

*QSPR моделиране
чрез адитивни схеми*

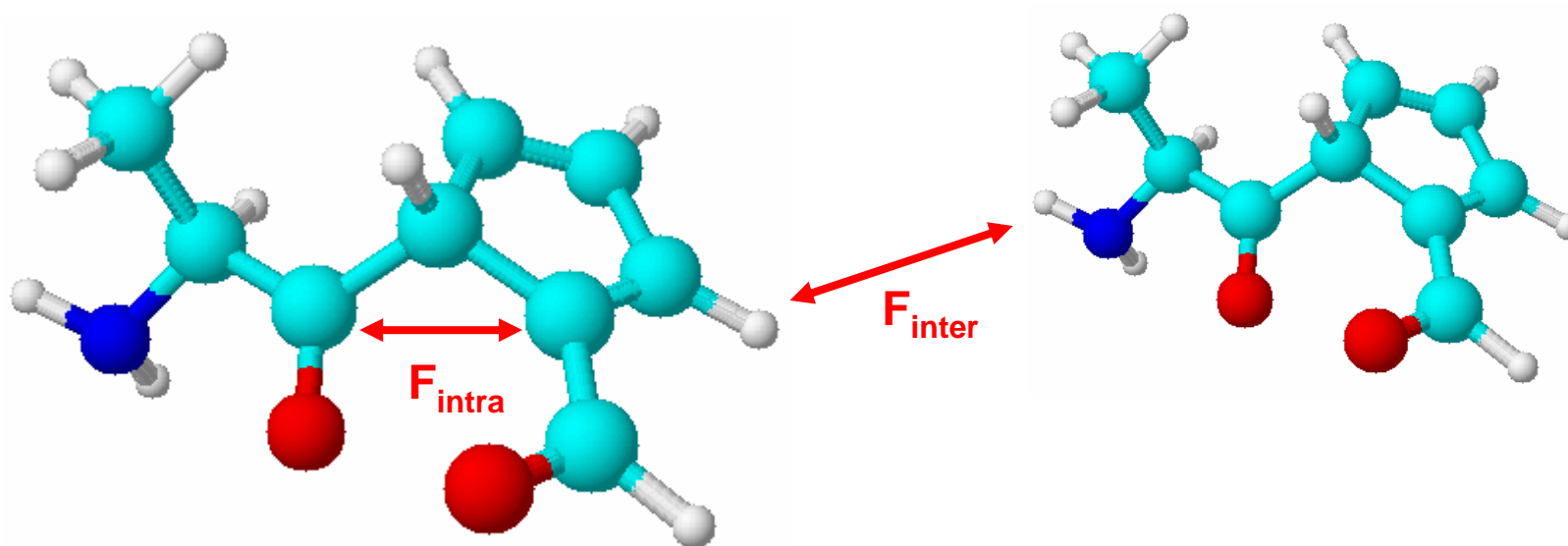
Адитивни схеми от нулев ред

Адитивни схеми от първи ред

Адитивни схеми от втори ред

Базово молекулно моделиране

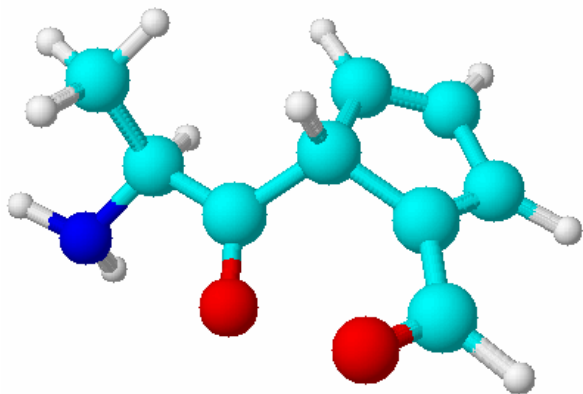
В основното си представяне молекулата се състои от атоми свързани по между с химични връзки.



Силите на междумолекулно взаимодействие са с един/два порядъка по-малки от силите, които държат атомите в молекулата: $F_{\text{intra}} \gg F_{\text{inter}}$

Базово молекулно моделиране

Взаимодействието на молекулата с други обекти (химични, физични или биологични) определя нейните свойства.



От основното молекулно представяне интуитивно възникват следните въпроси:

1. Дали атомите с тяхните състояния на свързаност в молекулата определят някакви молекулни свойства ?
2. Може ли да се разпознае това влияние и да се охарактеризира количествено ?

Атомна адитивна схема

Основно допускане: междуатомното взаимодействие в молекулата, което определя дадено свойството е на къси разстояния.

За молекулно ядро S, към което може да се свързват атоми X и Y, адитивният закон може да се изрази количествено:

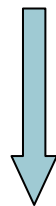
$$P(X-S-X) + P(Y-S-Y) = 2 \cdot P(X-S-Y)$$

Сумата на свойствата на молекулите X-S-X и Y-S-Y е равна на 2 пъти свойството на X-S-Y. **Сумата на свойствата не се променя при преразпределяне на атомите.**

Атомна адитивна схема

Когато адитивният закон е валиден, основното ядро S може да се премахне от двете страни на уравнението:

$$P(X-S-X) + P(Y-S-Y) = 2 \cdot P(X-S-Y)$$



$$P(X-X) + P(Y-Y) = 2 \cdot P(X-Y)$$

Това е схема на приближение на молекулно свойство от нулев ред.

Йерархия на адитивните схеми

Бенсън, Алън и др. обобщават идеята за атомните приноси, като в адитивната схема може да се разглеждат по-големи групировки от атоми: химични връзки и химични групи

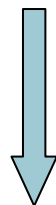
Когато изчисляването на свойство е базирано на приносите на отделните **химични връзки**, схемата се нарича адитивна схема от **първи ред**.

Когато изчисляването на свойство е базирано на приносите на по-голями атомни групи (3 и повече атома), схемата се нарича адитивна схема от **втори ред**.

Адитивна схема от първи ред

Например, ако молекулното ядро CH_2 е свързано с атомите X и Y , то адитивният закон може да се изрази количествено:

$$P(X-\text{CH}_2-X) + P(Y-\text{CH}_2-Y) = 2 \cdot P(X-\text{CH}_2-Y)$$

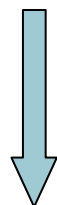


$$2 \cdot P(X-C) + 2 \cdot P(Y-C) = 2 \cdot (P(X-C) + P(Y-C))$$

Адитивна схема от втори ред

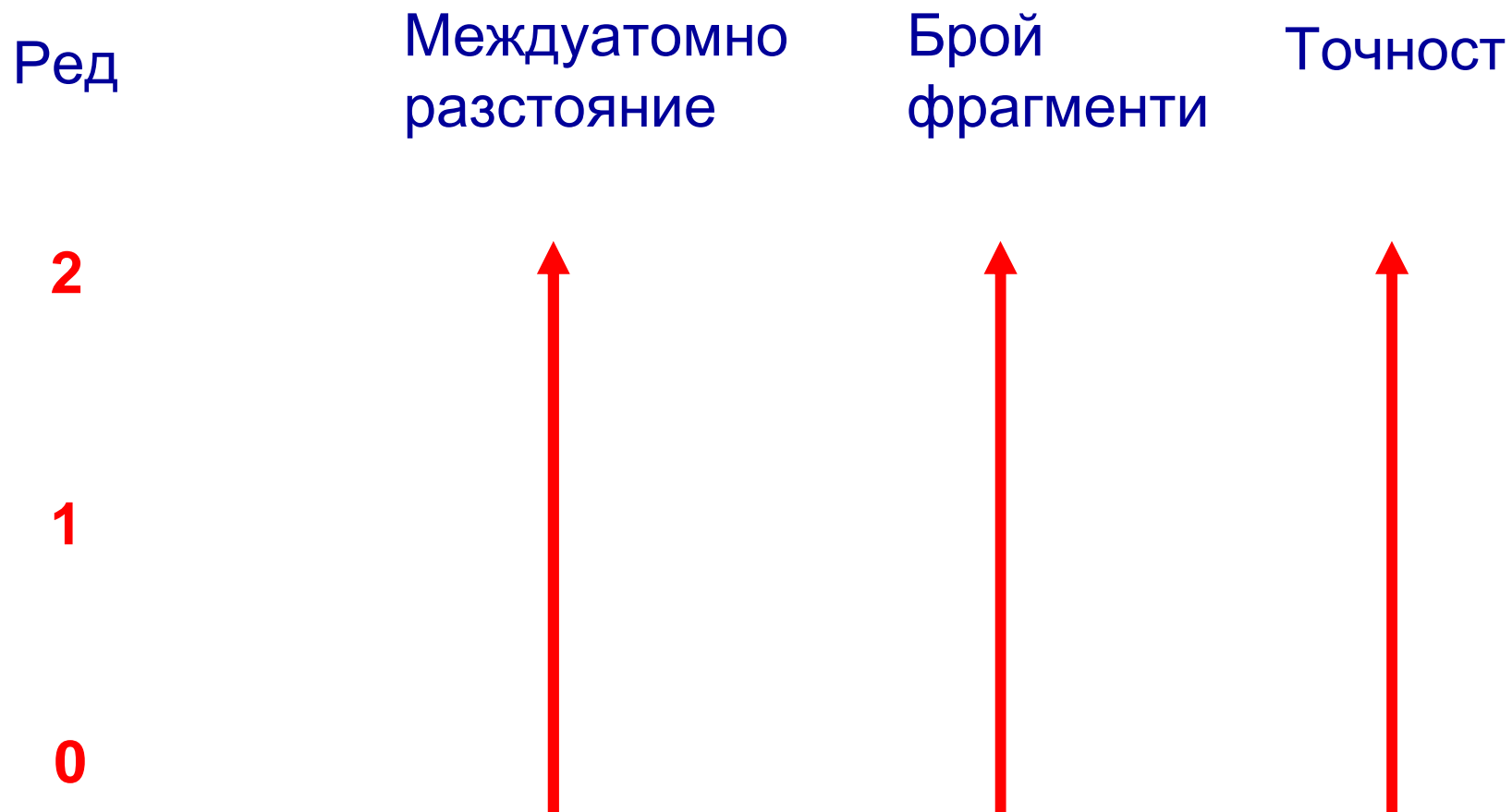
Например, молекулното ядро $\text{CH}_2\text{-CH}_2$ има групи атоми свързани с атоми X и Y . Адитивният закон може да се изрази количествено:

$$\begin{aligned} P(X\text{-CH}_2\text{-CH}_2\text{-X}) + P(Y\text{-CH}_2\text{-CH}_2\text{-Y}) &= \\ = 2 \cdot P(X\text{-CH}_2\text{-CH}_2\text{-Y}) \end{aligned}$$



$$\begin{aligned} 2 \cdot P(X\text{-C-C}) + 2 \cdot P(Y\text{-C-C}) &= \\ = 2 \cdot (P(X\text{-C-C}) + P(Y\text{-C-C})) \end{aligned}$$

Йерархия на адитивните схеми



Общ вид на адитивна схема

Адитивната схема се състои от множество атомни групировки $\{G_k\}$. На всяка групировка G_k е съпоставен инкремент I_k , който е приносът на тази групировка към молекулното свойство:

$$\begin{array}{ll} G_1 & I_1 \\ G_2 & I_2 \\ \dots & \\ G_m & I_m \end{array}$$

Абсолютната стойност на инкрементите е пропорционална на влиянието на атомните групи върху изследваното молекулно свойство.

Общ вид на адитивна схема

За да се приложи адитивната схема $\{G_k, I_k\}$ за структурата S е необходимо да се определи, колко пъти всяка група G_k се среща в структурата S . $N_k = f(S)$ е броят на появяванията (честотата) на групировката G_k в S .

Получава се вектор с дескриптори $N = (N_1, N_2, \dots, N_m)$.

Свойството P се изчислява като функция на N

$$P = \sum_k N_k \cdot I_k$$

Изчисляване на адитивна схема

Моделът на адитивната схема е известен, когато са определени инкрементите $\mathbf{l} = (l_1, l_2, \dots, l_m)$.

Векторът с инкрементите се определя чрез многопроменлива линейна регресия.

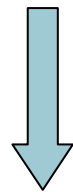
За обучаваща извадка съединения $(\mathbf{S}_1, \mathbf{S}_2, \dots, \mathbf{S}_n)$ се определя стойностите на свойството \mathbf{P} - вектор-стълб и стойностите на дескрипторите матрица \mathbf{N} . $[N_{ij}]$ е честотата на поява на група j в съединение i

$$\begin{pmatrix} P_1 \\ P_2 \\ \cdot \\ \cdot \\ P_n \end{pmatrix} \begin{pmatrix} N_{11} & N_{12} & \dots & N_{1m} \\ N_{21} & N_{22} & \dots & N_{2m} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ N_{n1} & N_{n2} & \dots & N_{nm} \end{pmatrix}$$

Изчисляване на адитивна схема

За да се определи векторът I се решава преопределена система по метода на най-малките квадрати от :

$$N \cdot I = P$$



$$I = (N' \cdot N)^{-1} N' P$$

Проблеми с преносимостта

При теоретично изчисляване (предсказване) на свойство чрез адитивна схема е възможно да се появи проблема с 'липсващите инкременти'.

Когато изследваното съединение съдържа фрагмент (атомна групировка), който не е участвал в линейната регресия с обучителната извадка, то този фрагмент има неизвестен инкремент.

На неизвестните инкременти по подразбиране се дава стойност 0. Те са основен източник на грешки в моделите чрез адитивни схеми.

Приложимост на адитивните схеми

При създаване на модел с адитивна схема трябва да се определи областта на приложимост на този модел.

Принципно още преди създаване на модела трябва да се реши дали изследваното свойство има адитивен характер спрямо приносите на различни атомни групи.

Когато върху свойството влияят междуатомни взаимодействия на по-големи разстояния, ефекти на спрегнати ситеми, ароматност и т.н. ефективното прилагане на адитивна схема е затруднено.

В този случай може да се приложи адитивна схема само за по-тясен клас съединения или адитивната схема да се коригира с друг метод.

Приложимост на адитивните схеми

Друг параметър, който трябва да се оптимизира е редът на адитивната схема и броят на инкрементите (класовете атомни групи).

При работа с адитивна схема от по-висок ред данните от обучителната извадка се апроксимират по добре, но преносимостта на модела намалява.

При схема от по-висок ред, броят на възможните комбинации атоми за даден тип атомно групиране е по-голям, което означава, че моделът се прави с по-голям брой дескриптори, като и вероятността за липсващи фрагменти при предсказването е по-голяма.

Аддитивна схема от нулев ред

Класическо приложение на адитивна схема с много висока точност - изчисляване на молекулната маса.

$$MW = \sum_k N_k \cdot A_k$$

Инкрементите са атомните маси $I_k = A_k$

Аддитивна схема от нулев ред

Аддитивните схеми от ред 0 може да се приложат ефективно за предсказване с добра точност на:

моларен обем

магнитна чувствителност

моларна рефракция

Подобрение на адитивните схеми от ред 0

За подобряване на точността на атомната адитивна схема, атомите се класифицират не само според тяхните типове.

Към описанието на атомния клас може да се включи хибридизацията, броят на водородните атоми, участие в пи-системи и други атомни свойства.

Допълнителната информация в описанието на атомния клас обикновено включва данни за първата топологична околност на атома, затова такива схеми може да се обозначат от ред 0.5

Приложение на подобрена схема от ред 0.5

Изчисляване на средна молекулна поляризуемост, с отчитане на атомната хибридизация:

$$\overline{\alpha} = \sum_k N_k \cdot \alpha_k$$

H ₂	0.79	0.77
CH ₄	2.60	2.61
n-C ₅ H ₁₂	9.95	9.95
neo-C ₅ H ₁₂	10.20	9.95
cyclo-C ₆ H ₁₂	10.99	11.01
CH ₂ =CH ₂	4.26	4.25
C ₆ H ₆	10.39	10.43
CH ₃ F	2.62	2.52
CF ₄	2.92	2.25
CCl ₄	10.47	10.32
NH ₃	2.26	2.13
Aniline	11.58–12.12	11.91
Acetic acid	5.05–5.15	5.17
\overline{T} Pyridine	9.14–9.47	9.72

Адитивна схема от ред 0.5

Фрагментите представляват класове атоми, които се формализират в следния общ вид:

$$A [d_1, d_2, \dots, d_m]$$

Атомният клас се описва чрез типа на атома **A** и няколко локални атомни дескриптора **d₁, d₂, ...**

Като локални атомни дескриптори може да се използват: броят на съседите, броят на двойните връзки, броят на тройните връзки, участие в пи-системи и т.н.

Коефициент на разпределение $\log P$

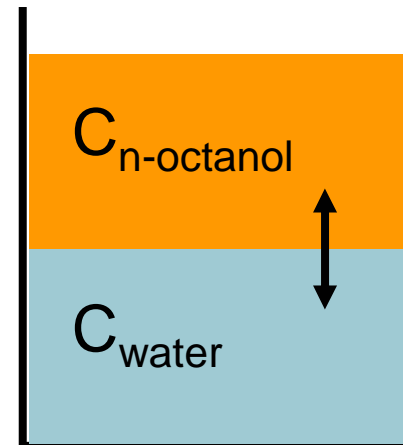
Един от най-изследваните параметри в QSAR и QSPR е коефициентът на разпределение на веществата в системата n-октанол/вода, означаван като $P_{o/w}$. Много често той се представя и чрез десетичен логаритъм – $\log P_{o/w}$. Причината за този интерес към него е, че биологичната активност на повечето лекарства зависи от $\log P$.

Коефициентът на разпределение $\log P_{o/w}$ се използва като количествена мярка за липофилността на веществата, която характеризира способността им лесно или по-трудно да преминават през клетъчната мембрана.

Коефициент на разпределение $\text{Log}P$

Експериментално определяне: веществото се разтваря в система от несмесващи се течности – **n-октанол** и **вода**. След разпределянето на съединението в двете фази, съдържанието му във всяка от тях се определя чрез даден аналитичен метод. Съотношението между концентрациите на веществото в двете фази се изчислява чрез израза:

$$P = \frac{C_{\text{n-octanol}}}{C_{\text{water}}}$$



Коефициент на разпределение $\text{Log}P$

Процедурата по експерименталното определяне на $\text{log}P$ също така се отличава с някои недостатъци: времето, необходимо за определяне коефициента на разпределение отнема над едно денонощие, затруднения при повърхностно-активни вещества, чувствителност към замърсвания в пробатите, невъзможност за прилагането на процедурата за вещества, които още не са синтезирани с цел предварителна оценка на техните свойства, и други.

Всичко това налага необходимостта от създаване на методи за теоретично изчисляване на коефициента разпределение за системата n -октанол-вода.

Адитивна схема за определяне на $\log P$

Избор на обучителна извадка съединения

Основен клас съединения	Подклас	Брой	Общо
Въглеродороди	Ациклични въглеродороди	20	40
	Бензоидни	20	
Кислородсъдържащи съединения	Алкохоли	31	82
	Етери	11	
	Карбонилни съединения	15	
	Карбоксилни киселини	9	
	Естери	16	
Азотсъдържащи съединения	Мастни амини	23	40
	Ароматни амини	17	

Адитивна схема за определяне на $\log P$

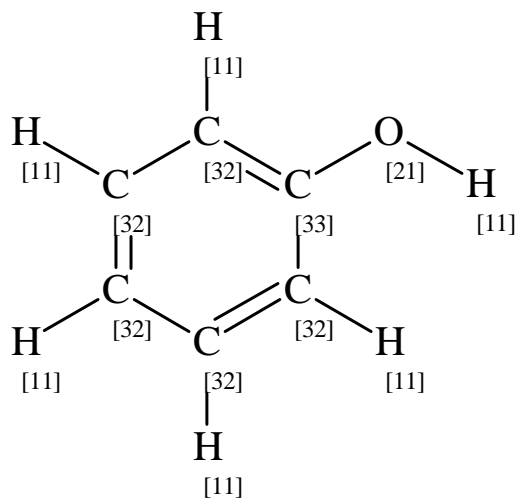
Прилага се адитивна схема от ред 0.5

$$\log P_{o/w} = \sum N_{A[n,\pi]} I_{A[n,\pi]}$$

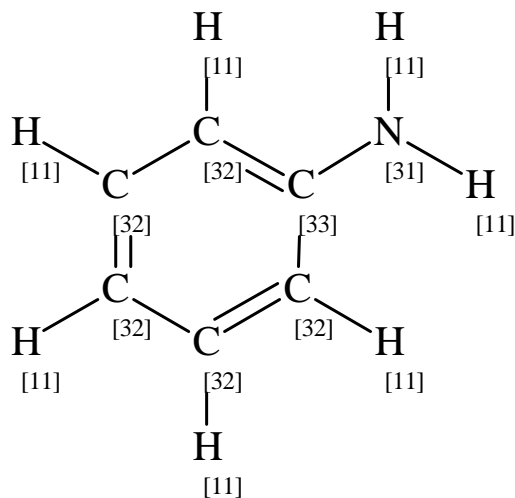
където **A** представлява видът на атома, **n** – общият брой на съседните атоми, **π** – общият брой на π -атомите, свързани с **A**, **$N_{A[n,\pi]}$** – броят на атомите от тип **$A[n,\pi]$** , **$I_{A[n,\pi]}$** – инкрементът за атомите от тип **$A[n,\pi]$** .

Атомите се класифицират според вида им, и според хибридно-то им състояние. Освен това се описва и частична информация както за връзките, в които атома **A** участва, така и за типа на атомите, намиращи се в съседство с атома **A**

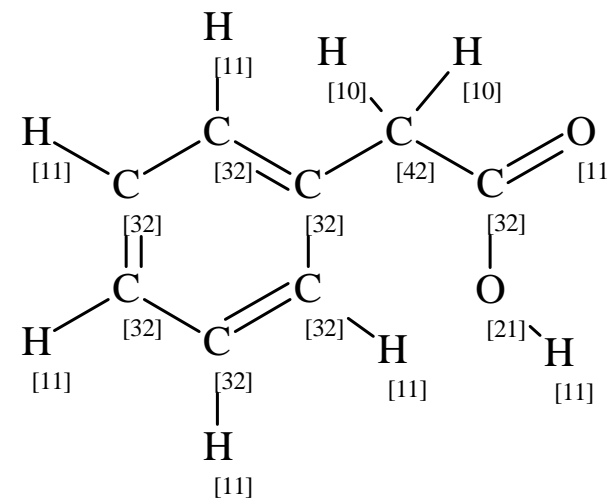
Адитивна схема за определяне на $\log P$



Фенол			
H[11]	C[32]	C[33]	O[21]
6	5	1	1



Анилин			
H[11]	C[32]	C[33]	N[31]
7	5	1	1



Фенилоцетна киселина					
H[10]	H[11]	C[32]	C[42]	O[21]	O[11]
2	6	7	1	1	1

Адитивна схема за определяне на $\log P$

След прилагане на линейна регресия се получават следните инкременти:

H[10]	0.283345	C[42]	-0.20447
H[11]	0.231257	N[30]	-2.03133
C[31]	0.133614	N[31]	-1.38215
C[32]	0.132817	O[11]	-0.960592
C[33]	0.293351	O[20]	-1.6672
C[40]	-0.0786977	O[21]	-0.440429
C[41]	-0.172448	O[22]	-0.409521

Аддитивна схема за определяне на $\log P$

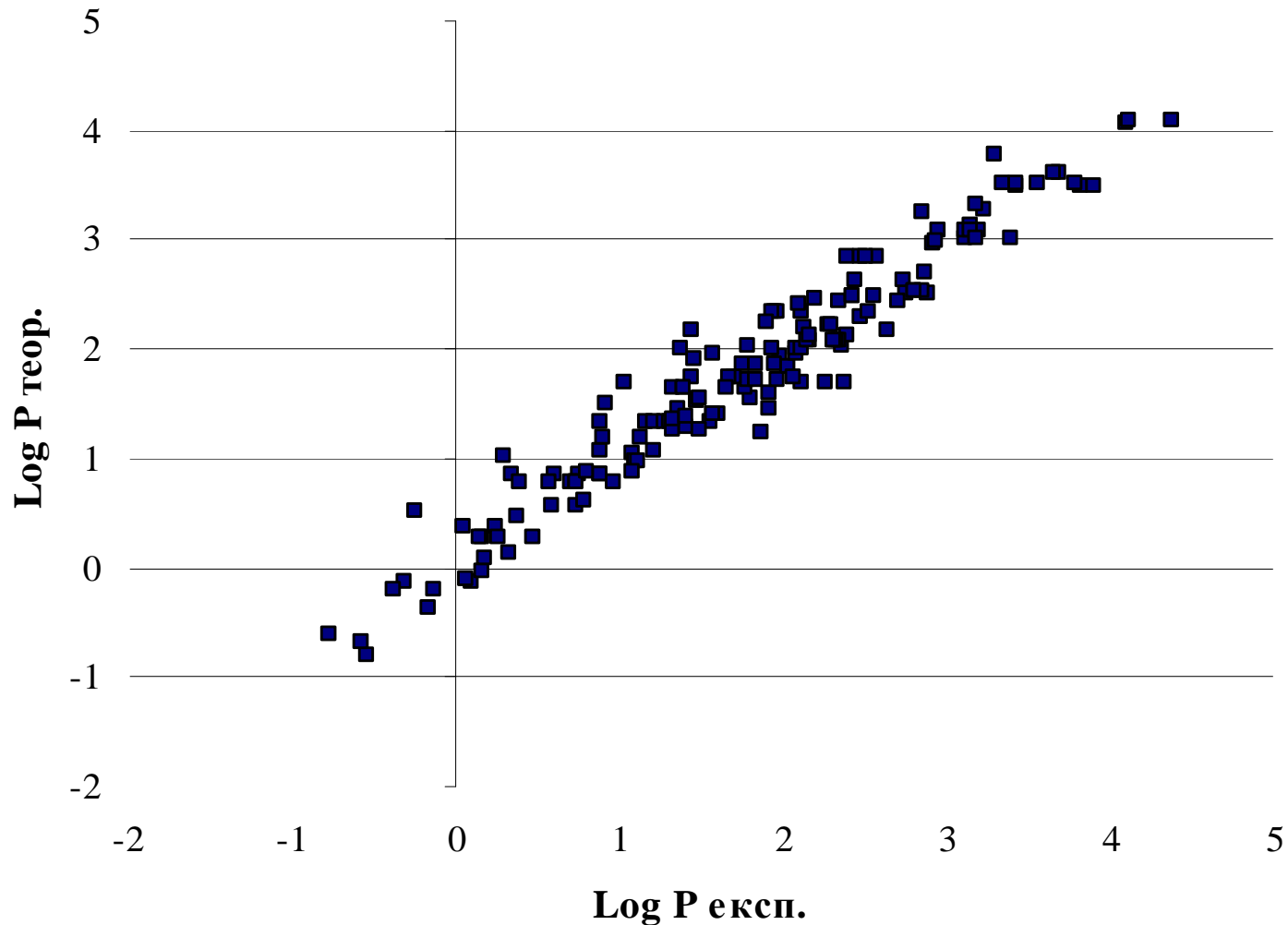
Прилагане на модела на адитивната схема за фенол-оцетната киселина:

H[10]	H[11]	C[32]	C[42]	O[21]	O[11]
2	6	7	1	1	1

$$\text{Log } P = 2 \cdot I_{\text{H}[10]} + 6 \cdot I_{\text{H}[11]} + 7 \cdot I_{\text{C}[32]} + 1 \cdot I_{\text{C}[42]} + 1 \cdot I_{\text{O}[11]} + 1 \cdot I_{\text{O}[21]}$$

Адитивна схема за определяне на $\log P$

Сравняване на стойностите на модела с експерименталните стойности



Адитивна схема от първи ред

Моделът се получава чрез сумиране на приносите (инкрементите) на всяка връзка.

Връзките формират така наречените В групи **A — B**

Адитивната схема може да се подсили (ред1.5) като се включат допълнителни атомни дескриптори

A[d₁,d₂,...] - B[d₁,d₂,...]

Аддитивна схема от първи ред

Аддитивните схеми от първи ред се използват ефективно за моделиране на термодинамични величини:

топлинен капацитет C_p (точност ± 4.0 J/mol)

ентропия S (точност ± 3.0 J/mol)

топлина на образуване ΔH_f (точност ± 10.0 J/mol)

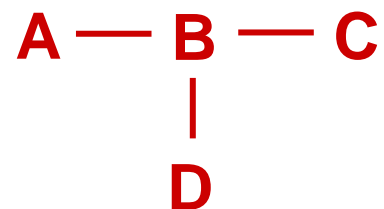
Адитивна схема от втори ред

Атомните групи се състоят от повече от два атома. Обикновено има един централен и прикрепени към него атоми.

G група

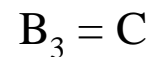
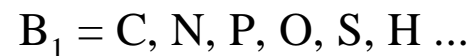
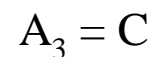
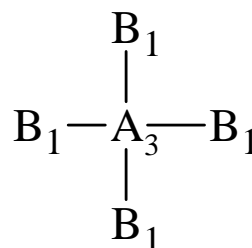
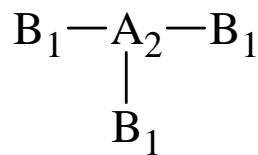
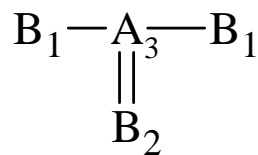
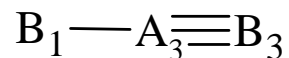
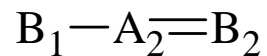


D група



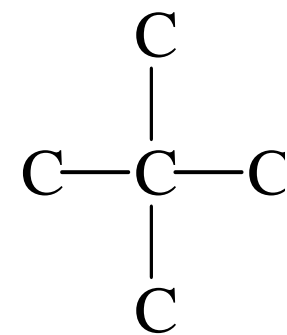
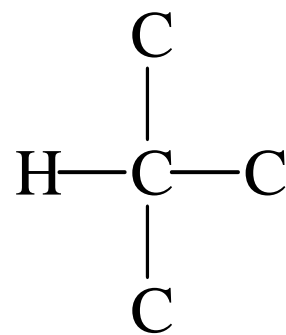
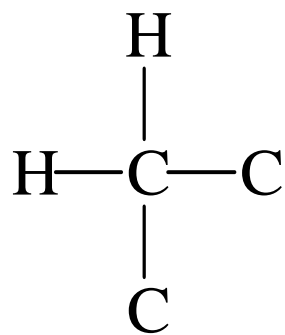
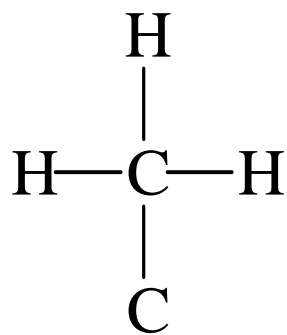
Адитивна схема от втори ред

Атомни групировки с 3, 4 и 5 атома:



Адитивна схема от втори ред

Възможни 5-атомни групировки за алканите



	C_p [J/mol K]	S [J/mol K]	ΔH_f° [kJ/mol]
C-(H) ₃ (C)	25.95	127.30	-42.19
C-(H) ₂ (C) ₂	22.81	39.43	-20.72
C-(H) (C) ₃	18.71	-50.53	-6.20
C-(C) ₄	18.21	-146.93	8.16

Адитивна схема от втори ред

Адитивната схема от втори ред не се използва самостоятелно, а като коригиращ фактор на адитивна схема от първи ред.

След корекция с инкременти от втори ред термодинамичните величини подобряват точността си:

топлинен капацитет C_p (точност ± 1.2 J/mol)

ентропия S (точност ± 1.2 J/mol)

топлина на образуване ΔH_f (точност ± 2.0 J/mol)

Адитивна схема от втори ред

Адитивните схеми допълнително се коригират с инкременти за отчитане на спреженията в циклите.

Циклите в структурата се разпознават чрез SSSR алгоритъм

Адитивната схема може да се подобри и чрез комбинация с други адитивни величини.

Примери:

(1) комбиниране със стеричната енергия на полето MMX за получаване на топлини на образуване;

(2) Подобряване на модела за LogP с добавяне на инкременти за вътрешномолекулни водородни връзки.

Енергия за разкъсване на връзка

Енергията за разкъсване на връзка BDE може да се моделира с помощта на адитивни схеми, като се използват моделите за топлини на образуване.



$$BDE(A - B) = \Delta H_f^0(A \cdot) + \Delta H_f^0(B \cdot) - \Delta H_f^0(A - B)$$

В обучаващата извадка за изчисляване на модела за топлини на образуване на органични съединения се включват и експериментални данни за радикали.