# QSAR: Discovery and First Steps

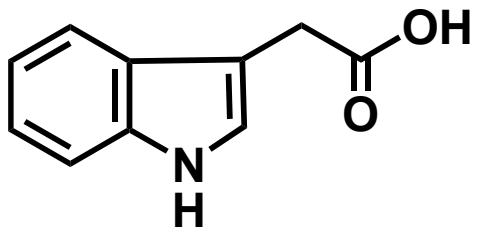
**Toshio Fujita**

**Kyoto University**

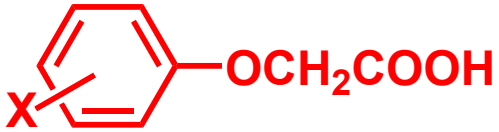
**Kyoto, Japan**

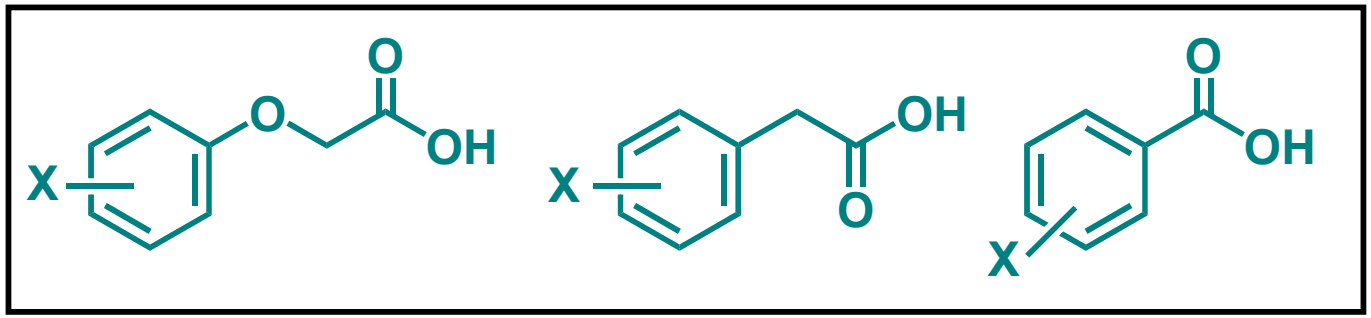
# Origin of Classical QSAR is from the SAR Studies of Agrochemicals—— Plant Growth Regulators/Herbicides

**Indole-3-acetic acid**

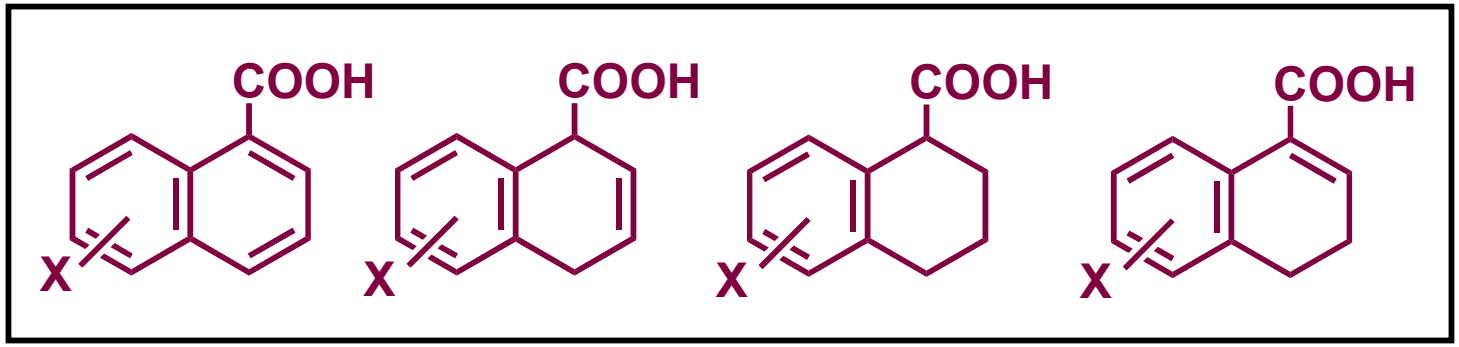



**Phenoxyacetic acids**



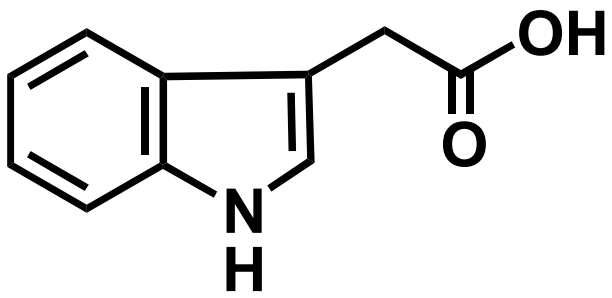




**Hansch and Muir at Pomona (1946 ~ )**




**Mitsui, Fujita and others at Kyoto (1950 ~ )**

# Identification of Natural Plant Growth Regulator




**Indole-3-acetic acid (Auxin)**

**Isolated from Human Urine**

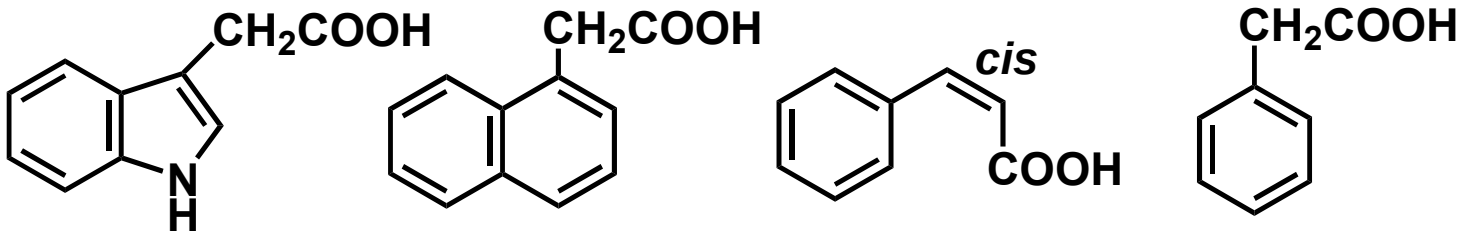
**F. Kögl, A. J. Haagensmit, H. Erxleben,**
**Z. Physiol. Chem. 228, 90 (1934)**

**Receptor Protein Identified Recently**

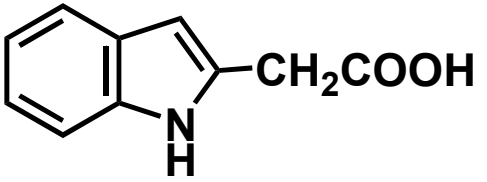
**N.Dharmasiri, S.Dharmasiri, M.Estelle,**
**Nature 435, 441 (2005)**

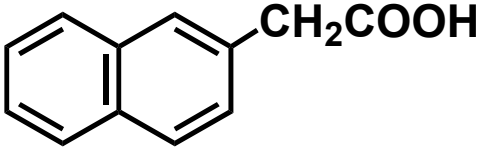
# Structure and Plant Growth Activity

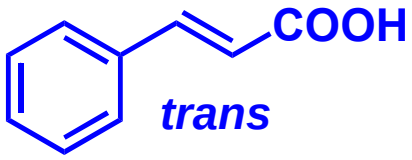
## Highly Active and Active

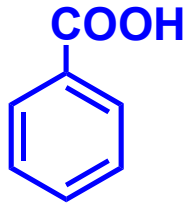



## Weakly active and Inactive



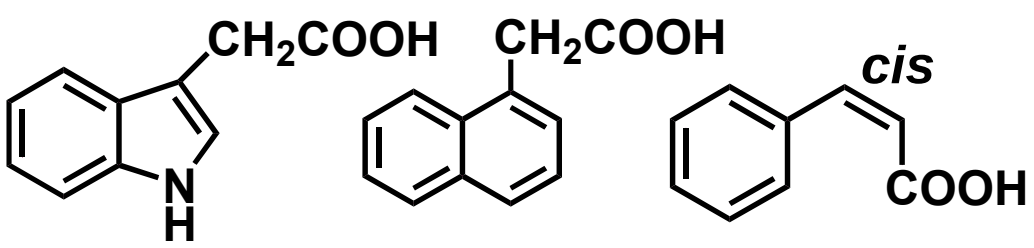



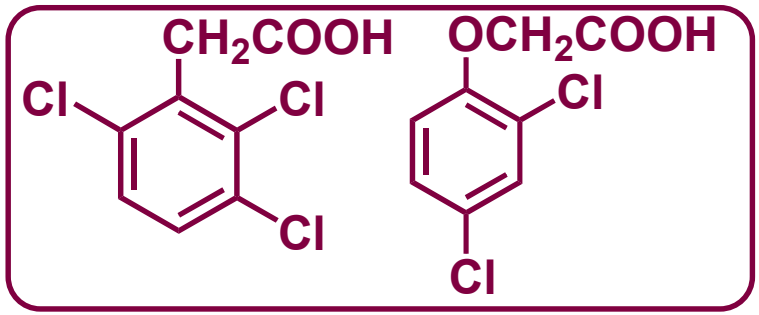




(COOH and Ring : Nearly Coplanar)

# SAR (Qualitative "Rules" ) for Plant Growth Regulators/Herbicides




1) Ring Structure with Unsaturated Bonds

*2) Carboxyl Group Separated by at least One Carbon Atom from the Ring

*3) A Free Hydrogen Atom at the $\alpha$-Position to the Carboxyl Group

4) Specific Spatial Relationship between the Ring and the Carboxyl Group

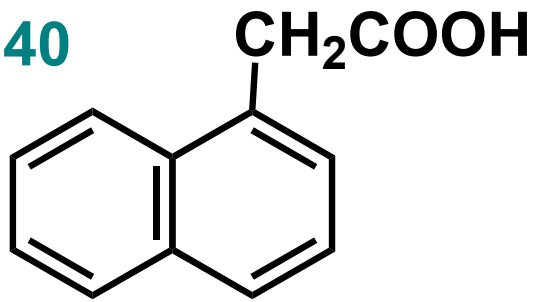
(Koepfli, Thimann, Went 1938)

* ammended subsequently

**Substituted phen(ox)yl acetic acids as potent regulators/herbicides were discovered after 1940.**

# Discovery of Agricultural Herbicides at ICI

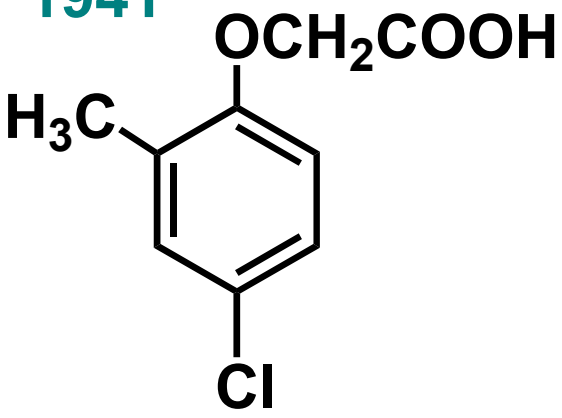
1940



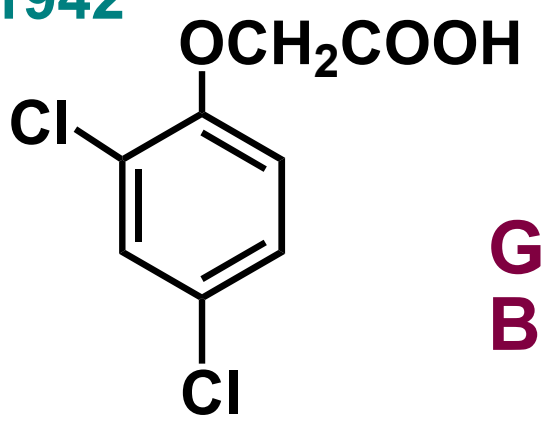

27 kg/ha

Oats *vs*. yellow charlock

1941




1942



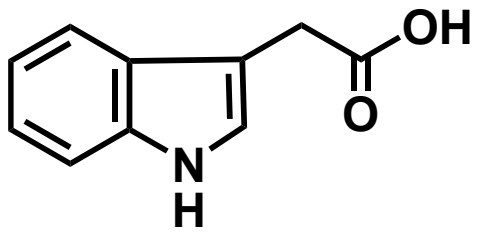

ca. 0.5 kg/ha

Grass crops *vs.*
Broad leaf weeds

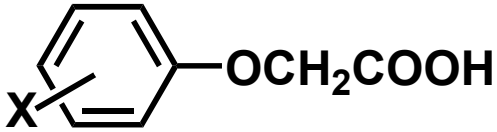
**Selective Toxicity eradicating weeds to protect crops.**

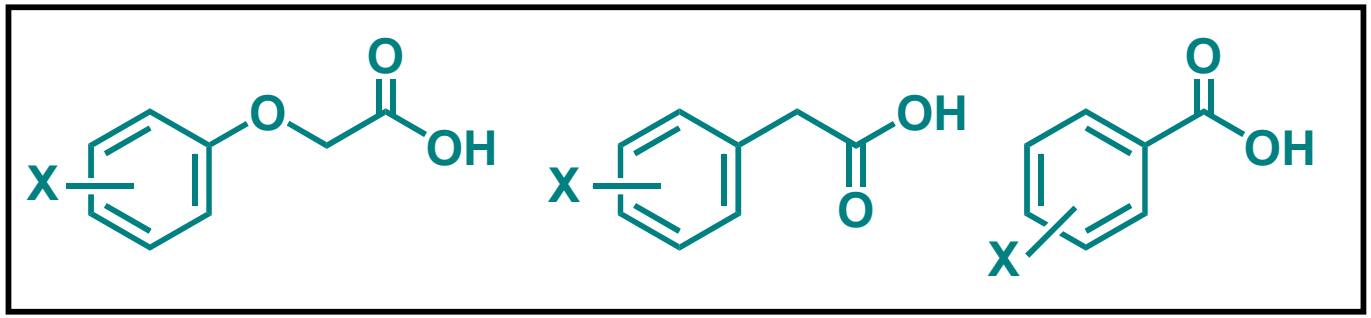
# Origin of Classical QSAR is from the SAR Studies of Agrochemicals—— Plant Growth Regulators/Herbicides

**Indole-3-acetic acid**

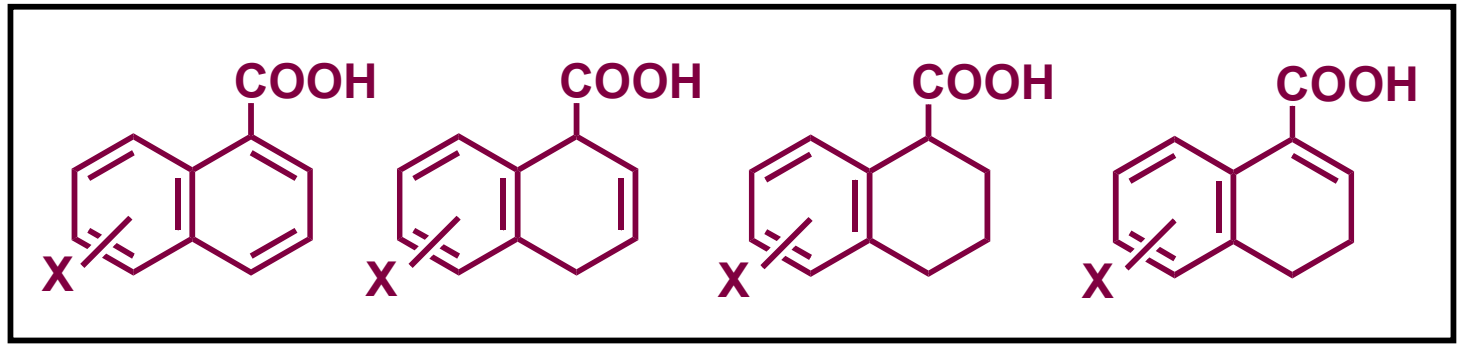



**Phenoxyacetic acids**




**Hansch and Muir at Pomona (1946 ~ )**




**Mitsui, Fujita and others at Kyoto (1950 ~ )**

# Substitution Patterns and the Activity

| | Very Active ~ Active | Inactive |
|---|---|---|
| Benzoic Acids | 2,3,5-$I_3$, 2,3,6-$Cl_3$, 2,5-$Cl_2$-6-OMe 2,5-$Cl_2$-3-$NH_2$, 2,5-$Cl_2$, 2,4,5-$Cl_3$-3-$NH_2$-6-Aza, 2,3-$Cl_2$, 2-Br, 2-Cl, 2-$NO_2$, 2-F-6-Cl, 2,6-$Cl_2$, 2-Cl-5-F. | H, 3- and 4-Halogeno, 3- and 4-$NO_2$, OH, $NH_2$, 2,4-$Cl_2$, 3,4-$Cl_2$, 3,5-$Cl_2$, 2,6-$OMe_2$, 2-F-5-Cl, and many more. |
| Phenyl-acetic Acids | 2,3,6-$Cl_3$, 3-Halogeno, 2,3-$(CH)_4$, 3-$CF_3$, 2- and 3-Br, 2- and 3-Cl, 2- and 3-Me, 2- and 3-CN 2- and 3-OMe, 3-$NO_2$, H. | 4-CN, I, $NO_2$, 4-$COCH_3$, 4-$NHCOCH_3$ and many more. |
| Phenoxy-acetic Acids | 2,4-$Cl_2$, 2-Me-4-Cl, 3,4-$(CH)_4$, 2,4,5-$Cl_3$, 3-$CF_3$, 4-Cl, 3-I, 4-F, 3-Br, 2,5-$Cl_2$, 3-Cl, 3-OMe, 3-Me, 3-Et, H. | 2,6-$Cl_2$, 4-Me, I, $NO_2$, 4-Ac, $COCH_3$, Et, CN, 3-$NHCOCH_3$, Bu, 2,3-$(CH)_4$, and many more. |

# Structure-Activity Patterns

**Importance of 2- and 2,6-Substitution Patterns favorable to the High Activity.**

**2,6-Disubstitution in Benzoic and Phenylacetic but not in Phenoxyacetic Acids, 2,6-$(OMe)_2$ being unfavorable.**

**Electron-attracting Cl substitutions are favorable to the activity depending on structures and ring positions.**

**Importance of 2- and 2,6-Substitution Patterns favorable for High Activity**

## Proposal of H. Veldstra (Leiden University)

(1) "Lipophilic" Ring with Multiple Substitutions
(2) Steric Repulsion of "Proximity Substituents" toward Carboxyl group so as to deviate from the Coplanarity

⇩

(1) A Basal Ring System (Nonpolar L-moiety) with a High Interface Activity.
(2) A Carboxyl Group (or surrogates) (Polar H-moiety) situated with respect to the Ring System as Peripheral and Perpendicular as Possible with an Optimum H/L Balance.

**Reasonably Hydrophobic Ring System (with Substituents) and Steric Conditions (Noncoplanar) of the COOH Group**

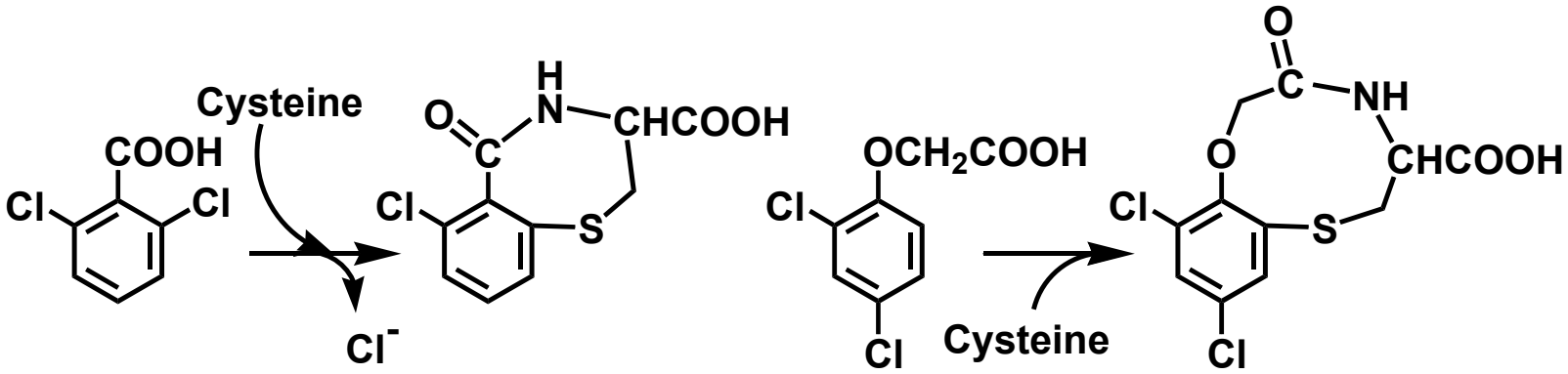
Veldstra (1953)

## Importance of 2- and 2,6-Substitution Patterns favorable for High Activity

# Proposal of C. Hansch (Pomona College)

(1) Activity Enhancements due to "Electron-Withdrawing" Substituents could be elucidated by Nucleophilic Attack of Biological Components.
(2) Importance of the "Reactivity" at One of the Ortho Positions.

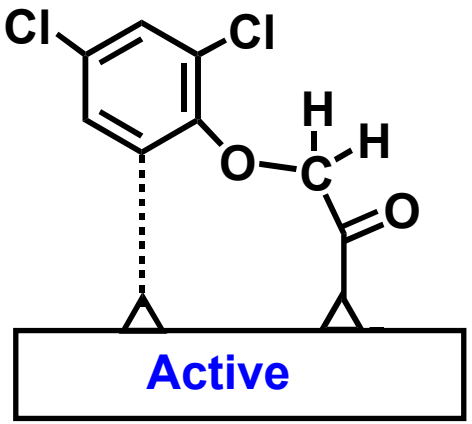
### Nucleophilic Susceptibility at the Ortho Position



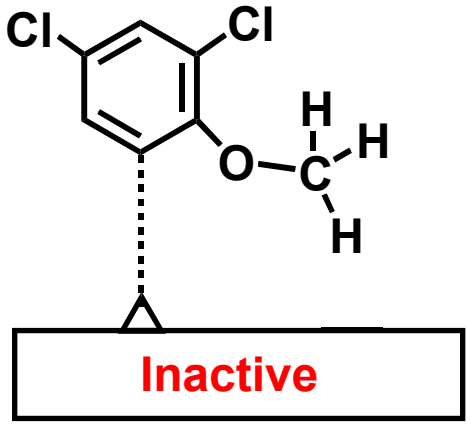

## Two point Attachment Hypothesis: The COOH Group and one of the Ortho Positions

Hansch et al. (1951)

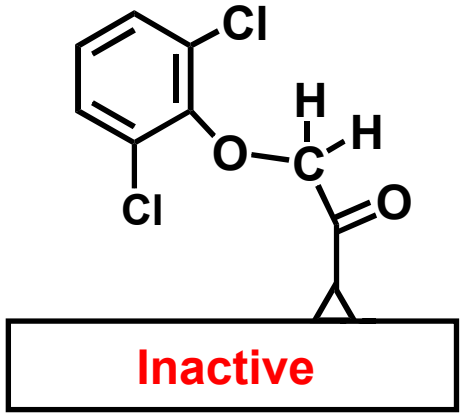
# Two Point Attachment Hypothesis McRae and Bonner (1952)



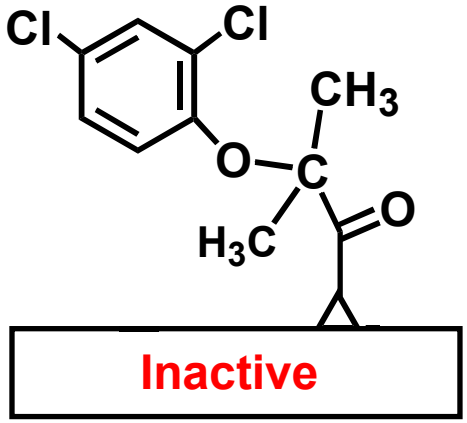

2,4-$Cl_2$ Phenoxyacetic Acid (2 Point Attachment)




2,4-$Cl_2$ Anisole (Ortho only)




2,6-$Cl_2$ Phenoxyacetic Acid (COOH only)




2,4-$Cl_2$ Phenoxy-*i*-butyric Acid (Blockage of Attachment)

## Susceptibility Index to Nucleophilic Attack to the Ortho Position and Plant Growth Activity of Substituted Benzoic Acids

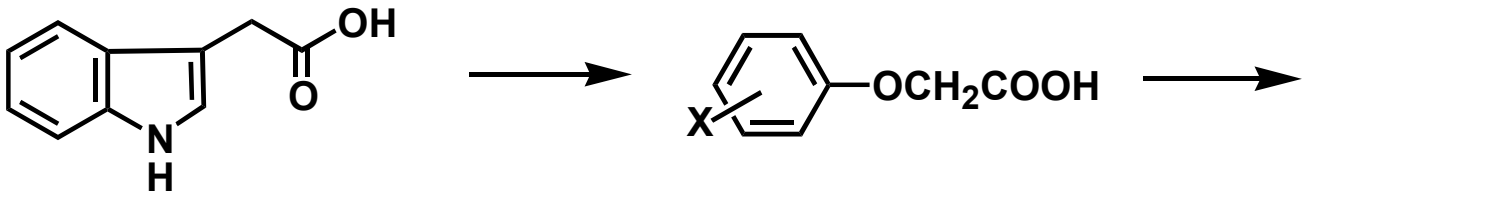
| | $S'_r(N)$ | Activity | | $S'_r(N)$ | Activity |
|---|---|---|---|---|---|
| 2,5-$Cl_2$ | 0.777 | ++ | 3-Cl | 0.643 | Inactive |
| 2,4-$Cl_2$ | 0.719 | Inactive | 3-F | 0.636 | Inactive |
| 2,3,6-$Cl_3$ | 0.359 | ++++ | 3-OH | 0.635 | Inactive |
| 2-Cl | 0.709 | + | 3-Br | 0.633 | Inactive |
| 2,3,5-$Cl_3$ | 0.707 | ++++ | 3-I | 0.631 | Inactive |
| 2,4,6-$Cl_3$ | 0.692 | Inactive | 4-Cl | 0.620 | Inactive |
| 2-Br | 0.683 | + | 4-I | 0.616 | Inactive |
| 2-I | 0.681 | + | 4-Br | 0.614 | Inactive |
| 2-Br-6-Cl | 0.678 | + | 2-OH | 0.613 | Inactive |
| 3-Me | 0.671 | Inactive | 2-F | 0.610 | Inactive |
| 2,6-$Cl_2$ | 0.669 | + | 3,4,5-$I_3$ | 0.608 | Inactive |
| 2,6-$Br_2$ | 0.650 | Inactive | H | 0.602 | Inactive |
| 2-Me | 0.648 | Inactive | 2-$NH_2$ | 0.601 | Inactive |

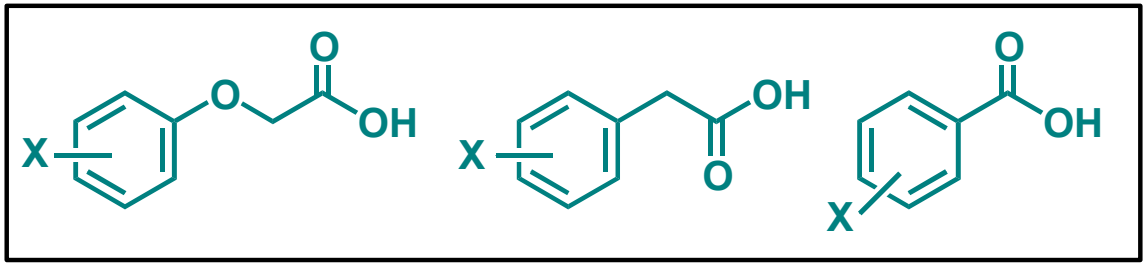
$S'_r(N)$ : Superdelocalizability of the frontier MO for susceptibility to "nucleophilic attack".

Fukui, Nagata et al. (1958)

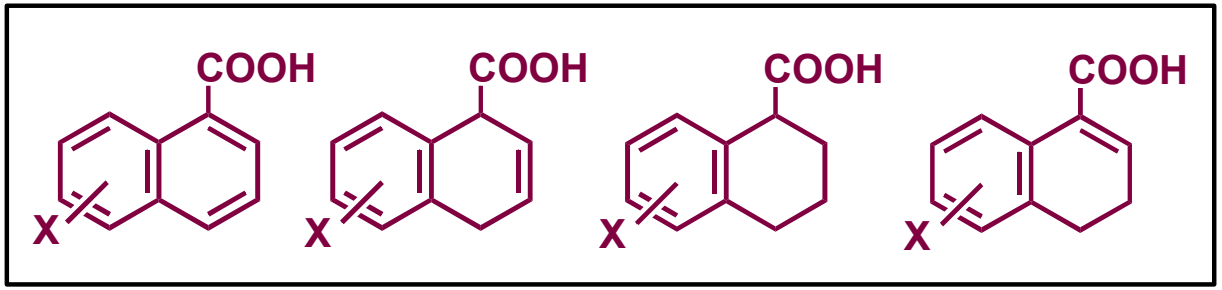
# Origin of Classical QSAR is from the SAR Studies of Agrochemicals—— Plant Growth Regulators/Herbicides

**Indole-3-acetic acid** → **Phenoxyacetic acids**



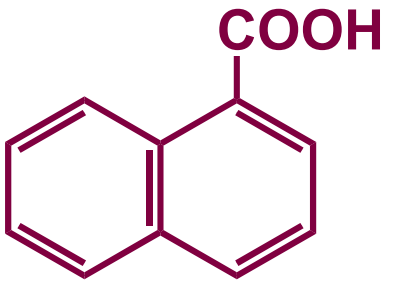



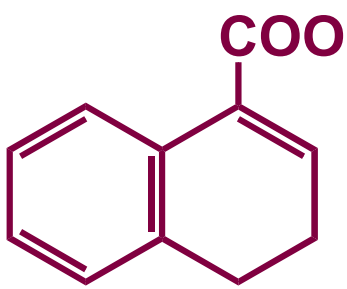

Hansch and Muir at Pomona (1946 ~ )




Mitsui, Fujita and others at Kyoto (1950 ~ )

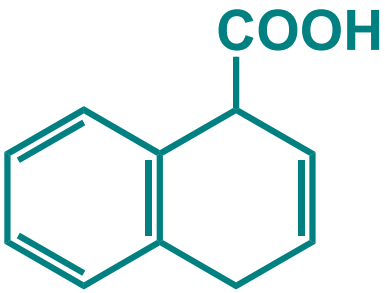
# Activity Profile of (Hydro)-1-Naphoic Acids

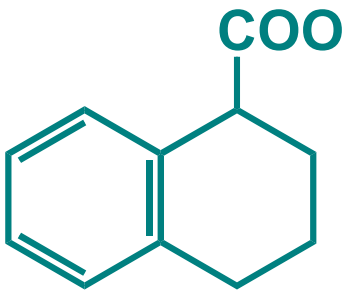


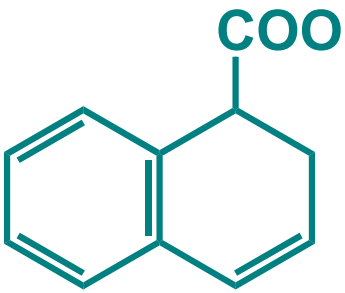




"Violation" of Rules #2 and 3.

Only weakly active.




Being regarded as $\alpha$-alkyl substituted phenylacetic acids in which the side chain is cyclized.

Highly ~ Moderately Active

Effect of substituents on the aromatic moiety.

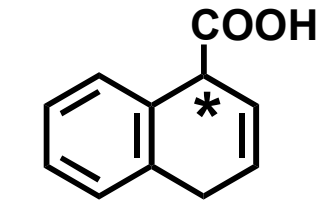
# Plant Growth Activity of Partially Hydrogenated 1-Naphthoic Acids

Fujita, Mitsui et al. (1951~ )

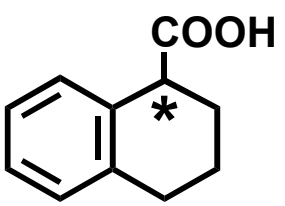
Bioassay: Straight Growth of Pea Sprout Stem

The S form (*) is more Active than L.

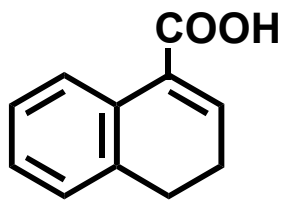
Activity seems to parallel with the angle (or Non-coplanarity) of the "COOH" relative to the ring-plane.



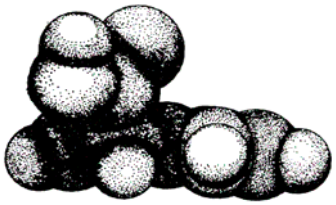

1,4-$H_2$

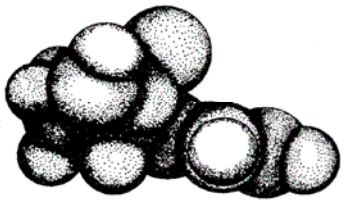


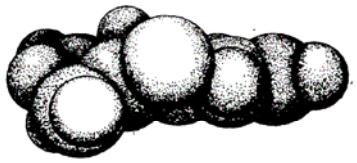

1,2,3,4-$H_4$

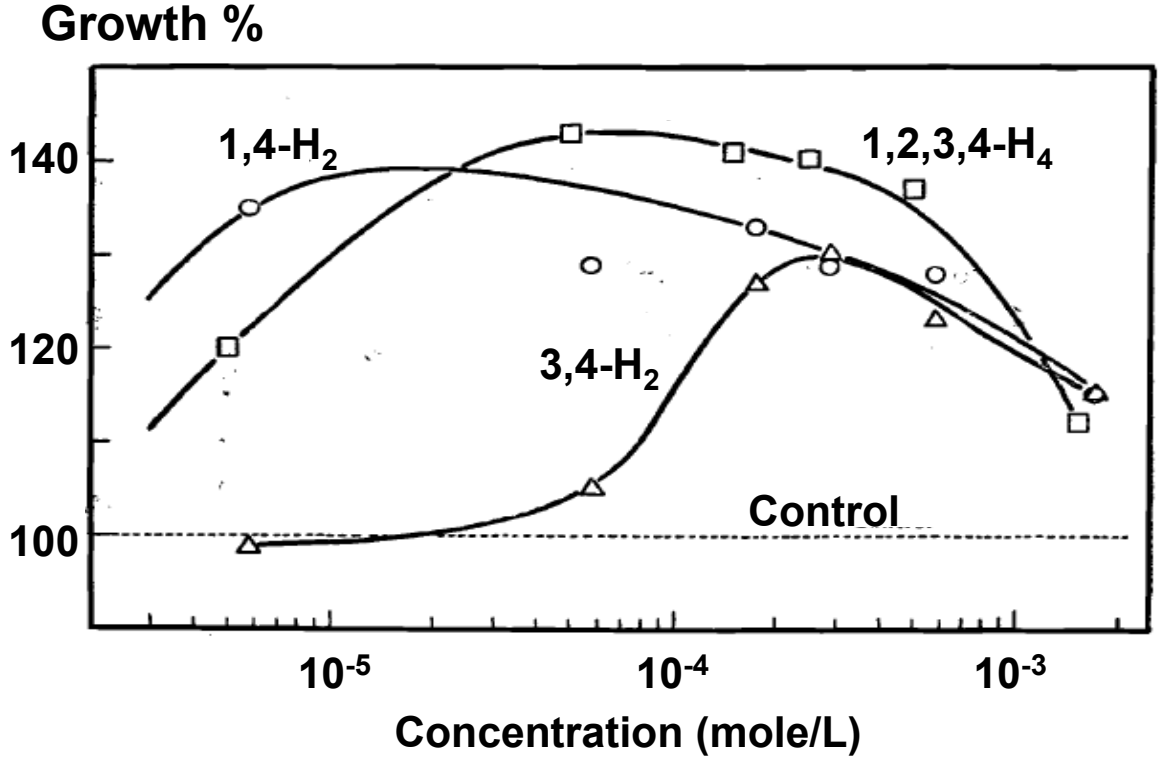



3,4-$H_2$





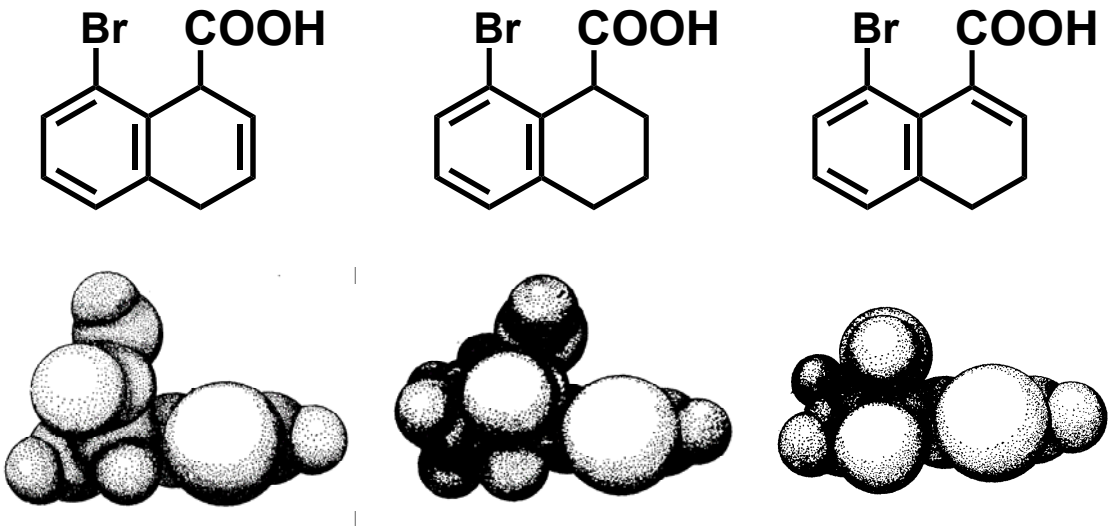


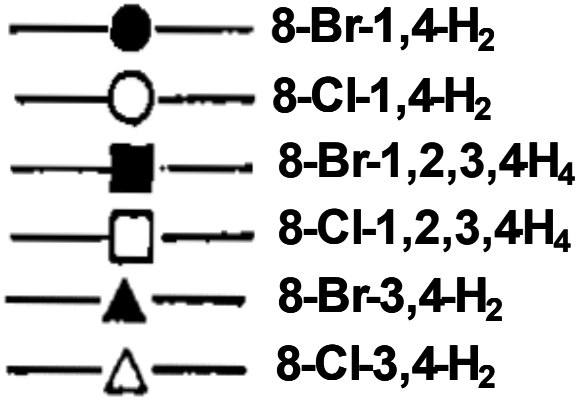


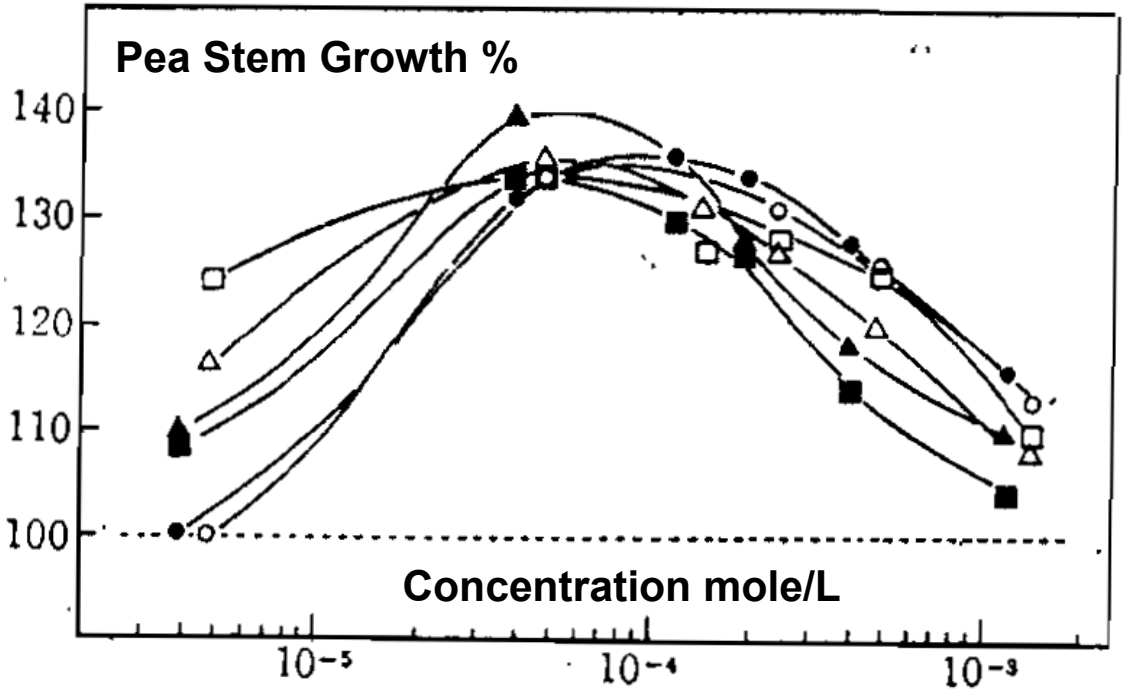
# Plant Growth Activity of 8-Halogeno-Hydro-1-Naphthoic Acids

Fujita et al. (1961 )

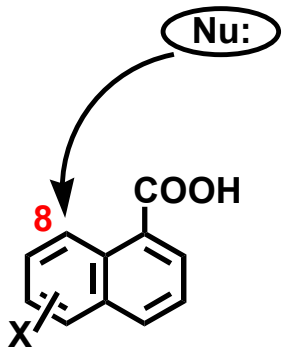
**Substitution at peri (8)-position causes repulsion or twisting of COOH.**



## Susceptibility Index to Nucleophilic Attack and Plant Growth Activity of Substituted 1-Naphthoic Acids




| | $S'_r(N)$ | Activity | | $S'_r(N)$ | Activity |
|---|---|---|---|---|---|
| 5-$NO_2$ | 1.384 | + | 6-Cl | 0.216 | + |
| 4-$NO_2$ | 0.699 | + | 6-Br | 0.214 | + |
| 8-Cl | 0.359 | +++ | 8-Me | 0.213 | +++ |
| 8-Br | 0.345 | +++ | 3-Cl | 0.212 | Inactive |
| 2-Cl | 0.331 | +++ | 3-Br | 0.212 | Inactive |
| 5-Br | 0.279 | Inactive | H | 0.211 | ++ |
| 5-Cl | 0.278 | Inactive | 3-Me | 0.207 | Inactive |
| 2-Me | 0.254 | + | 6-Me | 0.201 | Inactive |
| 4-Cl | 0.244 | + | 4-Me | 0.148 | Inactive |
| 4-Br | 0.234 | + | 6-$NO_2$ | 0.139 | Inactive |
| 8-$NO_2$ | 0.217 | + | 3-$NO_2$ | 0.110 | Inactive |

S'r(N) : Superdelocalizability for Nucleophilic Susceptibility in the frontier orbital; Calculated with Simple Hueckel LCMO method at the 8-position, The result was taken to indicate not necessarily nucleophilic substitution but an interaction due to charge-transfer is possible.
Twisted conformation ($60^o$) of COOH and $NO_2$ for 2- and 8-positions, and coplanar conformation for others.

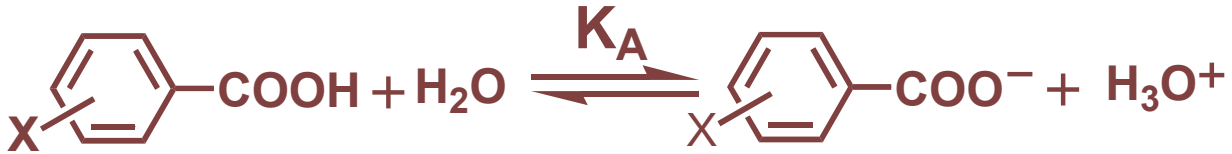
Koshimizu and Fujita (1960)

# The "Birth" of the Multi-variable Approach (I)

(1) Variations in the activity are obviously governed by not a single but plural physicochemical parameters at the same time.

(2) Most of growth regulators are aromatic, and the potency varies depending upon the substituent effects. Then, one of parameters determining the potency variations can be the Hammett σ constant representing the electron withdrawing character of substituents.

Definition of the Hammett σ :




$$\sigma = \log K_A^X - \log K_A^H$$

(3) Other possible parameters to be considered are those for steric and hydrophobic (lipophilic).

# The "Birth" of the Multi-variable Approach (II)

**(1) To represent the hydrophobicity of the entire molecule, log P, P being the partition coefficient measured with the 1-octanol/water system was selected.**

**(2) In aliphatic systems, variations in the reactivity (k) of a series of compounds is often formulated in a form of linear combination of free-energy-related parameters.**

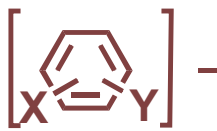
$$\log k = \rho\sigma^* + \delta Es \ ,$$

**Es: Taft steric parameter,**

**$\sigma^*$ : Aliphatic counterpart of the Hammett $\sigma$,**

**$\rho$, $\delta$ are the coefficient of respective terms.**

**(3) Definition of Hydrophobicity Substituent Constant $\pi$:**

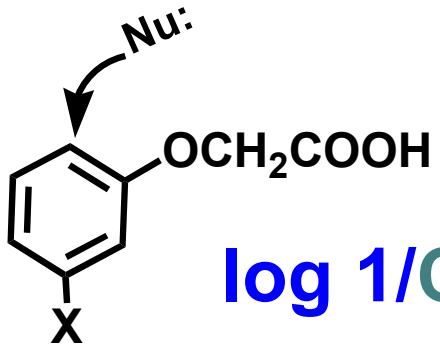
$\pi$ = log P of  − log P of 

**$\pi$ of certain X substituents varies with Y to some extent.**

**(4) Linear combination of $\sigma$ and $\pi$ parameters for the analysis.**

$$\log 1/C = a\pi + \rho\sigma + \text{constant}$$

# Plant Growth Activity of "*m*"-Substituted Phenoxyacetic Acids




$$\log 1/C = -1.97\,\pi^2 + 3.24\,\pi + 1.86\,\sigma_p + 4.16$$

$$s = 0.484,\ r = 0.881,\ r^2 = 0.776$$

**The first QSAR equation in JACS (1963)**

*n* = 21 including 17 meta-substituted derivatives ($CF_3$, 4-Cl, I, 4-F, Br, $SF_5$, Cl, $NO_2$, SMe, Et, $SCF_3$, 3,4-$(CH)_4$, OMe, Me, CN, Pr, 4-OMe, Ac, F, H)

C: Molar concentration producing 10% increase above control in the length of 3mm oat seedlings

**Hansch et al. (1963)**

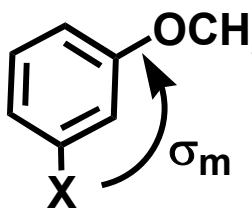
## Features of the QSAR

(1) The $\sigma_p$ parameter works much better than $\sigma_m$.
(2) There is "an optimum" in the $\pi$ parameter for substituents expresssible by the parabola model.

## Conditions for the QSAR

(1) 4-Substituted derivatives with substituents larger than Cl is inactive.
(2) 3-Substituted derivatives with substituents larger than Pr is inactive.
(3) There is a limit for 3,5-disubstitution in "lateral" width.
(4) The proximity effects of ortho-substituents were not estimable at that time in terms of Hammett-Taft type parameters.
(5) The set of meta plus a few para substituted derivatives was only a set of compounds of which activity was accurately measurable.

# Plant Growth Activity of *m*-Substituted Phenoxyacetic Acids (II)




Revision by Verloop (1981)

$$\log 1/C = 1.04\pi + 0.59\sigma_m - 0.67(L)^2 + 4.78\ L - 3.87$$

*n* = 19 (not including 4-isomers), *s* = 0.376, $r^2$ = 0.874

Revision by Hansch (1995)

$$\log 1/C = 1.25\pi + 0.97\sigma_m + 0.95\ L - 5.54 \log(\beta\ x10^{L} + 1) + 1.39$$

*n* = 19 (not including 4-isomers), *s* = 0.242, $r^2$ = 0.951, L(opt) = 3.74

(1) The $\sigma_m$ parameter works much better than $\sigma_p$ by considering the L parameter along with its optimum for meta substituents.
(2) The optimum can be represented either by parabola or the Kubinyi bilinear model.
(3) The hypothesis of the two-point attachment involving one of the ortho positions was abondoned.
(4) Hydrophobic, electronic, and steric effects of meta substituents are nicely separated in this revised equation.
(5) 'L' is the length parameter, one of the Verloop STERIMOL parameters.
(6) L(opt) = 3.74 corresponds with OMe, beyond which activity falls off sharply.

# Classical QSAR for Series of Substituted Analogs

$\Delta(\text{Biological Response}) = f(\Delta E_1, \Delta E_2, \Delta E_3, \bullet\bullet\bullet\bullet)$

$E_n$: Various "Free-Energy Related" Physicochemical Parameters

$$\log(1/C) = a\pi + \rho\sigma + \delta E_s + \bullet\bullet\bullet\bullet + \text{constant}$$

C : $EC_{50}$, $LD_{50}$, $I_{50}$, etc. in Molar basis.

When an optimum value exists for certain parameters, parabola or bilinear model can be used.

# Process of the Emergence of Biological Activity (1)

**When the rate of emergence is slow and the drug concentration in the transport and receptor-binding processes is lower than that (C) at the site of administration:**

$$d(\text{Response})/dt = BKkC$$

**C : concentration applied (almost unchanged)**

**K : “equilibrium” constant within transport process**

**k : the rate constant of the rate-determining step**

**B : proportional factor**

# Process of the Emergence of Biological Activity (2)

**When the rate of transport is quick and the drug concentrations at the site of administration and the "receptor" is in a state of (pseudo)equilibrium :**

**Response = BKC**

**C : "equilibrium" concentration**

**K : "equilibrium" constant**

**B : proportional factor**

## Biological Endpoint as $I_{50}$, $LD_{50}$, $EC_{50}$, MIC, etc.

**Defined as the "concentration" inducing a constant response a certain time after the onset of the activity test.**

**For series of compounds:**

$$d(\text{Response})/dt = BKkC = \text{constant}$$

$$\text{Response} = BKC = \text{constant}$$

$$BK'C = \text{constant} \quad (K' = K \text{ or } Kk)$$

$$\text{Log } (1/C) = \log K' + \text{constant}$$

**Log K' : Linear combination of terms attributed to effects of electronic, steric, hydrophobic, etc.**

# Early Trials of the Quantitative Approach (1)

## "Narcotic" Activity of Tadpoles (Meyer, Hemmi 1935)

## "Narcotic" : Paralyzation without Motion

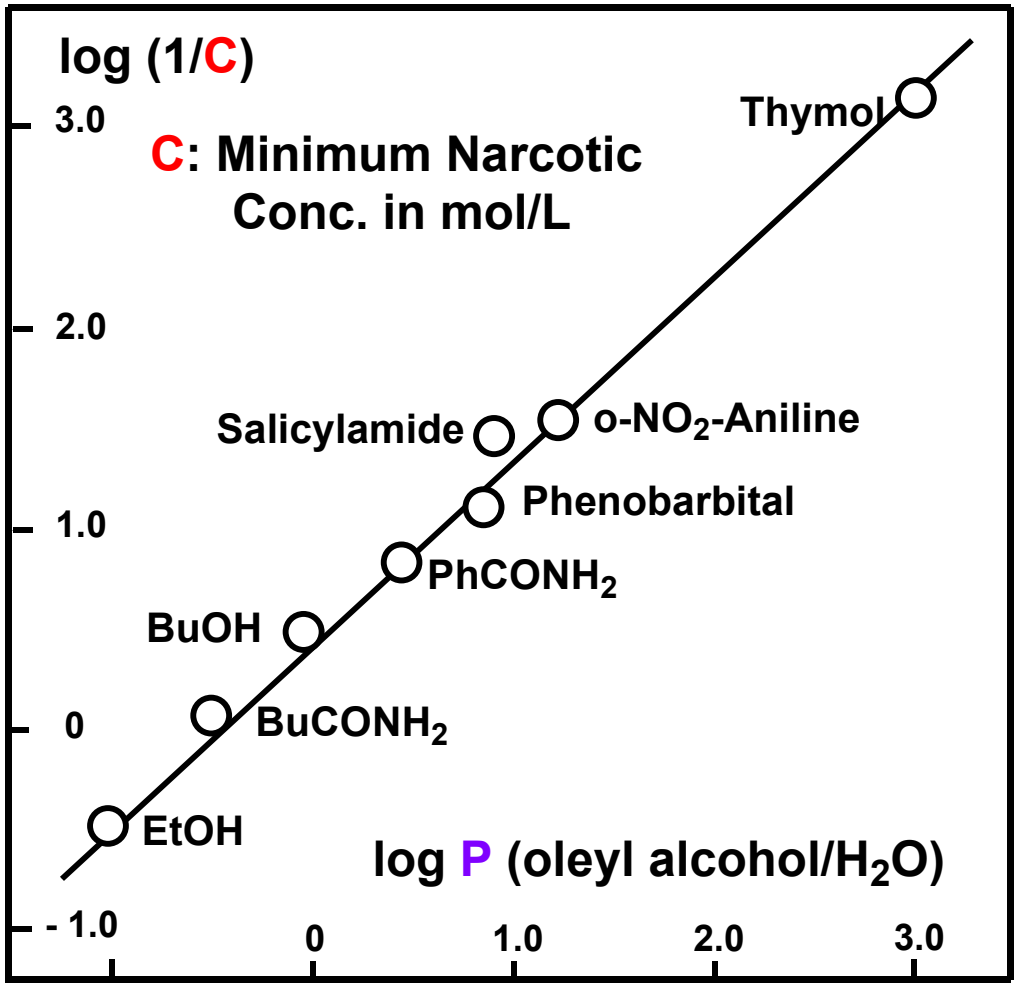
**Nonspecific Toxicity of Miscellaneous Compounds:**

**C x P = constant**

⇩

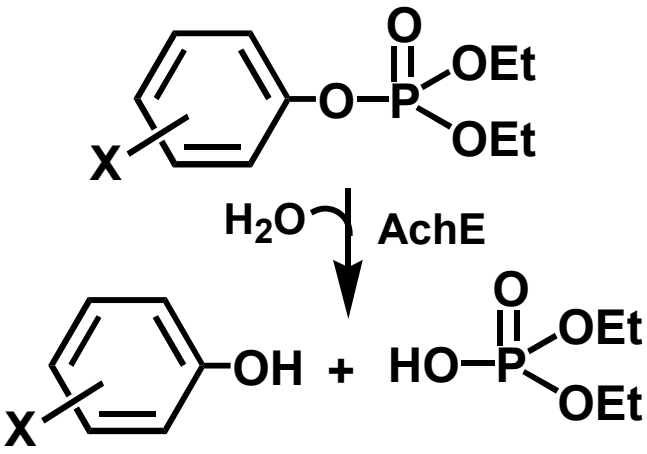
**log(1/C) = log P + constant'**

**Slope of log P ≈ 1.0**



# Early Trials of the Quantitative Approach (2)

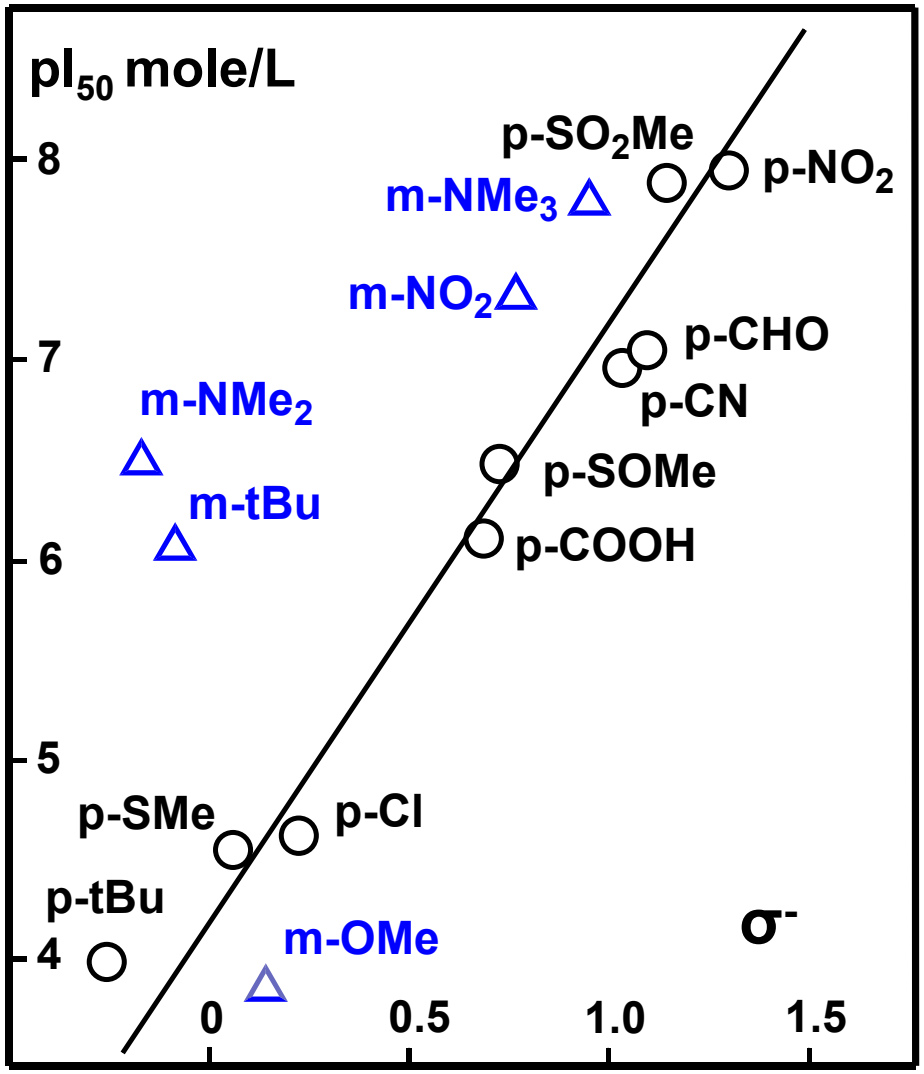
## Anti-ACh-Esterase Activity of Phenyl Diethyl Phosphates



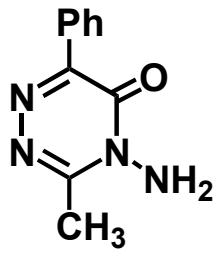

**Effect of Para Substituents:**
**Linear to $\sigma^-$ Parameter**

**Effect of Meta Substituents:**
**Participation of Others**

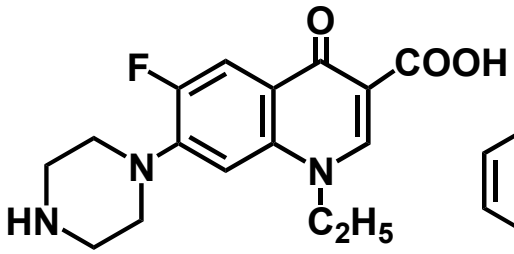
**Fukuto and Metcalf (1956)**



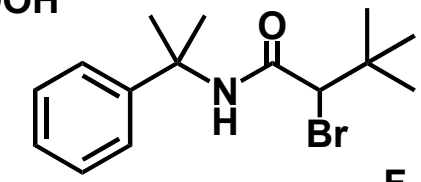
# Commercialized Drugs developed with the Aid of Classical QSAR



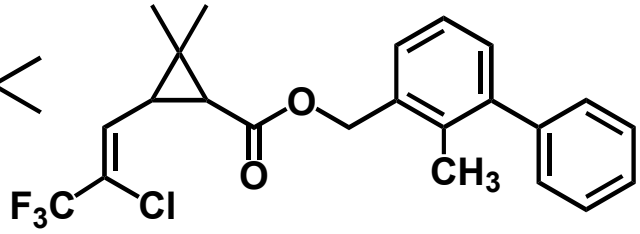

**Metamitron**
(Sugar-beet Herbicide)
Bayer 1975



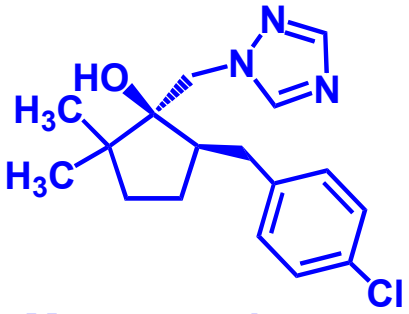

**Norfloxacin**
(Antibacterial)
Kyorin 1983



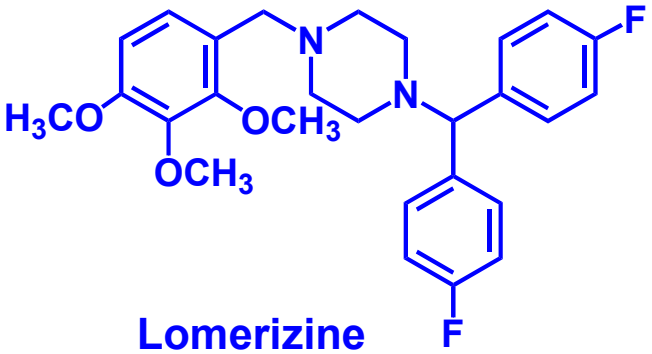

**Bromobutide**
(Paddy Field Herbicide)
Sumitomo 1984



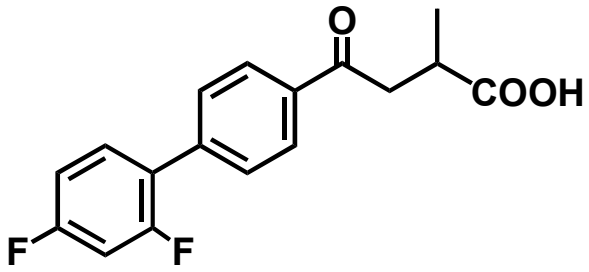

**Bifenthrin**
(Foliar Insecticide)
FMC 1984




**Metconazole**
(Wheat Fungicide)
Kureha 1994




**Lomerizine**
(Antimigrane, Antiglaucoma)
Organon Japan-Upjohn 1999




**Flobufen**
(Long-acting Antiinflammatory)
Kuchar et al.-Virbac 2000