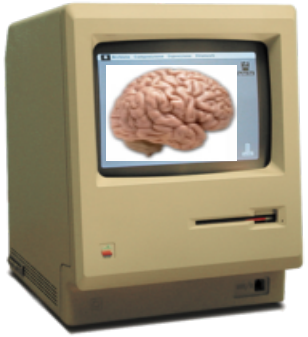
Descriptors for MAGICIAN

MAterials **G**enome/**I**nformatics and **C**hemo-**I**nformatics
Activate **N**etworks

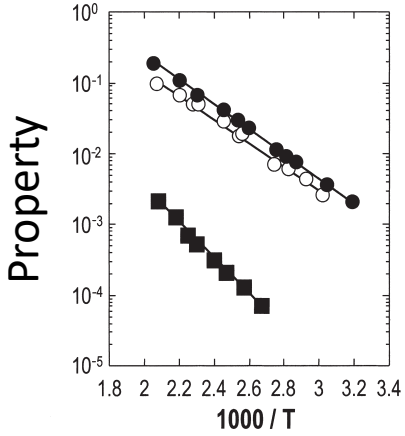
We have so small data for Materials

Machine Learning



Accurate
Estimated Value

Optimization by genome analysis

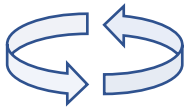


Evaluate Gene

Build Property Estimation Scheme

Genetic Algorithm method

Database



No data!

Second Generation

A	B	C	O
A	B	C	O
A	B	C	O
A	D	C	O

Who is Winner?

Winner

First generation

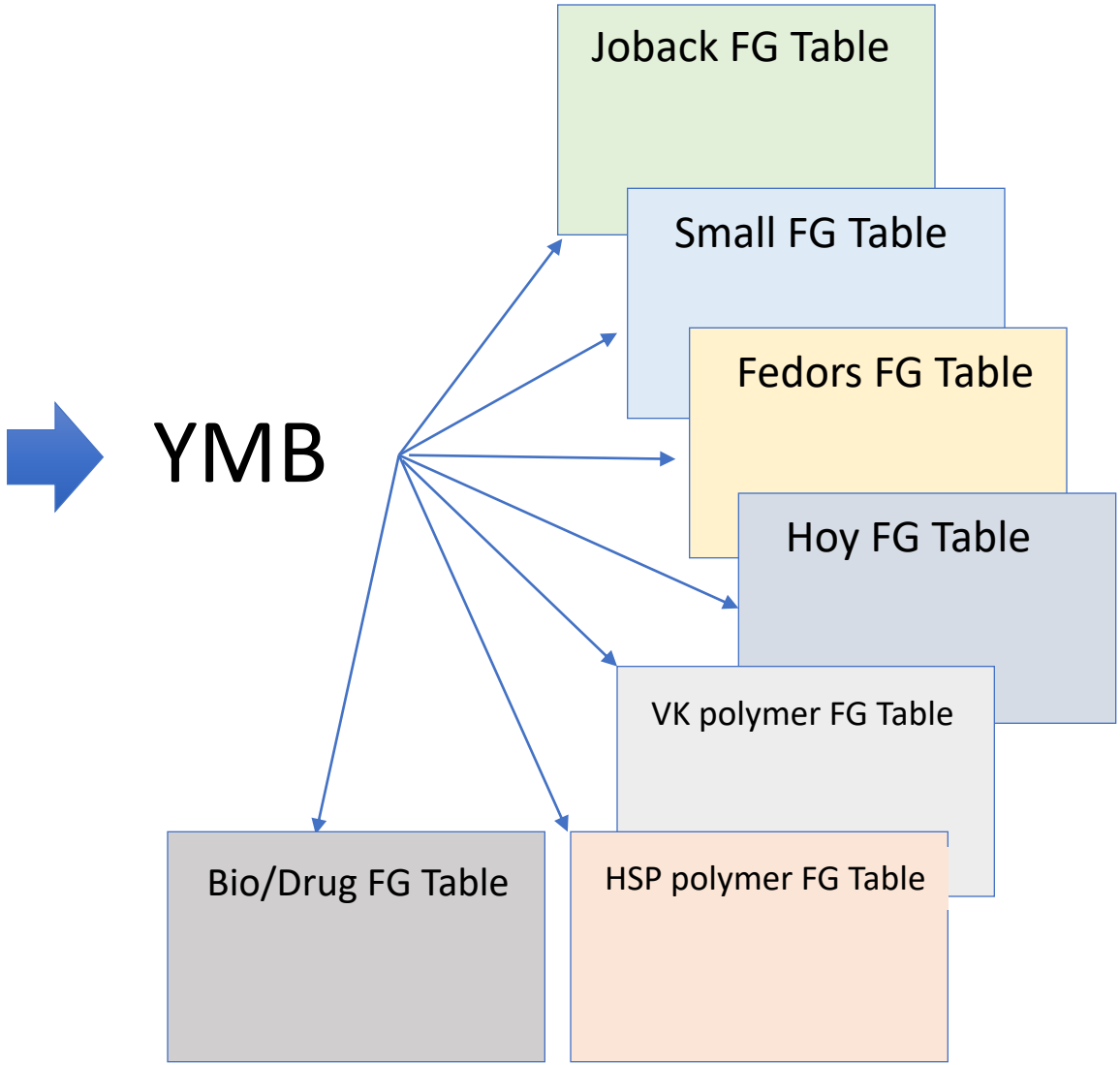
A	B	C	O
A	B	C	O
A	D	C	O
A	D	E	O

Have Property

Dynamic “Functional Groups” Generator -YMB-

Smiles
C(F)(F)C(F)(F)OCF
FC(F)C(F)(F)OCF
FCOC(F)(F)C(F)(F)
O(CF)C(F)(F)C(F)F

Polymer Smiles
XCCX
XC(F)(F)C(F)(F)X



If we have FG Table with Properties,

```
278.74=1*CH3+                    1*CF3        +1*CF2
326.79=        1*CH2F+1*CHF2                  +1*CF2
299.32=                1*CHF2+1*CF3 +1*CH2
300.15=                1*CHF2+1*CF3 +1*CH2
316.20=         1*CH2F+1*CHF2                 +1*CF2
```

Multiple Regression
Neural Network method
Support Vector Machine
Deep Learning

$$\begin{pmatrix} 278 \\ 326 \\ 299 \end{pmatrix} = \underbrace{\begin{pmatrix} 1, 0, 0, 1, 0, 1 \\ 0, 1, 1, 0, 0, 1 \\ 0, 0, 1, 1, 1, 0 \end{pmatrix}}_{\text{Matrix A}} * \begin{pmatrix} CH3 \\ CH2F \\ CHF2 \end{pmatrix}$$

Express simultaneous equations as matrix

$$A^{-1} * \begin{pmatrix} 278 \\ 326 \\ 299 \end{pmatrix} = A^{-1} A \begin{pmatrix} CH3 \\ CH2F \\ CHF2 \end{pmatrix}$$

Find the inverse matrix A^{-1} of the A matrix

Contributions for Property by each FG

$$= E \begin{pmatrix} CH3 \\ CH2F \\ CHF2 \end{pmatrix}$$

$$A^{-1}*A=E \qquad E= \begin{pmatrix} 1, 0, 0 \\ 0, 1, 0 \\ 0, 0, 1 \end{pmatrix}$$

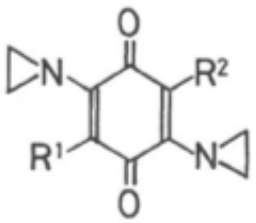
Bio/Drug properties Estimation

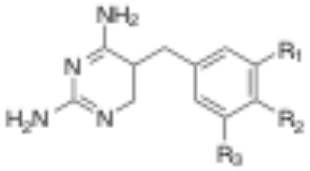
log BCF (Bio Concentration Factor)
log Kow(Octanol/Water Partition Coef.)
log S (Solubility to Water g/100g)
LD50, LC50

} Same with Group Contribution method

Drug Design

- Design of Lead compounds → difficult
- Modify lead compounds





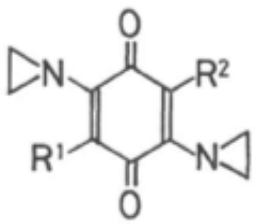
	R1			R1			
	CH3	OCH3	NH2	・・・	CH3	OCH3	NH2	・・・
	1	0	0	・・・	0	0	0	・・・
	0	1	0	・・・	0	0	0	・・・
	・・・・・・・・				・・・・・・・・			

Huge Column data!

Hansch Method reduce Column

π , MR, Hammet(Field, Resonance)

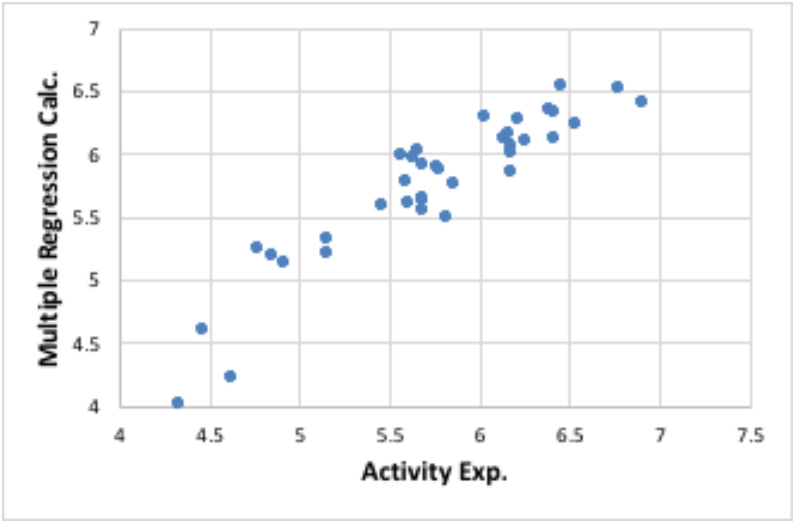
QSAR of carboquinone derivatives



NO	Activity	MR1,2	Pi1,2	Pi2	MR1	F	R	R1	R2
1		1.69	−0.05	−0.55	0.57	0.28	0.07	CH3	COCH3
2	4.33	5.08	3.92	1.96	2.54	0.16	−0.16	C6H5	C6H5
3	4.47	4.5	3.66	3.16	0.57	−0.08	−0.26	CH3	CH2CH2CH2Ph
4	4.63	4.86	5	2.5	2.43	−0.08	−0.26	C5H11	C5H11
5	4.77	3	2.6	1.3	1.5	−0.08	−0.26	CH(CH3)2	CH(CH3)2
6	4.85	3.57	2.51	2.01	0.57	−0.12	−0.14	CH3	CH2Ph
7	4.92	3	3	1.5	1.5	−0.08	−0.26	C3H7	C3H7
8	5.15	3.79	2.16	1.66	0.57	−0.04	−0.13	CH3	CH2OPh
9	5.16	6.14	0.72	0.36	3.07	−0.08	−0.26	CH2CH2OCON(CH3)2	CH2CH2OCON(CH
10	5.46	2.06	2	1	1.03	−0.08	−0.26	C2H5	C2H5
37	6.54	2.42	−0.32	−0.16	1.21	−0.08	−0.26	CH2CH2OH	CH2CH2OH
38	6.77	2.13	0.68	0.18	0.57	0.06	−1.05	CH3	N(CH3)2
39	6.9	2.47	−0.13	−0.63	0.57	−0.04	−0.13	CH3	CH(OCH3)CH2OH

Make table and solve with Multi Regression Method (You can do with Excel)

Activity
$= -0.02169*MR_{1,2} - 0.09473*\pi_{1,2}$
$-0.4152480*\pi_2 \; -0.274901*MR_1$
$-1.771798*F - 0.749842*R$
$+ 6.174982$



Quantitative
Structure Activity
Relationship

If you have new modifier's Hansch parameters, you can search more active Drug.

Descriptors for Drug design

Hansch Method

MO Base

ClogP Calc.
logKow

MO Calc.
HOMO
LUMO
Dipole moment
±Surface charge

Dragon
Descriptors Generator

Hydrophobicity (π)
Mol Refraction(MR)
F(Hammet Field)
R (Hammet Reso.)

1,2 Interaction
π_{12}, MR_{12}

These 2 parameters are
difficult to obtain

HSP Method

δD (Dispersion)
δP (Polality)
δH (Hydrogen Bonding)

δH_{edo} (Donnor)
δH_{eac} (Acceptor)

Mol Volume

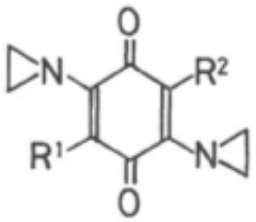
Mixing rule exist

Hansch method, HSP method

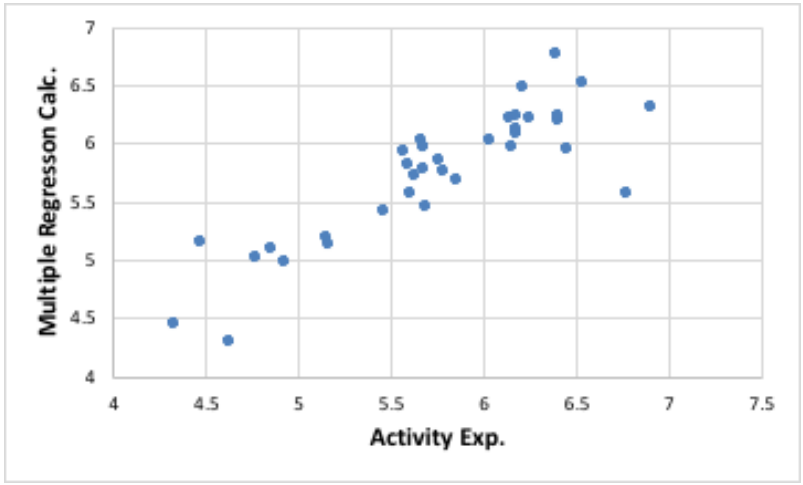
Activity is determined by Solubility
+
Fine tuning by Hammet or Do/Ac

You can calculate EVERY
Structure.
But with meaningless parameters

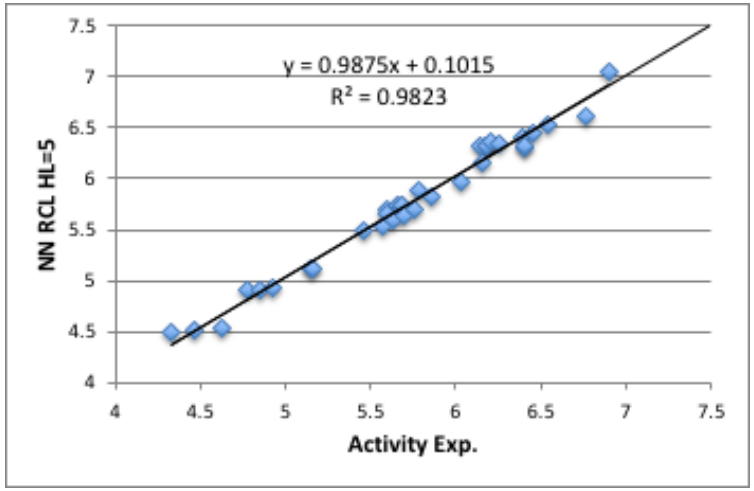
QSAR of carboquinone derivative with HSP



NO	Activity	Mol_Volume1	dD-1	dP-1	dH-1	Mol_Volume-2	dD-2	dP-2	dH-2
2	4.33	72.73	18.2	2.6	5.1	72.73	18.2	2.6	5.1
3	4.47	22.7	11.2	0.1	0.1	120.73	17.3	2.4	3.4
4	4.63	92.28	15	0.1	0.1	92.28	15	0.1	0.1
5	4.77	61.4	13.7	0.1	0.1	61.4	13.7	0.1	0.1
6	4.85	22.7	11.2	0.1	0.1	87.48	17.8	2.3	4.6

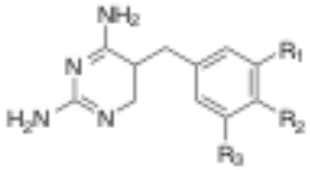


Multiple Regression



Neural Network method

QSAR of Pyrimidine derivatives

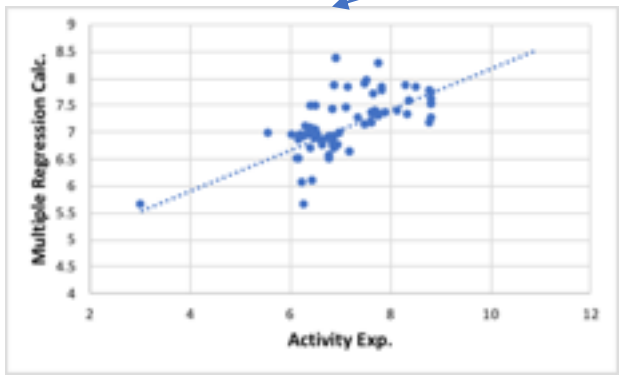


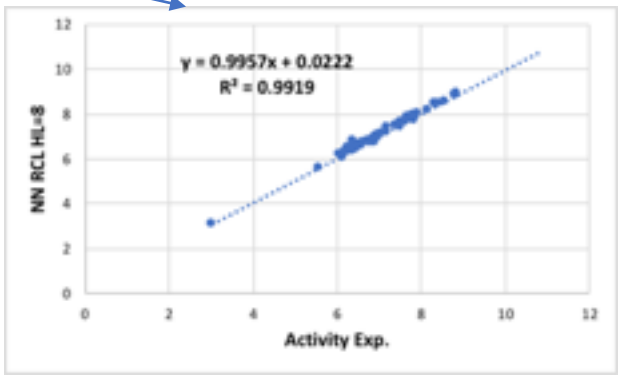
		R1				R2				R3							
No	Activity	Volume	dD33	dP33	dH33	Volume	dD33	dP33	dH33	Volume	dD33	dP33	dH33	MR	3	4	5
34	6.96	32.2	20.5	8.9	9	13.8	17.1	0.1	0.1	13.8	17.1	0.1	0.1	6.73611169	Br		
34	6.96	13.8	17.1	0.1	0.1	13.8	17.1	0.1	0.1	32.2	20.5	8.9	9	6.73611169			Br
53	8.85	32.2	20.5	8.9	9	105.9	18.5	4.5	5.3	32.2	20.5	8.9	9	7.22720083	Br	NH2	Br
37	7.02	44.98	10.8	4.2	4.2	13.8	17.1	0.1	0.1	13.8	17.1	0.1	0.1	6.93260575	CF3		
56	6.45	13.8	17.1	0.1	0.1	16.5	21.1	21.9	45.6	13.8	17.1	0.1	0.1	7.45572032		OH	
57	6.6	13.8	17.1	0.1	0.1	66	18	19.8	9.5	13.8	17.1	0.1	0.1	6.90903596		OSO2CH3	
59	6.89	13.8	17.1	0.1	0.1	105.9	18.5	4.5	5.3	13.8	17.1	0.1	0.1	6.70822422		OCH2Ph	
60	6.93	13.8	17.1	0.1	0.1	72.73	18.2	2.6	5.1	13.8	17.1	0.1	0.1	6.64600956		Ph	

Make Table

Multivariate analysis method

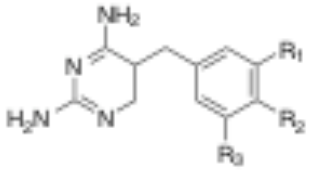
Multiple Regression
Neural Network Method
Support Vector Machine
Deep Learning





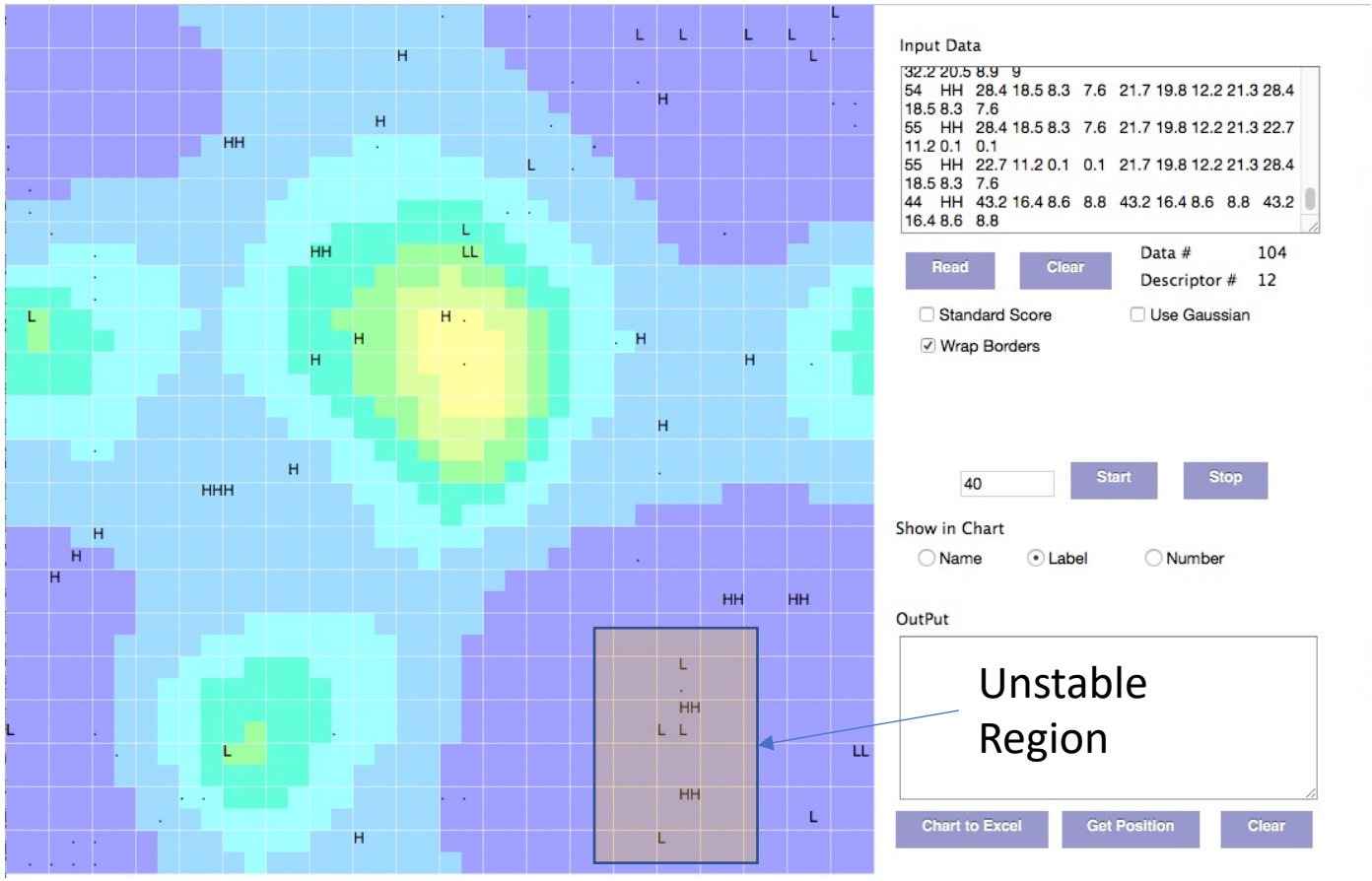
Every analysis method needs Descriptors.
If the modifier base descriptors are available,
"**Reverse Design**" of Drug become easy.

Self Organization Map (SOM) Neural Network



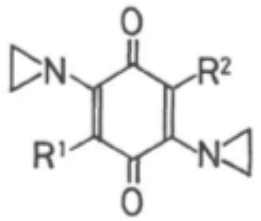
No	Activity	Volume	dD33	dP33	dH33	Volume	dD33	dP33	dH33	Volume	dD33	dP33	dH33	MR	3	4	5
34	6.96	32.2	20.5	8.9	9	13.8	17.1	0.1	0.1	13.8	17.1	0.1	0.1	6.73611169	Br		
34	6.96	13.8	17.1	0.1	0.1	13.8	17.1	0.1	0.1	32.2	20.5	8.9	9	6.73611169			Br
53	8.85	32.2	20.5	8.9	9	105.9	18.5	4.5	5.3	32.2	20.5	8.9	9	7.22720083	Br	NH2	Br
37	7.02	44.98	10.8	4.2	4.2	13.8	17.1	0.1	0.1	13.8	17.1	0.1	0.1	6.93260575	CF3		
56	6.45	13.8	17.1	0.1	0.1	16.5	21.1	21.9	45.6	13.8	17.1	0.1	0.1	7.45572032		OH	
57	6.6	13.8	17.1	0.1	0.1	66	18	19.8	9.5	13.8	17.1	0.1	0.1	6.90903596		OSO2CH3	
59	6.89	13.8	17.1	0.1	0.1	105.9	18.5	4.5	5.3	13.8	17.1	0.1	0.1	6.70822422		OCH2Ph	
60	6.93	13.8	17.1	0.1	0.1	72.73	18.2	2.6	5.1	13.8	17.1	0.1	0.1	6.64600956		Ph	

Activity<6.0
LL
6.0<Activity<6.5
L
6.5<Activity<7.72
“ ”
7.72<Activity<8.8
H
8.8<Activity
HH



Qualitative Analysis

Modifiers' descriptors generator



Modifyer
X-(CH2)2OCONH2
X-(CH2)3OCONH2
X-C2H5
X-C3H7
X-C5H11
X-C6H5
X-CH(C2H5)CH2OCONH2
X-CH(CH3)2
X-CH(CH3)CH2OCONH2
X-CH(OC2H5)CH2OCONH2
X-CH(OCH2CH2OCH3)CH2OCONH2
X-CH(OCH3)C2H5
X-CH(OCH3)CH2OCONH2
X-CH(OCH3)CH2OH
X-CH2CH(CH3)OCONH2
X-CH2CH2CH2Ph

X:CH3, Phenyl

Node base
YMB.js

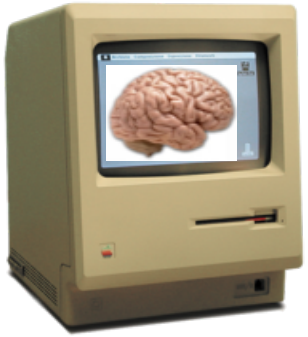
Calculate as Molecule.
Subtract CH3 or Phenyl group parameters

Modifier's HSP

Mol_Volume	dD	dP	dH
67.78	17.4	20.7	17.5
84.3	17.1	17.1	13.8
42.28	13.1	0.1	0.1
59.41	14.1	0.1	0.1
92.28	15	0.1	0.1
72.73	18.2	2.6	5.1
101.54	16.7	13.7	11.8
61.4	13.7	0.1	0.1
84.87	17.1	16.1	14.1
109.31	16.8	14.6	12
131.89	17	13.9	11.3
82.98	14.7	4.6	4.9
91.75	17.2	17.1	14.3
64.14	16.3	10.8	19.7
83.98	17.1	16.2	14.2
120.73	17.3	2.4	3.4

We have so small data for Materials

Machine Learning



Accurate
Estimated Value

Node base
YMB.js

Generate
modifier's
Structure

Generate
modifier's
Descriptors

Build
Property
Estimation
Scheme

Evaluate
Gene

Genetic Algorithm method

Database

Next Generation

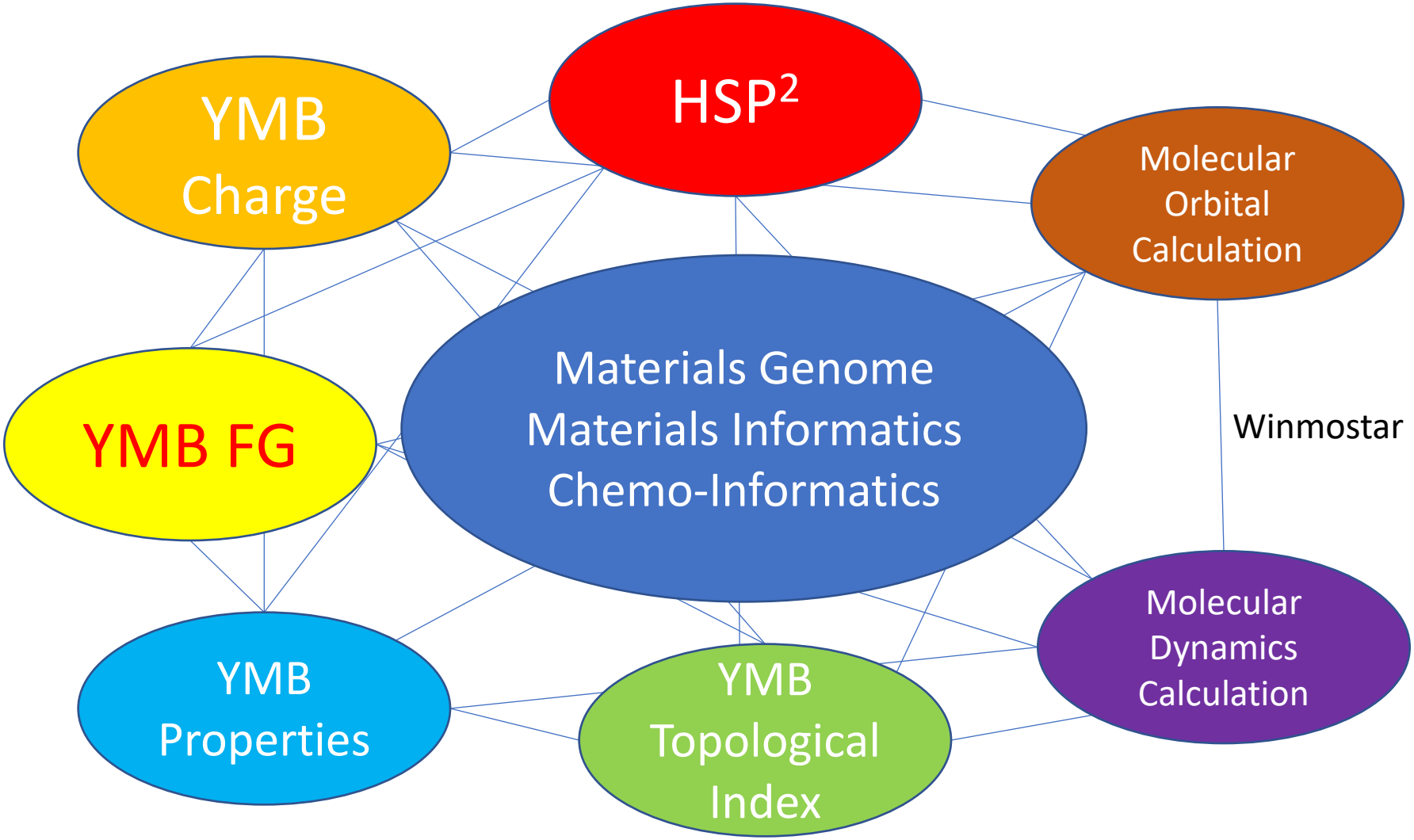
A	B	C	O
A	B	C	O
A	B	C	O
A	D	C	O

Winner

First generation

A	B	C	O
A	B	C	O
A	D	C	O
A	D	E	O

Descriptors for Magician



Not only properties Network but Researcher's Network