

# ◇本日の講義スケジュール

## 序章 13:00~13:20

- ・自己紹介
- ・時代の流れと、研究環境および研究内容や研究スタイルの変化

## 第一章 13:20~13:40 (10分休憩)

化学データサイエンスおよび人工知能の適用イメージ

## 第二章 13:50~15:00 (10分休憩)

化学データサイエンスおよび人工知能の適用事例：様々なアプローチが可能

## 第三章 15:10~15:50 (10分休憩)

化学データサイエンスおよび人工知能の適用手順やパターン

## 第四章 16:00~16:50

化学データサイエンスおよび人工知能の適用上での留意事項

- ①化学分野の留意点
- ②データサイエンス実施上での留意点

## まとめと提案 16:50~17:00

- ①化学データサイエンスおよび人工知能のまとめ
- ②今後の展開についての「オートノマス創薬」の提案
  - ・KY法の展開（クラス分類と重回帰型）
- ③自由討論

# 化学データサイエンスおよび人工知能 適用事例や可能性

何ができるのか？ 事例を知って、実施分野／手法や自分の研究への適用を極める。

## 適用パターンや実施手法

### ○分野共通

#### \* 要因解析

- ・ 構造活性／毒性／物性相関解析

#### \* 予測・評価

- ・ 薬理活性／毒性／物性

### ○手法の展開

#### \* 新規化合物デザイン

- ・ 高活性化／脱毒性／高機能物性へのデザイン
- ・ リアルタイムデザイン

### ○何ができるか、したいのか？

#### \* スクリーニング

- ・ 化合物データベース検索
- ・ 仮想スクリーニング
- ・ ドラグリポジショニング

# □適用分野：化学及び化合物関連分野

◆創薬関連分野

◆スペクトル解析関連分野

◆安全性評価関連分野

◆法医学／法化学関連分野

◆物性関連分野

◆動物実験代替法関連分野

◆法規制関連分野

◆環境関連分野

◆その他の関連分野

# 適用項目

## ◆創薬関連

- ・薬理活性向上
- ・毒性低下／脱毒性
- ・スクリーニング
- ・適用外適用

## ◆物性関連

- ・高機能化
- ・光透過度
- ・種々耐久性強化

## ◆分析

- ・メタボローム解析
- ・スペクトル解析
- ・種々スペクトルの適用解析

## ◆その他

- ・遺伝子関連  
遺伝子解析、 SNP's 解析、  
発現プロファイル解析、その他
- ・医療関連解析  
画像処理による、種々診断等

ほぼ全ての  
分野での  
適用が実現  
されている

## ◆分解関連

- ・光分解
- ・酸化分解
- ・土中分解速度

## ◆耐性関連

- ・耐温関連（高温、低温）
- ・酸化耐性

## ◆機器分析データ関連

- ・法医学／法化学関連
- ・麻薬依存性の有無
- ・シーバスリーガル本物/偽物判定
- ・玄武岩の土産物の生産地判定

## ◆その他

- ・死後経過時間評価
- ・癌の術後再発予測

# 時代の変化によるインシリコ（コンピュータ）関連技術の導入と重要性の増大：時代は常に変化している

様々な外的要因により研究スタイルが変化、あるいは追加の作業が生じる機会が大きくなっている。

例（**化合物関連規制**）：化合物に関する様々な政府関連規制が導入されており、特にEUでの化合物関連規制は強力で、起業の様々な活動に大きな影響を及ぼしている。対応を誤る、あるいは対応が遅れる等すれば企業活動自体が出来なくなることとなる。

背景には、環境関連への関心の高まりと、動物愛護の精神の拡大等が大きく影響している。

今後の企業活動等において、これらの規制のクリアが必須であり、個々の国単位の規制をクリアしなければ当該国での企業活動は出来なくなる。特に国際的に活動している企業は、これらの規制のクリアが無ければ、撤退の憂き目を見ることとなる。

新たな化合物関連規制等は**インシリコ（コンピュータ）関連技術**が関与、基本、適用することが多くなっている。特に化学データサイエンスやAIの適用が多くなっている。この観点からも、化学データサイエンスやAI関連技術に対応出来るようにすることが、極めて重要となる。

**今後の時代的な化学関連周辺の環境変化は急速に進展しており、企業や研究所内部で同じ仕事や研究を5年、10年、20年と継続できる可能性は小さくなっている。**

## ◆安全性評価関連分野

### ICH-M7 潜在的発がんリスクを低減するための医薬品中DNA反応性（変異原性）不純物の評価及び管理

\* ICHとは、International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use（**医薬品規制調和国際会議**）

ICH-M7にて、**コンピュータによるAmes試験予測（データサイエンス手法とAI手法）**が、化合物の安全性評価の参考情報として利用可能というお墨付きが与えられた。

#### ICHの目的と役割

ICHの目的は、新医薬品を時宜に即し、また継続的に患者が利用できるようにすること、ヒトにおける不必要な臨床試験の重複を避けること、安全性、有効性及び品質の高い医薬品が効率的に開発、登録及び製造されること、及び安全性及び有効性が損なわれることなく動物試験が軽減されることに資する技術的要件における国際調和を促進することで公衆衛生を促進することです。

PMDAは、日本の医薬品規制当局代表の一員として参加しています。薬学・医学・獣医学・理学・生物統計学等の専門性を有するPMDA職員が、ICHの専門家作業部会に参加し、ガイドラインの作成・改訂を行うとともに、ICHでの合意形成に関わっています。

PMDAホームページより：<https://www.pmda.go.jp/int-activities/int-harmony/ich/0014.html>

## ◆動物実感代替法関連分野

EU では**化粧品等の開発過程で実施される動物実験を禁止**し、化粧品の政府申請に動物由来の実験データを使うことは禁止された。

従ってEUにて化粧品の製造販売する時は、従来の動物実験由来の実験データを使ったデータを使うことはできない。すなわち、EU圏での化粧品販売が出来なくなることを意味する。

## ◆化合物登録関連分野

REACH規則：EU 域内での化合物の開発／製造／販売を行う時は、**定められた安全性評価に関する基本データの登録**が義務づけられた。

これら安全性に関する登録が無ければEU圏での化合物関連活動が基本的にできなくなる。

Registration, Evaluation, Authorisation, Restriction and Chemicalsの略で、2007年6月1日からスタートした欧州の化学物質管理における法規制です。EU域内で製造・使用される化学物質はRegistration(登録)、Evaluation(評価)、Authorisation(認可)、Restriction(制限)の義務が課されることとなります。

<https://www.chemical-substance.com/reach/gaiyo.html>

# 適用分野・項目・事象・他

## ◆創薬関連

- ・薬理活性向上
- ・毒性低下／脱毒性
- ・スクリーニング
- ・適用外適用

## ◆物性関連

- ・高機能化
- ・光透過度
- ・種々耐久性強化

## ◆分析

- ・メタボローム解析
- ・スペクトル解析
- ・種々スペクトルの適用解析

## ◆その他

- ・遺伝子関連  
遺伝子解析、SNP's 解析、  
発現プロファイル解析、その他
- ・医療関連解析  
画像処理による、種々診断等

ほぼ全ての  
分野での  
適用が実現  
されている

## ◆分解関連

- ・光分解
- ・酸化分解
- ・土中分解速度

## ◆耐性関連

- ・耐温関連（高温、低温）
- ・酸化耐性

## ◆機器分析データ

- ・法医学／法化学関連
- ・麻薬依存性の有無
- ・シーバスリーガル本物/偽物判定
- ・玄武岩の土産物の生産地判定

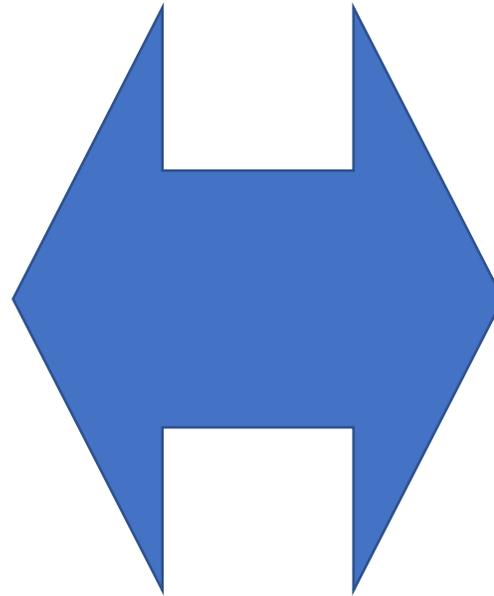
## ◆その他

- ・死後経過時間評価
- ・癌の術後再発予測

## 適用技術 (MV, PR, AI)

- 予測／評価
- 要因解析
- データベース連携
- マッピング
- クラスタリング
- 次元変換
- グラフ

- ◇ 動物実験代替法研究分野
- ◇ バイオ関連分野
- ◇ 医療関連分野



## 適用分野

- 毒性
- 薬物デザイン
- スクリーニング
- \*ドラグリポジショニング
- 機能性化合物デザイン
- 環境毒性
- 化合物規制

## □湯田による化学データサイエンス適用事例（創薬関連分野）

1. 種々要因による発がん性の変化（構造毒性相関、要因解析）
2. キノロン系抗菌性化合物の機能向上（リードリコンストラクション）  
・部分構造を利用した新規化合物設計
3. ドラグリポジショニングの適用事例（複数薬効の高速チェック）
4. 人工知能の適用とデータサイエンスによるドラグリポジショニング
5. リアルタイム創薬（実際に化合物を変化させながら対話的に創薬する）
6. インテグレートッド概念導入によるドラグデザイン
7. その他（メタボロミクスの実証実験：死後経過時間予測）

## 1. 種々要因による発がん性の変化

解析目的：発がん性は同一化合物であっても様々な要因で変化する  
要因の変化が発がん性に及ぼす要因をパターン認識  
による解析で明確にする

解析事例：

### 芳香族アミン化合物の発癌性予測

Kohtaro Yuta and Peter C. Jurs, *J.Med.Chem.*, **24**, 241 (1981)

## 要因解析を主体とした毒性解析

参考資料：ケモメトリックスによる発がん性評価（世界で3番目）

http://pubs.acs.org/doi/abs/10.1021/jm00135a003

Journal of Medicinal Chemistry

Computer-assisted structure-activity studies of chemical carcinogens. Aromatic amines

Kohtaro Yuta, Peter C. Jurs

*J. Med. Chem.* 1981, 24 (3), pp 241-251  
DOI: 10.1021/jm00135a003  
Publication Date: March 1981

ACS Legacy Archives

Note: In lieu of an abstract, this is the article's first page.

Articles

Computer-Assisted Structure-Activity Studies of Chemical Carcinogens. Aromatic Amines

Kohtaro Yuta and Peter C. Jurs\*

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802.  
Received August 18, 1980

Studies of molecular structure-carcinogenicity relations for a set of 157 aromatic amines are reported. A computer-assisted approach using pattern-recognition methods was used to develop a series of discriminants for aromatic amino carcinogenic potential. The 157 compounds were divided into subsets according to tumor site, route of administration, and activity. Sets of calculated molecular structure descriptors were generated that could support linear discriminant functions able to separate sets of active carcinogens from inactive compounds. Prominent among the important structural descriptors were those coding sizes and shapes of the amines. The pattern-recognition results were not strongly affected by differences in active site, and the study showed that mixed data sets could be used in computer-assisted structure-carcinogenicity studies.

Computer-Assisted Structure-Activity Studies of Chemical Carcinogens. Aromatic Amines

Kohtaro Yuta and Peter C. Jurs\*

Department of Chemistry, The Pennsylvania State University, University Park, Pennsylvania 16802.  
Received August 18, 1980

Studies of molecular structure-carcinogenicity relations for a set of 157 aromatic amines are reported. A computer-assisted approach using pattern-recognition methods was used to develop a series of discriminants for aromatic amino carcinogenic potential. The 157 compounds were divided into subsets according to tumor site, route of administration, and activity. Sets of calculated molecular structure descriptors were generated that could support linear discriminant functions able to separate sets of active carcinogens from inactive compounds. Prominent among the important structural descriptors were those coding sizes and shapes of the amines. The pattern-recognition results were not strongly affected by differences in active site, and the study showed that mixed data sets could be used in computer-assisted structure-carcinogenicity studies.

Aromatic amino compounds have been studied and a great deal of data accumulated over the years. The first report on the carcinogenicity of aromatic amines was by Rehn<sup>1</sup> in 1895 who described bladder cancers among employees in a Swiss dye factory. Many studies have followed, in part because aromatic amines are strongly related to industrialization and humans have been exposed to them widely. For example, aniline is employed world-wide in numerous industrial processes, fluorene compounds were originally developed as insecticides, and compounds such as butter yellow have been used as food additives. Primary aromatic amines are commonly used in industrial azo dye syntheses and as antioxidants in rubber products. Thus, the study of the carcinogenic potential of this class of compounds is of widespread interest.

A variety of hypotheses regarding the relationships between structure and activity of aromatic amines have been advanced. These compounds usually produce tumors at a site remote from the site of administration, including the liver, intestine, and bladder. The effects of aromatic amine exposure are species dependent. Activation through metabolic change is an essential step in aromatic amine carcinogenesis. The most prominent theory involves N-hydroxylation of aromatic amines.<sup>2-6</sup> Conversion to N-hydroxy compounds appears to be a requirement for carcinogenic activity. However, not all N-hydroxy compounds are carcinogenic. Although N-hydroxylation seems to be necessary for carcinogenic activity, it is not sufficient to explain the effect. The "para principle" involves the proposition that an amine should have a long, uninterrupted conjugated system with the amino group attached to one of the para carbons.<sup>7-9</sup> The presence of such a conjugated system lends stability to the ultimate reacting species. However, these theories do not explain all available biological data, and they are still under development. The carcinogenic activity of aromatic amines and some structure-activity relationships have been presented and discussed in several books and reviews.<sup>10-13</sup>

The present work involves the computer-assisted study of a set of 157 aromatic amines using chemical structure information handling and pattern-recognition methods to attempt to develop structure-activity relationships. The methodology of this approach to SAR studies has been described elsewhere and will be summarized here.

**Methodology**

The fundamental premises involved in applying pattern-recognition methods to SAR studies of chemicals with genetic toxicity are as follows: (1) Molecular structure and biological activity (genetic toxicity) are related. (2) The structures of compounds having genetic toxicity and compounds of similar structural classes that are nontoxic can be adequately represented by a set of molecular structure descriptors. (3) A relationship can be discovered between the structure and activity by applying statistical and/or pattern-recognition methods to a set of tested compounds. (4) The relation can be extrapolated to untested compounds to provide predictive ability.

The structure-activity studies were done using the ADAPT (automatic data analysis using pattern-recognition techniques) computer software system. This system has been developed over the past few years and has been described previously.<sup>14</sup>

The fundamental steps involved in performing an SAR study using the ADAPT system are as follows: (a) Identify, assemble, input, store, and describe a data set of structures for chemicals that have been tested for the biological activity of interest. (b) Develop computer-generated molecular structure descriptors for each of the members of the data set. The descriptors may be derived directly from the stored topological representations of

(1) L. Rehn, *Arch. Klin. Chir.*, **50**, 588 (1895).  
(2) J. L. Radomski and E. Brill, *Science*, **167**, 992 (1970).  
(3) J. L. Radomski and E. Brill, *Arch. Toxicol.*, **28**, 159 (1971).  
(4) G. M. Conzelmann, Jr., A. A. Rey, and E. Brill, *J. Natl. Cancer Inst.*, **50**, 989 (1973).  
(5) F. F. Kadlubar, J. A. Miller, and E. C. Miller, *Cancer Res.*, **37**, 805 (1977).  
(6) E. H. Weisburger, *Annu. Rev. Pharmacol. Toxicol.*, **18**, 395 (1978).  
(7) H. Druckrey, *Arzneim. Forsch.*, **2**, 503 (1952).  
(8) N. P. Bui-Hoi, *Arzneim. Forsch.*, **4**, 531 (1954).  
(9) J. C. Arcos and M. F. Argus, "Chemical Induction of Cancer", Volume 2B, Academic Press, New York, 1974.

(10) J. C. Arcos and M. Arcos, in "Progress in Drug Research", E. Jucker, Ed., Wiley-Interscience, New York, 1962.  
(11) W. C. Hoepfer and W. D. Conway, "Chemical Carcinogenesis and Cancers", Charles C. Thomas, Springfield, Ill., 1964.  
(12) D. B. Clayton and R. C. Garner, in "Chemical Carcinogens", C. W. Searle, Ed., American Chemical Society, Washington, D.C., 1976.  
(13) J. L. Radomski, *Annu. Rev. Pharmacol. Toxicol.*, **19**, 129 (1979).  
(14) A. J. Stuper, W. E. Brugger, and P. C. Jurs, "Computer Assisted Studies of Chemical Structure and Biological Function", Wiley-Interscience, New York, 1979.

0022-2623/81/1824-0241\$01.00/0 © 1981 American Chemical Society



## 「医薬品のインシリコによる毒性、副作用予測の現状」

湯田 浩太郎

CBI学会誌 2019年 第7巻 第2号

[http://cbi-society.org/home/documents/society\\_journal/CBI\\_2019\\_7\\_2.pdf](http://cbi-society.org/home/documents/society_journal/CBI_2019_7_2.pdf)

# 発癌性と解析目的、および解析上での留意点

同一化合物が実験動物、発ガン部位、投与法等の差により  
発ガン性に大きな差異がある

解析目的： 発がん医関する要因抽出を主目的とする

1. 発ガン部位や、投与手法の差異に影響されない  
発ガン要因の抽出 → 発ガンに極めて本質的な要因
2. 発ガン部位の差異や、投与手法の差異を反映した  
発ガン要因の抽出 → 特殊要因を反映した要因

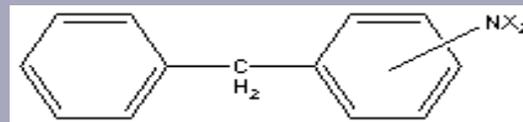
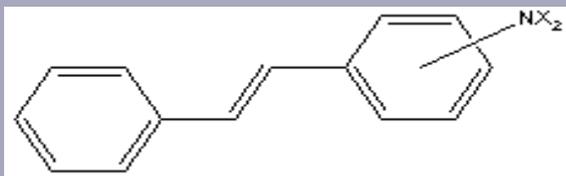
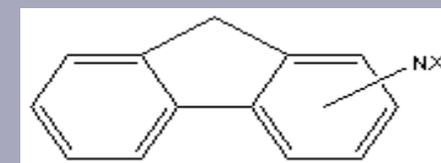
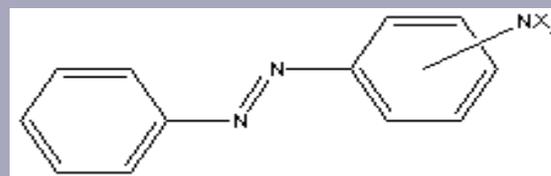
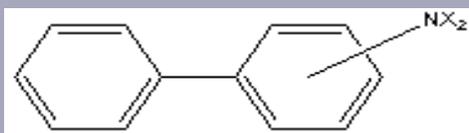
解析上での留意点：

1. 解析手法に起因する問題点をクリアする
  - ・ 細かな初期条件の差異による解析結果の変動のクリア
  - ・ サンプル数の変動による解析信頼性のクリア
2. その他の留意点
  - ・ 可能な限り総てのケースで比較的高い分類、予測率の達成
  - ・ 情報取りだしに有利なパラメータを優先する

# 解析に用いた芳香族アミン化合物データセット

## 1. 使用化合物群

総数：157化合物



others

## 2. 発癌性データ

実験動物； R a t

投与方法； oral, subcutaneous injection, topical,  
intraperitoneal injection, various routes, not available

発現部位； breast, ear duct, liver, other sites, all sites

# 使用データ例 (化合物 1 ~ 15)

compounds		breast	earduct	liver	other	all	mixed
1. 4-biphenylacetamide	po	+	+	-	+	+	+
2. 4-biphenyldimethylamine	po	+	+	-	+	+	+
3. 4-biphenylacethydroxamic acid	po	+	+	-	+	+	+
4. 2-fluoro-4-phenylaniline	po	+	-	-	-	-	+
5. 3'-fluoro-4-phenylaniline	po	+	-	-	-	-	+
6. 2-methyl-4-phenylaniline	po	-	+	-	-	+	+
7. 3-methyl-4-phenylaniline	po	-	-	-	-	-	-
8. 2'-methyl-4-phenylaniline	po	-	-	-	-	-	-
9. 4'-methyl-4-phenylacetanilide	po	-	-	-	-	-	-
10. 3,2'-dimethyl-4-biphenylamine	po	+	+	-	+	+	+
11. o,o'-dianisidine	po	-	na	-	-	-	-
12. 3,3'-dichlorobenzidine	po	+	+	-	+	+	+
13. 3,3'-dihydroxybenzidine	po	-	-	-	-	-	-
14. 2-methyldiacetylbenzidine	po	-	-	na	-	-	-
15. 4,4'-methylenedianiline	po	+	-	+	+	+	+
16. . . . .				. . . . .			

"Survey of Compounds Which Have Been Tested for Carcinogenic Activity", U.S. Department of Health, Education and Welfare, U.S. Public Health Service, National Institute of Health. D.B. Clayson and R.C. Garner, in "Chemical Carcinogenesis and Cancers", Charles C. Thomas, Springfield, Ill., 1964.

# 使用データセットのデータ構成（一部）

structure	active site administration	breast			earduct			liver		
		po	other	all	po	other	all	po	other	all
biphenyl	active	9	1	10	7	2	9	0	4	4
	not active	7	7	14	8	6	14	15	4	19
stilbene	active	4	0	4	7	4	11	1	4	5
	not active	6	11	17	1	6	7	9	7	16
azo	active	0	0	0	2	0	2	3	1	4
	not active	16	2	18	13	2	15	13	1	14
fluorene	active	18	10	28	16	4	20	15	3	18
	not active	10	4	14	10	8	18	13	11	24
methylene	active	1	0	1	0	0	0	3	0	3
	not active	3	0	3	4	0	4	0	0	0
misc	active	19	3	22	8	0	8	3	0	3
	not active	15	11	26	25	14	39	29	14	43
total		108	49	157	101	46	147	104	49	153

同じ化合物群でも、投与形体および発ガン部位により、発癌性に大きな差異が見られる。

# 使用システムと発癌性解析の流れ

芳香族アミン化合物群



初期パラメータセット



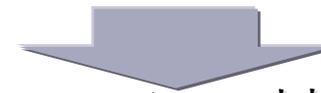
特徴抽出の実施



最終パラメータセット



分類／予測率の算出



パラメータの情報解析

**ADAPT**

(Automated **D**ata **A**nalysis by **P**attern recognition **T**oolkits)

# 初期パラメータセット

初期パラメータセット： 約200（種類／化合物）創出

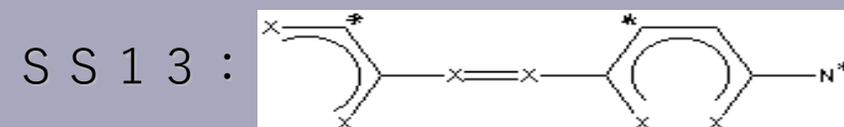
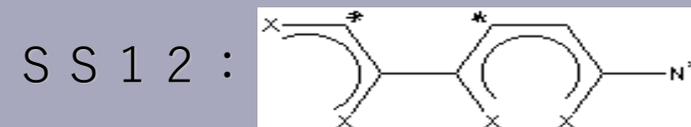
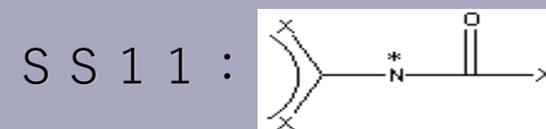
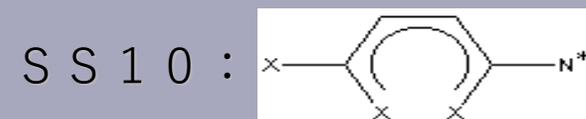
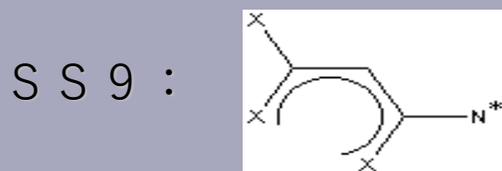
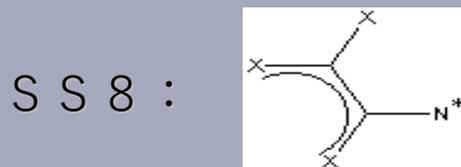
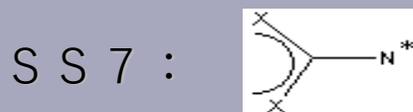
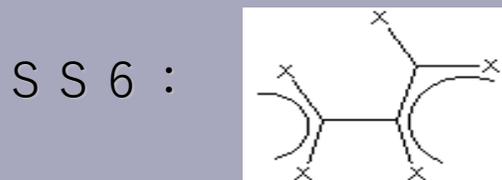
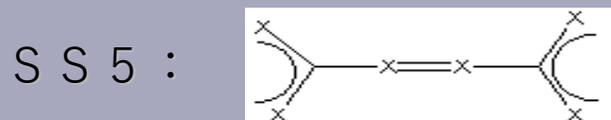
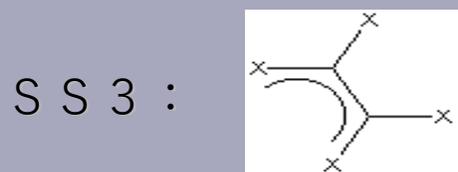
1. トポロジカルパラメータ（2次元構造情報）
  - ・分子パスパラメータ
  - ・分子結合パラメータ
2. トポグラフィカルパラメータ（3次元構造情報）
  - ・シャドウパラメータ
  - ・分子主成分パラメータ
3. 理論化学的パラメータ
  - ・分子軌道法関連パラメータ
  - ・分子力学関連パラメータ
4. 化学特性パラメータ
  - ・種々物性（分子屈折率、分子表面積、分子容、その他）
  - ・原子（総原子数、個々の原子数）、種々結合数、環数、その他
5. その他のパラメータ
  - ・演算パラメータ
  - ・部分構造パラメータ

# 解析初期パラメータセット

## 解析初期パラメータリスト (3次元データ)

- 
1. number of carbon atoms
  2. number of oxygen atoms
  3. number of nitrogen atoms
  4. number of single bonds
  5. number of double bonds
  6. number of aromatic bonds
  7. number of basis rings
  8. number of ring atoms
  9. path 3 molecular connectivity
  10. path 4 molecular connectivity
  11. total number of paths
  12. molecular connectivity environment: SS1
  13. molecular connectivity environment: SS2
  14. molecular connectivity environment: SS3
  15. molecular connectivity environment: SS4
  16. molecular connectivity environment: SS5
  17. molecular connectivity environment: SS6
  18. largest principal moment
  19. intermediate principal moment
  20. smallest principal moment
  21. ratio of largest to smallest principal moment
  22. ratio of intermediate to smallest principal moment
  23. molecular connectivity environment: SS7
  24. molecular connectivity environment: SS8
  25. molecular connectivity environment: SS9
  26. molecular connectivity environment: SS10
  27. molecular connectivity environment: SS11
  28. molecular connectivity environment: SS12
  29. molecular connectivity environment: SS13
  30. molecular volume
  31. number of F + Cl + I + S
-

# 使用部分構造リスト



既存仮説の検証目的

- 1. N-hydroxylation theory ; SS2, SS7
- 2. o-hydroxy theory ; SS3, SS8
- 3. co-planality of biphenyl structure ; SS6, SS12, SS13

# 発ガン部位別および投与方法別の最終パラメータセット

descr. no.	BO	EO	LO	OO	AO	B	E	L	O	A	M	mixed admin.	oral admin.	total
1			b					b	X		c	1	0	1
2				X	X	X	X	X	X	X	c	5	2	7
3		X	X	X	X	X	X	X	X	X		5	4	9
4			X	X		X	X	X		X	X	4	2	7
5			X				X	X	X	X		4	1	5
6		X	b	X	X		X		X	X	c	3	3	6
7	X	X	X		X	X	X	X	X	X	X	5	4	10
8	X	X	X	X		X	X	X	X			4	4	8
9	X	X	X			X	X	X	X	X		5	3	8
10				X				X	X	X	X	3	1	5
11	X		b	X		X	X		X	X		4	2	6
12	X		X		X		X	X	X		X	3	3	7
13	X	X		X	X			X	X	X	X	3	4	8
14	X	X	X	X	X	X		X	X			3	5	8
15		X	X		X		X	X	X	X		4	3	7
16			b	X		X		b			c	1	1	2
17	X			X		X		X	X	X		4	2	6
18		X		X	X	X	X	X	X		X	4	3	8
19	X		X	X			X	X	X	X	X	4	3	8
20	X		b	X	X	X	X		X	X		4	3	7
21	X	X	X	X		X	X	X		X	X	4	4	9
22	X		b	X	X	X				X	X	2	3	6
23	X	X	b	X	X	X	X			X		3	4	7
24	X		X	X		X	X	X	X	X		5	3	8
25			b	X	X			b			c	0	2	2
26	X	X	X		X			X	X	X	X	3	4	8
27			X	X	X	X		X	X		c	3	3	6
28		X	b				X		X		c	2	1	3
29	X	X	b		X		X		X	X	X	3	3	7
30	X	X			X	X		X				2	3	5
31	X	X	X		X	X		X		X	X	3	4	8
no. in final set	18	16	15	20	18	19	19	21	23	21	12			

## パラメータの出現パターン

descr. no.	mixed admin.	oral admin.	total
1	1	0	1
2	5	2	7
3	5	4	9
4	4	2	7
5	4	1	5
6	3	3	6
7	5	4	10
8	4	4	8
9	5	3	8
10	3	1	5
11	4	2	6
12	3	3	7
13	3	4	8
14	3	5	8
15	4	3	7
16	1	1	2
17	4	2	6
18	4	3	8
19	4	3	8
20	4	3	7
21	4	4	9
22	2	3	6
23	3	4	7
24	5	3	8
25	0	2	2
26	3	4	8
27	3	3	6
28	2	1	3
29	3	3	7
30	2	3	5
31	3	4	8

### パターン 1 : 総てに高い出現回数を持つ

[  $\cong 8$  ]      3, 7, 8, 9, 13, 14, 18,  
19, 21, 24, 26, 31

発ガン部位や投与方法の差異に影響されない、  
発ガンに本質的に重要な要因

### パターン 2 : グループ内で高い出現回数を持つ

[  $\cong 4$  ]

mixed admin. group; 2, 3, 4, 5, 7, 8, 9,  
11, 15, 17, 18, 19,  
20, 21, 24,  
oral admin. group; 3, 7, 8, 13, 14, 21, 23,  
26, 31

個々の投与形体に高い影響を及ぼす要因

# 最終パラメータセット(パターン1)

- 発ガン部位や投与法の差異に影響されない、  
**発ガンに本質的に重要な要因**

3. number of nitrogen atoms

7. number of basis rings

8. number of ring atoms

9. path 3 molecular connectivity

13. molecular connectivity environment: SS2

14. molecular connectivity environment: SS3

18. largest principal moment

19. intermediate principal moment

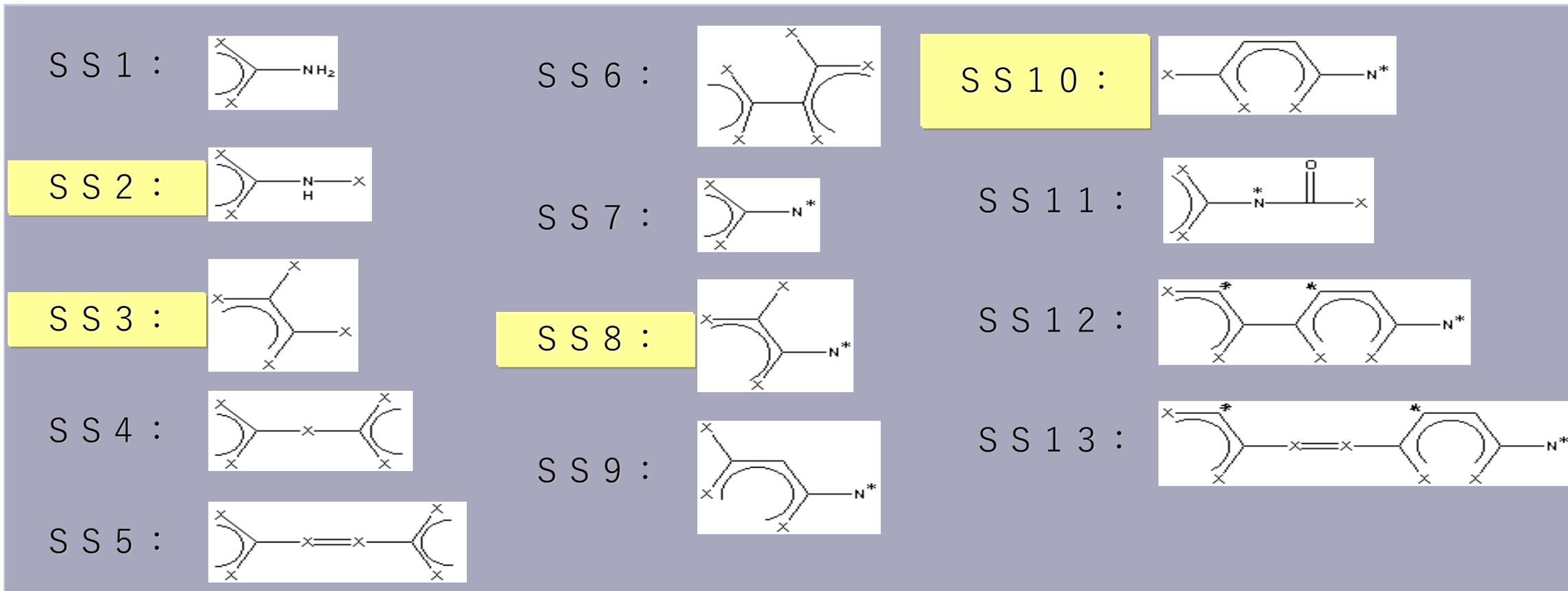
21. ratio of largest to smallest principal moment

24. molecular connectivity environment: SS8

26. molecular connectivity environment: SS10

31. number of F + Cl + I + S

# 使用部分構造リスト(パターン1)



既存仮説の検証目的 :

1. N-hydroxylation theory ; SS2, SS7
2. o-hydroxy theory ; SS3, SS8
3. co-planality of biphenyl structure ; SS6, SS12, SS13

# 最終パラメータセット(パターン2)

## グループ内で高い出現回数を持つ (投与形態の差異による影響の違い)

1. number of carbon atoms	16. molecular connectivity environment: SS5
2. number of oxygen atoms	17. molecular connectivity environment: SS6
3. number of nitrogen atoms	18. largest principal moment
4. number of single bonds	19. intermediate principal moment
5. number of double bonds	20. smallest principal moment
6. number of aromatic bonds	21. ratio of largest to smallest principal moment
7. number of basis rings	22. ratio of intermediate to smallest principal moment
8. number of ring atoms	23. molecular connectivity environment: SS7
9. path 3 molecular connectivity	24. molecular connectivity environment: SS8
10. path 4 molecular connectivity	25. molecular connectivity environment: SS9
11. total number of paths	26. molecular connectivity environment: SS10
12. molecular connectivity environment: SS1	27. molecular connectivity environment: SS11
13. molecular connectivity environment: SS2	28. molecular connectivity environment: SS12
14. molecular connectivity environment: SS3	29. molecular connectivity environment: SS13
15. molecular connectivity environment: SS4	30. molecular volume
	31. number of F + Cl + I + S

 :Mixed admin. group

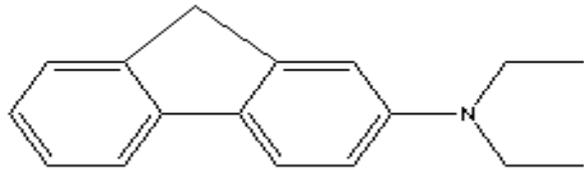
 : Oral administration group

## 最終パラメータセットを用いた分類率

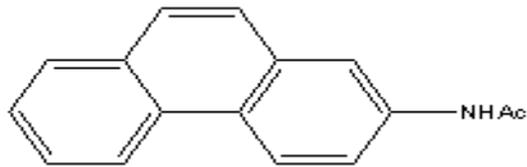
classifier	data set	% correctly classified		
		active	not active	total
Bayes (quadratic)	A	90.9	87.0	88.6
	B	96.7	93.3	94.7
Bayes (linear)	A	75.8	84.8	81.0
	B	83.3	91.1	88.0
KNN (k=1)	A	66.7	73.9	70.9
	B	73.3	82.2	78.7
KNN (k=3)	A	66.7	82.6	76.0
	B	73.3	86.7	81.3
iterative least square	A	84.9	97.8	92.4
	B	96.7	97.8	97.3
simplex	A	81.8	89.0	84.8
	B	93.3	93.3	93.3
linear learning machine	A	100.0	100.0	100.0
	B	100.0	100.0	100.0

set A: original data set    set B: pruned data set

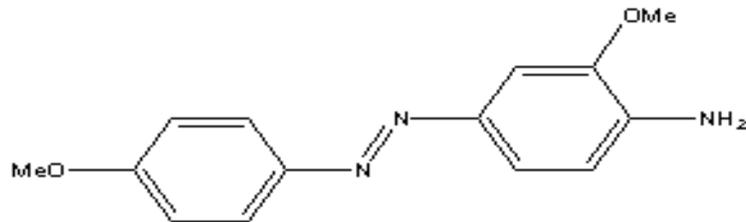
# 解析中に取り除かれた化合物群



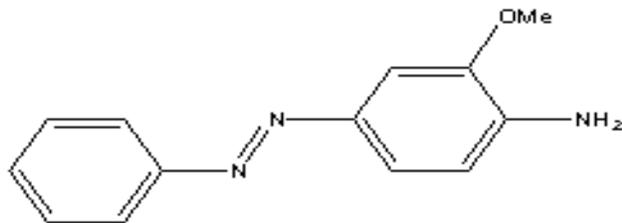
2-Fluorenyldiethylamine (4)



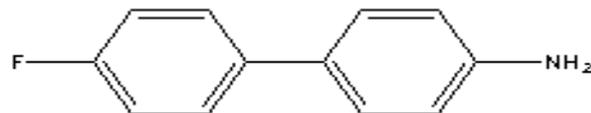
2-Phenanthrylacetamide (11)



4'[(p-Methoxyphenyl)azo]-o-anisidine (4)



4'-Fluoro-p-phenylaniline (4)



4-(Phenylazo)-o-anisidine (6)

# 人工知能の適用とデータサイエンスによる 「ドラグリポジショニング」

□本研究に適用された人工知能（EMIL）は、ルールベース型の人工知能である

創薬研究者が有する

化合物構造式の変換（バイオアナログウス）

ノウハウ活用

## 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）

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

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

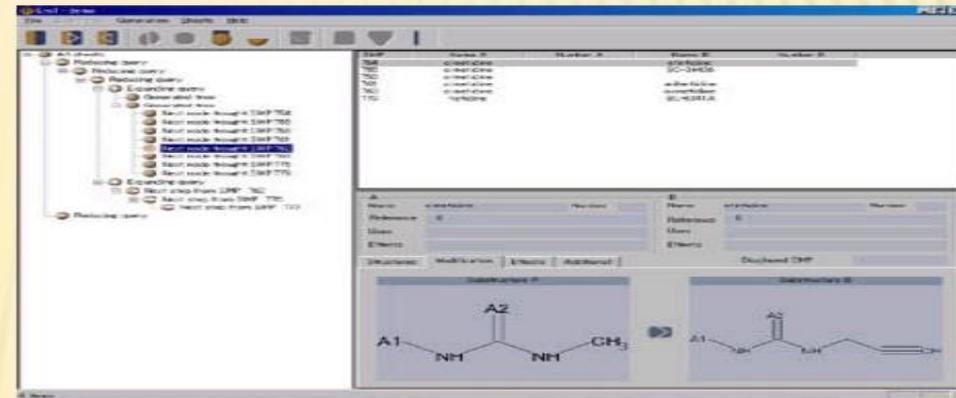
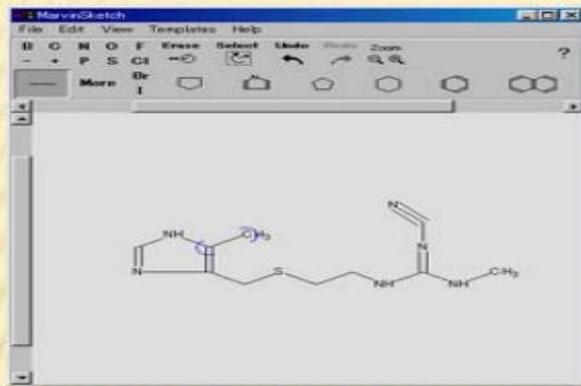
貴重なノウハウ：活用したい

# EMIL

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

## EMILシステムの流れ図

In **Silico** Data  
Miracles by the KY-methods



過去の構造変換ルールの適用  
医薬品／農薬

バイオアナログウス  
化合物群

化合物1

化合物2

化合物3

化合物4

化合物1-1

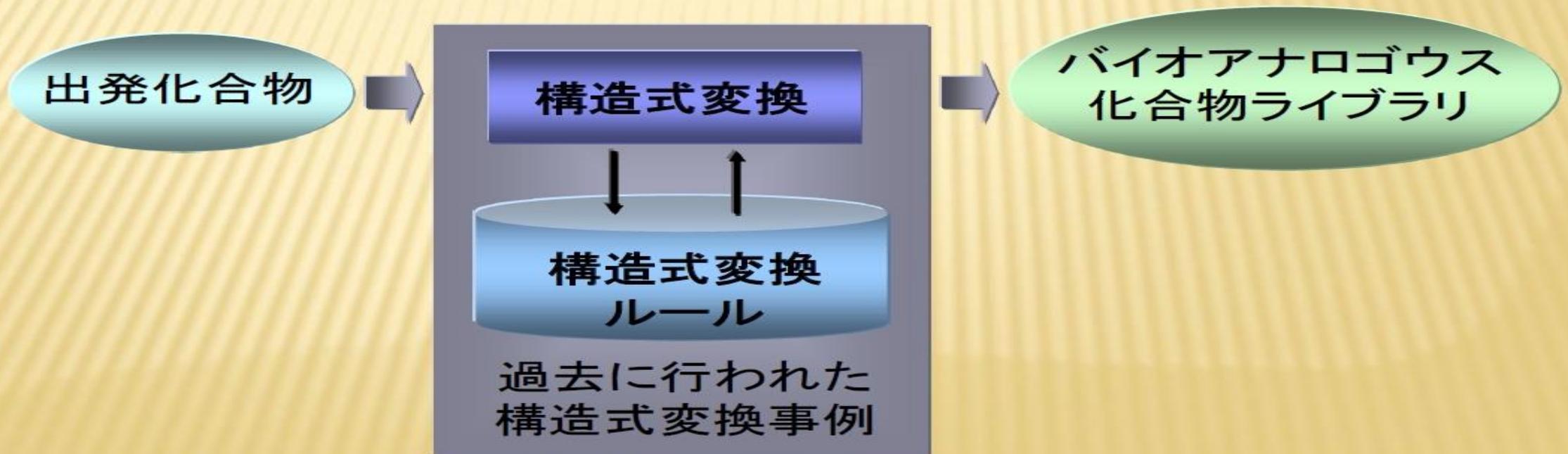
化合物1-2

# EMIL

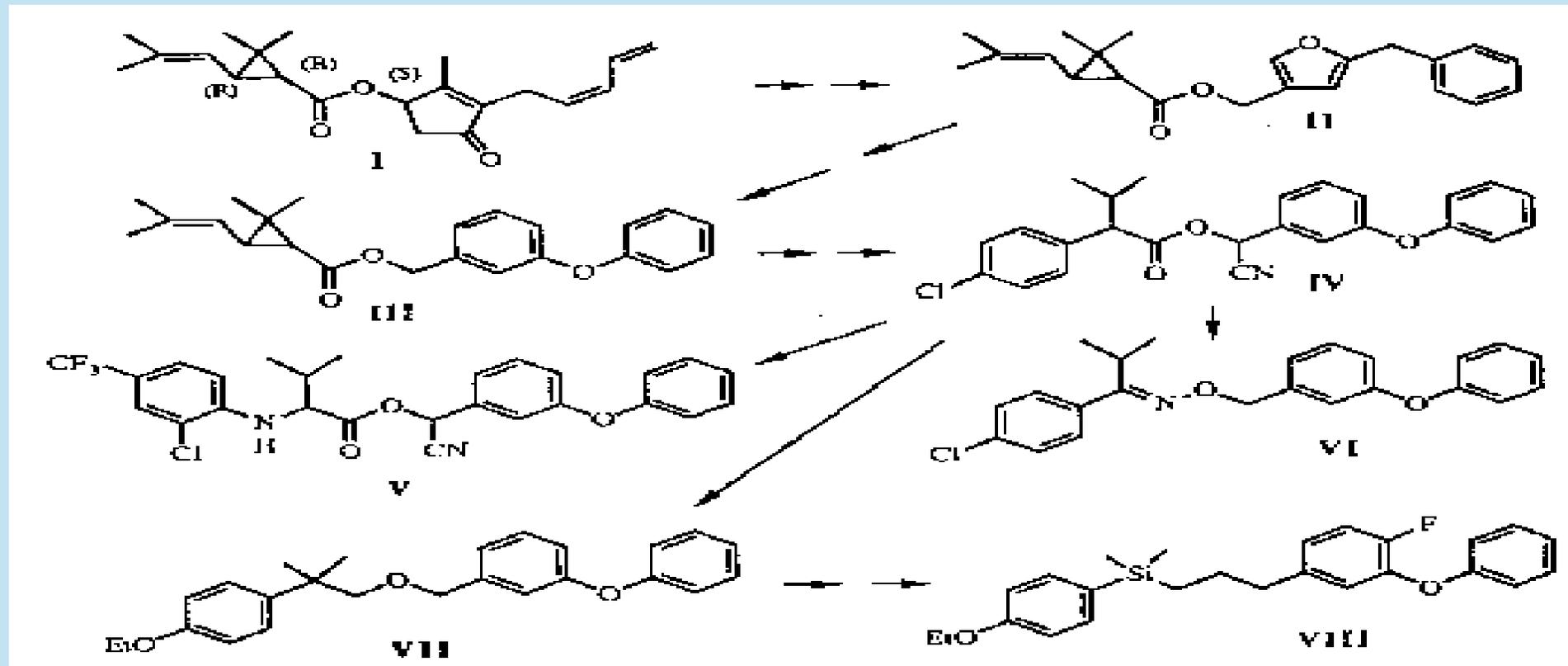
(Example - Mediated - Innovation-for - Lead-Evolution)

## 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



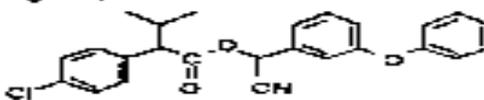
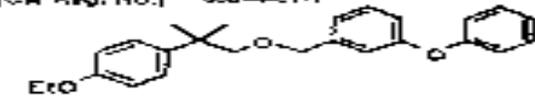
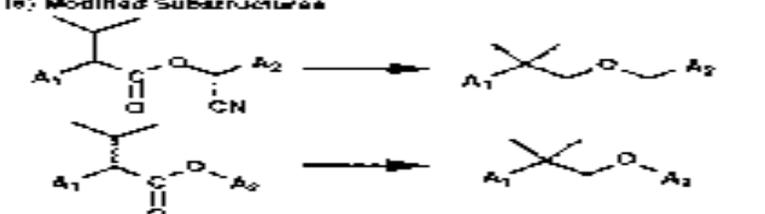
## 化合物構造のバイオアナログス変換事例



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

## 化合物構造変換シート:化学研究者のノウハウまとめ (創薬用と農薬用の二種類あり)

**EMIL DATA SHEET**

Structure A Med. Agro. Others.	Date	Ser. No.	Reg. No.	Prepared by	Structure B Med. Agro. Others.																																	
<b>Use Effect:</b> Insecticide <b>Insecticide Name:</b> Fenitrothion <b>Structural Type:</b> Phosphorothioates Name (General Name) Fenitrothion (Code No.) S-5602 (CA Reg. No.) 51630-56-1  (Institution) Sumitomo Chemical	10/10/05		219	Harosuke Yoshitaka	<b>Use Effect:</b> Insecticide <b>Insecticide Name:</b> Defenprophos <b>Structural Type:</b> Phosphorothioates Name (General Name) Defenprophos (Code No.) ACTLS500 (CA Reg. No.) 90844-07-1  (Institution) Mitsui-Toatsu Chemical																																	
<b>References:</b> Satoshi Numata, "Design of Bioactive Molecules", H. Yoshioka and K. Shirota, Eds., Soil Science Publications, Tokyo, 1986, p.338.																																						
<b>Structural Evolution</b>																																						
<b>(a) Modified Substructures</b> 																																						
<b>(b) Effects of the Modification</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Unchanged Activity</th> <th>Activity</th> <th>Losses</th> </tr> </thead> <tbody> <tr> <td>Selectivity</td> <td>+</td> <td>-</td> </tr> <tr> <td>Activity Spectrum</td> <td>0</td> <td>-</td> </tr> <tr> <td>Potency</td> <td>0</td> <td>-</td> </tr> <tr> <td>Toxicity</td> <td>+</td> <td>-</td> </tr> <tr> <td>Physical</td> <td>+</td> <td>-</td> </tr> <tr> <td>Chemical</td> <td>+</td> <td>-</td> </tr> <tr> <td>Stability</td> <td>+</td> <td>-</td> </tr> <tr> <td>Biodegradability</td> <td>+</td> <td>-</td> </tr> <tr> <td>Delivery</td> <td>+</td> <td>-</td> </tr> <tr> <td>Δ-transition</td> <td>+</td> <td>-</td> </tr> </tbody> </table>						Unchanged Activity	Activity	Losses	Selectivity	+	-	Activity Spectrum	0	-	Potency	0	-	Toxicity	+	-	Physical	+	-	Chemical	+	-	Stability	+	-	Biodegradability	+	-	Delivery	+	-	Δ-transition	+	-
Unchanged Activity	Activity	Losses																																				
Selectivity	+	-																																				
Activity Spectrum	0	-																																				
Potency	0	-																																				
Toxicity	+	-																																				
Physical	+	-																																				
Chemical	+	-																																				
Stability	+	-																																				
Biodegradability	+	-																																				
Delivery	+	-																																				
Δ-transition	+	-																																				
<b>(c) Physicochemical Effects</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Properties</th> <th>Activity</th> <th>Losses</th> </tr> </thead> <tbody> <tr> <td>Hydrophobicity</td> <td>0</td> <td>-</td> </tr> <tr> <td>Acid Solubility</td> <td>0</td> <td>-</td> </tr> <tr> <td>Basicity</td> <td>+</td> <td>-</td> </tr> <tr> <td>Conformational Flexibility</td> <td>+</td> <td>-</td> </tr> <tr> <td>Electron Donor</td> <td>+</td> <td>-</td> </tr> <tr> <td>Electron Acceptor</td> <td>+</td> <td>-</td> </tr> </tbody> </table>						Properties	Activity	Losses	Hydrophobicity	0	-	Acid Solubility	0	-	Basicity	+	-	Conformational Flexibility	+	-	Electron Donor	+	-	Electron Acceptor	+	-												
Properties	Activity	Losses																																				
Hydrophobicity	0	-																																				
Acid Solubility	0	-																																				
Basicity	+	-																																				
Conformational Flexibility	+	-																																				
Electron Donor	+	-																																				
Electron Acceptor	+	-																																				
<b>(d) Pro- and/or Epitope(s) for the Modification</b>  <p>Instead of the strong interaction between C1 and C1' for fenitrothion, that between C1 and C1' optimization of the C number and conversion of ester to other bonding types were tried.</p>																																						
<b>(1) Mode of Action</b> - Prolongation of the Open State of Sodium Channels - Block of the Nerve Conduction.					<b>(1) Mode of Action</b> - Prolongation of the Open State of Sodium Channels - Repetitive Excitation of the Nervous System																																	
<b>(2) Structural Requirements</b> For the Acid Moiety: S >> R Form, G-C can be replaced by E and L by, and p-C by E, R, and (Cis/trans) is at m- and p-positions					<b>(2) Structural Requirements</b> General Me group and the connecting CH2 are essential. Introduction of CN at the m-positions in the phenyl group is unfavorable.																																	
<b>(3) Activity (Potency and Features)</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th>LD50</th> </tr> </thead> <tbody> <tr> <td>Spodoptera littoralis</td> <td>0.23 mg/2</td> </tr> <tr> <td>Musca domestica</td> <td>0.12 mg/head</td> </tr> <tr> <td>Aphis gossypii</td> <td>0.15 mg/head</td> </tr> <tr> <td>L. striatellus</td> <td>4.1 mg/kg</td> </tr> <tr> <td>LD50 (sp):</td> <td>0.26 ppm (25 h)</td> </tr> </tbody> </table>		LD50	Spodoptera littoralis	0.23 mg/2	Musca domestica	0.12 mg/head	Aphis gossypii	0.15 mg/head	L. striatellus	4.1 mg/kg	LD50 (sp):	0.26 ppm (25 h)					<b>(3) Activity (Potency and Features)</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th></th> <th>LD50</th> </tr> </thead> <tbody> <tr> <td>Spodoptera littoralis</td> <td>0.15 mg/2</td> </tr> <tr> <td>Musca domestica</td> <td>0.21 mg/head</td> </tr> <tr> <td>Aphis gossypii</td> <td>0.09 mg/head</td> </tr> <tr> <td>LD50 (sp):</td> <td>&gt; 50,000 mg/kg</td> </tr> <tr> <td>LD50 (sp):</td> <td>5 ppm (25 h)</td> </tr> </tbody> </table>		LD50	Spodoptera littoralis	0.15 mg/2	Musca domestica	0.21 mg/head	Aphis gossypii	0.09 mg/head	LD50 (sp):	> 50,000 mg/kg	LD50 (sp):	5 ppm (25 h)									
	LD50																																					
Spodoptera littoralis	0.23 mg/2																																					
Musca domestica	0.12 mg/head																																					
Aphis gossypii	0.15 mg/head																																					
L. striatellus	4.1 mg/kg																																					
LD50 (sp):	0.26 ppm (25 h)																																					
	LD50																																					
Spodoptera littoralis	0.15 mg/2																																					
Musca domestica	0.21 mg/head																																					
Aphis gossypii	0.09 mg/head																																					
LD50 (sp):	> 50,000 mg/kg																																					
LD50 (sp):	5 ppm (25 h)																																					
<b>(4) Physicochemical Data</b> Vapor pressure: $2.4 \times 10^{-4}$ mmHg logP (octanol): 6.12 at 25°					<b>(4) Physicochemical Data</b> mp: 16-17.5° Vapor pressure: $1.4 \times 10^{-4}$ mmHg (100°) logP (octanol): 6.87																																	
<b>(5) Others</b> Commercialized for use as the insecticide.					<b>(5) Others</b> Commercialized as a paddy field insecticide.																																	

(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

EMIL

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

# インシリコ創薬への人工知能の部分適用事例

In Silico Data  
Miracles by the KY-methods

## 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、バイオアナログウス化合物群

### ■ 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

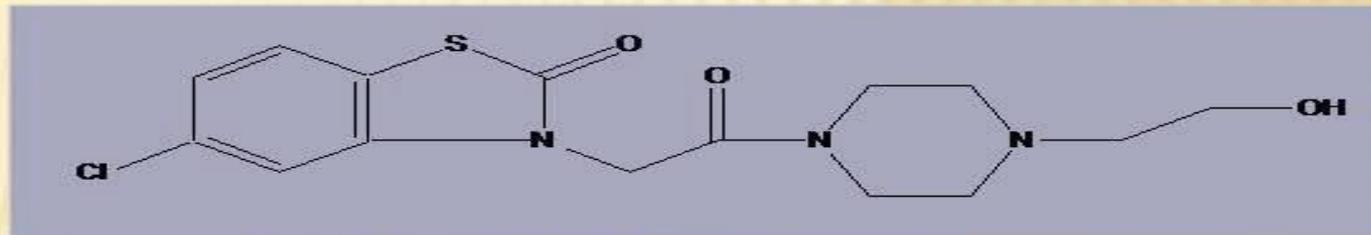
(Example - Mediated - Innovation-for - Lead-Evolution)

## 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)

# EMIL

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

## 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<sup>2+</sup>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	×	○	○	○	○	4
4	○	×	×	×	×	1
5	×	×	○	○	○	3
6	×	×	○	○	○	3
7	×	×	○	○	○	3
8	×	×	×	○	○	1
9	○	×	○	○	×	3
10	×	×	○	○	○	4
⋮						⋮

## □ 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

## ADMEWORKSを用いた安全性評価の実施

---

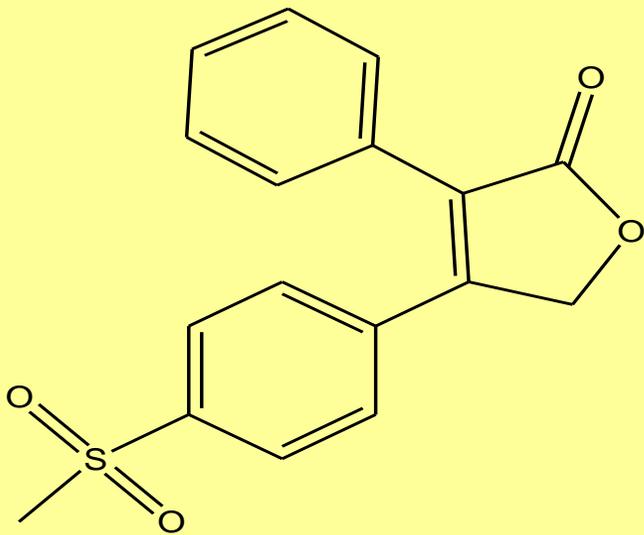
1. 化合物デザインにおける副作用の恐ろしさ
2. 副作用を有する化合物群
3. 副作用とCYP3A4特性
4. CYP3A4特性の改良を目指して



# 同系列化合物への対応

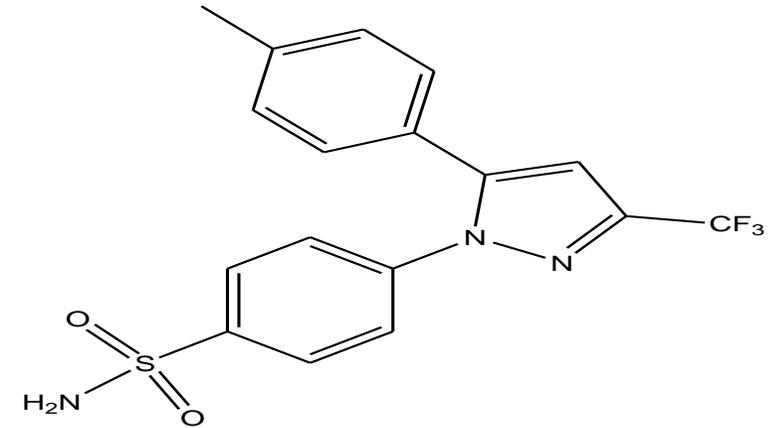
**VIOXX**  
消炎鎮痛剤  
Merck

心臓発作



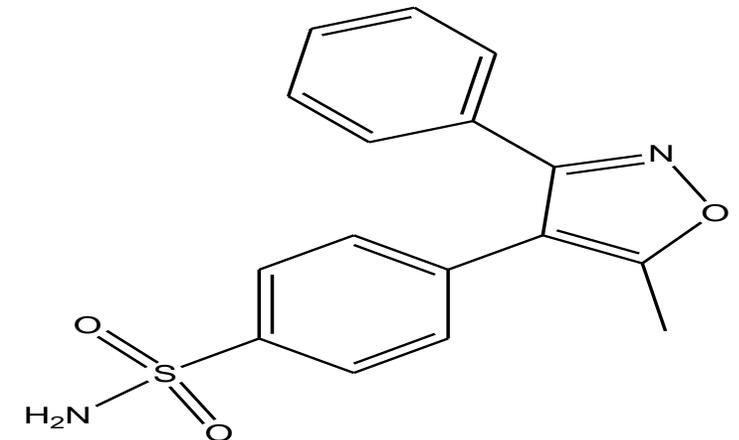
**CELEBREX**  
リウマチ・鎮痛剤  
Pfizer

心臓障害



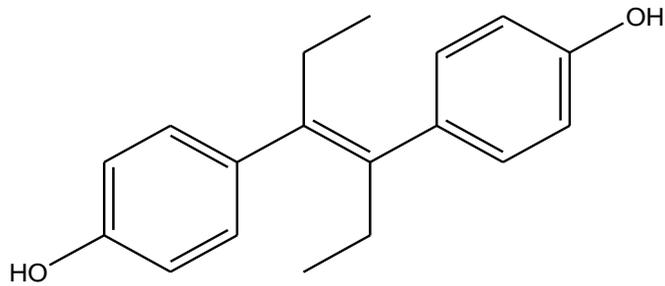
**BEXTRA**  
Pfizer

FDAによる  
自主回収勧告

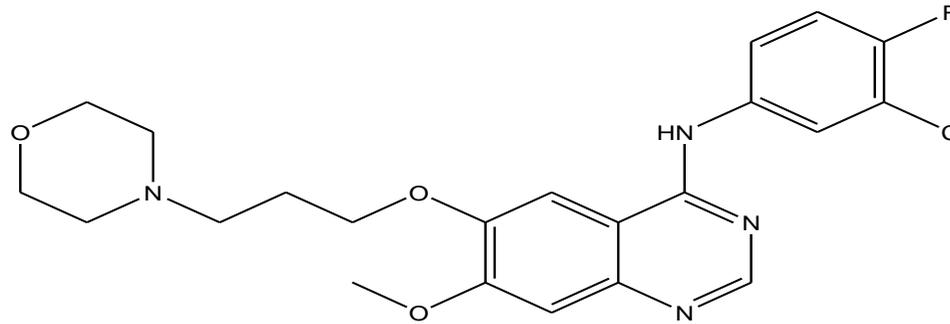


# 副作用を有する市販薬群

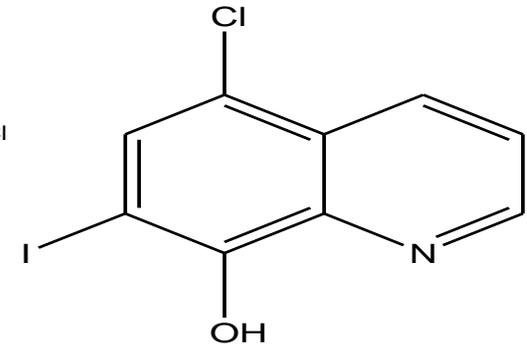
**DESPLEX**  
流産予防薬



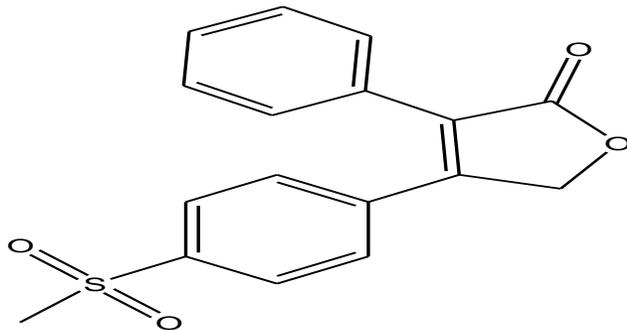
**IRESSA**  
肺ガン治療薬



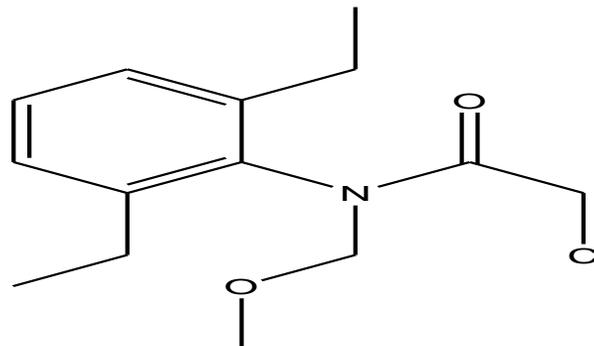
**CLIOQUINOL**  
殺菌薬



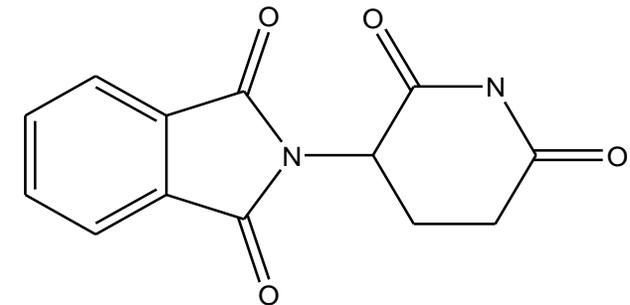
**VIOXX**  
消炎鎮痛剤



**ALACHLOR**  
農薬 (除草剤)



**ISOMIN**  
睡眠薬



# 副作用性化合物の種々特性予測の結果

ADMEWORKS: Worksheet - Microsoft Internet Explorer

ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 検索 お気に入り メディア

アドレス(D) http://localhost:8080/ADMEWORKS/Worksheet

ADMEWORKS

Molecules Properties Evaluation Tools Workspace Help Logout

Worksheet: adverse 7 comps

Selected molecules  
Page: 1/1

View: All

Display: None

副作用性化合物は **CYP3A4** 特性が良くない

**CYP3A4(INHIBITOR/LIGAND)**  
 CARCINOGENICITY(FN/FP MODELS)  
 AMES TEST (TA98 & TA100)

	<input type="checkbox"/>	<input checked="" type="checkbox"/>					
	ID	Name	3A4 Inhibitor	Carcinogenicity-FN	Carcinogenicity-FP	AMES-TA98	AMES-TA100
<input type="checkbox"/>	1	CLIOQUINOL	++	-*	-*	++	++
<input type="checkbox"/>	2	DESPLEX	+	+	-	-	-
<input type="checkbox"/>	3	IRESSA	+	-	-	+	+
<input type="checkbox"/>	4	ISOMIN	+	-	-	-	-
<input type="checkbox"/>	5	ALACHLOR	+	+	-	-	+
<input type="checkbox"/>	6	VIOXX	++	-*	-*	-*	++

イントラネット

# CYP3A4 予測時に判別関数中で用いられたパラメータ群

## □ 要因解析とドラッグデザインへのアプローチ

### ■ 係数が正：阻害剤になりやすい要因

1. Number of S atoms
2. 6th order cluster MC Simple
3. Average sigma charge ss1
4. 2nd order path MC valence
5. All-path ss2
6. Count of ss3
7. Balaban's topological index J
8. Average sigma charge atoms in ss4
9. Total electron density of atoms
10. Number of aromatic bonds
11. Count of ss5
12. Average sigma charge of ss6
13. Average sigma charge of ss7
14. Sum of hetero atoms

### ■ 係数が負：基質になりやすい要因

1. Number of N atoms
2. 6th order chain MC Simple
3. Average sigma charge of ss8
4. Absolute value of atomic charge
5. Most positive sigma charge of ss9
6. Number of double bonds
7. Average sigma charge of ss10
8. Average sigma charge of ss11
9. Number of basis rings
10. Count of ss12
11. 6th order cluster MC Valence

# 構造式修正でCYP3A4の目標特性は変化しないのに 発ガン性が悪化した例（予測によるシミュレーション）

ADMEWORKS: Worksheet - Microsoft Internet Explorer

ISOMIN.mol - Microsoft Internet Explorer

ISOMIN 2 - Microsoft Internet Explorer

構造式修正前

構造式修正後

ID	Name	3A4 Inhibitor	Carcinogenicity-FN	Carcinogenicity-FP	AMES-TA98	AMES-TA100
<input type="checkbox"/>	1 CLOQUINOL	++	-*	-*	++	++
<input type="checkbox"/>	2 DESPLEX	+	+	-	-	-
<input type="checkbox"/>	3 IRESSA	+	-	-	+	+
<input checked="" type="checkbox"/>	4 ISOMIN	+	-	-	-	-
<input type="checkbox"/>	5 ALACHLOR	+	+	-	-	+
<input type="checkbox"/>	6 VIOXX	++	-*	-*	-*	++
<input checked="" type="checkbox"/>	9 ISOMIN 2	+	+	-	-	-

イントラネット

# CYP3A4の目標特性変更の目的で基本骨格を固定し、 一連の仮想化合物創出（バーチャルスクリーニング）

ADMEWORKS: Virtual Library Generator - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 検索 お気に入り メディア

アドレス(D) http://localhost:8080/ADMEWORKS/VLGenerator

ADMEWORKS Fujitsu

Fragments Scaffold Generation Active Set Repository Worksheet Help Logo

Fragment Set	Set Size
R1	5
R2	5
R3	3
R4	3
R5	0
R6	0
R7	0
R8	0
R9	0

Max. no. of generated molecules 225

Scaffold: ISOMIN DERIV

Active Set: **R1** R2 R3 R4 R5 R6 R7 R8 R9

Structure display: Sn

アプレット DrawApplet started

イントラネット

# 創出された仮想化合物群

ADMEWORKS: Worksheet - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 検索 お気に入り メディア

アドレス(D) http://localhost:8080/ADMEWORKS/Worksheet 移動 リンク

**ADMEWORKS** FUJITSU

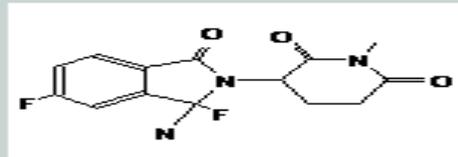
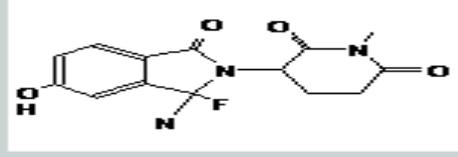
Molecules Properties Evaluation Tools Workspace Help Logout

Worksheet: **ISOMIN**

Selected molecules: 225/225  
Page: 1/3

View: All Selected Print Status Structure display: Small

Go to Page

ID	Name
<input checked="" type="checkbox"/> 	ISOMIN DERIV.methyl.-F (fluorides).-CN (cyanides).-F (fluorides)
1	
<input checked="" type="checkbox"/> 	ISOMIN DERIV.methyl.-F (fluorides).-CN (cyanides).-OH (alcohols)

アプレット DrawApplet started イントラネット

# 創出された仮想化合物群に対する予測結果

ADMEWORKS: Worksheet - Microsoft Internet Explorer

ファイル(E) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 検索 お気に入り メディア

アドレス(D) http://localhost:8080/ADMEWORKS/Worksheet

ADMEWORKS Fujitsu

Molecules Properties Evaluation Tools Workspace Help Logout

Worksheet: **ISOMIN**

Selected molecules: 16/16  
Page: 1/1

View: All Selected Print Status Structure display: None

	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	
	ID	Name	3A4 Inhibitor	Carcinogenicity-FN	Carcinogenicity-FP	AMES-TA98	AMES-TA100
<input checked="" type="checkbox"/>	1	ISOMIN DERIV.methyl.-F (fluorides).-CN (cyanides).-F (fluorides)	++	-*	-*	-*	-
<input checked="" type="checkbox"/>	2	ISOMIN DERIV.methyl.-F (fluorides).-CN (cyanides).-OH (alcoholes)	++	-*	-*	-*	-
<input checked="" type="checkbox"/>	3	ISOMIN DERIV.methyl.-F (fluorides).-CN (cyanides).ethyl	++	-*	-*	-*	-
<input checked="" type="checkbox"/>	4	ISOMIN DERIV.methyl.-F (fluorides).ethyl.-F (fluorides)	+	-	-	-	-
<input checked="" type="checkbox"/>	5	ISOMIN DERIV.methyl.-F (fluorides).ethyl.-OH (alcoholes)	+	-*	-*	-	-

イントラネット

# PREDICTORによる 二次元インシリコスクリーニング（予測実施画面）

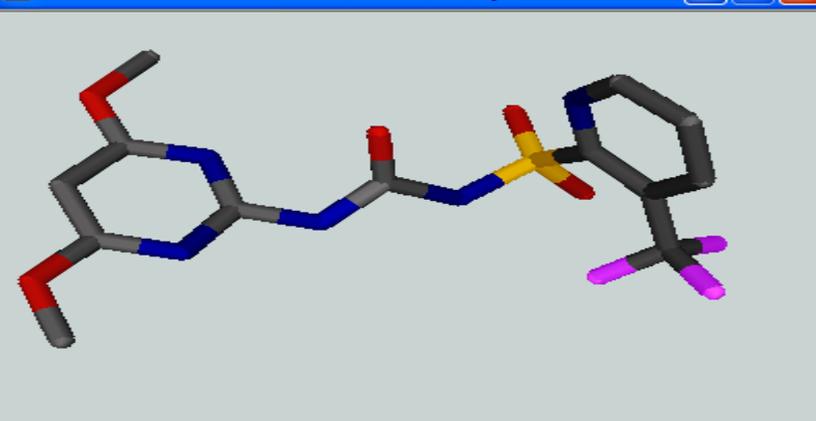
The screenshot displays the ADMEWORKS software interface within a Microsoft Internet Explorer browser window. The main window title is "ADMEWORKS: Worksheet - Microsoft Internet Explorer". The address bar shows "http://localhost:8080". The interface includes a navigation bar with "戻る" (Back), "移動" (Move), and "リンク" (Link) buttons, and a "Logout" button. Two smaller browser windows are open, showing molecular structures: "Mol\_1-2-4" and "Mol\_2-5".

Below the molecular structure windows is a table of prediction results. The table has columns for ID, Name, 3A4 Inhibitor, Carcinogenicity-FN, Carcinogenicity-FP, Solubility, AMES-TA100, AMES-TA98, and Rule Of Five. Each row represents a molecule (Mol\_1 to Mol\_9) with its corresponding predicted values. The "3A4 Inhibitor" column has a green checkmark icon above it, and the "Rule Of Five" column has a green checkmark icon and a pencil icon above it.

	<input type="checkbox"/>	<input checked="" type="checkbox"/>							
	ID	Name	3A4 Inhibitor	Carcinogenicity-FN	Carcinogenicity-FP	Solubility	AMES-TA100	AMES-TA98	Rule Of Five
<input type="checkbox"/>	1	Mol_1	+	-	-	-0.75	-	-	+
<input type="checkbox"/>	2	Mol_2	-	+	-	1.93	-	-	+
<input type="checkbox"/>	3	Mol_3	-	+	-	1.92	+	-	+
<input type="checkbox"/>	4	Mol_4	+	+	-	1.37	-	-	+
<input type="checkbox"/>	5	Mol_5	+	+	-	-0.64	-	-	+
<input type="checkbox"/>	6	Mol_6	+	+	-	-1.54	-	-	+
<input type="checkbox"/>	7	Mol_7	+	+	-	-1.26	-	-	+
<input type="checkbox"/>	8	Mol_8	-	-	-	1.73	-	-	+
<input type="checkbox"/>	9	Mol_9	+	+	-	1.67	-	-	+

# 予測信頼性チェック機能：パラメータ関連表示

4145.mol - Microsoft Internet Explorer



Internet Explorer

グループ(H)

気に入る メディア

details?MoleculeID=2&MethodID=4&trset=1&substr=1

移動 リンク >>

fujitsu

Space Worksheet Help Logout

ta

2

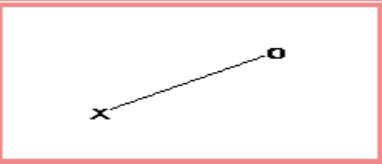
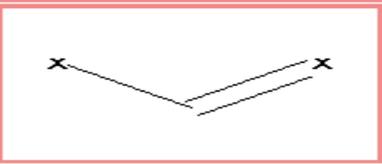
Predicted value

20.mol

3A4 Inhibitor

+

Training set statistics: On off

Descriptor	Hide Substructures	Value	Training Set			
			Min	Max	Mean	Standard deviation
Number of nonhydrogen atoms		14	2	87	24.35	15.51
6th order chain MC Valence		0.13	0	1.12	0.11	0.12
Count of substructure (DMPATH)		0	0	14	1.19	2
Count of substructure (DMPATH)		1	0	10	1.14	1.77

アプレット DrawApplet started

イントラネット

# 予測信頼性チェック機能： 類似化合物との比較チェック

ADMEWORKS: Find Similar - Microsoft Internet Explorer

4125.mol - Microsoft Internet Explorer

ADME Prediction

Worksheet Help Logout

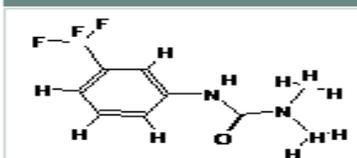
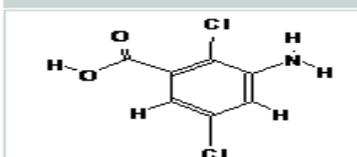
Sorted by similarity:

4

arcinogenicity-FN

Predicted value

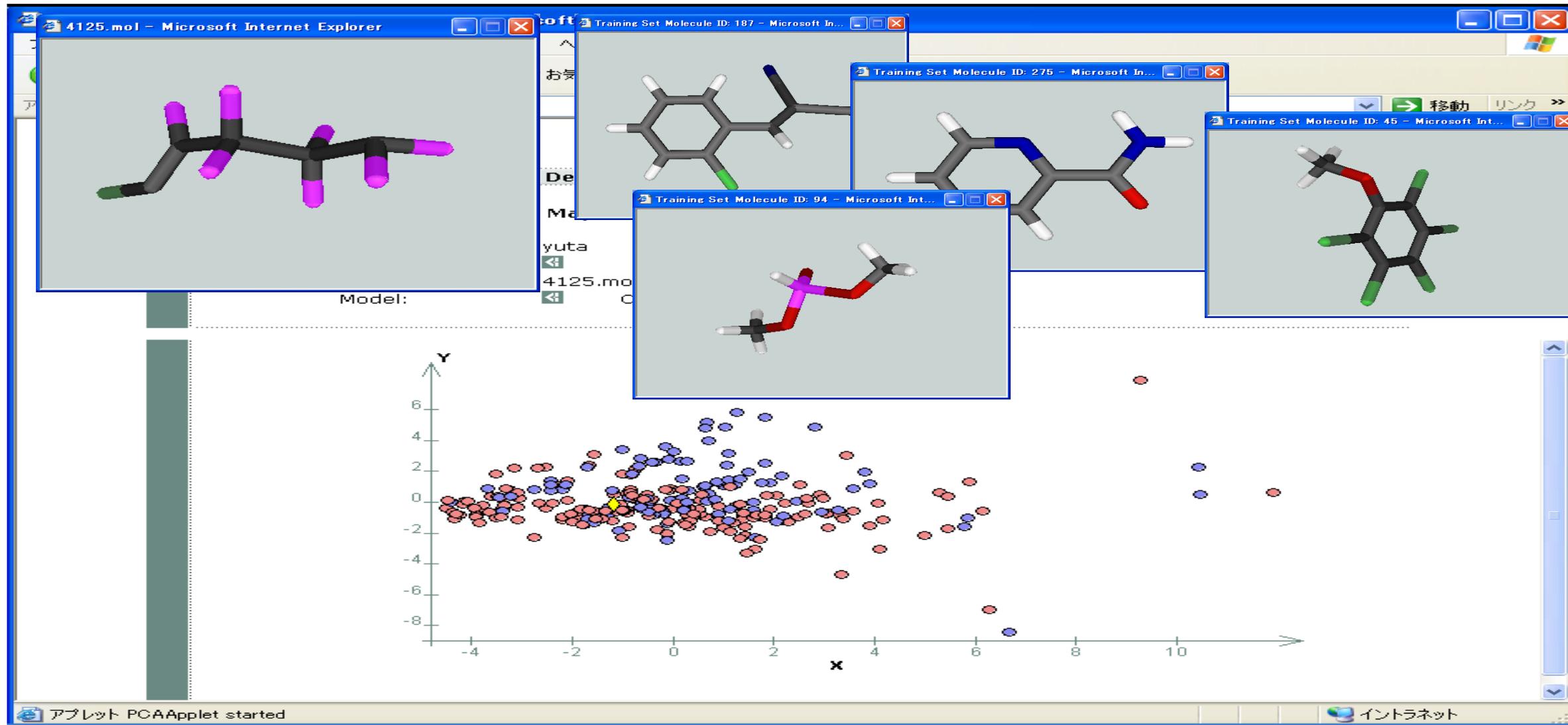
View: Top 10 Top 50 Top 100 All Structure display: Small

ID	Molecule Name	Distance	Predicted Value	Experimental Value
 180	Fluometuron	11	-	-
 161	Amiben	14.06	-	-

アプレット DrawApplet started

イントラネット

# 予測信頼性チェック機能 サンプル化合物母集団上での比較チェック機能



# 要因解析の実施：パラメータ情報の表示画面

The screenshot displays a web browser window with a 3D molecular model on the left and a data table on the right. The molecular model shows a benzene ring with a chlorine atom and a methyl group, with bond lengths labeled: 2.80, 1.42, 2.42, 1.40, 1.40, 1.39, and 2.73. The data table below is titled "Training set statistics:" and is currently set to "On".

Training set statistics:						
	On	Off				
Total electron density of atoms in structure	-0.29	-0.44	-1.5E-3	-0.27	6.71E-2	
Count of substructure (DMPATH)	0	0	7	0.64	1.11	
WTPT 1/number of atoms in the structure	1.89	1.62	2.02	1.88	8.22E-2	
3rd order cluster MC Simple	0.7	0	4.37	1.07	0.79	
Count of substructure (DMPATH)	1	0	6	0.78	1.09	
Number of Nitrogen atoms	0	0	7	0.99	1.27	
Number of Chlorine atoms	2	0	7	0.7	1.47	
Most negative sigma charge of atoms in the substructure	0	-8.23E-2	4.63E-2	-6.32E-3	2.21E-2	
Sum of absolute value of atomic charges	1.64	-99	13.71	-30.34	47.61	
Minimum bond order	0.26	-0.99	0.67	0.21	0.3	
Most positive sigma charge of atoms in the substructure	-2.36E-2	-8.8E-2	0.11	-3.52E-3	2.89E-2	

# 新規化合物のデザインと評価画面

ADMEWORKS: Worksheet - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 検索 お気に入り メディア

アドレス(D) http://localhost:8080/ADMEWORKS/Worksheet

ADMEWORKS Fujitsu

Molecules Properties Evaluation Tools Workspace Help Logout

Worksheet: altz-test2

Selected molecules: 1/10  
Page: 1/1

View: All Selected Print Status Structure display: None

	<input type="checkbox"/>	<input checked="" type="checkbox"/>								
	ID	Name	3A4 Inhibitor	Carcinogenicity-FN	Carcinogenicity-FP	Solubility	AMES-TA100	AMES-TA98	Rule Of Five	
<input type="checkbox"/>	1	MOL-1	+	-	-	-0.75	-	-	+	
<input type="checkbox"/>	2	MOL-2	-	+	-	1.93	-	-	+	
<input type="checkbox"/>	3	MOL-1-5	-	-	-	-0.59	-	-	+	
<input type="checkbox"/>	4	MOL-1-2	+	-	-	-0.03	+	-	+	
<input type="checkbox"/>	5	MOL-2-3	+	+	-	0.38	-	-	+	
<input type="checkbox"/>	6	MOL-2-4	-	+	-	-0.09	+	-	+	
<input type="checkbox"/>	7	MOL-2-7	+	-	-	0.66	-	-	+	
<input type="checkbox"/>	8	MOL-2-10	-	-	-	0.42	-	-	+	

イントラネット

# □メタボロミクス実施事例

## □死後経過時間予測；骨格筋

日本医科大学様との共同研究

NMRのスペクトルデータを用いてメタボローム解析を実施

□ 使用データ

サンプル及びNMRデータ

15分 ⇒ クラス1 (57サンプル)

60分 ⇒ クラス2 (20サンプル)

240分以上 ⇒ クラス3 (17サンプル)

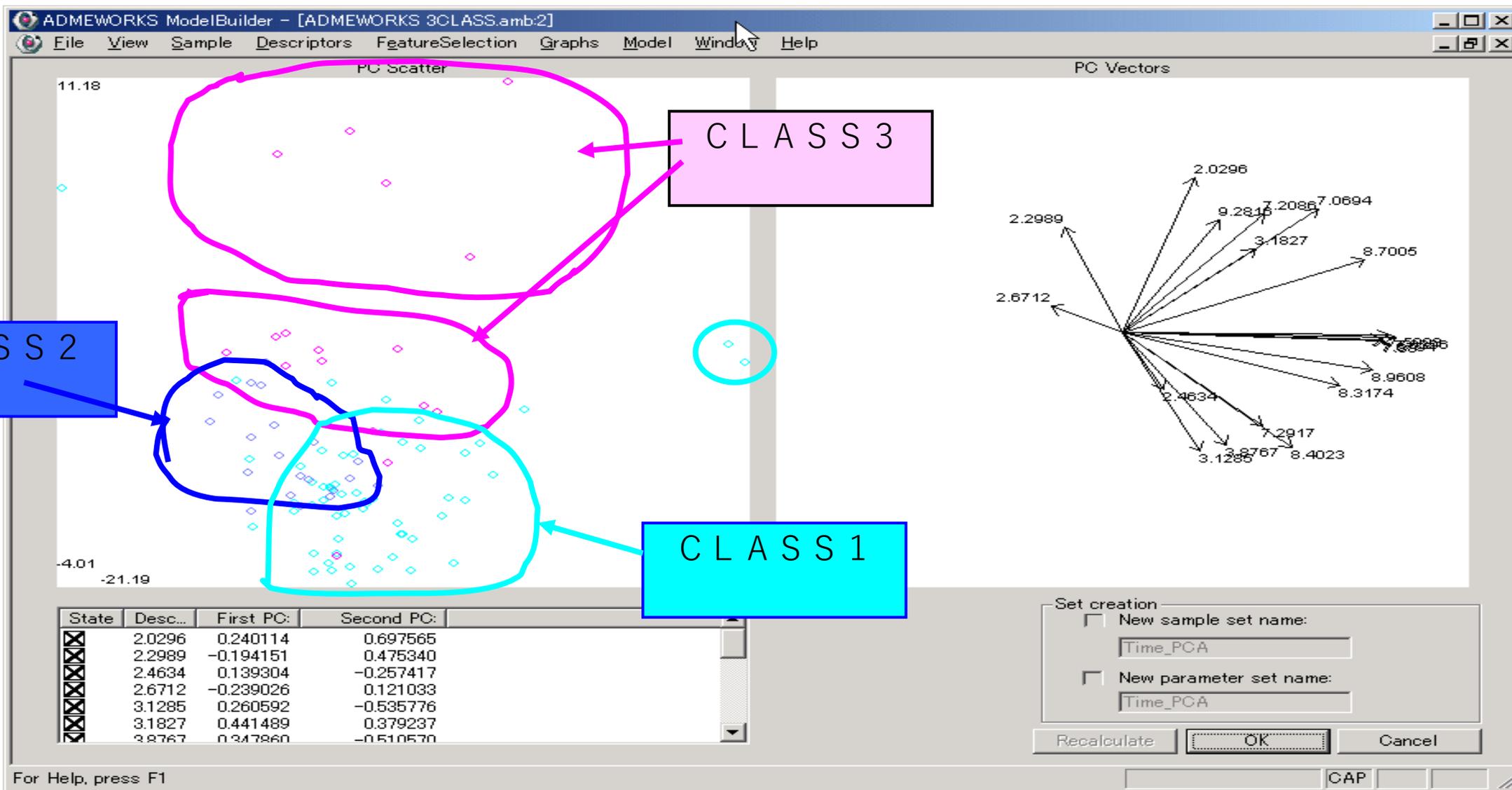
総サンプル数                    総数：94サンプル

初期パラメータ群：

140パラメータ (ケミカルシフト)

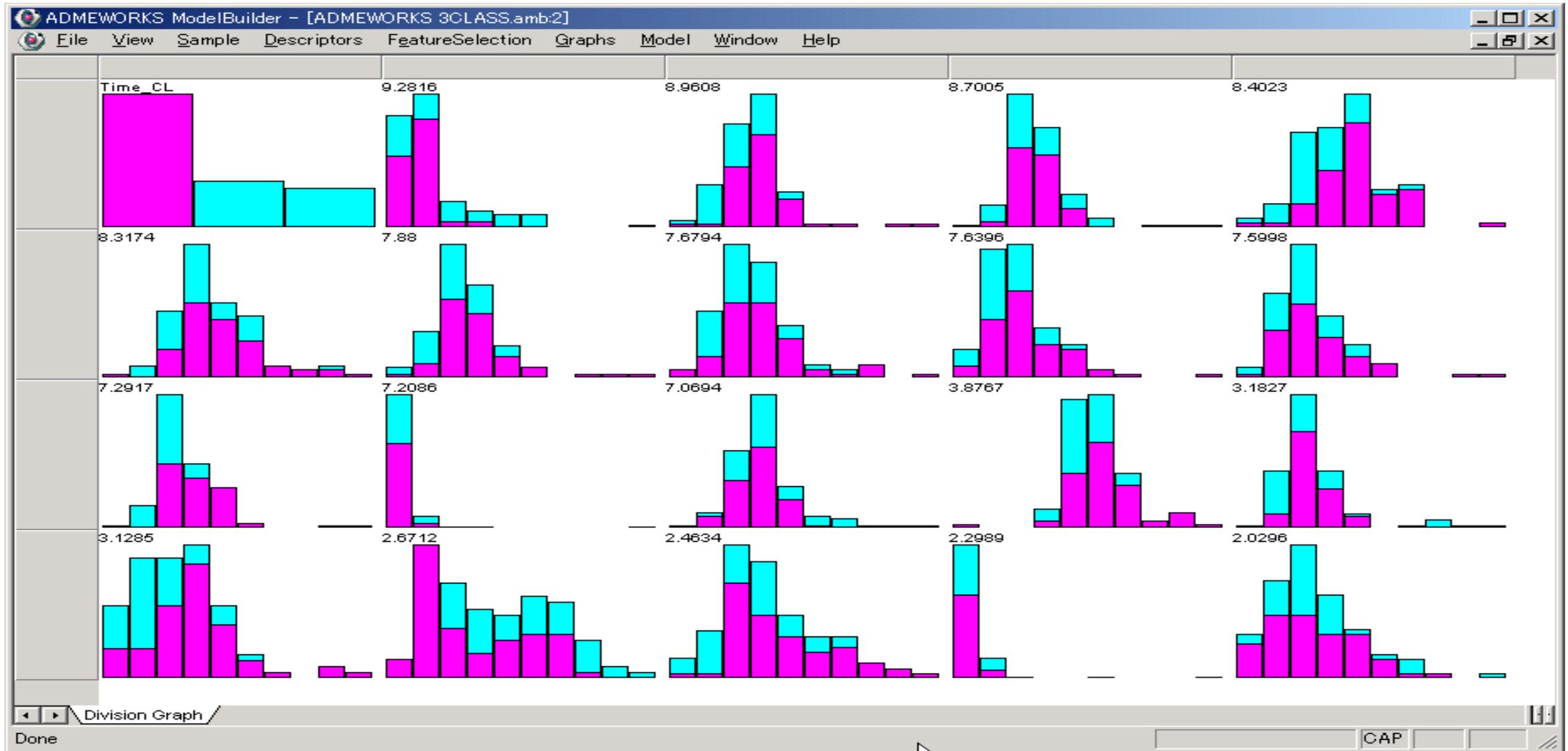
	0.67	140						
	94/94	Item	Time_CL	SampleID		o	Time	9.47
<input checked="" type="checkbox"/>	1	Item_1	Class 1	M1C01.als	cocaine	15	0.25	0
<input checked="" type="checkbox"/>	2	Item_2	Class 1	M1C02.als	cocaine	15	0.25	0
<input checked="" type="checkbox"/>	3	Item_3	Class 1	M1C03.als	cocaine	15	0.25	0
<input checked="" type="checkbox"/>	4	Item_4	Class 1	M1C04.als	cocaine	15	0.25	0
<input checked="" type="checkbox"/>	5	Item_5	Class 1	M1C05.als	cocaine	15	0.25	0
<input checked="" type="checkbox"/>	6	Item_6	Class 1	M1C06.als	cocaine	15	0.25	0
<input checked="" type="checkbox"/>	7	Item_7	Class 1	M1H01.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	8	Item_8	Class 1	M1H02.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	9	Item_9	Class 1	M1H03.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	10	Item_10	Class 1	M1H04.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	11	Item_11	Class 1	M1H05.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	12	Item_12	Class 1	M1H06.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	13	Item_13	Class 1	M1H07.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	14	Item_14	Class 1	M1H08.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	15	Item_15	Class 1	M1H09.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	16	Item_16	Class 1	M1H10.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	17	Item_17	Class 1	M1H11.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	18	Item_18	Class 1	M1H12.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	19	Item_19	Class 1	M1H13.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	20	Item_20	Class 1	M1H14.als	hypoxia	15	0.25	0
<input checked="" type="checkbox"/>	21	Item_21	Class 1	M1H15.als	hypoxia	15	0.25	0

# □ PCA (19パラメータ)

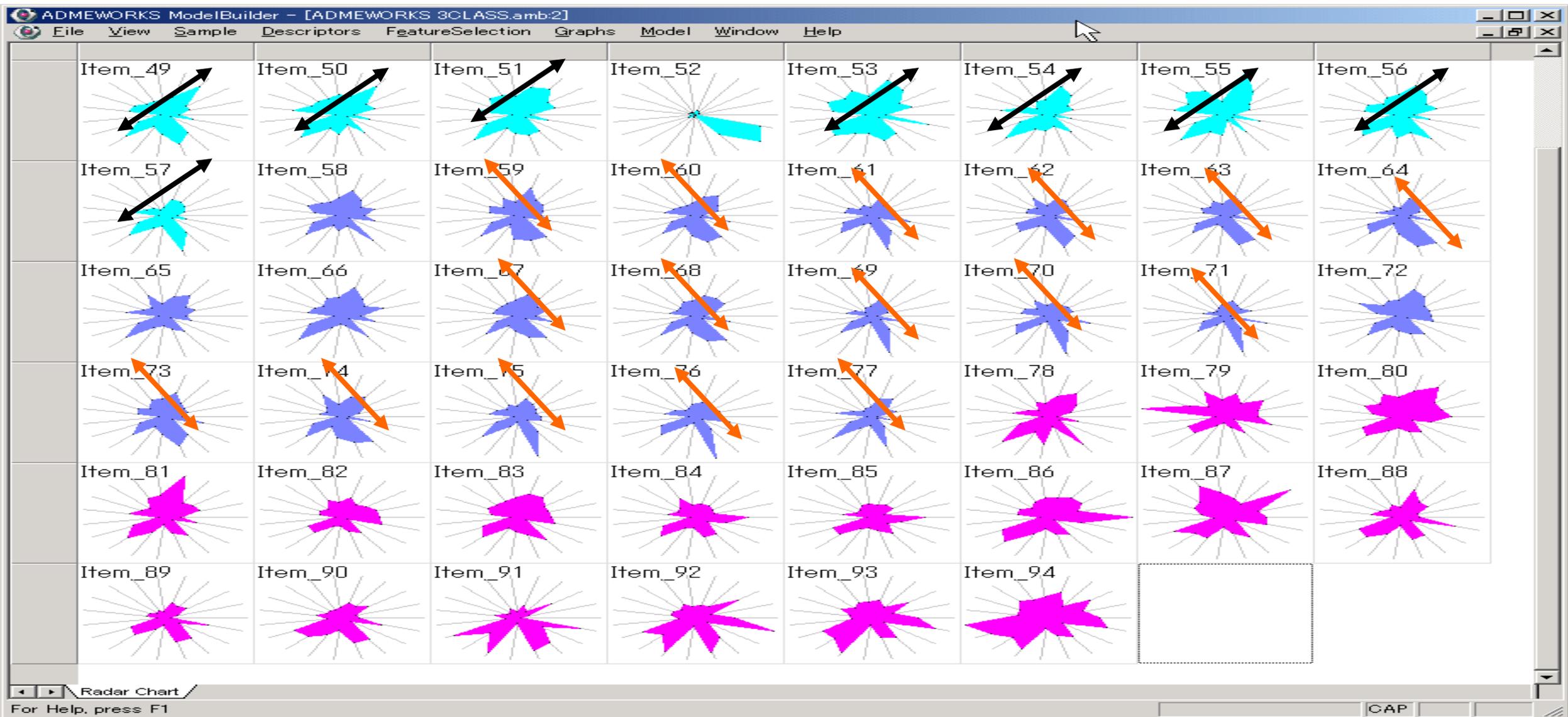




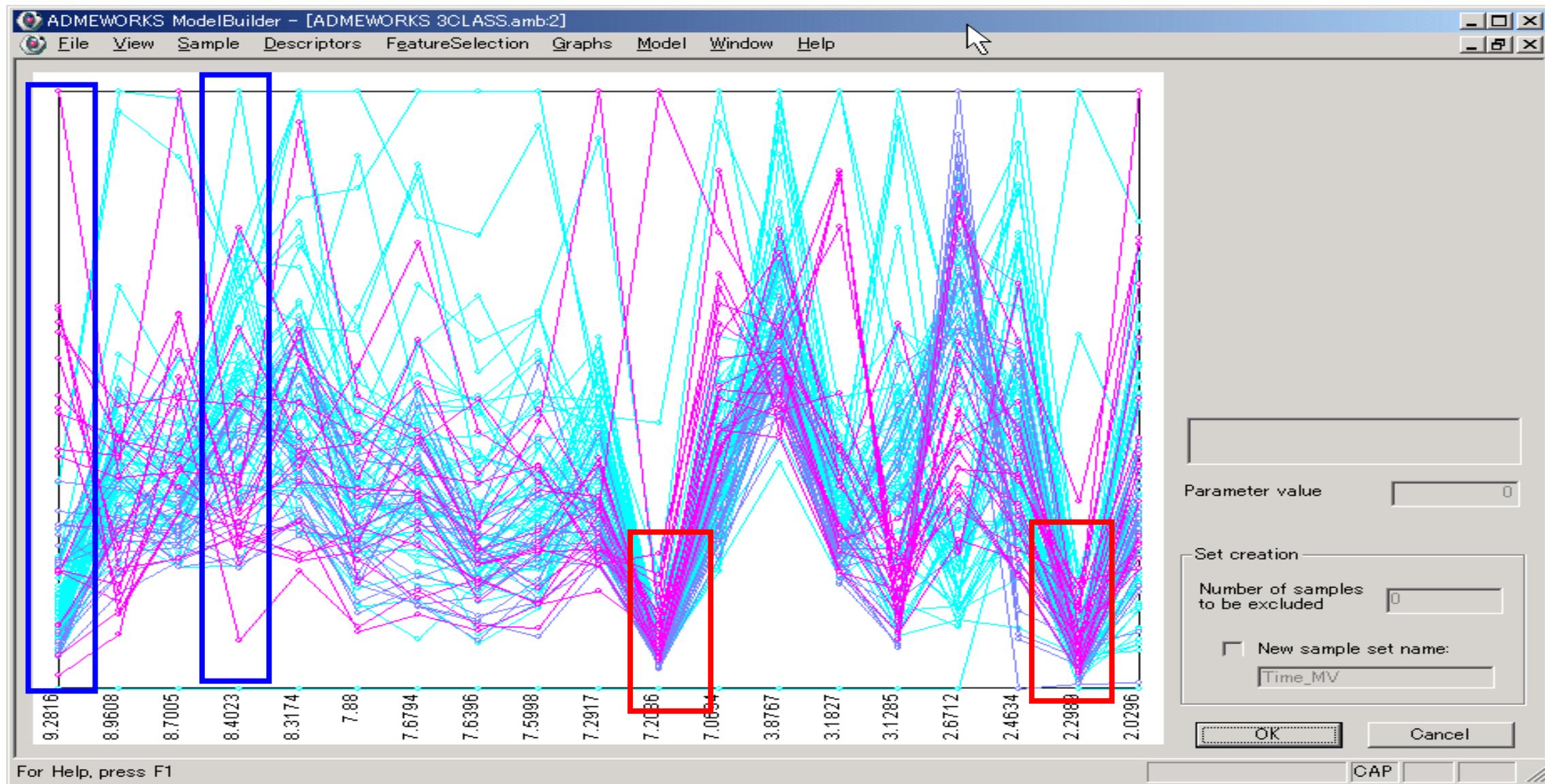
# □ クラス分割グラフ (19パラメータ)



# □ レーダーチャート (19パラメータ)



# □ ライングラフ (19パラメータ)



□ クラス分類手法（19パラメータ）

KNN Model Parameters

File Edit

Number of Samples = 94  
Number of Parameters = 19  
Samples/Parameters Ratio= 4.95  
Method used : KNN

Class	Members	Wrong	% Correct
Class 1	57	4	92.98
Class 2	20	3	85.00
Class 3	17	4	76.47
Total:	94	11	88.30

Validation ROC curve    CV ROC curve    Create new parameter set

Description:

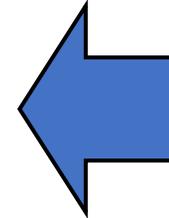
Version     Name

Calc. Prediction Ratio    Prediction

Calc. Random Test    Y-RAND     p-value

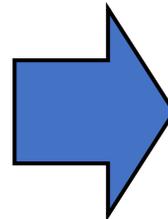
**KNN法 (K = 3)**

\* 分類率 : 88.3 %  
\* 予測率 : 76.6 %



**SVM**

\* 分類率 : 100.0 %  
\* 予測率 : 84.0 %



SVM parameters

File Edit

Number of Parameters = 19  
Samples/Parameters Ratio= 4.95  
Method used : SVM

Class	Members	Wrong	% Correct
Class 1	57	0	100.00
Class 2	20	0	100.00
Class 3	17	0	100.00
Total:	94	0	100.00

Validation ROC curve    CV ROC curve    Create new parameter set

Description:

Version     Name

Calc. Prediction Ratio    Prediction

Calc. Random Test    Y-RAND     p-value

# まとめ

## □多変量解析／パターン認識関連

1. 多変量解析／パターン認識はツールである
2. 原理よりも、利用／適用する時の方法を学ぶ
3. 解析目的にあわせて、使い分ける
4. 手法の限界を知る

## □データ解析実施関連

1. データ解析の標準的な流れを身につける
2. 解析結果の見方／解釈法を学ぶ
3. データ解析結果と、研究上での一般の知識の連携の重用さを知る

# TS-02 第三章 解析手法に続く