

Toxicity Control Map

Knowledge Acquisition from Global Database

AICHE 2004 Spring National Meeting
April 25-29

Yano Research Laboratory
Kenji Yano, Ph.D.

Sub-structural Balance Method (SBM)

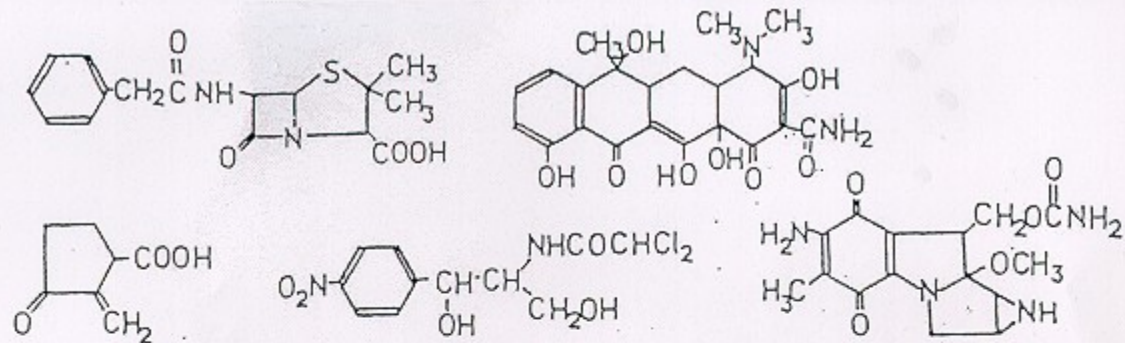
- A molecular design method
- New Theory: Molecule = System
- Acquires new knowledge from global database by data retrieval (not by probability models)

History

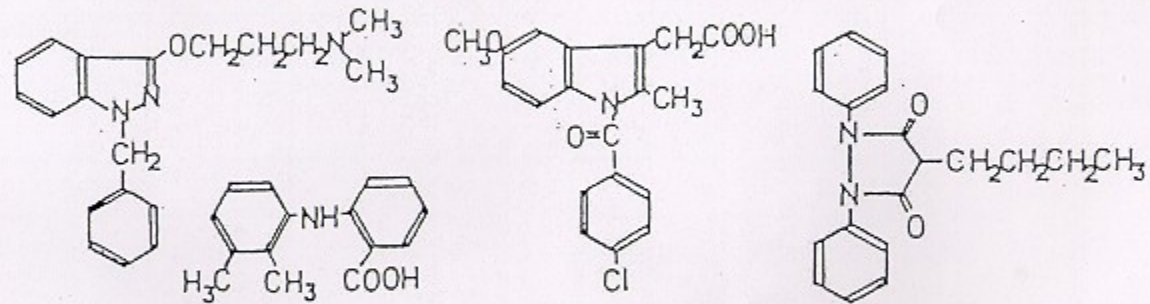
- Started SAR/SPR research in 1980s to develop a simple and practical method for predicting molecular properties
- Attempted to extract common structural features from compounds known to have excellent performances
- Made our first “Compound Map” to guide molecular design

Structure Samples Used

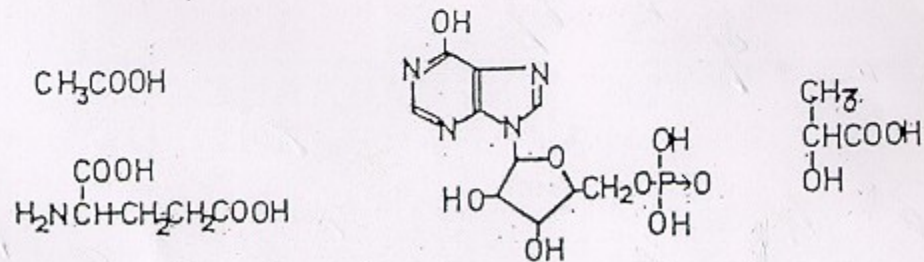
Anti-biotics
(42)



Anti-inflammatory agents (5)

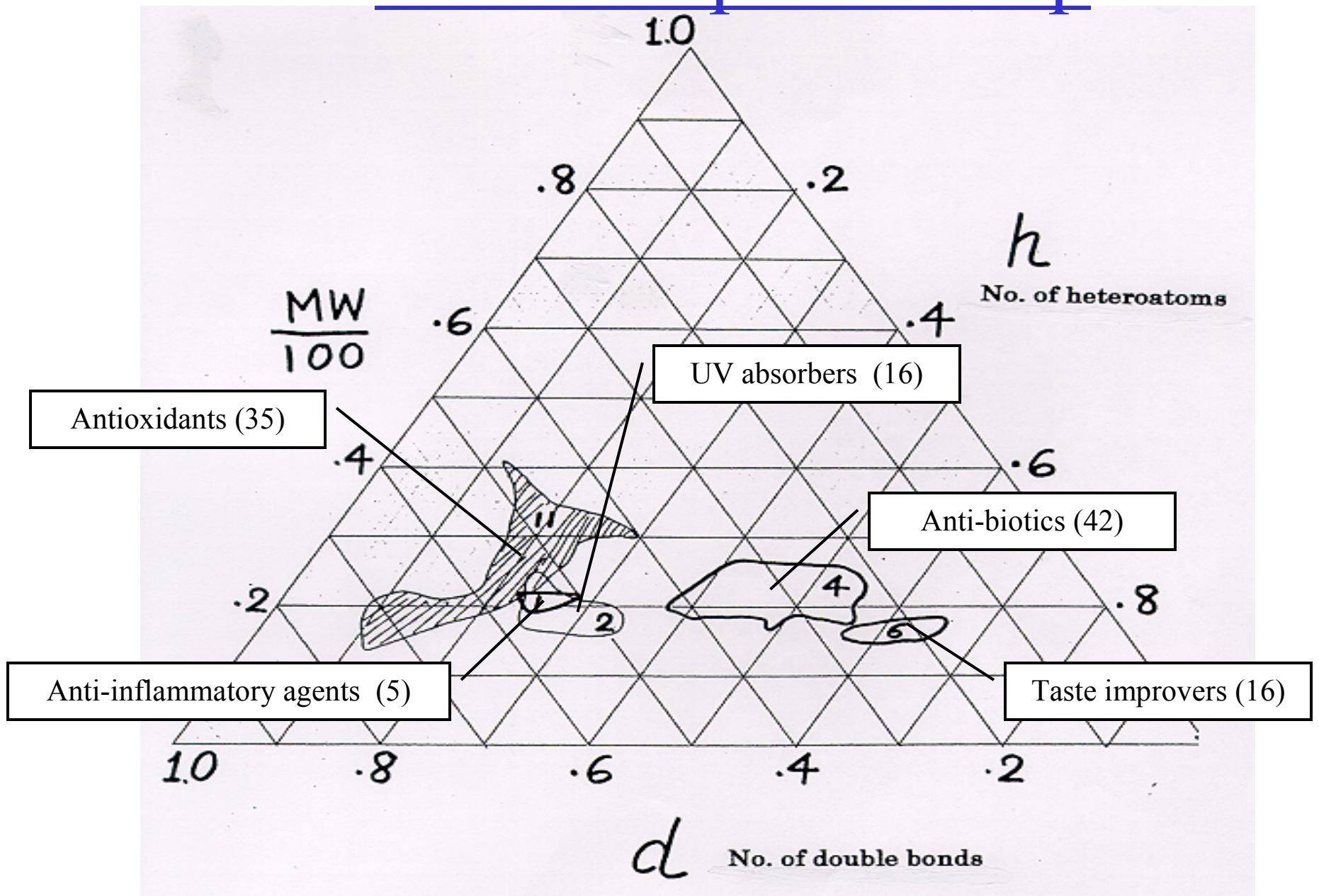


Taste-improvers
(16)

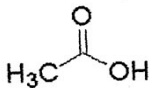
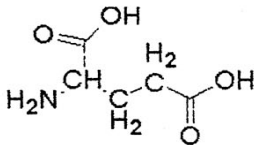
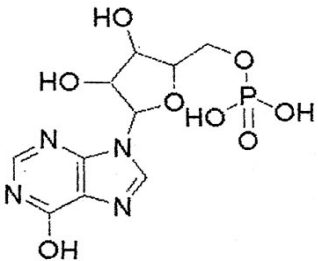


Examples of Structures

First Compound Map



Taste Improvers

	MW/100	d	h	Σ
	0.6 (0.17)	1 (0.27)	2 (0.56)	3.6
	1.47 (0.17)	2 (0.24)	5 (0.59)	8.47
	3.48 (0.17)	5 (0.24)	13 (0.60)	21.48

s-7. tab

Finding

Internal sub-structural ratios may be useful to express unique structural patterns corresponding to similar molecular performances

Hypothesis

An optimum sub-structural balance range exists for each molecular performance

Such a balance range may be demanded by the environment surrounding the molecule (e.g. receptors)

New Descriptor Concept

- Sub-structural Ratio: $P = (X_i + d) / (X_j + c)$
- Built a pool of 400 sub-structures by atom grouping
- 80,000 ($400C2$) possibilities of P

Applications

- Anti-allergic activity
SAR. QSAR. Environ. Res. 12,417-37 (2001)
- Tasty activity
Proceedings of the 13th Japanese Symposium on Taste and Smell (1979)
- Melting point
- Acute toxicity
SAR. QSAR. Environ. Res. 3,15-26 (1995)
SAR. QSAR. Environ. Res. 14,405-12 (2003)

SBM Procedure: Acute Toxicity

- Collected random data and made 2 groups (Learning Set)
 - Low Toxicity Group: $LD50 > 1000$ mg/kg (111)
 - High Toxicity Group: $LD50 < 100$ mg/kg (817)
- Counted the no. of every sub-structure of each molecule and computed their ratios ($400C2 = 80,000$)
- For each ratio, determined the range of Low Toxicity Group (100%) and its discrimination power against High Toxicity Group

Sub-structure Examples

C	No. of carbon atoms
H	No. of heteroatoms
(C+h)nr	No. of skeletal atom outside the ring
hCh	No. of hydrophilic carbon atoms
C-hCh	No. of lipophilic carbon atoms
D	No. of electron donors
h(-)	No. of electron-rich heteroatoms
X	No. of halogen atoms
B	No. of branches excluding halogen atoms
BX	No. of branches including halogen atoms

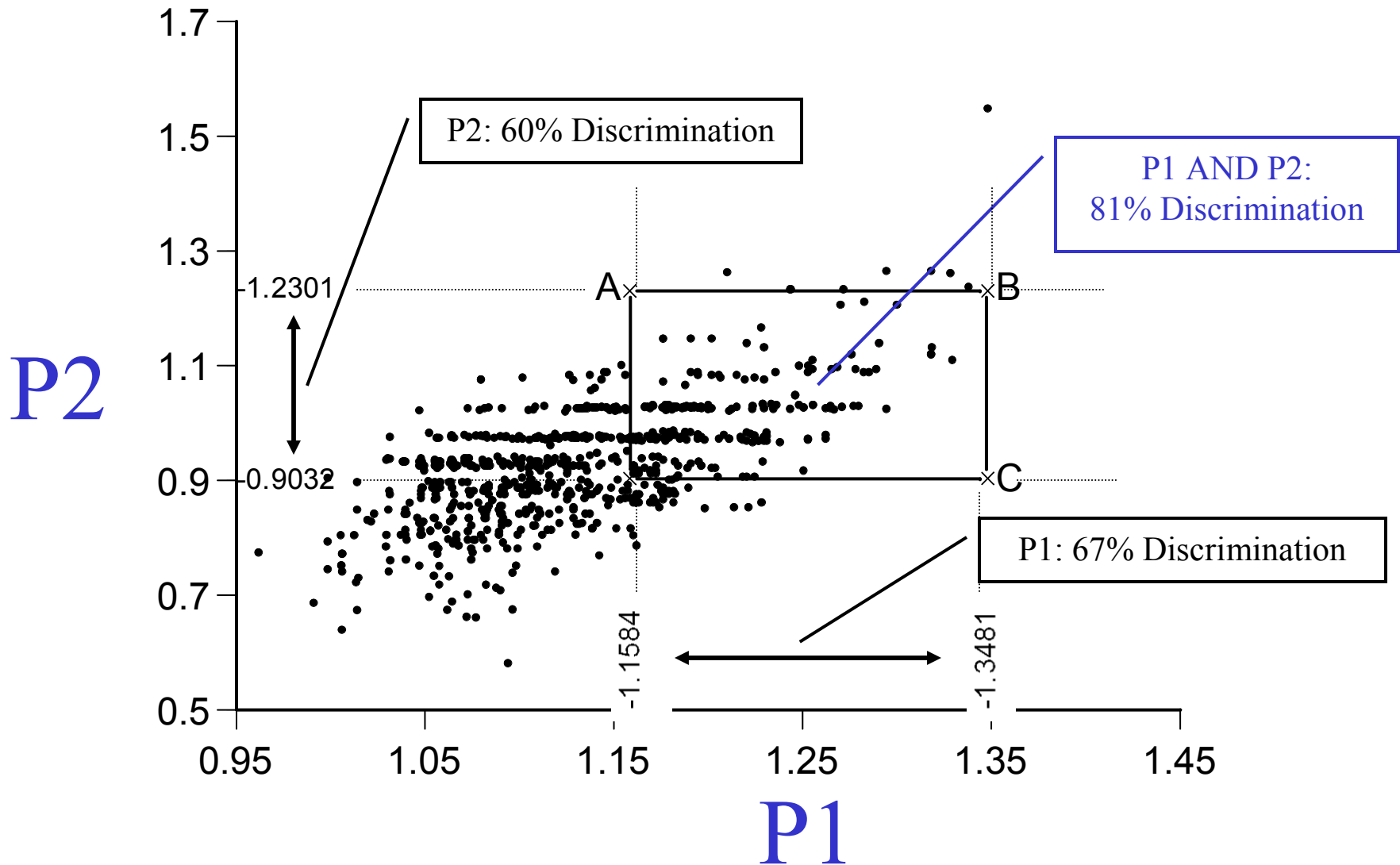
Best Descriptors

- $P1 = (C + h)nr + 122.9 / (C-hCh) + 100.0$
67% discrimination
- $P2 = D + 30.6 / BX + 30.0$
60% discrimination

Noise Reduction by Data Retrieval

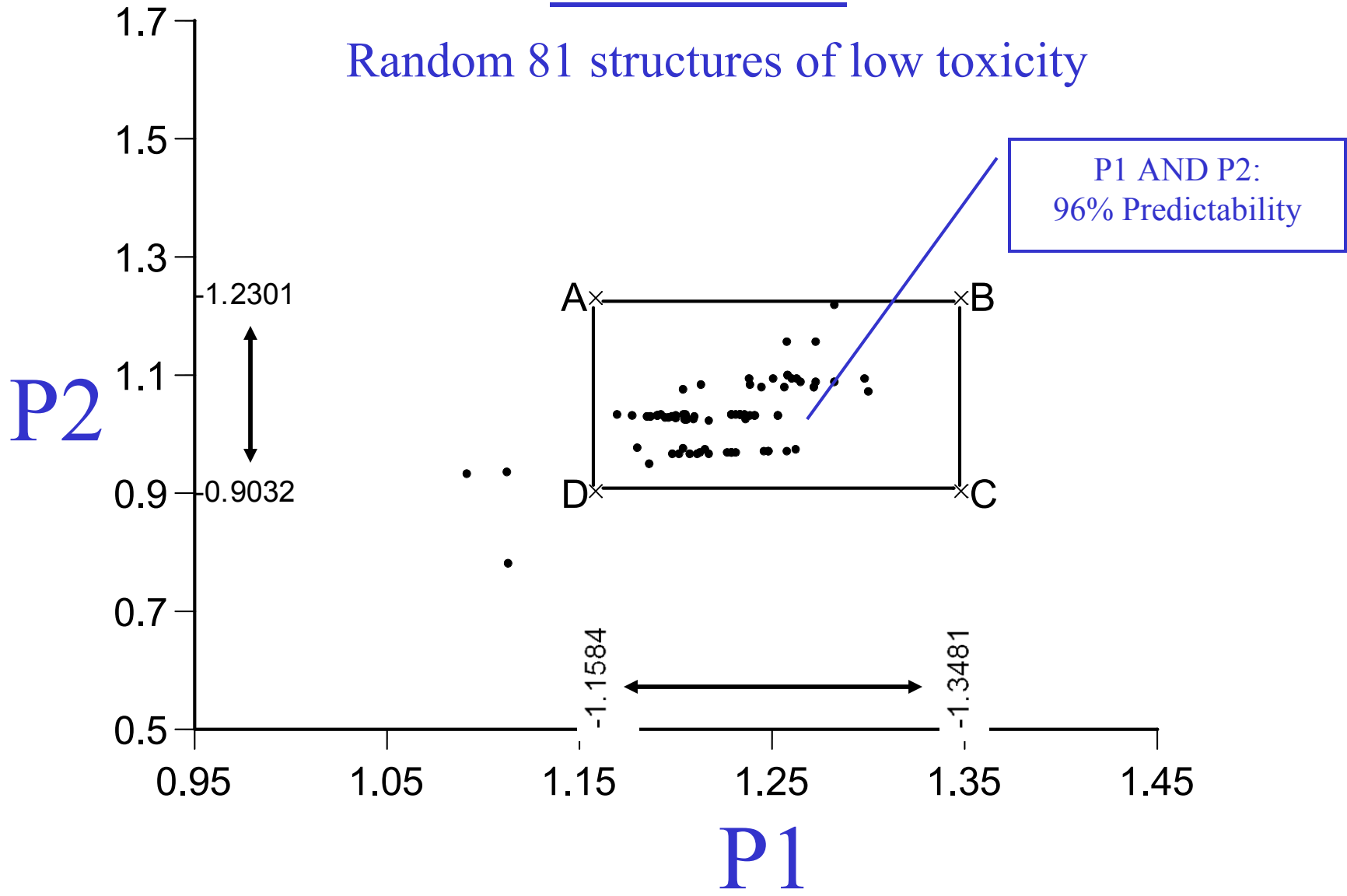
- Similar to Web search using key words
- P1 AND P2
- Low Toxicity Control Limits (Strike Zone)
81% discrimination

Low Toxicity Strike Zone

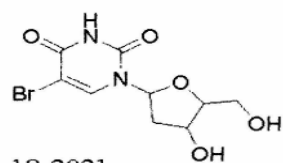


Test Set

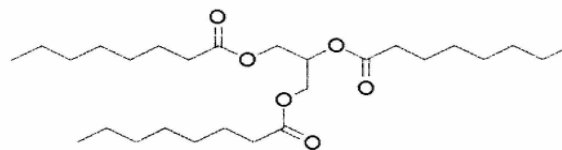
Random 81 structures of low toxicity



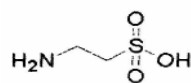
Low Toxicity Structures



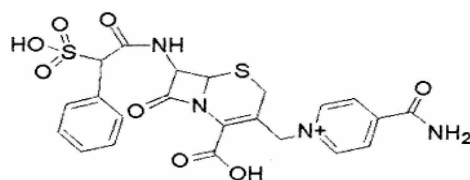
18-2021



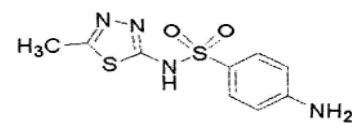
18-2046



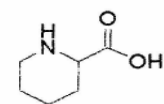
18-2116



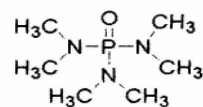
18-2262



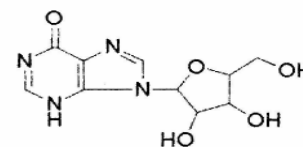
18-2128



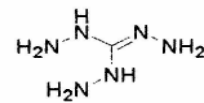
18-2378



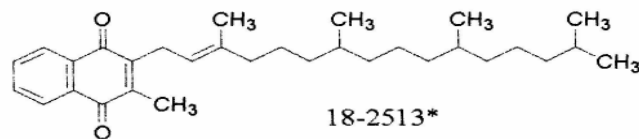
18-2399



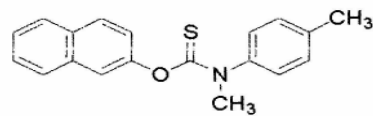
18-2573



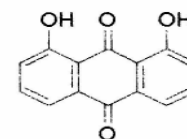
18-2611



18-2513*



18-2747*



18-2853*

Toxicity Control Map Features

- Global structure application
- 100% sensitivity for low toxicity group
- No Type-I statistical error
- Reduction of structure search space

Business Applications

- A screening level predictive tool for chemicals lacking toxicity data / hypothetical structures
- A decision support tool for designing safer chemicals
- An EMS tool: helps create a set of metrics that can be applied at the early stages in the product development process

Data Retrieval for Global Dataset

Local

Mechanistic approach
(bottom-up)

Probability models
(statistical method)

Global

Epidemiological
approach (top-down)

Data Retrieval
(noise reduction)

e.g. Google Search

General System Theory as a Corollary

A system is a set of variables in interaction. When a steady-state is reached between a system and the surrounding environment, any two variables of the system reach a constant ratio.

A system = a molecule

Variables = sub-structures

Environment = receptors, etc.

About Us

Kenji Yano, Ph.D., General Manager

Nick Yano, Ph.D., Research Director

Chiharu Tamura, Ph.D., Research Chemist

END