



спецкурс
“СУЧАСНІ КОМП’ЮТЕРНІ МЕТОДИ В ХІМІЇ”



ОСНОВИ QSAR

Дескрипторний метод опису молекулярної будови

В. В. ІВАНОВ

Materials Chemistry Department
V. N. Karazin National University,
61077, Kharkiv, Ukraine
vivanov@karazin.ua

For Educational Use Only



1KRU.PDB

Ліганд

(низькомолекулярна сполука, лікарська сполука)

Мішень

(рецептор, фермент)

Основні підходи до розв'язку задачі прогнозу біоактивності

1) Структура мішені невідома, але структура ліганду відома:

ligand-based design,

QSAR (Qualitative Structure-Activity Relationship)

2) Структура мішені (рецептора, фермента,...) і ліганду відомі:

Structure-based design (Docking)

3) Структура мішені відома, але структура ліганду невідома:

De-novo design

4) Структура мішені і ліганду невідомі:

Комбінаторна хімія, "high throughput screening"

QSAR

(Quantitative Structure-Activity Relationship)

QSAR це віртуальний (комп'ютерний, *in silico*) пошук закономірностей і математичних моделей, які кількісно або якісно описують будь-які властивості системи (біологічна активність, фізико-хімічні параметри, ефекти суб'єктивного сприйняття) в термінах дескрипторів хімічної структури.

Основні стадії побудови QSAR

1. Набір сполук з відомою активністю підрозділяється на **навчальну** та **тестову** вибірки.
2. Відбираються дескриптори, які характеризують системи (молекули).
3. Аналіз кореляції вибраних дескрипторів зі своєю властивістю на прикладі навчальної вибірки з використанням методів статистики.
4. Побудова математичної моделі «структура-властивість»
5. Оцінка прогностичної здатності побудованої QSAR-моделі. Аналіз результатів тестового вибору сполук з відомими властивостями.



$$\text{Властивість} = F(d_1, d_2, d_3, d_4, \dots)$$

$d_1, d_2, d_3, d_4, \dots$ – дескриптори – індекси які описують різні аспекти молекулярної структури

Індекси біологічної активності

LD_{50} Токсичність. Летальна доза
(мін. Кількість речовини яка веде до того що гинуть 50% організмів)

ED_{50} Ефективна доза (мін. кількість речовини яка 50% блокує дію ендогених сполук)

$\log \frac{C_{org}}{C_w}$ Фактор біоконцентрації

$Ib = \frac{\%_{ill}}{T_{latent}}$ Канцерогена активність (Індекс Айбола)

MBC Мін блокуюча концентрація (мін. кількість речовини що веде до блокування провідності *нервових* волокон)

$\log \frac{1}{C}$ Концентрація речовини (на одиницю біомаси) що веде до заданого біоефекту

Дескриптори молекулярної структури

- Елементарні (молекулярна маса, кількість атомів певного сорту)
- Топологічні (індекси Рандича, теоретико-інформаційні індекси,...)
- Електронні (заряди на атомах, індекси реакційної спроможності, дипольні моменти, поляризованості тощо)
- Геометричні (об'єм молекули, площа поверхні, асиметрія,...)
- Фізико-хімічні (рефракція, ліпофільність, ...)
- Хімічні (наявність або відсутність тієї чи іншої функціональної групи,,,,)

Відомо більш 3000 різноманітних молекулярних дескрипторів

Дескриптори молекулярної структури

1D - молекулярна формула

2D - молекулярна зв'язність (топологічні індекси)

3D - молекулярна геометрія (стереохімія)

4D-5D - конформаційний аналіз

Елементарні дескриптори

- Наявність або відсутність тих чи інших функціональних груп, або атомів
- Молярна маса
- Число атомів певного сорту

Приклад Теплоота сгорання насичених вуглеводнів
нормальної будови

$$\Delta H_{298}^0 = 237.9 + 660.0 \cdot N_c \quad (\text{кДж/моль})$$

Електронні дескриптори

Вони є результатом квантовохімічних розрахунків і представляють собою набір індексів, так чи інакше характеризуючих зарядовий розподіл у молекулі і її енергію.

Електронна густина $\rho_r = \sum_i n_i c_{ir}^2$

Заряд на атомі $q_r = Z_r - \rho_r$

Порядок зв'язку $\rho_{rs} = \sum_i n_i c_{ir} c_{is}$

Суперделокалізованість (сверхделокалізуемость) атома $S_e(r) = 2 \sum_a^{\text{vac.}} \frac{c_{ar}^2}{0.5(\epsilon_{(\text{HOMO})} + \epsilon_{(\text{LUMO})}) - \epsilon_a}$

Електронні дескриптори

Молекулярний дипольний момент (μ)

Електрична дипольна поляризованість (α)

$$\vec{\mu} = \alpha \cdot \vec{E}$$

$$\bar{\alpha} = \frac{1}{3} (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})$$

Самополяризованість атому (π)

$$\pi_r = \frac{\partial \rho_r}{\partial \alpha_r} = 4 \sum_i^{\text{occ.}} \sum_a^{\text{vac.}} \frac{c_{ir}^2 c_{ar}^2}{\epsilon_i - \epsilon_a}$$

Енергія системи (E)

$$E_b = E - E'$$

Енергія електронного переходу (λ)

$$\lambda = \epsilon_{(\text{LUMO})} - \epsilon_{(\text{HOMO})}$$

Електронні дескриптори

Енергія іонізації молекули

$$IP = -\epsilon_{(\text{HOMO})}$$

Енергія спорідненості до електронів

$$E_{\text{e}} = \epsilon_{(\text{LUMO})}$$

Енергія протонування (ΔH_{H})

Токсичність нітрилів

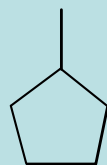
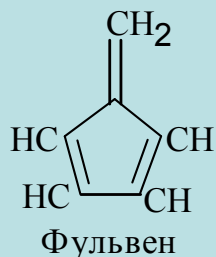
$$-\lg LD_{50} = -1.69 \frac{\alpha}{\Delta H_{\text{H}}} + 0.47, \quad (n = 20, r = 0.87, \sigma = 0.2)$$

**інгібування ацетилхолінестерази
(похідні бензилпиперидину)**

$$-\lg IC_{50} = 2.21C_4 - 6.65\mu + 1.18\mu^2 - 162.9\epsilon_{(\text{HOMO})} - 8.58\epsilon_{(\text{HOMO})}^2$$

$$(n = 16, r = 0.939, \sigma = 0.25)$$

Топологічні дескриптори



Граф фульвена

$$G = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$W = \frac{1}{2} \sum_{i,j}^n g_D(ij)$$

Вінера

$$M_1(G) = \sum_i v_i^2$$

$$\chi^{(1)} = \sum_{(i,j)} (v_i v_j)^{-1/2}$$

Рандіча

$$M_2(G) = \sum_{(i,j)} (v_i v_j)$$

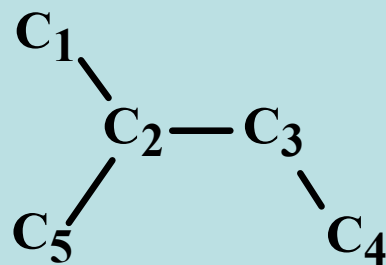
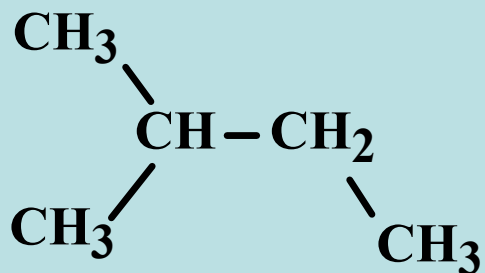
$$\chi^{(k)} = \sum_{(i,K,l,K,j)} (v_i^k v_l^k v_j^k)^{-1/2}$$

Узагальнений Рандіча

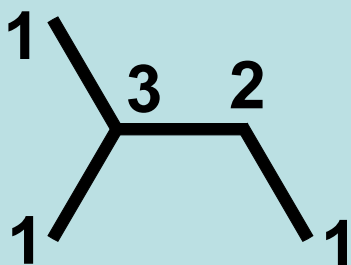
$$I = - \sum_i \frac{n_i}{n} \log_2 \frac{n_i}{n}$$

IC_n Теоретико-інформаційний індекс

Індекс Рандіча (с)



	C ₁	C ₂	C ₃	C ₄	C ₅
C ₁	1	0	0	0	0
C ₂	0	1	0	0	0
C ₃	0	0	1	0	0
C ₄	0	0	0	1	0
C ₅	0	0	0	0	1



$$\chi^{(1)} = \sum_{\substack{\text{СВЯЗИ} \\ (i,j)}} \frac{1}{\sqrt{v_i \cdot v_j}}$$

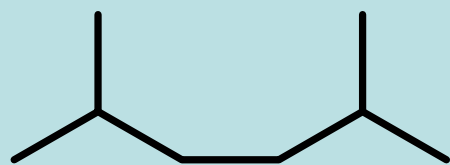
$$\chi^{(1)} = \frac{1}{\sqrt{1 \cdot 3}} + \frac{1}{\sqrt{1 \cdot 3}} + \frac{1}{\sqrt{2 \cdot 3}} + \frac{1}{\sqrt{1 \cdot 2}} = 2.270$$

Узагальнений індекс зв'язності

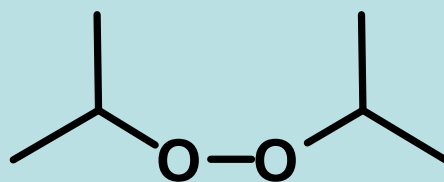
$${}^{(m)}\chi^S = \frac{1}{2^{m+1}} \sum_{(i-j-\dots-k)}^n \frac{Z_i \cdot Z_j \cdot \dots \cdot Z_k}{\sqrt{v_i \cdot v_j \cdot \dots \cdot v_k}}$$

$$\chi^S = 0.25 \sum_{\text{bonds}} Z_i \cdot Z_j / \sqrt{v_i \cdot v_j}$$

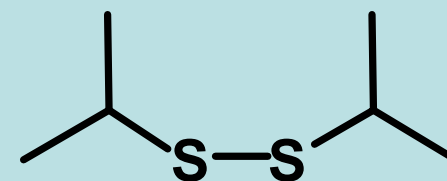
m - число вершин в маршруті
 n - число можливих маршрутів
 Z - номер елемента в періодичній системі



$$\begin{aligned} \chi &= 3.626 \\ {}^3\chi^S &= 1.321 \\ {}^3\chi^V &= 1.321 \end{aligned}$$



$$\begin{aligned} \chi &= 3.626 \\ {}^3\chi^S &= 1.321 \\ {}^3\chi^V &= 0.440 \end{aligned}$$



$$\begin{aligned} \chi &= 3.626 \\ {}^3\chi^S &= 2.973 \\ {}^3\chi^V &= 3.964 \end{aligned}$$

електротопологічний індекс

$$I_i = \frac{(\delta_i^v + 1)}{\delta_i}$$

δ_i – кол-во σ - зв'язків C–C атома i

δ_i^v – загальне кол-во електронів $\delta^v = \delta + \pi + n$

теплота атомізації алканів

$$\Delta H_a = 283.33 N - 6.321\chi^{(1)} + 115.72$$

$$r = 0.9999 \quad \sigma = 0.960$$

Температура кипіння алканів

$$T(^{\circ}\text{C}) = 57.85 \chi^{(1)} - 97.9$$

**мінімальна блокуюча
концентрація анестетиків**

$$\lg \text{MBC} = -0.762\chi^{(1)} + 3.55 \quad (n = 36, r = 0.98, \sigma = 0.39)$$

Теоретико-інформаційні дескриптори

$$\text{інформація} = -\sum_i \frac{n_i}{n} \log_2 \frac{n_i}{n} \text{ (bit)} \quad n = \sum_i n_i$$

- Неоднорідність розподілу відстаней між вершинами в графі

I_D , n_i – кількість зв'язків довжини i

$TI_D = n \cdot I_D$ Загальна кількість інформації

$$IC_k = -\sum_i p_i \log_2 p_i$$

$$TIC_k = n \cdot IC_k$$

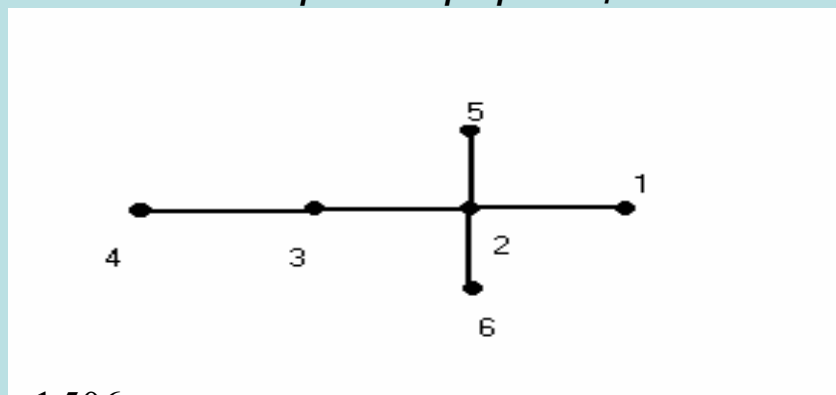
$SIC_k = IC_k / \log_2 n$ *структурний інформаційний склад*

$BIC_k = IC_k / \log_2 N_b$ *інформаційний склад зв'язування*

$CIC_k = \log_2 n - IC_k$ *комплементарне інформаційний склад*

Приклад (2,2-диметилбутан)

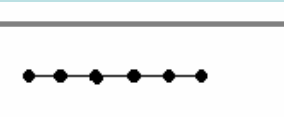
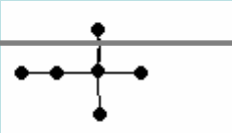
$$n = P_1 + P_2 + P_3 = 5 + 7 + 3 = 15$$

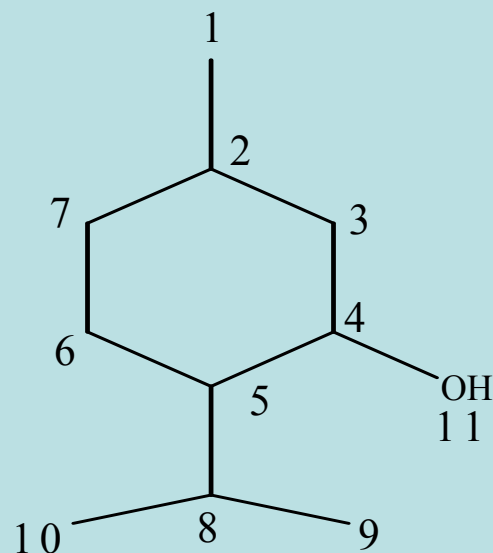


$$I_D = -\left(\frac{5}{15} \log_2 \frac{5}{15} + \frac{7}{15} \log_2 \frac{7}{15} + \frac{3}{15} \log_2 \frac{3}{15}\right) = 1,506$$

$$IC_0 = -\frac{6}{6} \log_2 \frac{6}{6} = 0 \quad IC_1 = -\left(\frac{4}{6} \log_2 \frac{4}{6} + \frac{1}{6} \log_2 \frac{1}{6} + \frac{1}{6} \log_2 \frac{1}{6}\right) = 1,252$$

$$IC_2 = -\left(\frac{3}{6} \log_2 \frac{3}{6} + \frac{1}{6} \log_2 \frac{1}{6} + \frac{1}{6} \log_2 \frac{1}{6} + \frac{1}{6} \log_2 \frac{1}{6}\right) = 1,792$$

Индекс		
N_C	6	6
P_1	5	5
P_2	4	7
P_3	3	3
W	35	28
I_D	2,149	1,506
IC_0	0	0
IC_1	0,918	1,252
IC_2	1,585	1,792
$\chi^{(1)}$	2,914	2,561



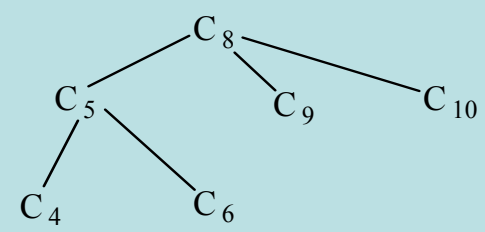
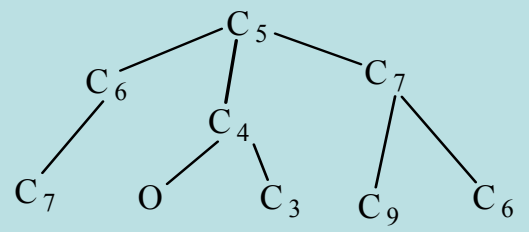
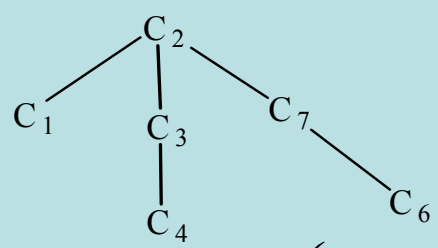
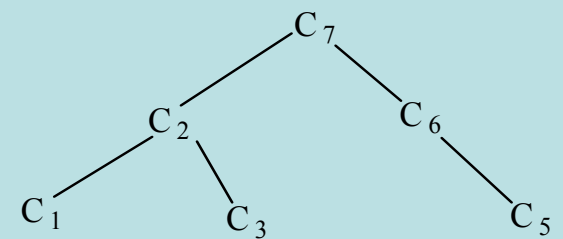
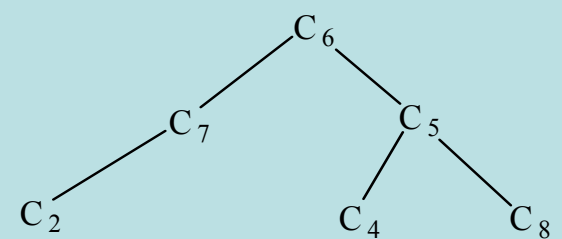
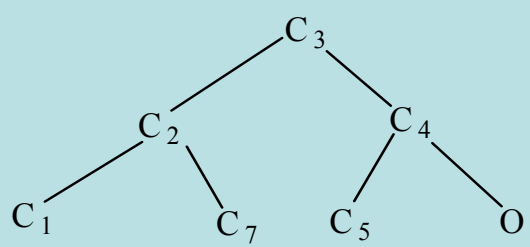
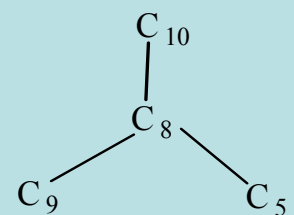
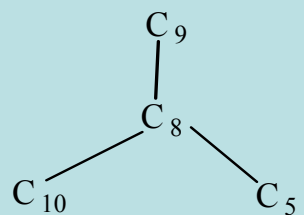
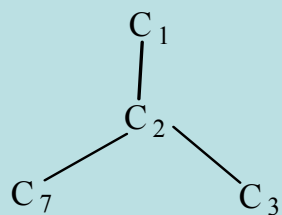
$$IC_0 = -\left(\frac{10}{11} \log_2 \frac{10}{11} + \frac{1}{11} \log_2 \frac{1}{11}\right) = 0,439$$

$$TIC_0 = 4,829 \quad SIC_0 = 0,127 \quad CIC_0 = 3,020 \quad BIC_0 = 0,127$$

$$IC_1 = -\left(\frac{3}{11} \log_2 \frac{3}{11} + \frac{3}{11} \log_2 \frac{3}{11} + \frac{3}{11} \log_2 \frac{3}{11} + \frac{1}{11} \log_2 \frac{1}{11} + \frac{1}{11} \log_2 \frac{1}{11}\right) = 2,163$$

$$TIC_1 = 23,793 \quad SIC_1 = 0,625 \quad BIC_1 = 0,625 \quad CIC_1 = 1,296$$

Другий порядок



$$IC_2 = -\left(\frac{3}{11} \log_2 \frac{3}{11} + \frac{2}{11} \log_2 \frac{2}{11} + 6 \cdot \frac{1}{11} \log_2 \frac{1}{11} \right) = 2,845$$

$TIC_2 = 31,295$

$SIC_2 = 0,822$

$BIC_2 = 0,822$

$CIC_2 = 0,614$

Фізико-хімічні дескриптори

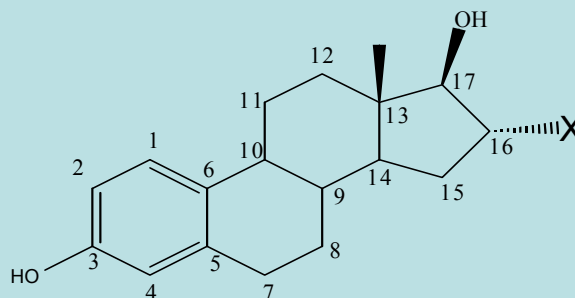
Молярна рефракція (MR, AMR)

$$MR = \frac{n^2 - 1}{n^2 + 2} \frac{MW}{d} = \frac{4}{3} \pi N_A \bar{\alpha}$$

для насичених вуглеводнів



$$MR = n \cdot MR(C) + (2n + 2) \cdot MR(H)$$



$$\lg RBA = -0.48 MR + 2.08, \quad r^2 = 0.84$$

Сучасні Комп'ютерні Методи

Хімії

Ліпофильність

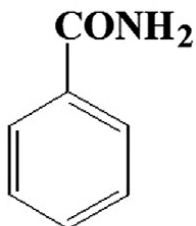
$$\lg P = \lg \frac{C_{1-\text{oc tan ole}}}{C_{\text{water}}} = \lg C_{1-\text{oc tan ole}} - \lg C_{\text{water}}$$

$$\lg \frac{P_x}{P_H} = \lg P_x - \lg P_H = \rho \cdot \pi_x$$

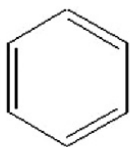
$$\pi_X = \lg P_{C_6H_5X} - \lg P_{C_6H_6}$$

$$\pi_X = \lg P_{RX} - \lg P_{RH}$$

$$\begin{aligned} \pi_{CONH_2} &= \log P(\text{benzamide}) - \log P(\text{benzene}) \\ &= 0.64 - 2.13 \\ &= -1.49 \end{aligned}$$



Benzamide ($\log P = 0.64$)



Benzene ($\log P = 2.13$)

Зам.	Аром.	Аліф.
-NH ₂	-1.23	-1.19
-I	1.12	1.0
-SCH ₃	0.61	0.45
-COCH ₃	-0.55	-0.71
-CONH ₂	-1.49	-1.71
-COOCH ₃	-0.01	-0.27
-Br	0.86	0.60
-CN	-0.57	-0.84
-F	0.14	-0.17
-Cl	0.71	0.39
-COOH	-0.28	-0.67
-OCH ₃	-0.02	-0.47
-N(CH ₃) ₂	-0.18	-0.30
-OH	-0.67	-1.16
-NO ₂	-0.28	-0.85

$$\lg P = \lg P_H + \sum_{X=1}^N \pi_X$$

модель *Моригучи*

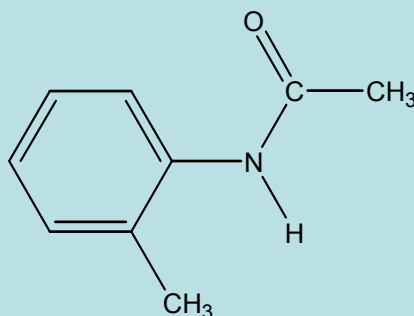
$$\lg P = -1.06 + 1.9 \cdot S - \sum_k S_K$$

S_k – міра площіни поверхні полярних груп

S – площа молекулярної поверхні, доступної розчиннику

CLOGP

$$\lg P = \sum_{k=1}^N f_k \cdot N_k + \sum_{j=1}^C F_j \cdot N_j$$



о-метилацетанилида

Компонента ліпофільності	Внесок
Фрагмент CO-NH	-1.510
2 ізольованих групи CH ₃	0.39
6 ароматичних атомів С	0.78
10 атомів водню зв'язаних з вуглецем	2.270
1 ланцюг зв'язаних атомів	-0.120
1 спряжене бензольне кільце-ланцюг	-0.150
1 орто замісник	-0.76
Ітого CLOGP	0.90
експериментально	0.86

Якими мають бути дескриптори ?

1. Мають бути простими;
2. Бажано, щоб вони мали структурне пояснення;
3. Має бути можливість застосувати до локальної структури (фрагментів);
4. Повинен мати добру кореляцію принаймні з однією властивістю;
5. Бажано розрізняти ізомери;
6. Має бути можливість побудувати ієрархію дескрипторів;
7. Бажано, щоб дескриптори були лінійно незалежними;
8. Бажано, щоб вони не базувалися на властивостях;
9. Має бути можливість їх ефективно розрахувати;
10. Повинні мати коректну залежність від розміру;
11. Мають змінюватися поступово з поступовою зміною молекулярної структури.

Running the program

Calculate descriptors

Load descriptors

Load responses

View descriptors

Save descriptors



Descriptor blocks

0D 1D 2D 3D Others

- | | |
|-------------------------------|------------------------------|
| 1. constitutional descriptors | 2. topological descriptors |
| 3. walk and path counts | 4. connectivity indices |
| 5. information indices | 6. 2D autocorrelations |
| 7. edge adjacency indices | 8. Burden eigenvalues |
| 9. topological charge indices | 10. eigenvalue-based indices |
| 11. Randic molecular profiles | 12. geometrical descriptors |
| 13. RDF descriptors | 14. 3D-MoRSE descriptors |
| 15. WHIM descriptors | 16. GETAWAY descriptors |
| 17. functional group counts | 18. atom-centred fragments |
| 19. charge descriptors | 20. molecular properties |

Descriptor list | Descriptor search | Bibliography

About



Address



Handbook



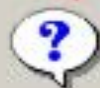
Tools



Thanks



Help



Example Data



Weightings



Comments



WHIM and GETAWAY



Versions



Tips of the day

