
5. **適用範囲** そのモデルが適しているのはどの医薬品カテゴリーか

Outline

- Who we are and what we do?
- How to develop a good QSAR model?
- Our lessons-learned for predicting endocrine disruptors
  - How the models were developed and why they were developed in such a way?
  - Regulatory evaluation of our model by EPA
- The difference between in-house tools and commercial tools
  - Introduce Decision Forest and Mold2
- What's the best practice for regulatory application with QSARs

アウトライン

- 私達は何者で何をしているか？
- 良いQSARモデルの開発方法は?
- 内分泌かく乱物質の予測で私達が学んだこと
  ーどのようにモデルを開発し、なぜそのように開発したか？
  ー EPAによる私達のモデルの規制評価
- 組織内開発ツールと市販ツールの違い
  ーDecision Forest とMold2の紹介
- 化学物質の規制のためのQSARsのベストプラクティス（成功例）

Endocrine Disruptors

- An international issue
- Two laws passed by US congress require evaluation of chemicals found in foods and water for endocrine disruption.
- Similar regulation is also implemented in Europe and Asia
- ~ 90,000 commercial chemicals needs to be screened
- EPA has identified ~58,000 eligible chemicals
- A minimum of 8,000 of the 58,000 chemicals are FDA-regulated, including cosmetic ingredients, drug products …

内分泌かく乱物質

- 国際的な問題である
- 米国議会を通過した2つの法律は、食品および水中に認められた内分泌かく乱化学物質の評価を要求している。
- 欧州やアジアでも同様に規制されている
- ~ 90,000の工業用化学物質のスクリーニングが必要
- EPAは~58,000の該当する化学物質を同定した
- 58,000 物質中最低8,000物質（化粧品原材料や医薬品製品を含む）を FDAは規制している…

Overview of NCTR's Endocrine Disruptor Knowledge Base (EDKB)

- Objective: A resource that contains both experimental data and predictive models for endocrine disrupting compounds

- Overview:
  - The >200 same chemicals were assayed at NCTR for their binding affinities to estrogen (ER), androgen (AR), and serum protein (AFP and SHBG) receptors
  - A database containing various assay data (~10 different *in vitro* and *in vivo* assays) for ~8000 chemicals with estrogenic activity data
    - The ER data was used by Tox21 to validate the ER assay results
  - Many SAR/QSAR models have been developed for ER and AR binding using EDKB
  - Published 46 articles

- Webpage: (http://www.fda.gov/nctr/science/centers/toxicoinformatics/edkb/)

NCTRの内分泌かく乱知識ベース (EDKB)の概要

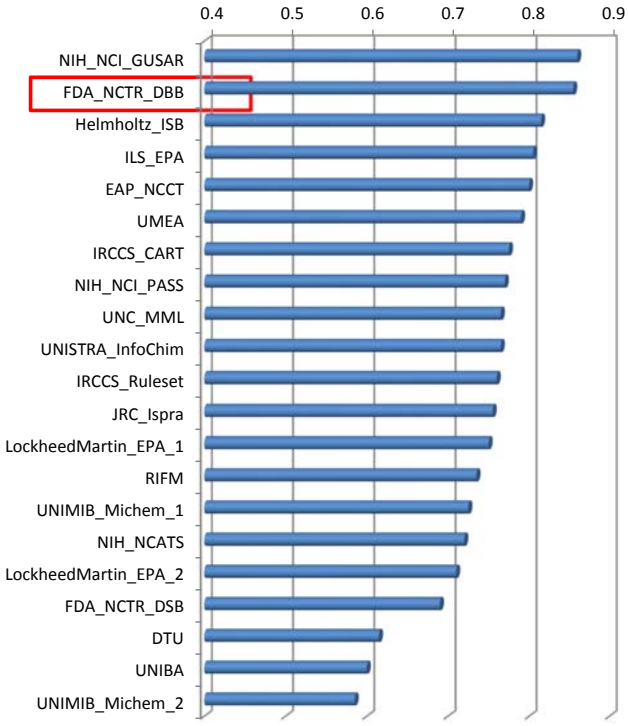
- 目的:内分泌かく乱物質の実験データ及び予測モデルを含む情報源

- 概要:
  - 200を超える化学物質のエストロゲン受容体(ER)、アンドロゲン受容体 (AR)及び血清タンパク質(AFP, SHBG)受容体との結合親和性をNCTRで解析した。
  - エストロゲン活性データを有する~8000物質の様々な試験データ(~10種のin vitro及びin vivo試験)を含むデータベース
  - ERデータは、ER試験結果を検証するため、Tox21プロジェクトに用いられた。
  - EDKBを用いて、多くのSAR/QSARモデルが、ER及びAR結合性を評価するため開発された。
  - 46報の論文を発表した

- ウェブページ: (http://www.fda.gov/nctr/science/centers/toxicoinformatics/edkb/)

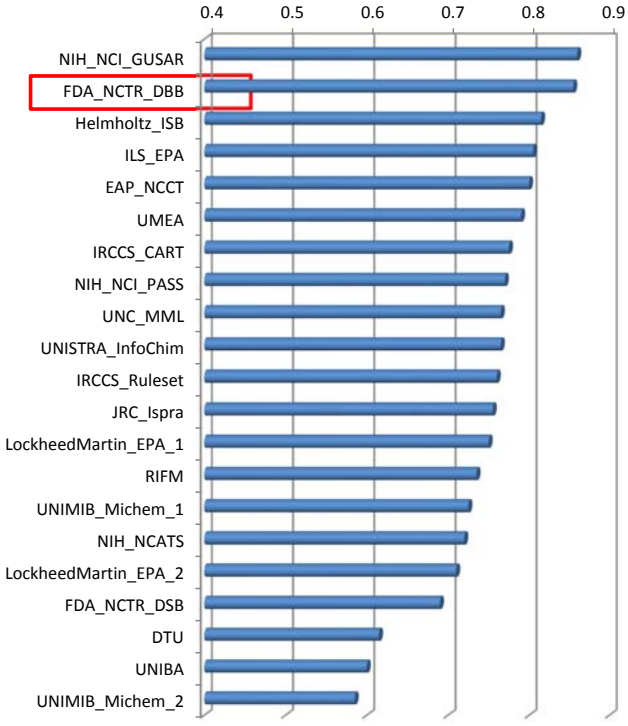
Community-Wide QSAR Exercise

- Project name: Collaborative Estrogen Receptor Activity Prediction Project (CERAPP)
- Objective: Predicting ER binding by chemical structures
- Organizer/Duration: EPA and 2013.9-2014.12
- Participants: 21 teams
- Method: 1529 chemicals for model development and the model was then challenged by the blind dataset with 7283 chemicals to assess performance

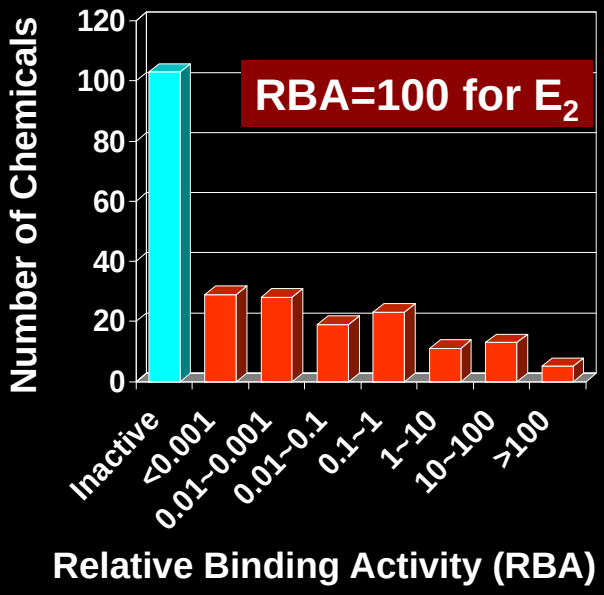


組織的な QSAR の演習

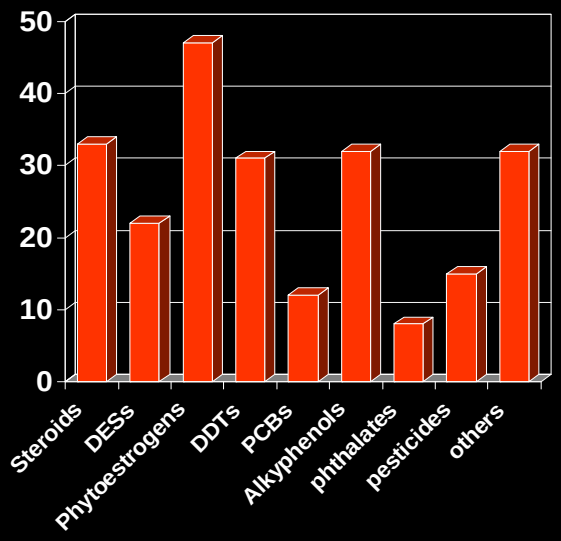
- プロジェクト名: エストロゲン受容体活性予測共同プロジェクト (CERAPP)
- 目的: 化学構造によりER結合能を予測すること
- 世話人/期間: EPA /2013.9-2014.12
- 参加者: 21チーム
- 方法: 1529物質をモデル開発のため用い、そのモデル開発は動作確認するため7283物質の名称を隠したデータセットを用いて行った。



Both ER and AR Datasets Cover a Wide Range of Chemical Classes and Activity



Relative Binding Activity (RBA)

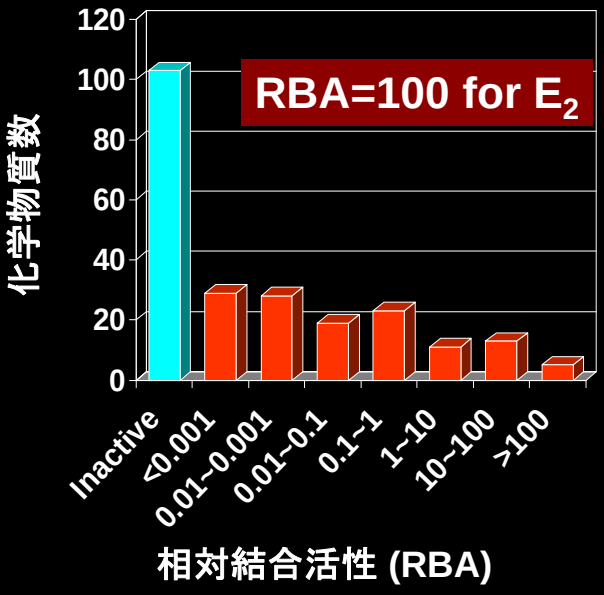


Chemical classes

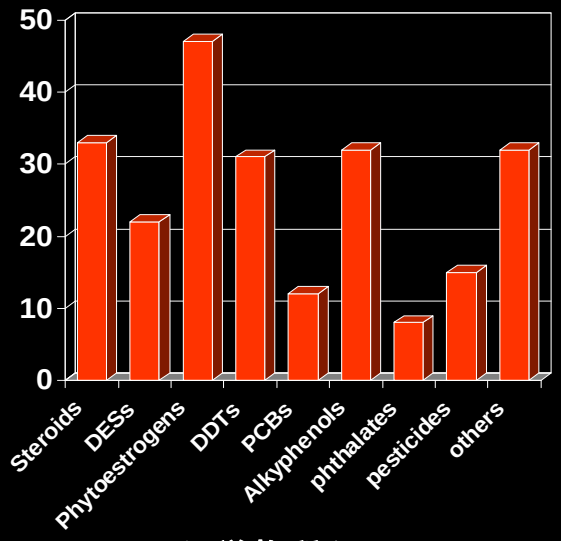
Over six orders of magnitude of RBA range

A wide range of chemical classes

ER及びARのデータセットは広範囲の化学物質クラス及び活性を網羅していた



相対結合活性 (RBA)

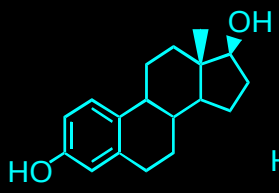


化学物質クラス

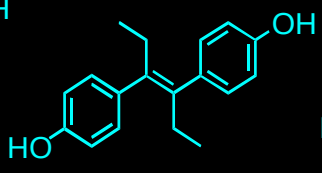
6桁を超えるRBA範囲

広範囲の化学物質クラス

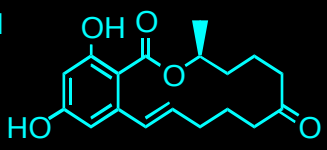
Structural Diversity of NCTR Estrogen Dataset



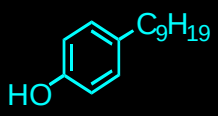
Estradiol



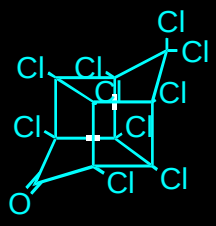
DES



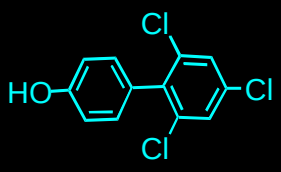
Zearalenone



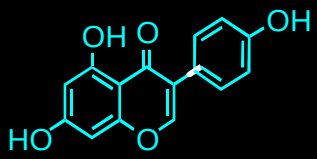
Nonylphenol



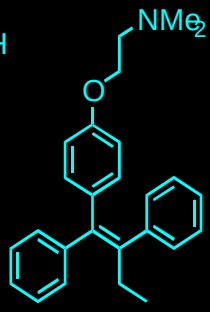
Kepone



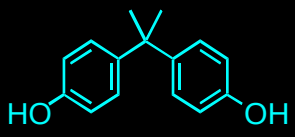
4-OH-2',4',6'-PCB



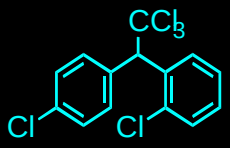
Genistein



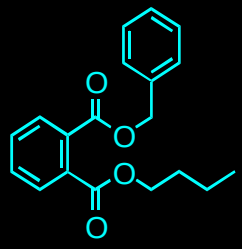
Tamoxifen



Bisphenol A

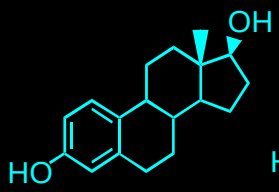


o,p'-DDT

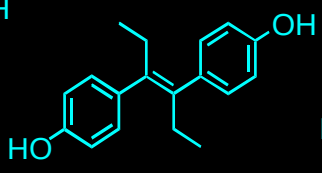


Butylbenzylphthalate

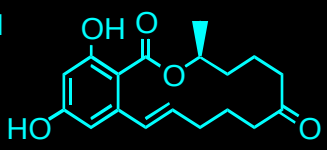
NCTRエストロゲンデータセットの構造多様性



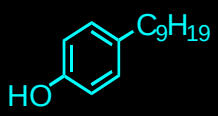
Estradiol



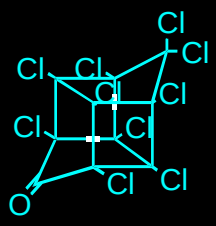
DES



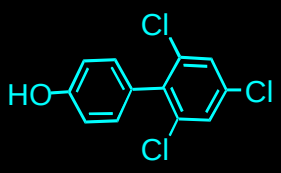
Zearalenone



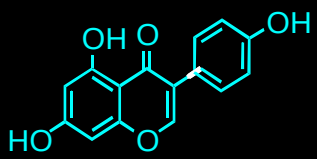
Nonylphenol



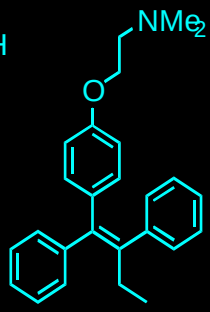
Kepone



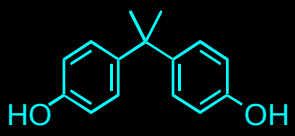
4-OH-2',4',6'-PCB



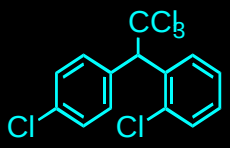
Genistein



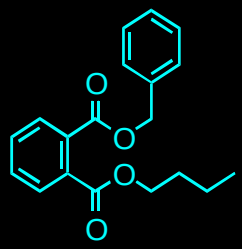
Tamoxifen



Bisphenol A



o,p'-DDT



Butylbenzylphthalate

EDKB – Modeling component

Assessing various SAR/QSAR approaches for ER

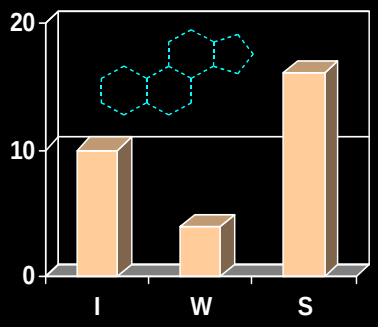
- Rules or filters (e.g. if MW>1000, it is inactive)
- Presence/absence of structural features
  - Structural alerts - 2D substructure
  - Pharmacophores - 3D structural features
- Classification models (YES/NO) - KNN, SIMCA, ANN, Decision Tree
- Quantitative structure-activity relationships (QSARs) – Classical QSARs, HQSAR, 3D QSAR/CoMFA

EDKB – モデリング構成

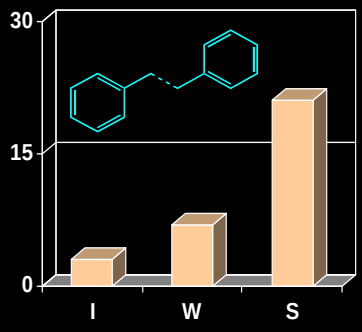
ERのための様々なSAR/QSARアプローチの評価

- ルール又はフィルター(例:分子量>1000の場合、不活性とする)
- 構造的特徴の有無
  - 構造アラート - 2D 部分構造
  - ファーマコフォア - 3D 構造的特徴
- 分類モデル(YES/NO) - KNN, SIMCA, ANN, Decision Tree
- 定量的構造活性相関 (QSARs) – 古典的QSARs, HQSAR, 3D QSAR/CoMFA

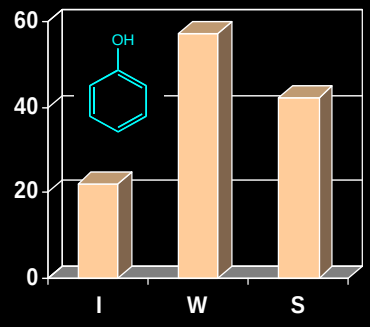
2D - Structural Alerts



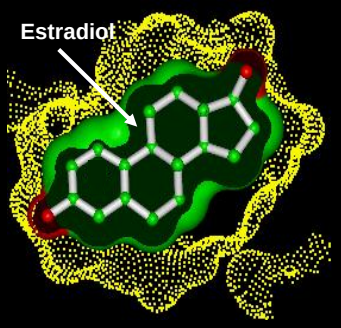
Steroid Skeleton

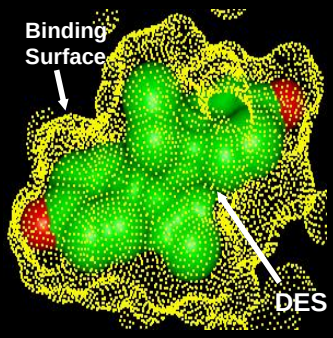


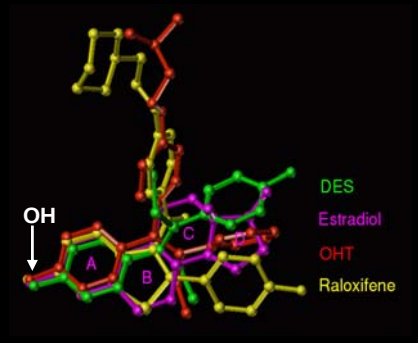
DES Skeleton



Phenolic Ring

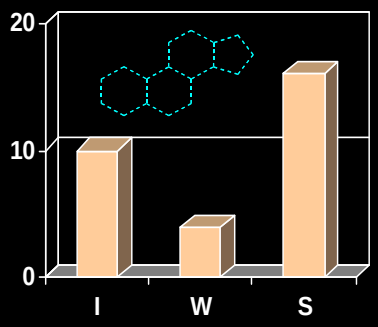




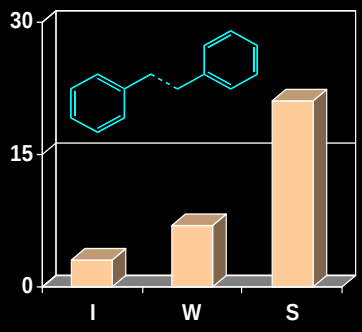


* Strong (> 0.1), Weak (0.1 ~ 10^{-4}), Inactive (< 10^{-4}); RBA for E_2= 100

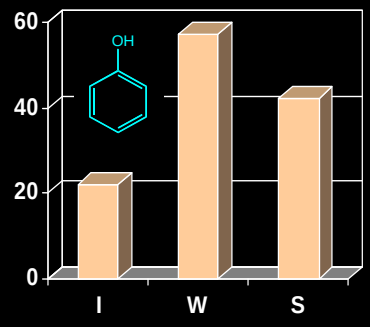
2D – 警告構造



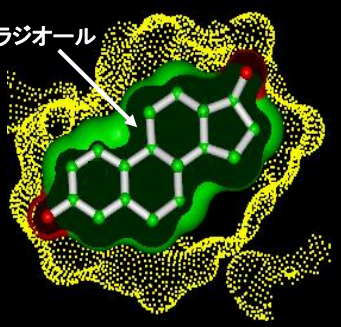
ステロイド骨格

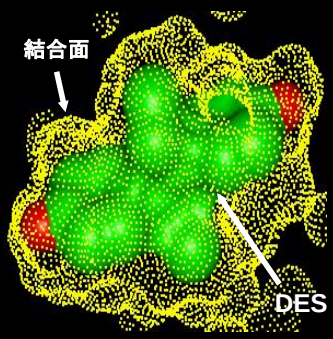


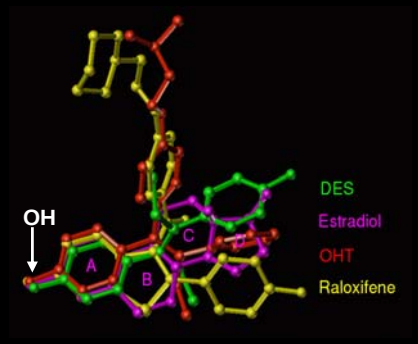
DES骨格



フェノール環

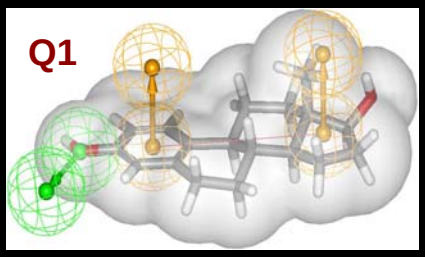


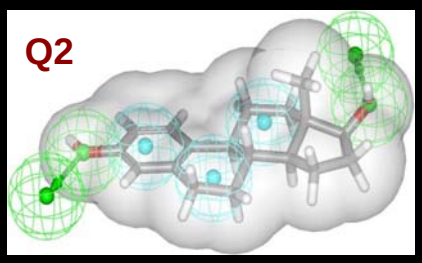


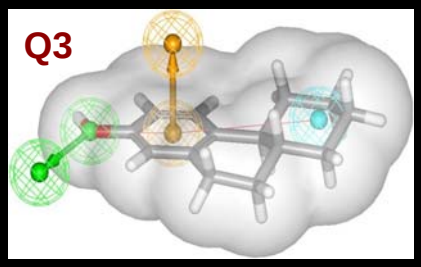


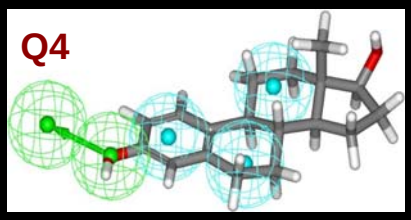
* 強い (> 0.1), 弱い (0.1 ~ 10^{-4}), 不活性 (< 10^{-4}); RBA for E_2= 100

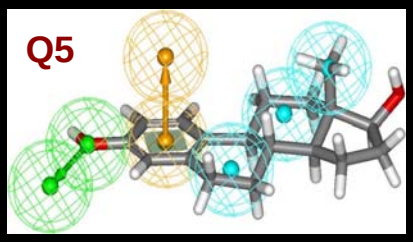
Seven Pharmacophore Queries

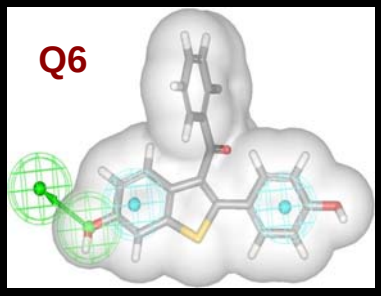


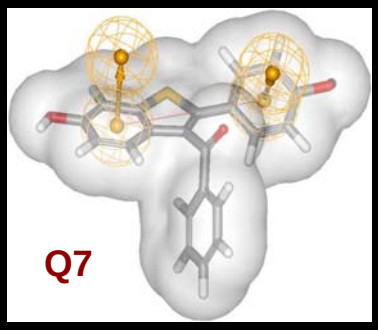




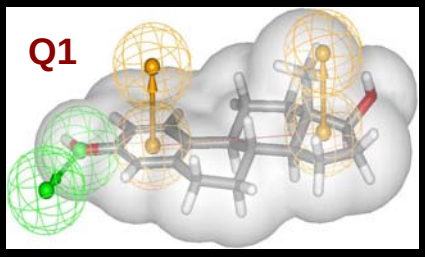


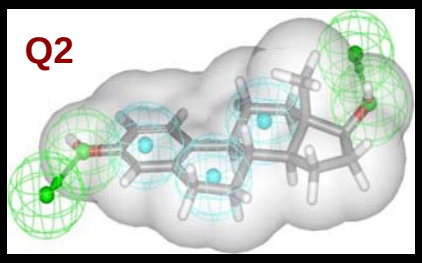


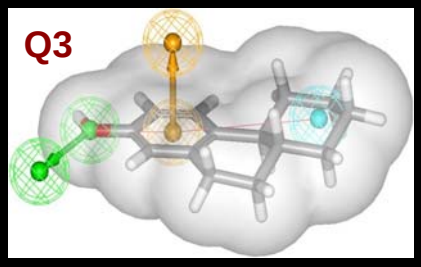


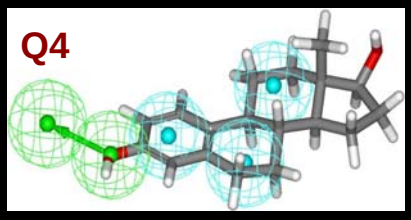


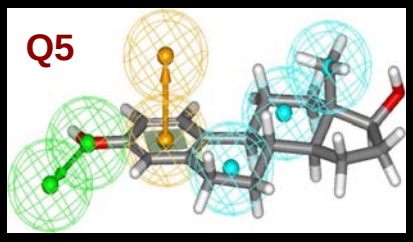
7つのファーマコフォアクエリー

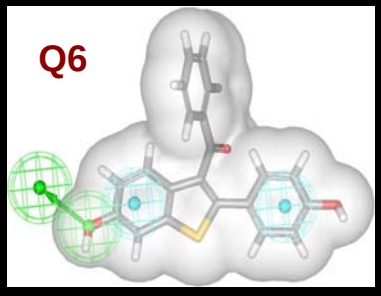


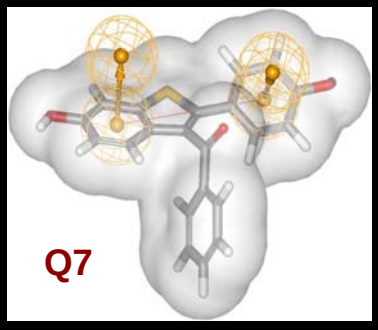












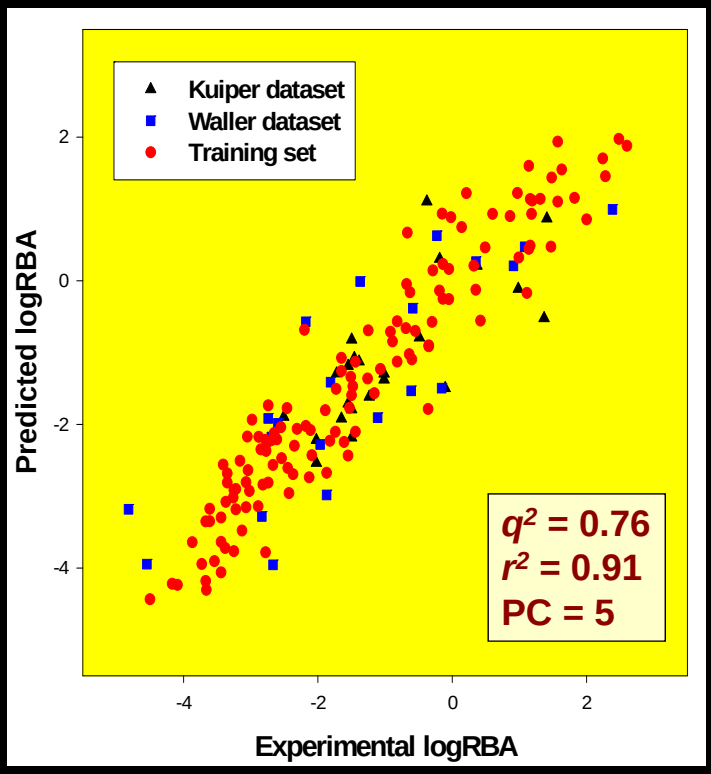
Validation

1. Model's Performance in the Training Set
   - Classification: overall predictivity, sensitivity and specificity
   - QSAR: q^2 between calculated and exp results
2. Internal Validation (within the sampe population)
   - Cross validation process (Leave-One-Out (LOO), 10-fold …)
3. External Validation
   - Test sets: chemicals with known activity but from different laboratories

検　証

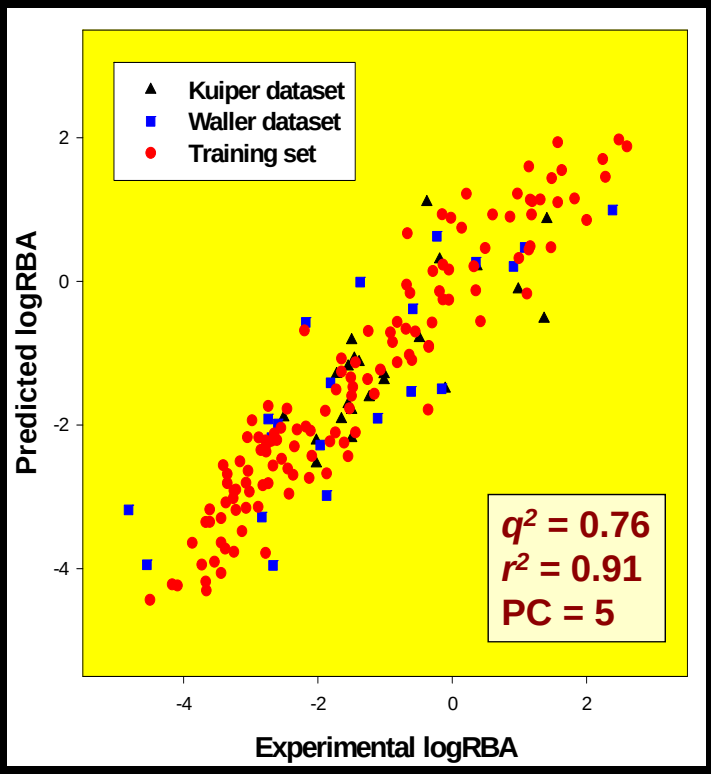
1. トレーニングセットによるモデルの性能
   - 分類: 全体の予測性、感度、特異性
   - QSAR: 計算結果と試験結果間のq^2
2. 内部検証(標本群内)
   - 交差検証プロセス (Leave-One-Out (LOO, 標本群から1つの事例を抜き出してテスト事例とする), 10倍 …)
3. 外部検証
   - テストセット: 異なる実験施設から提供された既知の活性を有する化学物質

3D QSAR/CoMFA Results



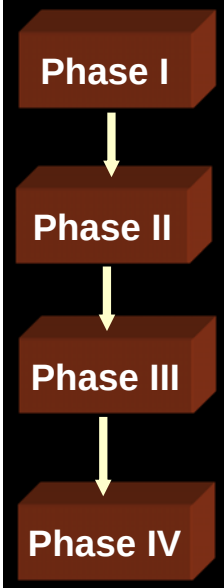
Inactives	Pred.
4,4'-biphenol (<-2.51)	-2.6
Ipriflavone (<-2.51)	NA
DACT	NA
Hydroxyflutamide	NA
M1	NA
M2	NA

3D QSAR/CoMFA 結果



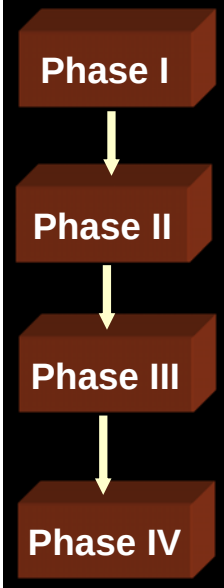
Inactives	Pred.
4,4'-biphenol (<-2.51)	-2.6
Ipriflavone (<-2.51)	NA
DACT	NA
Hydroxyflutamide	NA
M1	NA
M2	NA

“4-Phase” Approach for ER



Rejection Filters	- MW<94 or >1000 - No ring
Active/Inactive Assignment	- 3 structural alerts - 7 pharmacophores - 1 Decision Tree model
Quantitative Predictions	- CoMFA model
Knowledge-Base Approach	- Other priority setting factors - Re-evaluating Phases II & III - Human knowledge

ERのための“4-段階” アプローチ



却下フィルター	- 分子量<94 または >1000 - 環状構造なし
活性／不活性の決定	- 3 構造アラート - 7 ファーマコフォア - 1 決定樹モデル
定量的予測	- CoMFA モデル
知識ベースのアプローチ	- 他の優先順位決定因子 - Phases II & IIIの再評価 - ヒトの知識

Outline

- Who we are and what we do?
- How to develop a good QSAR model?
- Our lessons-learned for predicting endocrine disruptors
  - How the models were developed and why they were developed in such a way?
  - Regulatory evaluation of our model by EPA
- The difference between in-house tools and commercial tools
  - Introduce Decision Forest and Mold2
- What's the best practice for regulatory application with QSARs

アウトライン

- 私達は何者で何をしているか？
- 良いQSARモデルの開発方法は?
- 内分泌かく乱物質の予測で私達が学んだこと
  ーどのようにモデルを開発し、なぜそのように開発したか？
  ー EPAによる私達のモデルの規制評価
- 組織内開発ツールと市販ツールの違い
  ーDecision Forest とMold2の紹介
- 化学物質の規制のためのQSARsのベストプラクティス（成功例）

How QSARs be used in regulatory decision-making ?

Utility of SAR/QSARs for Prescreening potential EDCs by EPA:

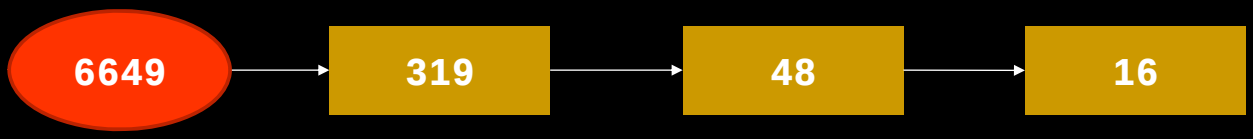
- **Requirement 1 – Enrichment**
  - Narrow down the number of chemicals for testing
  - Or enhance the chance to find actives in a reduced population prioritized by the models
- **Requirement 2 – Low false negatives**
  - 4-Phase model are used in early stage of priority setting
  - High false positives are less concern, which could be removed in experiment
  - High false negatives are of more concerns

QSARsは規制当局の政策決定にどのように活用されているか？

EPAによるSAR/QSARsを用いた内分泌かく乱物質の予備スクリーニング:

- **要求 1 –絞り込み**
  - 試験のために化学物質数を絞る
  - またはモデルにより優先順位付けして減らした化学物質から活性があるものを見つける見込みを高める
- **要求 2 – 偽陰性の確率を低くする**
  - 4-段階モデルは優先順位付けの初期段階で実施する
  - 偽陽性については懸念が少ない。試験で排除することができる。
  - 高率な偽陰性はより懸念が大きい

Enrichment (6 fold increase)



EPA's list; estimated that 5% are active

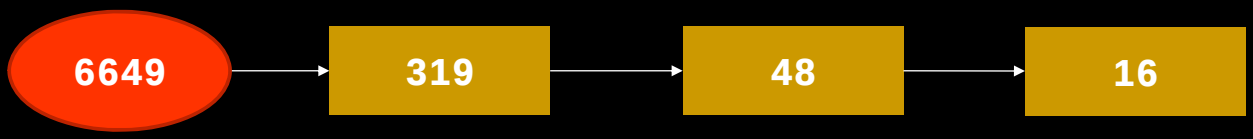
Predicted to be active by QSARs (5.8%)

Randomly select 48 cmpds for assaying (15%)

Active in exp. (33.3%)

$$\text{Enrichment} = \frac{16/48}{5\%} > 6$$

絞り込み (6倍増加)



EPAのリスト; 5%が活性ありと予測

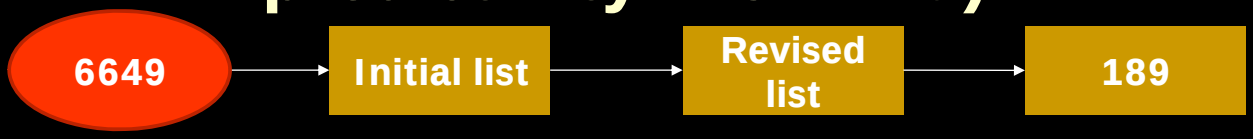
QSARsによる活性予測 (5.8%)

アッセイのためにランダムに選別した48物質 (15%)

試験で活性あり (33.3%)

$$\text{絞り込み} = \frac{16/48}{5\%} > 6$$

Low False Negatives (Negative predictivity = 97.2%)



EPA's list; estimated that 5% are active

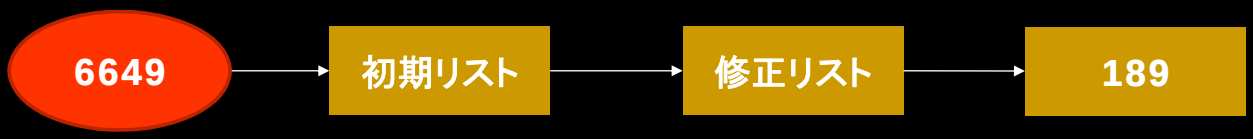
Randomly select 200

Removed the "unavailable", insoluble, and failed chemicals

Assayed

Total 189 cmpds	Inactive	Active (RBA, E_2 = 1.0)		
		10^{-5} ~ 10^{-6}	10^{-4} ~ 10^{-5}	10^{-3} ~ 10^{-4}
Exp	178	6	4	1
4-Phase	189	0		

低い偽陰性率(陰性予測性 = 97.2%)



EPAのリスト; 5%が活性ありと予測

ランダムに200物質選別

"入手できない"、不溶性、試験できない化学物質を除いた

解析したもの

合計 189 物質	不活性	活性 (RBA, E_2 = 1.0)		
		10^{-5} ~ 10^{-6}	10^{-4} ~ 10^{-5}	10^{-3} ~ 10^{-4}
試験	178	6	4	1
4-段階	189	0		

Outline

- Who we are and what we do?
- How to develop a good QSAR model?
- Our lessons-learned for predicting endocrine disruptors
  - How the models were developed and why they were developed in such a way?
  - Regulatory evaluation of our model by EPA
- The difference between in-house tools and commercial tools
  - Introduce Decision Forest and Mold2
- What's the best practice for regulatory application with QSARs

アウトライン

- 私達は何者で何をしているか？
- 良いQSARモデルの開発方法は?
- 内分泌かく乱作用の予測で私達が学んだこと
  －どのようにモデルを開発し、なぜそのように開発したか？
  － EPAによる私達のモデルの規制評価
- 組織内開発ツールと市販ツールの違い
  －Decision Forest とMold2の紹介
- 化学物質の規制のためのQSARsのベストプラクティス（成功例）

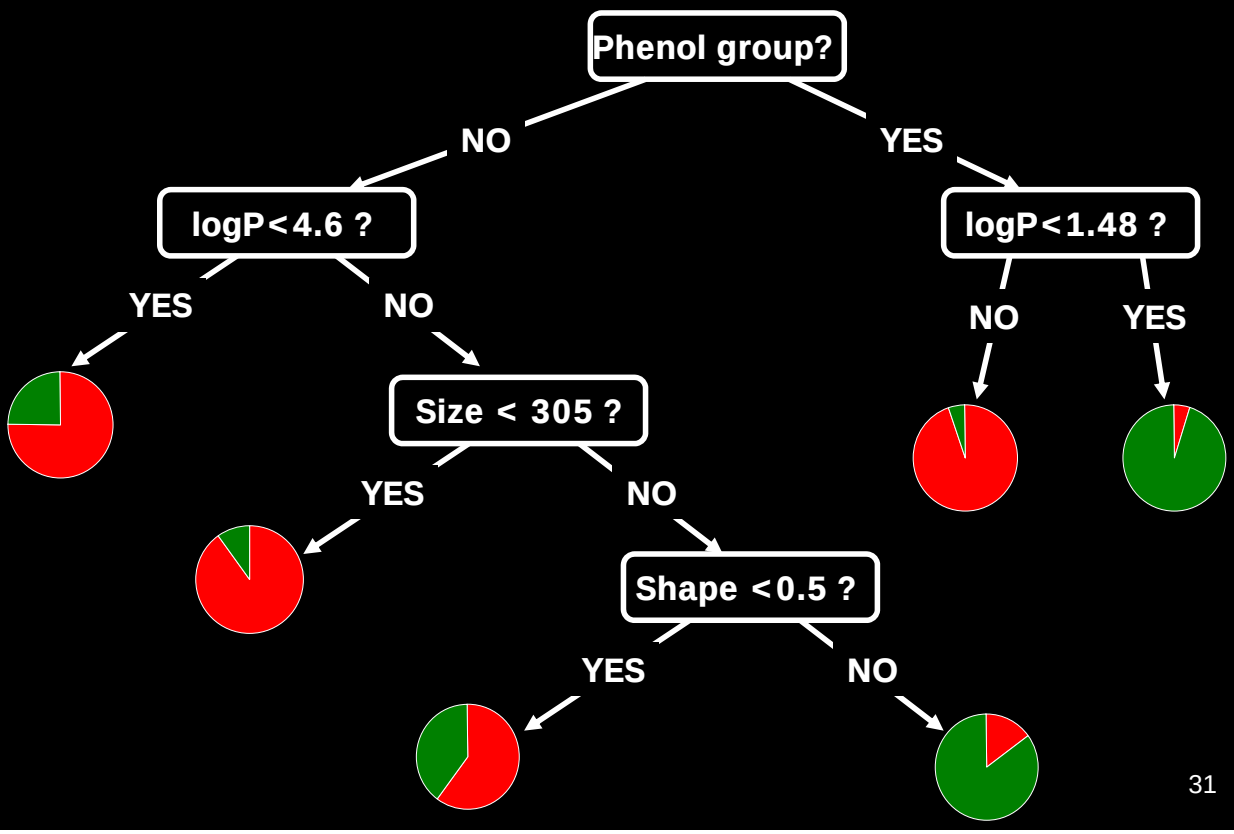
Are We There Yet

- Drawbacks for the 4-Phase models:
  - Multi-software and platform
  - Expensive, difficult to be adopted (or used) by others
  - Maintain a validated model become difficult because of periodically updating in commercial software
  - Expert-dependent (CoMFA)
  - No confidence associated with each prediction
- An ideal SAR approach
  - Accurate and reproducible
  - Can be independently operated by regulators
  - Applicability domain should be well defined
- **Decision Forest – A consensus approach**

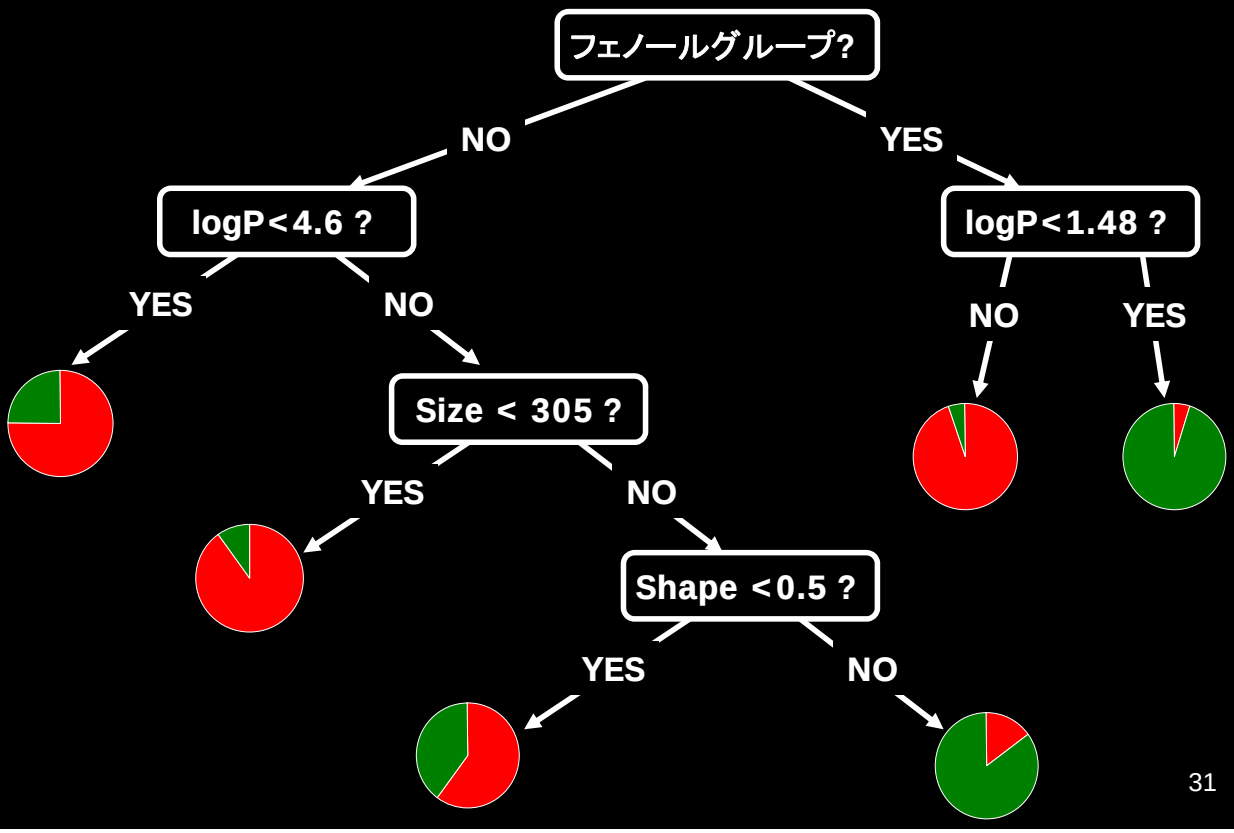
まだ道半ば

- **4-段階モデルの欠点:**
  - **マルチソフトウェアとプラットフォーム**
  - **高価で、外部者は改良・利用しづらい**
  - **市販ソフトウェアが定期的に更新されるため、検証モデルの維持が難しい**
  - **専門家依存性 (CoMFA)**
  - **各予測に関する信頼性がない**
- **理想的なSARアプローチ**
  - **正確性と再現性**
  - **規制当局が独立して運用できる**
  - **適用範囲が十分定義されている**
- **ディシジョンフォレスト– コンセンサスアプローチ**

Decision Tree

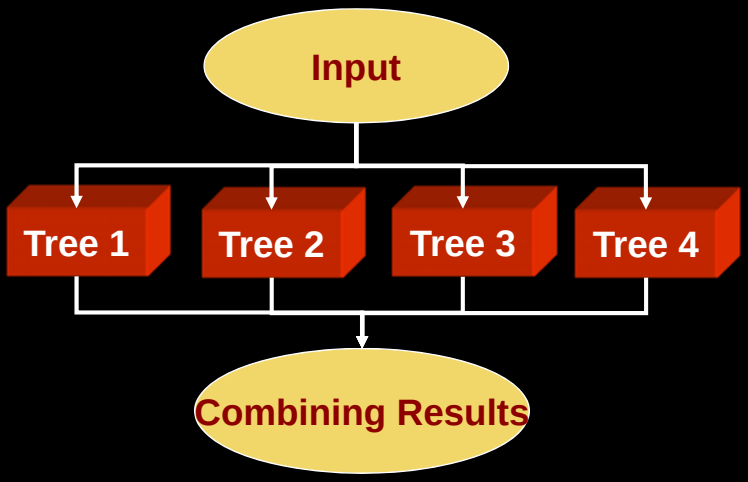


ディシジョンツリー（決定樹）



Decision Forest

Assumption: A better classification can be reached by combining the results from several individual models.

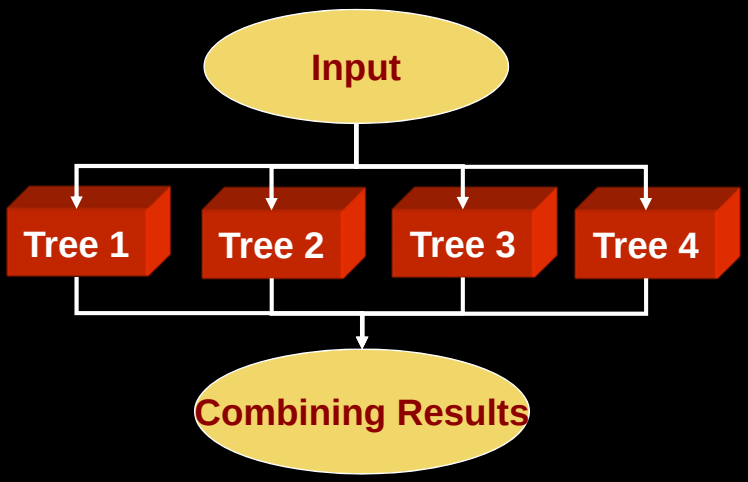


Key points

- Combining several identical trees produce no gain
- Combining several highly correct trees that disagree as much as possible

ディシジョンフォレスト

仮説: より良い分類は複数のモデルから得た結果を合わせることにより得られる



キーポイント

- いくつかの同一のツリーを組み合わせても利益はない
- できる限り一致しない複数の非常に正しいツリーを組み合わせる

Decision Forest - Two Premises

- Each tree was developed using a distinct set of descriptors that was explicitly excluded from other trees to ensure its *unique* contribution in prediction
- All trees were statistically comparable to ensure their *equal* weight in combining prediction

1. Tong et al. Decision Forest – Combining multiple independent models for prediction, JCICS, 43(2):525-531, 2003
2. Tong et al. Assessment of prediction confidence and domain extrapolation of two structure-activity relationship models for predicting estrogen receptor binding activity, EHP Tox, 112(12):1249-1254, 2004

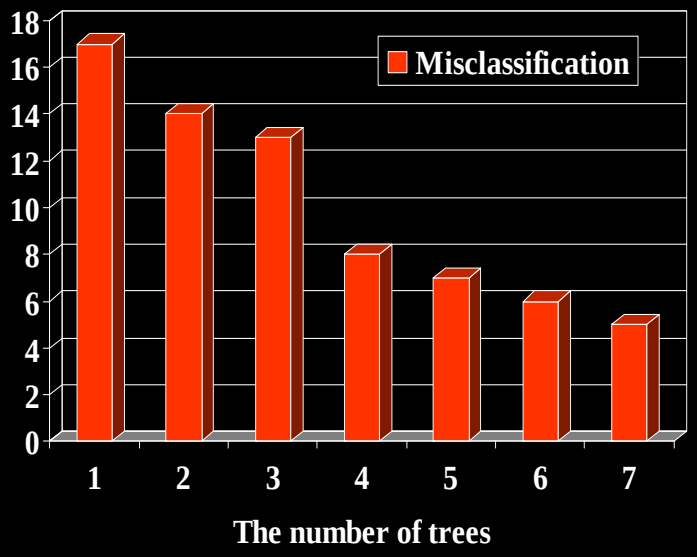
ディシジョンフォレスト - 2つの前提

- 各ツリーは、予測への唯一の寄与のため、他のツリーから確実に除外された異なるセットの記述子を用いて構築された
- 全てのツリーは、予測結果を結合する際その重みが等しいことを保証するため、統計学的に同等である

1. Tong et al. Decision Forest – Combining multiple independent models for prediction, JCICS, 43(2):525-531, 2003
2. Tong et al. Assessment of prediction confidence and domain extrapolation of two structure-activity relationship models for predicting estrogen receptor binding activity, EHP Tox, 112(12):1249-1254, 2004

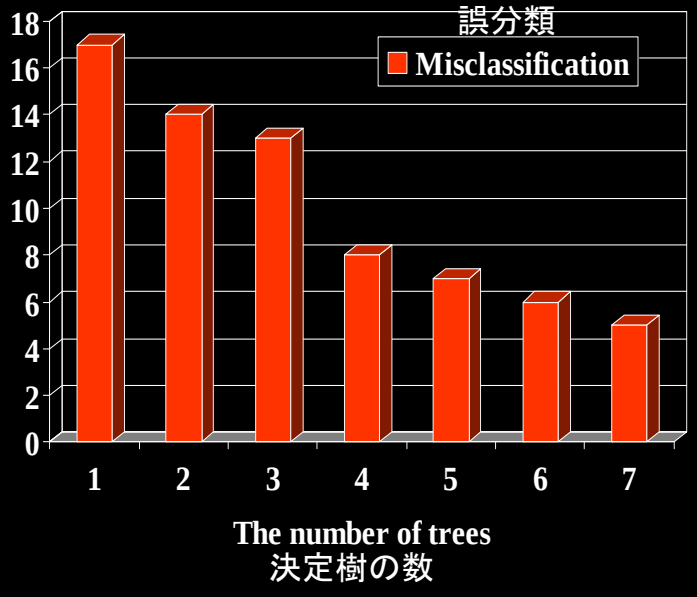
Combining Decision Trees – Decision Forest

Trees	Des.	Mis.
Tree 1	10	17
Tree 2	10	19
Tree 3	12	17
Tree 4	12	17
Tree 5	15	19
Tree 6	16	20
Tree 7	13	18



決定樹の結合 – ディシジョンフォレスト

Trees	Des.	Mis.
Tree 1	10	17
Tree 2	10	19
Tree 3	12	17
Tree 4	12	17
Tree 5	15	19
Tree 6	16	20
Tree 7	13	18



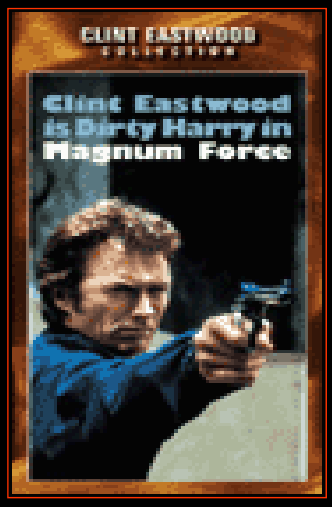
Decision Forest – Advantages

- There is always a certain degree of noise in the biological data (e.g. HTS data)
- Optimizing a model is at risk to over-fit the noise
- Each tree construction is a fitting process
- The combination scheme is not a fitting process
- The overfitted noise in each tree might be *cancelled out* in the combining process
- Results are reproducible
- Computational inexpensive
- Easy to be operated by the non-expert users

ディシジョンフォレスト – 長所

- 生物学的データに常にある程度のノイズがある
  （例：**HTS data)**
- モデルを効率的に利用することはノイズにオーバーフィットするため危険
- 各ツリーの組み立てはフィッティング過程である
- 結合の枠組みはフィッティング過程ではない
- 各ツリーのオーバーフィットさせたノイズは結合過程で帳消しになる
- 結果に再現性がある
- 計算的に費用が掛からない
- 専門家でなくても操作しやすい

“QSARs has to know its limitations”



"Man got to know his limitations" Dirty Harry - Magnum Force (circa 1973)

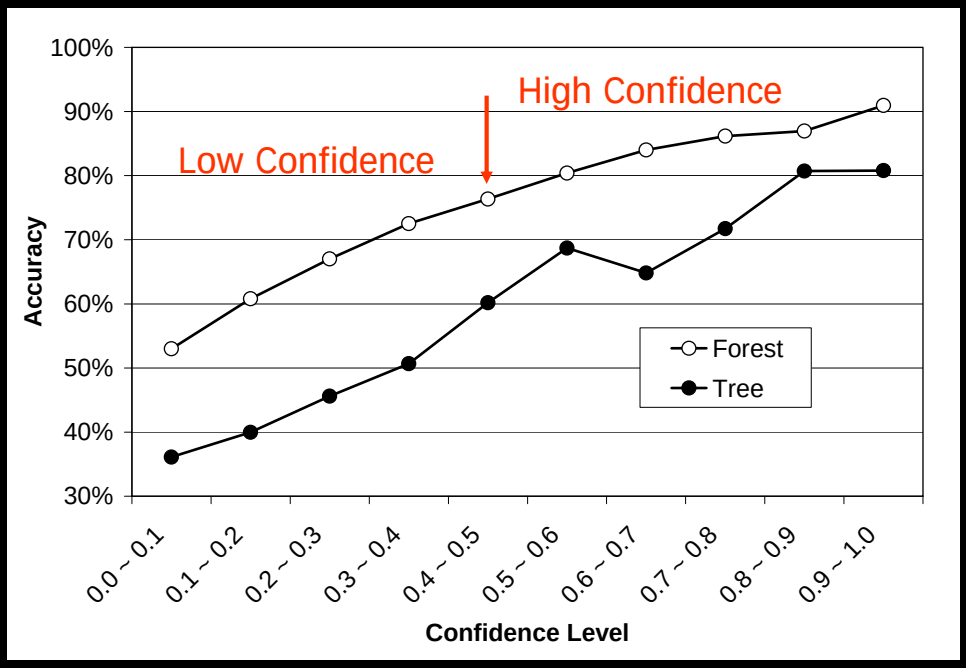
- *A QSAR model has its limitation (i.e., applicability domain)*
- The current challenge is not to develop a good fitted model
- IT IS how to assess the limitation and applicability domain of the model

“QSARsには限界があることを知るべきだ”

“人は自分の限界を知らなければならない” ダーティーハリー2 (1973年公開)

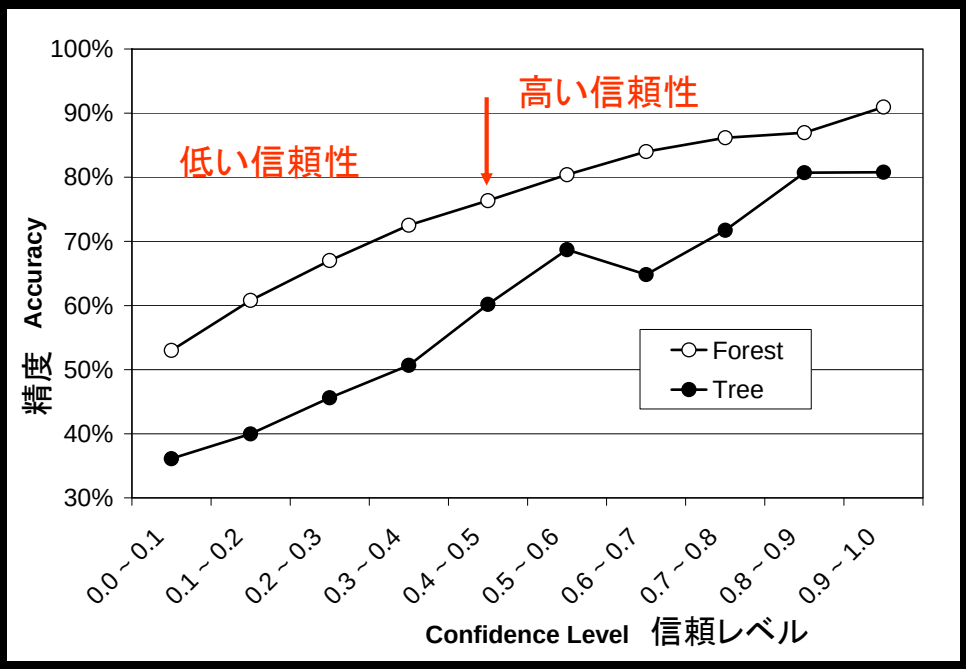
- *QSAR モデルには限界がある (例、適用範囲)*
- 現在の挑戦は良くフィットしたモデルを開発することではない
- どのようにモデルの限界や適用範囲を評価するか、だ。

Prediction Accuracy vs Confidence Level



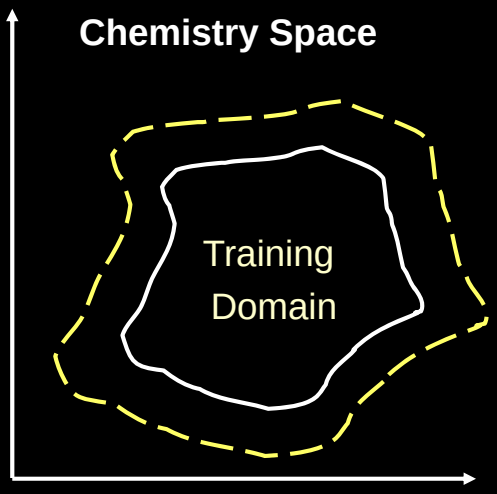
Confidence Level = $|P - 0.5| / 0.5$

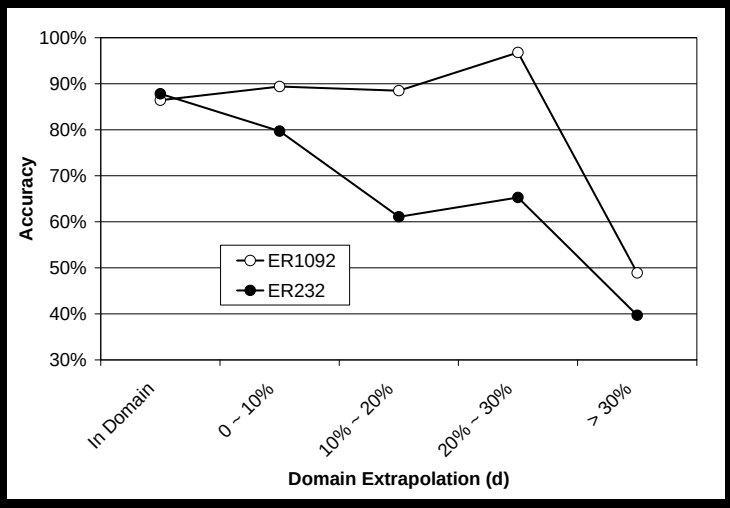
予測精度　vs 信頼レベル



Confidence Level = $|P - 0.5| / 0.5$

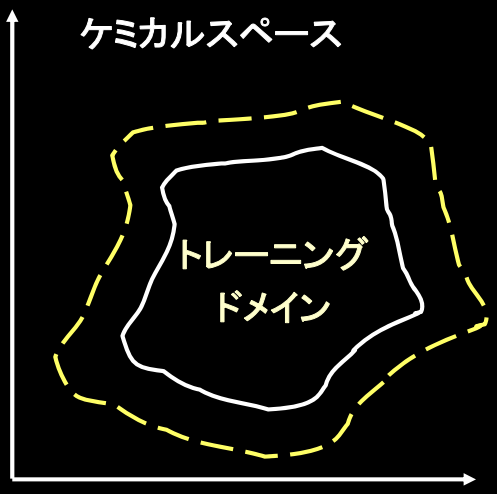
Domain Extrapolation

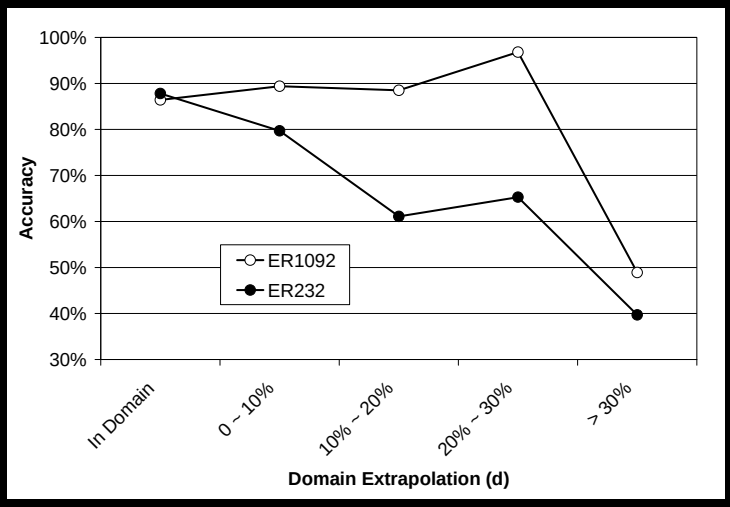




- The farther away from the training domain, the more loss in prediction accuracy
- The larger the dataset, the farther the prediction extrapolates

ドメインの外挿性





- トレーニングドメインから離れるほど、予測の精度は低下する
- データセットが大きいほど、予測の外挿性は高くなる

Summary – What's the Best Practice in Regulatory Application of QSARs?

- Commercial tools vs in-house tools:
  - Commercial tools: quick jump start, custom support, but no control of the tools (can't be GLP-like) and the company might go away
  - In-house tools: easy for version control (thus GLP-like), but require internal investment and mechanism to keep the talent
    - We are using in-house Decision Forest and MOLD2

- Need to define clearly how QSARs will be used in decision-making process
  - Focused on fit-for-purpose application, not one-fits-all
  - Understand the limitation of a QSAR model is more important than its accuracy, thus applicability domain is crucial
  - Combination of several models (consensus approach) is always perform better than a single model

要約 – 規制当局によるQSARs適用のベストプラクティスは？

- 市販ツール vs 組織内開発ツール:
  - 市販ツール: すぐ始められカスタムサポートがある、しかしツールは管理されてなく(GLP様ではない)、企業は消え失せてしまうかもしれない
  - 組織内ツール: バージョン管理が容易(GLP様)、しかしその性能を保つため内部投資や機構が必要である
    - 私達は組織内ツールであるDecision Forest and MOLD2を使用している

- 政策決定過程においてどのようにQSARsを活用するか明確に定義しておく必要がある
  - 全てに合うものではなく、目的に合った適用に焦点を絞る
  - QSARsモデルに限界があることを理解することがその精度よりも重要である。つまり適用範囲が極めて重要である。
  - 複数のモデルの結合(コンセンサスアプローチ)は単一モデルの適用より常に良い結果を出す