

# 計算毒性学における 人工知能の歴史と現状

株式会社 インシリコデータ  
湯田 浩太郎

<http://www.insilicodata.com>

# 人工知能に注目させたトピックス

## □ 成功事例

AlphaGo(アルファ碁)が人間に打ち勝って世界一になった  
人間がコンピュータに勝てる最後の分野の神話が崩れた



**碁の学習アルゴリズムに深層学習を適用していた**

## ■ 失敗事例

人工知能のチャットボット「Tay」がヘイト発言を乱発した



Twitter等の内容からヘイト発言を学習してしまった

# 人工知能の歴史

## 人工知能言語

1958:LISP(List Processor)

1972:Prolog

1994:Python

## ルールベース型人工知能

\* 第五世代コンピュータ(日本)

## 多変量解析/パターン認識

重回帰、パーセプトロン、PCA、  
クラスタリング、他

## 機械学習発展・新アプローチ

ニューラルネットワーク、  
遺伝的アルゴリズム、ファジィ、

## 深層学習開発／展開

新世代  
人工知能

## 過古の多変量解析/パターン認識と人工知能との関係

多変量解析/パターン認識と人工知能は  
互いに独立している

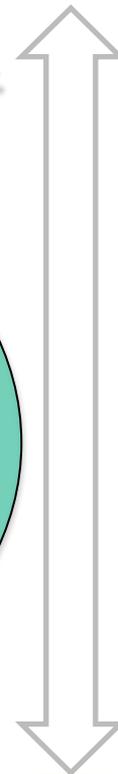
### 多変量解析/パターン認識

様々な  
データ解析手法

クラス分類  
フィッティング  
マッピング  
クラスタリング  
他

### 人工知能

既存のルール  
ルールベース型



# 現在における多変量解析/パターン認識と人工知能との関係

多変量解析/パターン認識と人工知能は  
機械学習により繋がっている

## 多変量解析/パターン認識

## 人工知能

様々な  
データ解析手法

クラス分類  
フィッティング  
マッピング  
クラスタリング  
他

既存のルール  
(特化型、汎用型)

発見された  
新ルール

機械学習

ビッグデータ

深層学習  
(DNN, CNN, RNN)

# 現在における多変量解析/パターン認識と人工知能との関係

多変量解析/パターン認識と人工知能は  
機械学習により繋がっている

## 多変量解析/パターン認識

## 人工知能

様々な  
データ解析手法

クラス分類  
フィッティング  
マッピング  
クラスタリング  
他

既存のルール  
(特化型、汎用型)

機械学習

発見された新  
ルール

ビッグデータ

深層学習  
(DNN, CNN, RNN)

# 計算毒性学(Computational Toxicology)と人工知能

# 計算毒性学(安全性評価)への 二大適用手法

化学多変量解析  
パターン認識

人工知能

化合物安全性評価

# 世界最初のシステム

化学多変量解析/パターン認識



人工知能



# 化学分野での人工知能システム

化学合成研究を人工知能で置き換えようとしたシステム

**LHASA**

有機合成設計支援システム



化学(化合物)

## DENDRAL および LHASAシステムの構成技術の比較

比較項目	DENDRAL	LHASA
基本形式	←ルールベース型→	
元データ形式	機器スペクトルデータ	化合物構造式
適用言語	LISP	FORTRAN
展開分野	人工知能分野	化合物合成分野
関連技術	ルール開発等	化合物表示／数値化等

## システム開発でのチームリーダー:

**DENDRAL** (**D**endritic **A**lgorithm)

→ **Edward A. Feigenbaum** → 人工知能の父

**LHASA** (**L**ogic and **H**euristics **A**ppplied to **S**ynthetic **A**nalysis)

→ **E. J. COREY** → 合成の秀才

→ ノーベル化学賞受賞(1990年)

授賞理由: 新たな合成概念(逆合成)導入にて

参考: Robert B. Woodward → 合成の天才(芸術家)

→ ノーベル化学賞受賞(極めて複雑な天然化合物の合成)

**注意!** 圧縮/解凍ソフトとは全く異なる

# 化学分野の人工知能の歴史と種類

## LISP および FORTRAN のプログラムコード事例

LISP(人工知能言語)  
DENDRALにて採用

```
SCOPE AND EXTENT 47

(defun contorted-example (f g x)
  (if (= x 0)
      (funcall f)
      (block here
        (+ 5 (contorted-example g
                                #'(lambda ()
                                   (return-from here 4))
                                (- x 1))))))

Consider the call (contorted-example nil nil 2). This produces the result
4. During the course of execution, there are three calls on contorted-example,
interleaved with two establishments of blocks:

(contorted-example nil nil 2)

  (block here1 ...)

    (contorted-example nil #'(lambda () (return-from here1 4)) 1)

      (block here2 ...)

        (contorted-example #'(lambda () (return-from here1 4))
                           #'(lambda () (return-from here2 4))
                           0)

          (funcall f)
            where f => #'(lambda () (return-from here1 4))

        (return-from here1 4)

      (return-from here2 4)

    (return-from here1 4)

  (return-from here 4)

(return-from here 4)
```

FORTRAN  
(科学用言語)  
LHASAにて採用

```
1 module simpleOps_m
2 contains
3   at tributes(global) subroutine increment(a, b)
4     implicit none
5     integer, intent(inout) :: a(:)
6     integer, value :: b
7     integer :: i, n
8
9     i = blockDim%x*(blockIdx%x-1) + threadIdx%x
10    n = size(a)
11    if (i <= n) a(i) = a(i)+b
12
13  end subroutine increment
14 end module simpleOps_m
15
16
17 program incrementTest
18 use cudafor
19 use simpleOps_m
20 implicit none
21 integer, parameter :: n = 1024*1024
22 integer, allocatable :: a(:)
23 integer, device, allocatable :: a_d(:)
24 integer :: b, tPB = 256
25
26 allocate(a(n), a_d(n))
27 a = 1
28 b = 3
29
```

## その後のDENDRALおよびLHASA

**DENDRAL**は  
metaDENDRAL等の関  
連システムに展開され  
た。  
→役割果たして終了

**LHASA**は合成研究支援システム  
として完了した。  
その後、合成研究の学習システム  
として展開されており、EUに渡っ  
たLHASAは適用分野を毒性評価  
に変え、名前も**DEREK**として最  
前線で活躍中である。  
化合物の情報学的操作技術も多  
数開発された。

# 現在までに展開された毒性予測関連システム

**ADAPT:** Automated **D**ata **A**nalysis by **P**attern recognition **T**echniques

**TOPKAT:** **T**Oxicity **P**rediction by **K**omputer **A**ssisted **T**echnology

**MULTICASE:** **M**ULTIple **C**omputer **A**utomated **S**tructure **E**valuation

**DEREK:** **D**eductive **E**stimation of **R**isk from **E**xisting **K**nowledge

**HazardExpert:**

**RIPT:** **R**ule **I**nduction for **P**redictive **T**oxicology

**COMPACT:**

Computer-**O**ptimized **M**olecular **P**arametric **A**nalysis of **C**hemical **T**oxicity

**Toxtree:**

**TOX-MATCH:**

**DART:** Decision **A**nalysis by **R**anking **T**echniques

**And Many Other Systems**

**ADMEWORKS, Chemotargets CLARITY, OFF-X,  
Derek, ChemTunes & ToxGPS, ACD/Percepta,**

**TIBCO Spotfire, ModelBuilder**

## ◆代謝予測関連支援システム(人工知能)

---

**COMPACT, MetaSite, SPORCalc,  
XENO, MetabolExpert, META, Meteor**

## ◆生分解性予測関連支援システム(人工知能)

---

**BESS, CATABOL, UMBBD,  
PPS, META**

# 化学分野への人工知能適用上での留意点

# 化学分野特有の問題

## 化合物構造式に始まり、化合物構造式に終わる

化学研究者の思考過程は化合物構造式で考え、  
相互コミュニケーションし、化合物構造式で答える。



人工知能システムが利用者である研究者と、  
化合物構造式で対話できることが必要

例：創薬研究者

薬理活性を強くするには、化合物構造式のどの部分を  
どのように変化させればいいのか？⇒研究者との対話必要

チェス、将棋、碁のように、盤上の座標を指定するようにはゆかない  
勝つだけで良いというわけでもない

# 化合物や安全性(毒性)データの取り扱い技術

## ◇ 化合物関連操作技術

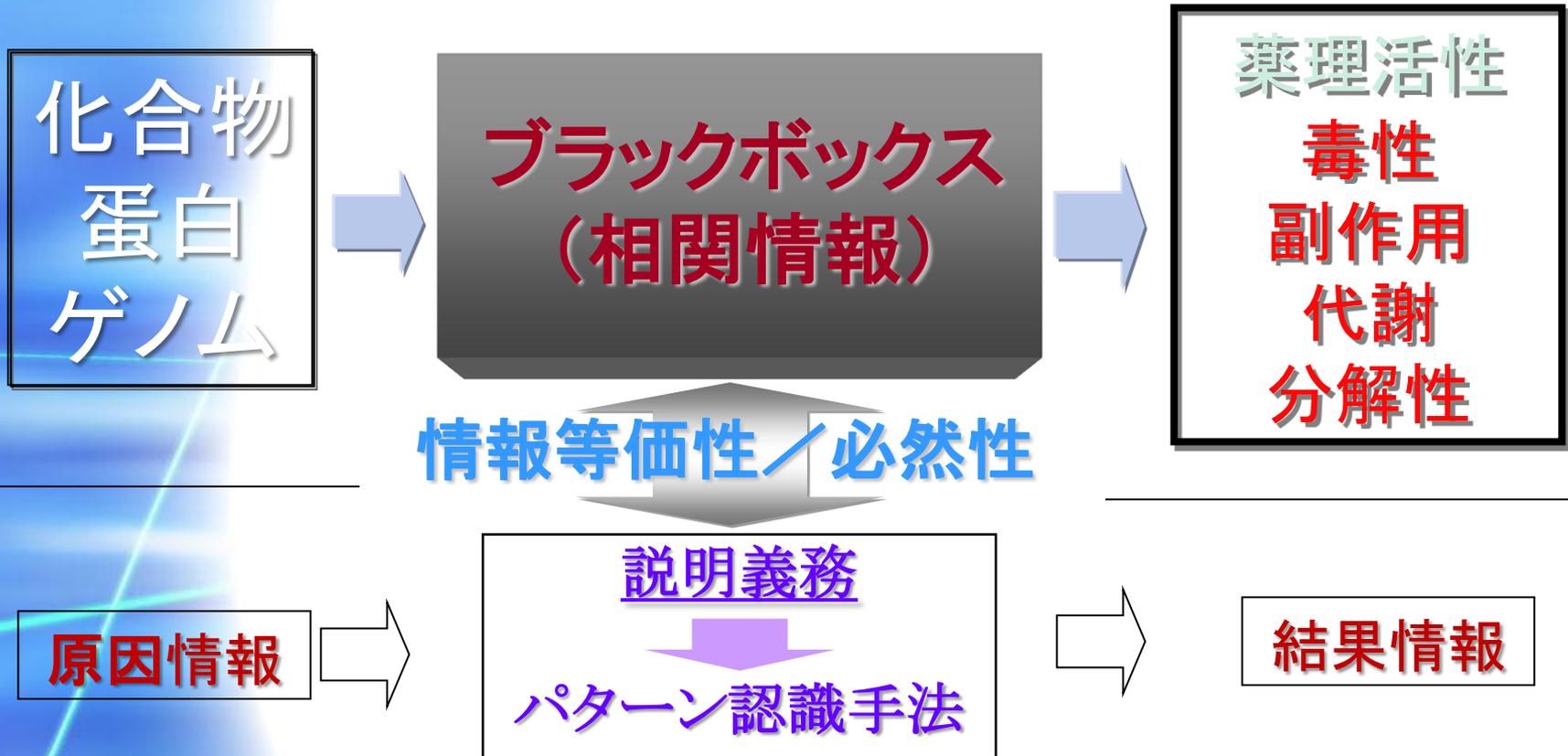
- ・1/2/3次元化合物構造式関連
- ・2次元→3次元化合物構造式変換技術
- ・化合物検索/変換関連技術
- ・幾何/立体異性等の取り扱い技術
- ・化合物及び数値データ扱いファイル関連

## ◇ 安全性(毒性)データ関連

- ・実験の種類を選択、決定
- ・実験プロトコルの吟味(実験条件、実験動物種、他)
- ・実験データの統一や標準化
- ・データ収集時のデータの吟味(解析目的、条件の統一、他)

# 多変量解析やパターン認識適用の基本原則

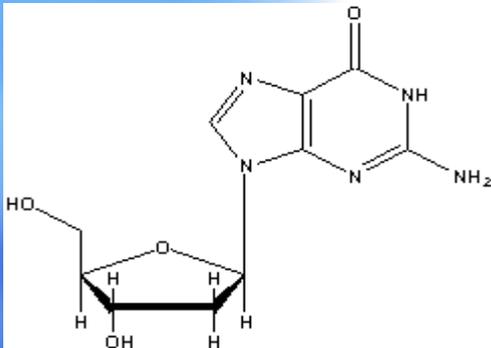
## 情報等価原理



# 多変量解析やパターン認識適用の流れ(1)

(化合物構造式と目的活性データを用いた予測)

## 化合物構造式



ブラックボックス  
(相関情報)

薬理活性  
毒性  
副作用  
代謝  
分解性

種々パラメータ発生

初期パラメータ群  
分子量  
原子/結合数  
HOMO/LUMO  
その他

特徴抽出

情報の等価性/必然性

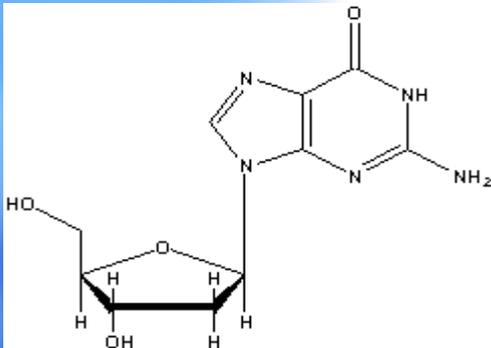
最終  
パラメータ群

種々の解析手法  
・判別分析  
・クラスタリング  
・フィッティング  
・その他

# 多変量解析やパターン認識適用の流れ(1)

(化合物構造式と目的活性データを用いた予測)

化合物構造式



薬理活性  
毒性  
副作用  
代謝  
分解性

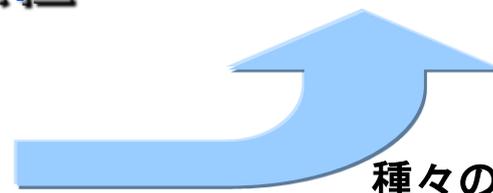
種々パラメータ発生

初期パラメータ群  
分子量  
原子/結合数  
HOMO/LUMO  
その他

特徴抽出

情報の等価性/必然性

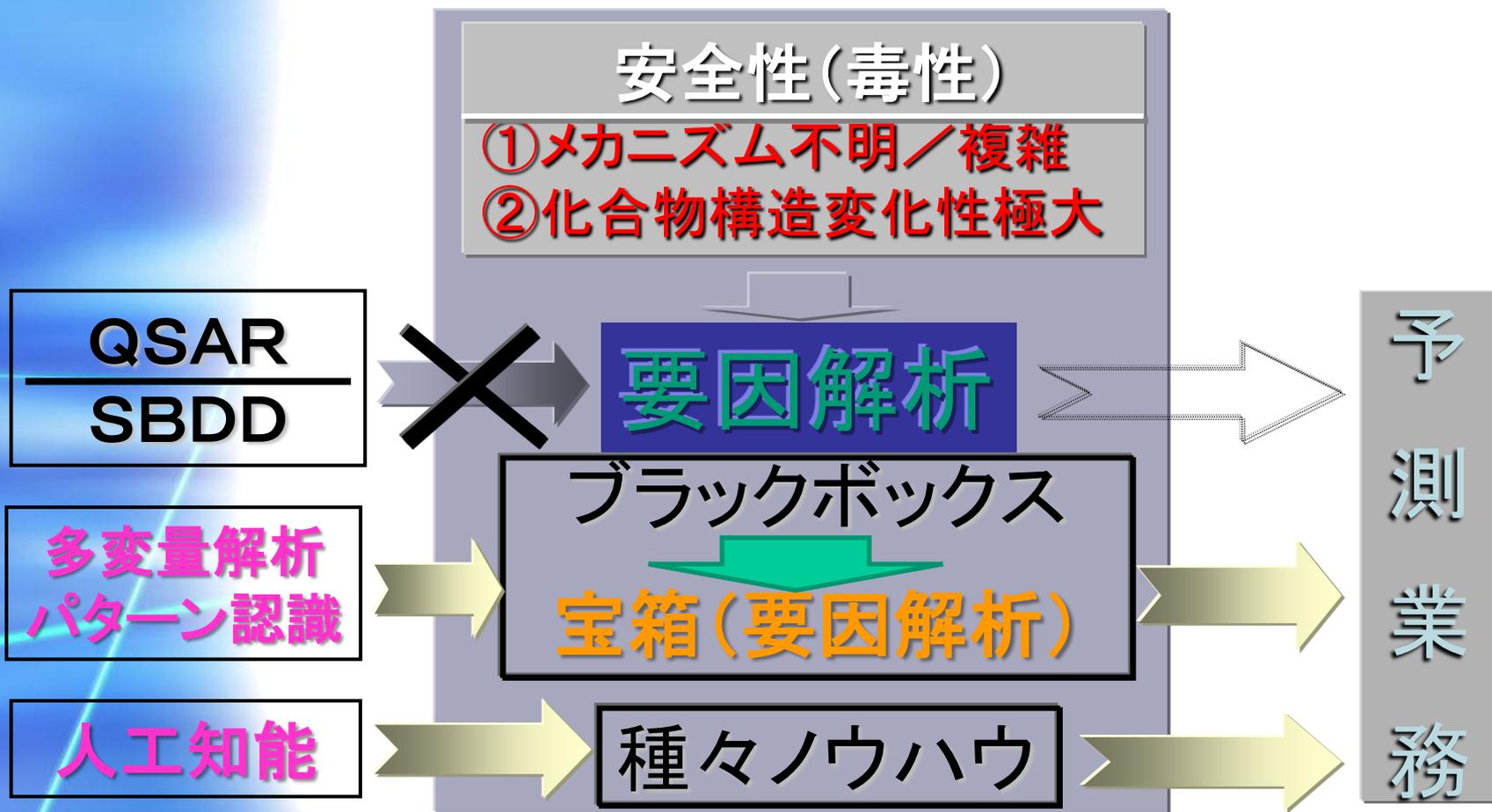
最終  
パラメータ群



種々の解析手法  
・判別分析  
・クラスタリング  
・フィッティング  
・その他

# 化合物毒性評価分野へのインシリコ技術の適用

## コンピュータによる安全性予測へのアプローチ



## 化学分野、特に安全性、代謝、分解性や合成、創薬等の 多くの分野にて人工知能システムが多い理由

- 安全性、代謝、分解性等は法則やメカニズム等が明確にならない事が多く、法則化や数式化が困難で、基本的に論理的なアプローチを取りにくい
- 研究者が積み重ねた経験や観察に基づいた規則的なものが比較的多数存在する
- いわゆる、その道のプロが有するノウハウの活用が極めて有効と考えられ、人工知能への期待が大きくなった

# 化学分野で現在展開されている人工知能システム

□歴史的に化学関連分野への人工知能適用の歴史は長い  
化学分野では数式に乗らない事項が多く、経験則が重要となることが多い⇒人工知能が活躍する地盤がある

□適用事例は多い

- ・機器スペクトルデータの解析支援システム
- ・有機合成支援システム
- ・毒性予測システム
- ・構造-活性相関支援システム
- ・創薬化学者支援システム
- ・その他

---

従来より展開されてきた化学分野の人工知能システムは、その展開上化学的なノウハウや考え方等のアナログ的な内容を、デジタルに変換する事が必要

# ルールベース型人工知能および 化学多変量解析/パターン認識支援システム開発事例

# □過去に開発した人工知能システム

第五世代コンピュータ時代(日本)に開発に関与したシステム

- **EMIL(京都大学);創薬支援システム**  
創薬研究者の化合物変換ノウハウをルールとして用いる
- **CASINO(産総研);有機合成設計支援システム**  
スーパーコンピュータを用いて反応の組み合わせを計算
- **LogP推算システム(北里大学);化学者のノウハウを適用**  
LogP計算するパラメータにノウハウを展開
- **毒性予測システム(CELI);WEB上での毒性予測**  
化学者のノウハウと多変量解析のハイブリッドシステム

## □過去に開発したケモメトリックス支援システム

インシリコデータの湯田は過去に以下のシステム開発に関与

### \* **ADAPT** : ペンシルバニア州立大学のJurs教授開発

(**A**utomated **D**ata **A**nalysis by **P**attern recognition **T**echniques)

世界初のケモメトリックス研究支援システム

米国ミニコン上で稼働⇒国内汎用機に移植

### \* **Model Builder** : (株)富士通九州システムズ開発に関与

・上記ADAPTシステムのGUIを強化開発し、

PC上に移植した

・新たに開発した最新の開発手法であるKY法等が  
組み込まれている

(**E**xample - **M**ediated - **I**nnovation-for - **L**ead-Evolution)

## EMILとは:

- ・京都大学教授(現名誉教授)発案による、ルールベース型人工知能システム

## 実施目的:

- ・メディシナルケミストが有する化合物変換ノウハウを 用いて  
目的薬理活性向上を効率的に実現する化合物構造式変換を行う  
人工知能システム
- \* 変更された化合物は「バイオアナログウス(生物学的等価)」化合物と  
命名された

## 変換ルールの取り出し:

- ・製薬および農薬関連企業における最先端のメディシナルケミストが召集され、  
変換ルール取り出しチームを結成して実施

## 開発言語(当初):ハイブリッド形態

- 化合物関連操作等 ⇒ FORTRAN(科学関連分野言語)
- ルール稼働用言語 ⇒ LISP(人工知能関連分野言語)

## 開発目的:

### 創薬研究者が有する化合物構造式の変換ノウハウ活用

- ・創薬研究者(特にメディシナルケミスト)は薬理活性の最適化を目指して化合物構造式のデザインが必要
- ・化合物構造式のデザイン可能性はほぼ無限にある
- ・開発目的に従った結果となるようにデザインしたい
- ・薬理活性のみならず、毒性や副作用回避も必要
- ・QSARやドッキングが適用困難な場合も多い

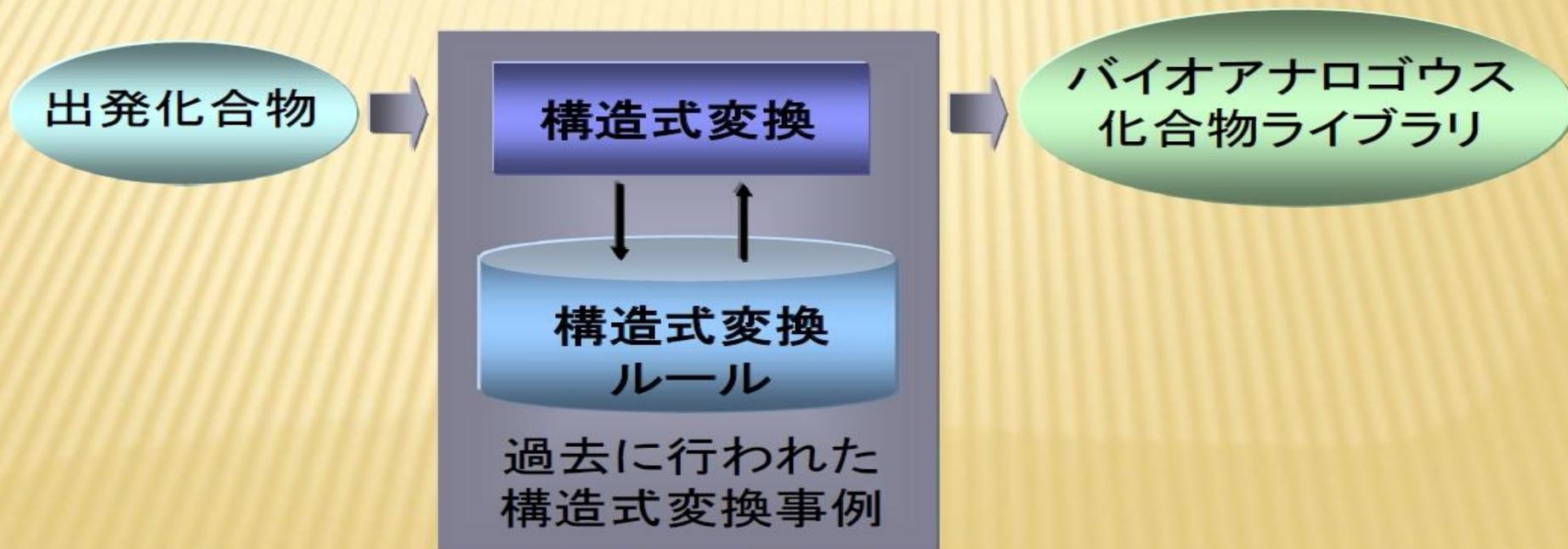
□プロの創薬研究者は、過去の豊富なデザイン経験より、効率的なデザイン手法についての様々な知識を持つ

貴重なノウハウ:活用したい



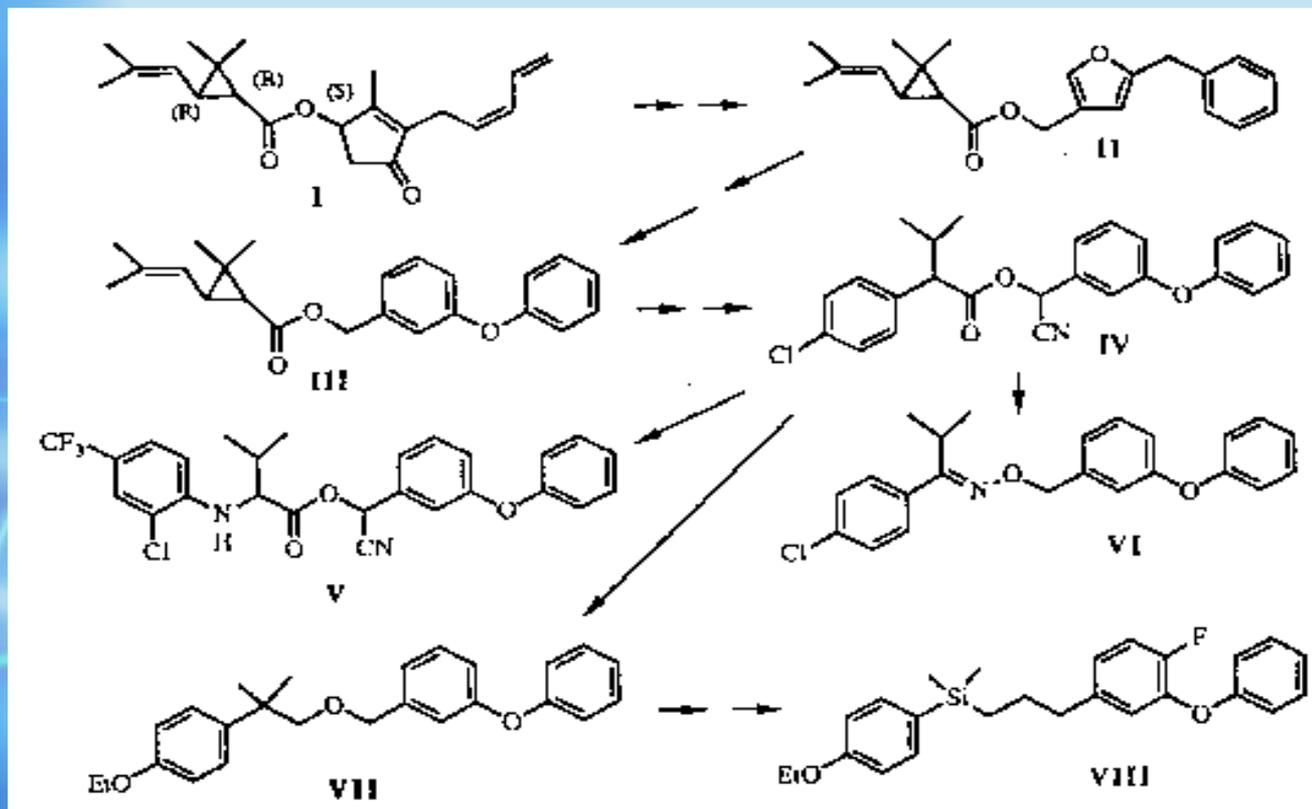
## EMILシステム(人工知能システム)の機能と流れ

- 過去に行われた医薬品や農薬等の構造変換事例に従って、新たにバイオアナログウス化合物群を創出するプログラム



(Example - Mediated - Innovation-for - Lead-Evolution)

## 化合物構造のバイオアナログス変換事例

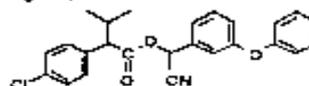
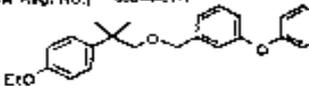
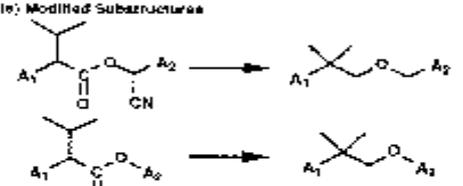


[http://cicsj.chemistry.or.jp/14\\_1/14\\_1emil.html](http://cicsj.chemistry.or.jp/14_1/14_1emil.html)より引用

(Example - Mediated - Innovation-for - Lead-Evolution)

化合物構造変換シート  
(創薬用と農薬用の二種類あり)

**EMIL DATA SHEET**

Structure A	Med. Para. Effects	Use Effect	Insecticide Name	Date	Ser. No.	Reg. No.	Prepared by	Structure B	Med. Para. Effects	Use Effect	Insecticide Name																																																	
<p><b>Name (General Name)</b> Permethrin (Code No.) S-5602 (CA Reg. No.) 51636-56-1</p>  <p>(Institution) Sumitomo Chemical</p>			Pyrethroids	9/1/2005		29	Haruhide Yoshitaka	<p><b>Name (General Name)</b> Deltamethrin (Code No.) MET500 (CA Reg. No.) 80844-07-1</p>  <p>(Institution) Nitto-Tokai Chemical</p>			Pyrethroids																																																	
<b>Structural Evolution</b>																																																												
<p>(b) Modified Substructures</p> 																																																												
<p>(1) Mode of Action - Prolongation of the Open State of Sodium Channels - Block of the Nerve Conduction.</p> <p>(2) Structural Requirements For the Acid Moiety: B and R, Form, G, H can be replaced by E and I, D, and J, G by F, R, and C (Cycloalkenyls at m- and p-positions)</p> <p>(3) Activity (Potency and Features) LUs Spectroscopic 0.21 mg/kg Infrared 0.17 mg/kg Mass 0.15 mg/kg Lethal dose 4.1 mg/kg LD50 (rat): 0.25 mg/kg (i.p.)</p> <p>(4) Physicochemical Data Vapor pressure: 2.4 x 10<sup>-10</sup> mmHg lgP (octanol): 6.12 at 25°</p> <p>(5) Others Commercialized first as the insecticide.</p>				<p>(b) Effects of the Modification</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Parameter</th> <th>Value</th> <th>Change</th> <th>Impact</th> </tr> </thead> <tbody> <tr> <td>Unsaturation</td> <td>1</td> <td>-</td> <td>+</td> </tr> <tr> <td>Electrofile</td> <td>1</td> <td>-</td> <td>+</td> </tr> <tr> <td>Activity Spectra</td> <td>0</td> <td>-</td> <td>+</td> </tr> <tr> <td>Potency</td> <td>+</td> <td>-</td> <td>+</td> </tr> <tr> <td>Taste</td> <td>+</td> <td>-</td> <td>+</td> </tr> <tr> <td>Stability</td> <td>+</td> <td>-</td> <td>+</td> </tr> </tbody> </table> <p>(c) Physicochemical Effects</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Property</th> <th>Value</th> <th>Change</th> <th>Impact</th> </tr> </thead> <tbody> <tr> <td>Stability</td> <td>+</td> <td>-</td> <td>+</td> </tr> <tr> <td>Hydrophobicity</td> <td>+</td> <td>-</td> <td>+</td> </tr> <tr> <td>Aq. Solubility</td> <td>-</td> <td>-</td> <td>+</td> </tr> <tr> <td>Biological Flexibility</td> <td>+</td> <td>-</td> <td>+</td> </tr> </tbody> </table> <p>(d) Pro- and/or Epimer(s) for the Modification</p>  <p>Instead of the ring B on between C1 and C1 for (invariant), that between C1 and C1, optimization of the C number and conversion of ester to other bonding types were used.</p>					Parameter	Value	Change	Impact	Unsaturation	1	-	+	Electrofile	1	-	+	Activity Spectra	0	-	+	Potency	+	-	+	Taste	+	-	+	Stability	+	-	+	Property	Value	Change	Impact	Stability	+	-	+	Hydrophobicity	+	-	+	Aq. Solubility	-	-	+	Biological Flexibility	+	-	+	<p>(1) Mode of Action - Prolongation of the Open State of Sodium Channels - Repetitive Excitation of the Nervous System</p> <p>(2) Structural Requirements General Me group and the connecting CH are essential. Introduction of CN at the m-positions in the (phenyl)ethyl is unfavorable.</p> <p>(3) Activity (Potency and Features) LUs Spectroscopic 0.15 mg/kg Infrared 0.21 mg/kg Mass 0.09 mg/kg LD50 (rat) &gt; 5000 mg/kg LD50 (mouse): 5 ppm (i.p.)</p> <p>(4) Physicochemical Data mp: 16.4-17.5° Vapor pressure: 1.4 x 10<sup>-10</sup> mmHg (100°) lgP (octanol): 6.87</p> <p>(5) Others Commercialized as a pyrethroid insecticide.</p>			
Parameter	Value	Change	Impact																																																									
Unsaturation	1	-	+																																																									
Electrofile	1	-	+																																																									
Activity Spectra	0	-	+																																																									
Potency	+	-	+																																																									
Taste	+	-	+																																																									
Stability	+	-	+																																																									
Property	Value	Change	Impact																																																									
Stability	+	-	+																																																									
Hydrophobicity	+	-	+																																																									
Aq. Solubility	-	-	+																																																									
Biological Flexibility	+	-	+																																																									

[http://cicsj.chemistry.or.jp/14\\_1/14\\_1emil.html](http://cicsj.chemistry.or.jp/14_1/14_1emil.html)より引用

(Example - Mediated - Innovation-for - Lead-Evolution)

## □ EMILシステムを用いたドラグリポジショニング

### 実施目的:

既存薬物(Tiaramide)誘導体の新規薬理活性の探索

### 手順と基本技術:

1. 探索する新規薬理活性を5種類選択する
2. 薬理活性ごとに判別関数作成(高速スクリーニング用)
3. Tiaramideの誘導体をEMILを用いて生成
4. 生成された誘導体化合物について薬理活性チェック

## インシリコ創薬への人工知能の部分適用事例

### マルチカテゴリースクリーニング実施事例

#### 5種類の薬理活性に関する同時スクリーニング

1. CCK-A (Cholecystinin-A) 受容体拮抗剤; Malonamide誘導体、急性膵炎治療薬
2.  $\text{Ca}^{2+}$  遊離抑制作用; 1,5-benzothiazepine誘導体、降圧薬、抗狭心症薬
3. 抗潰瘍作用; 4-Bis[(aminoalkyl)amino]-9,10-anthracenediones誘導体
4. 抗菌作用; Quinolone誘導体
5. 抗アレルギー作用; Imidazo誘導体

#### 薬理活性スクリーニング対象化合物群 (化合物ライブラリー)

人工知能システムEMILを用いて創出された、バイオアナログス化合物群

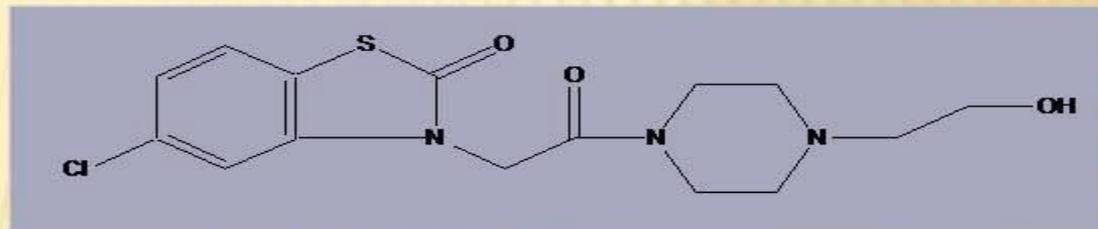
#### 薬理活性スクリーニング (多変量解析/パターン認識手法)

データ解析支援システムADAPTを用いて5種類の活性スクリーニングを行った

## EMIL実行上での初期化合物と創出化合物群

### ■ 出発化合物構造式

Tiaramide  
(Antiasthmatic, Anti-inflammatory)



### ■ 適用した構造式変換ルール

医薬品関連変換ルール

### ■ 変換結果

第一次変換出力構造式数      78化合物

第二次変換出力構造式数      76化合物

(第一次変換化合物中の1化合物を出発化合物とした)

## 化合物活性スクリーニング結果 (1)

1. CCK-A (Cholecystinin-A) 受容体拮抗剤; Malonamide誘導体、急性膵炎治療薬
2.  $Ca^{2+}$ 遊離抑制作用; 1,5-benzothiazepine誘導体、降圧薬、抗狭心症薬
3. 抗潰瘍作用; 4-Bis[(aminoalkyl)amino]-9,10-anthracenediones誘導体
4. 抗菌作用; Quinolone誘導体
5. 抗アレルギー作用; Imidazo誘導体

化合物ID	活性1	活性2	活性3	活性4	活性5	活性総数
1	×	○	○	○	○	4
2	×	○	○	○	○	4
3	○	×	×	×	×	1
4	×	×	○	○	○	3
5	×	×	○	○	○	3
6	×	×	○	○	○	3
7	×	×	×	○	○	1
8	○	×	○	○	×	3
9	○	×	○	○	○	4
10	×	×	○	○	○	3
⋮						⋮

## □ EMILシステムを用いたドラグリポジショニング

### 実験結果:

新規薬理活性可能性ありとされる化合物を  
効率よく探索できた

### 結果の考察:

1. 複数薬理活性候補化合物が頻度高く発見できた  
EMILにより創出された段階で薬理活性候補化合物が  
多数含まれた→EMILの構造変換ルールの効果を実証
2. 化学多変量解析／パターン認識(ケモメトリクス)により、  
5種類の薬理活性について効率よくチェック出来た
3. 全てバーチャル化合物を用いて短時間で実施された

# 今後の化合物関連人工知能システム構築上での留意点

# 化学分野特有の問題

## 化合物構造式に始まり、化合物構造式に終わる

化学研究者の思考過程は化合物構造式で考え、  
相互コミュニケーションし、化合物構造式で答える。



人工知能システムが利用者である研究者と、  
化合物構造式で対話できることが必要

例：創薬研究者

薬理活性を強くするには、化合物構造式のどの部分を  
どのように変化させればいいのか？⇒研究者との対話必要

チェス、将棋、碁のように、盤上の座標を指定するようにはゆかない  
また、勝つだけで良いというわけでもない

## □最近の人工知能は機械学習がメインである

### 利点:

- ・大量のデータを扱える
- ・従来は人工知能で展開出来なかった内容を展開できる
- ・ノウハウ(ルール)等を必要としない:データがあれば良い  
ノウハウがない分野での展開が可能となる
- 新たな知識を発見出来る可能性がある

### 欠点:問題点

- ・化学的な知見をシステムに理解させられるか?
- ・結果のフィードバックが手法的に困難(要因解析が困難)
- 新たな知見を人間が解釈できるレベルへの具象化が困難

## □化合物構造式に始まり、化合物構造式に終わる

・研究者の思考過程は総て化合物構造式で終始する

### ・化合物の表現の問題:

化合物名、分子式、二次元構造式、3次元構造式、等々  
同じ化合物が表現系により様々な形式を取り、それぞれの  
表現系が持つ情報の内容や情報量も異なる。



### ・入力の問題:

Journal や一般の化学文献が膨大な量あっても、単に  
文字情報を読み込ませただけで、化学情報を正確に人工  
知能システムに理解させることは困難

### ・結果の問題:

結果が出ても化学情報への変換が困難

## □大量データの効率的な使用が可能か

- ・データの形式や、情報内容が不ぞろいであり、機械学習で扱えるようにすることが重要
- ・量があっても、人工知能の実施目的に必要な情報を有する事が必要

## □化学構造式の正確な理解を機械学習で行う技術

- ・化学特有の様々な問題を、大量のデータから自動的に学習することには困難が予想される

創薬や安全性等の化合物を解析対象とする場合、機械学習のみならず、既存のノウハウ導入を行い、多変量解析/パターン認識(ケモメトリックス)技術との連携を念頭に、総合的なアプローチを考えるべき

# 人工知能システム構築上での留意点

## ■ 知識整理および適用型

ルールベース型人工知能

解決すべき問題点:

- ・ 目的解決に適したルール作成
- ・ ルール間の階層、衝突回避
- ・ エキスパートの存在必要

## □ 発見型および要因解析型

機械学習型人工知能  
ニューラルネットワーク  
深層学習

解決すべき問題点:

- ・ データ解析上の問題点  
過剰適合、偶然相関、クラス分布、欠損データ、他
- ・ 解析手法の特性／限界
- ・ 解析結果の解釈

# 人工知能トピックスに隠れた事実

## □ 成功事例

AlphaGo(アルファ碁)が人間に打ち勝って世界一になった  
人間がコンピュータに勝てる最後の分野の神話が崩れた



**碁の学習アルゴリズムに深層学習を適用していた**

## ○ 注意点: 過剰適合の回避

・学習用サンプルの数を極めて多数利用

\* ユニット数が多くなると過剰適合の回避困難

・ネットワークの解析が極めて困難となる

## ■ 失敗事例

人工知能のチャットボット「Tay」がヘイト発言を乱発した



Twitter等の内容からヘイト発言を学習してしまった

○ 注意点: 学習の偏りをなくす

・学習用サンプルやデータの品質・内容が重要

\* サンプルの偏りに注意する

・クラスポピュレーションに注意

・文献データは成功事例。失敗事例も必要

**Thank you for your attention**

株式会社 インシリコデータ  
湯田 浩太郎

<http://www.insilicodata.com>