

# 药物分子设计中定量结构-活性关系计算方法的研究\*

## Study of Mathematical Method in the Quantitative Structure-activity Relationship for Drug Design

王青艳, 谢能中, 许晓东\*\*

WANG Qing-yan, XIE Neng-zhong, XU Xiao-dong

(广西科学院, 非粮生物质酶解国家重点实验室, 国家非粮生物质能源工程技术研究中心, 广西生物质产业化工程院, 广西生物炼制重点实验室, 广西南宁 530007)

(Guangxi Academy of Sciences, State Key Laboratory of Non-Food Biomass and Enzyme Technology, National Engineering Research Center for Non-food Biorefinery, Guangxi Biomass Industrialization Engineering Institute, Guangxi Key Laboratory of Biorefinery, Nanning, Guangxi, 530007, China)

**摘要:**【目的】发展一种新的定量结构-活性关系(简称“构效关系”QSAR)计算方法。【方法】双层 QSAR 预测模型的第 1 层是分子的理化性质参数, 第 2 层是分子片段结构参数。两组权重系数  $\{a_k\}$  和  $\{b_l\}$  分别指派给理化参数和片段参数。在双层 QSAR 预测模型中引入迭代的双最小二乘(IDLS)计算方法, 把单层次的构效关系发展成二层次、双方向的预测网络; 在训练集中两组系数用最小二乘法(或偏最小二乘法、主成分分析等)交替求解, 直至收敛。【结论】二层次、双方向的 QSAR 预测模型不仅能够提高预测能力, 具有一定的信息反馈和学习功能, 而且赋予 QSAR 更多的信息分析能力。

**关键词:** 药物设计 构效关系 分子片段 理化性质

**中图分类号:** Q74, R966 **文献标识码:** A **文章编号:** 1005-9164(2014)01-0006-06

**Abstract:** 【Objective】Quantitative structure-activity relationship (QSAR) is a widely used approach in chemistry, biology, and environment science. In recent years the iterative double least square (IDLS) method was applied in the QSAR by authors. 【Method】The simple and single direction prediction method was reformed to a two-level and double direction prediction network. In the two-level QSAR model one level is physicochemical parameters  $\{a_k\}$ , and the other is the molecular fragments  $\{b_l\}$ . In a training set the two coefficient sets are solved by using least square (or partial least square and principal component analysis) iteratively. 【Conclusion】The two-level and two-direction QSAR method not only enhances the prediction power, but also endows more information analysis ability to QSAR.

**Key words:** drug design, QSAR, molecular fragments, physicochemical properties

【研究意义】分子的定量结构-活性关系(简称“构效关系”QSAR)是应用最为广泛的药物分子设计的

预测计算方法<sup>[1~4]</sup>, 在化学、生物学和环境科学中有广泛应用。世界上各种药物设计和生物信息学商品软件中(如 SYBYL<sup>[5]</sup> 和 MOE<sup>[6]</sup>), QSAR 是必备的计算分析工具, 往往是最贵和最具卖点的模块。【前人研究进展】传统 QSAR 的最大缺点是使用分子整体的理化参数建立预测模型, 而同系列的药物的发展是从一个先导化合物母体出发, 具有相同的分子骨架, 只在少数取代基上做微小改进。因此同系衍生物的分子整体的性质参数的差别往往很小, 无法准确描述取代基的微小变化造成的生物活性的改变。【本研究切入点】本文发展的二层次 QSAR 预测模型<sup>[7,8]</sup>把

收稿日期: 2013-09-04

修回日期: 2013-10-11

作者简介: 王青艳(1972-), 女, 副研究员, 博士研究生, 主要从事微生物学生物技术研究。

\* 国家自然科学基金项目(项目代号: 31370716 和 31360207)资助。

\*\* 通讯作者: 许晓东(1974-), 男, 研究员, 主要从事组合数学研究。

分子分解成几个片段(每个片段包含一个取代基),用分子片段的性质参数建立二层次的 QSAR 预测模型。

## 1 双层 QSAR 模型与计算方法

传统的构效关系算法采用数学模型:

$$a_h f_{\text{hydr}}(x_{i,h}) + a_e f_{\text{elec}}(x_{i,e}) + a_{st} f_{st}(x_{i,s}) + \dots = \sum_{k=1}^K a_k f_k(x_{i,k}) = \log A_i, i=1,2,\dots,N, \quad (1)$$

式中,  $\log A_i$  是药物分子  $i$  的生物活性  $A_i$  的对数,  $f_{\text{hydr}}(x_{i,h})$  是疏水作用函数,  $x_{i,h}$  是分子  $i$  的疏水参数;  $f_{\text{elec}}(x_{i,e})$  是静电作用函数,  $x_{i,e}$  是分子  $i$  的静电参数;  $f_{st}(x_{i,s})$  是立体排斥函数,  $x_{i,s}$  是分子  $i$  的立体参数;  $K$  是分子作用项的数目;  $N$  是数据集中分子的个数。(1)式还可以包含更多的作用项,  $f_k(x_{i,k})$  代表一般的作用函数和分子  $i$  的参数,  $\{a_k\}$  是各作用项的系数。

传统的 QSAR 在训练集里用最小二乘(least square),或偏最小二乘(partial least square)、主成份分析(principal component analysis)解出预测系数  $\{a_k\}$ ,再用这组系数去预测新设计分子的生物活性:

$$\log A_i^{\text{Pred}} = \sum_{k=1}^K a_k f_k(x_{i,k}). \quad (2)$$

其最大缺点是使用分子整体的参数  $\{x_{i,k}\}$  建立预测模型,很难反映取代基的微小变化造成的生物活性的改变。

新发展的二层次 QSAR 预测模型<sup>[7,8]</sup>把分子分解成几个片段(每个片段包含一个取代基),用分子片段的性质参数建立二层次的 QSAR 预测模型。把分子分解为分子片段的思路还可以推广到短肽,短肽的每个氨基酸可以作为一个分子片段。药物分子和短肽的分子片段的示意图 1,其中图 1(a)是抗 A 型流感病毒达菲类药物母体和 3 个取代基;图 1(b)中短肽的每个氨基酸侧链可视为一个分子片段。

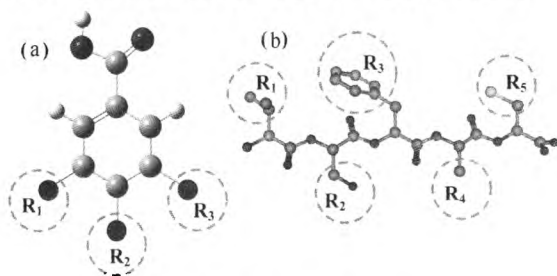


图 1 分子片段示意图

Fig. 1 Illustration of molecular fragments

(a)取代基是药物分子的片段;(b)氨基酸是短肽的片段。

(a)Molecular fragments,(b)Peptide fragments

在二层次 QSAR 模型里,分子的生物活性是各分子片段的贡献的总和:

$$\sum_{l=1}^L b_l \Delta g_{i,l} = (Act)_i. \quad (3)$$

式(3)中,  $\Delta g_{i,l}$  是分子  $i$  的片段  $l$  对生物活性  $\log A_i$  的贡献,  $\{b_l\}$  是分子片段的权重系数,  $(Act)_i$  是分子  $i$  的生物活性。分子片段  $l$  对生物活性的贡献  $\Delta g_{i,l}$  是分子片段性质参数的线性函数:

$$\Delta g_{i,l} = \sum_{k=1}^K a_k x_{i,l,k}. \quad (4)$$

式(4)中,  $x_{i,l,k}$  是分子  $i$  的片段  $l$  的第  $k$  个分子参数。把式(4)代入式(3)得到方程:

$$\sum_{l=1}^L b_l \left( \sum_{k=1}^K a_k x_{i,l,k} \right) = (Act)_i, i=1,2,\dots,N. \quad (5)$$

式(5)是二层次 QSAR 预测模型的基本方程,其中有两组待定系数  $\{a_k\}$  和  $\{b_l\}$ ,前者是分子理化参数的权重系数,后者是分子片段的权重系数。

方程组(5)用线性代数的方法表达为

$$X_{N,L,K} B_L A_K = (Act)_N, \quad (6)$$

$(Act)_N$  是分子生物活性组成的列矩阵。方程组(6)用迭代的双最小二乘法(IDLS)<sup>[7,8]</sup>求解:先指定一组系数  $\{b_l\}$  的初始值,如  $\{b_l^{(0)} = 1\}$ ,得到样品-性质矩阵  $G_{N,K}^{(1)}$ ,

$$X_{N,L,K} B_L^{(0)} = G_{N,K}^{(1)}, \quad (7)$$

再把矩阵  $G_{N,K}^{(1)}$  代入式(6),得到一般线性方程组:

$$G_{N,K}^{(1)} A_K^{(1)} = (Act)_N. \quad (8)$$

式(8)用最小二乘法求解,得到性质系数矩阵的第一组解  $A_K^{(1)}$ 。由系数矩阵  $A_K^{(1)}$  可以得样品-片段矩阵  $H_{N,L}^{(1)}$ ,

$$X_{N,L,K} A_K^{(1)} = H_{N,L}^{(1)}. \quad (9)$$

把矩阵  $H_{N,L}^{(1)}$  代入式(6)后得到一般线性方程组:

$$H_{N,L}^{(1)} B_L^{(1)} = (Act)_N. \quad (10)$$

式(10)用最小二乘法求解,得到片段系数矩阵的解  $B_L^{(1)}$ 。由片段系数矩阵  $B_L^{(1)}$  又可以得性质系数矩阵的第 2 组解  $A_K^{(2)}$ ,……。重复式(7)~(10)式的步骤,直到迭代达到收敛,

$$\begin{aligned} |Q^{(n+1)} - Q^{(n)}| &= \sqrt{\frac{1}{N} \sum_{i=1}^N (Act_i^{\text{expt}} - Act_i^{(n+1)})^2} \\ &- \sqrt{\frac{1}{N} \sum_{i=1}^N (Act_i^{\text{expt}} - Act_i^{(n)})^2} \leq \epsilon (10^{-6}). \end{aligned} \quad (11)$$

式(11)中  $\epsilon = 10^{-6}$  是收敛判据。双最小二乘迭代解法(IDLS)是一个双向的过程,有一定的信息反馈和学习功能。收敛后的系数  $\{a_k^{(n)}\}$  和  $\{b_l^{(n)}\}$  可用于预测新设计药物分子的生物活性:

$$Act_i^{\text{Pred}} = \sum_{l=1}^L b_l^{(n)} \left( \sum_{k=1}^K a_k^{(n)} x_{i,l,k} \right). \quad (12)$$

以上迭代的双最小二乘(IDLS)计算过程可以用图 2 来说明。

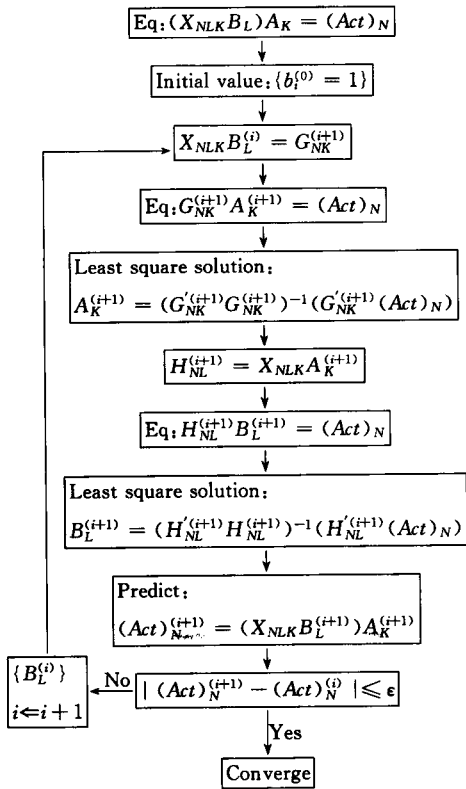


图 2 迭代的双最小二乘(IDLS)计算程序  
Fig. 2 Flow chart of IDLS procedure

## 2 应用举例

MHC-I 肽是 9 个氨基酸组成的短肽,具有调节

表 1 20 种天然氨基酸侧链的 8 种理化性质参数

Table 1 Eight properties of 20 natural amino acids

A. A.	<i>Lip</i>	<i>Hyd</i>	$S_L (\text{\AA}^2)$	$S_H (\text{\AA}^2)$	$P_\alpha$	$P_\beta$	$P_c$	$V (\text{\AA}^3)$
Leu (L)	1.2906	0.0000	84.5476	0.0000	1.21	1.30	0.68	166.7
Ile (I)	1.1046	0.0000	88.6055	0.0000	1.08	1.60	0.66	166.7
Val (V)	0.5324	0.0000	77.8108	0.0000	1.06	1.70	0.62	140.0
Phe (F)	0.4412	-0.1195	105.7054	11.2472	1.13	1.38	0.71	189.9
Met (M)	1.0768	-0.3068	70.3631	23.2299	1.45	1.05	0.58	162.9
Trp (W)	0.8364	-0.4310	133.6980	14.8820	1.08	1.37	0.75	227.8
Ala (A)	0.1744	0.0000	34.7760	0.0000	1.42	0.83	0.70	88.6
Cys (C)	0.2479	-0.2402	23.5563	30.4540	0.70	1.19	1.18	108.5
Gly (G)	0.0208	0.0000	3.7616	0.0000	0.57	0.75	1.50	60.1
Tyr (Y)	0.4534	-0.5896	80.9646	42.7160	0.69	1.47	1.06	193.6
Thr (T)	1.4265	-0.4369	46.7285	16.0490	0.83	1.19	1.07	116.1
Ser (S)	0.2346	-0.6040	26.0681	15.9613	0.77	0.75	1.32	89.0
His (H)	0.8124	-0.7766	82.1701	13.8631	1.00	0.87	1.06	153.2
Gln (Q)	1.0036	-0.7211	70.0876	17.8662	1.11	1.10	0.86	143.9
Lys (K)	1.4600	-0.6229	97.7144	8.0786	1.16	0.74	0.98	168.7
Asn (N)	0.6396	-0.7211	50.5075	17.7804	0.67	0.89	1.35	117.7
Glu (E)	1.0315	-0.9298	57.1582	25.5726	1.51	0.37	0.84	138.4
Asp (D)	0.6058	-0.9298	37.4173	25.2736	1.01	0.54	1.20	111.1
Arg (R)	1.2424	-1.4797	90.8008	35.3095	0.98	0.93	1.04	173.4
Pro (P)	0.3226	0.0000	69.2297	0.0000	0.57	0.55	1.59	122.7

生物体免疫系统的功能<sup>[9,10]</sup>。短肽的每个氨基酸可以认为是一个分子片段。我们以 MHC-I 短肽的生物活性预测为例,证明基于分子片段的双向 QSAR 预测网络的功能。氨基酸的理化参数见表 1。*Lip*: 疏水指标; *Hyd*: 亲水指标;  $S_L$ : 疏水表面积;  $S_H$ : 亲水表面积;  $P_\alpha$ :  $\alpha$ -螺旋倾向因子;  $P_\beta$ :  $\beta$ -带倾向因子;  $P_c$ : 卷曲倾向因子;  $V$ : 侧链体积。

训练集的 90 个 MHC-I 短肽的序列,实验生物活性 (Expt pIC<sub>50</sub>) 和预测计算的生物活性 (Pred pIC<sub>50</sub>) 见表 2。用 90 个 MHC-I 短肽为训练集,建立 9 个分子片段 (氨基酸残基) 和 8 个理化参数的 QSAR 预测模型。QSAR 预测模型的系数 { $a_k$ } 和 { $b_l$ } 列于表 3。训练计算的相关系数为  $R = 0.8871$ , 预测残差为  $Q = 0.0387$ 。

表 4 列出了测试集的 40 个 MHC-I 短肽的序列,实验生物活性 (Expt pIC<sub>50</sub>) 和预测生物活性 (Pred pIC<sub>50</sub>), 可以看出预测计算的相关系数为  $R = 0.8679$ , 预测残差为  $Q = 0.074271$ 。图 3 是 IDLS 迭代计算中相关系数  $R_A$  和  $R_B$  随迭代的变化, 可以看出, 经过 98 次迭代计算, 相关系数  $R$  由最初的  $R_A^1 = 0.4167$  上升到  $R_A^{98} = 0.8817$ 。图 4 是 IDLS 迭代计算中预测残差  $Q_A$  和  $Q_B$  随迭代的变化, 结果显示, 预测计算残差平方和由最初的  $Q_A^1 = 0.7223$  下降到  $Q_A^{98} = 0.3668$ 。

表 2 训练集的 90 个 MHC-I 短肽序列和生物活性及训练计算结果

Table 2 Calculation results of 90 MHC-I training peptides

No.	Peptide Sequence	Expt pIC <sub>50</sub>	Pred pIC <sub>50</sub>	pIC <sub>50</sub> Diff	No.	Peptide Sequence	Expt pIC <sub>50</sub>	Pred pIC <sub>50</sub>	pIC <sub>50</sub> Diff
1	VALVGLFVL	5.148	5.7543	-0.6063	46	VVMGTLVAL	7.174	7.3163	-0.1423
2	GTLVALVGL	5.342	5.9368	-0.5948	47	YLEPGVPTI	7.187	7.1654	0.0216
3	LQTTIHDII	5.501	5.8143	-0.3133	48	GLSRVVARL	7.248	7.4620	-0.2131
4	SLHVGQTCA	5.842	6.1580	-0.3160	49	LLAQFTSAI	7.301	7.4302	-0.1292
5	ALPYWNFAT	5.869	6.6416	-0.7726	50	VLLDYQGML	7.328	7.5911	-0.2631
6	SLNFMGYVI	5.881	5.9560	-0.0750	51	YLEPGPVTV	7.342	7.4078	-0.0658
7	NLQSLTNLL	6.000	6.6992	-0.6992	52	ILSFMPLL	7.3470	7.1400	0.2070
8	FVTWHRHYHL	6.025	5.7230	0.3020	53	YLSPGPVTA	7.383	7.5610	-0.1780
9	DPKVKQWPL	6.176	5.7407	0.4354	54	IIDQVPFSV	7.398	7.6528	-0.2548
10	ITSQVPFSV	6.196	6.5888	-0.3928	55	SVYDFVWL	7.444	7.3654	0.0786
11	ALAKAAAI	6.211	6.2433	-0.0323	56	ITWQVPFSV	7.463	7.4417	0.0213
12	GLGQVPLIV	6.301	6.5651	-0.2641	57	ITYQVPFSV	7.480	7.6613	-0.1813
13	MLDLQPETT	6.335	6.8570	-0.5220	58	GLYSTVPV	7.481	7.6303	-0.1493
14	LLSSNLSWL	6.342	6.3502	-0.0082	59	VMGTLVALV	7.553	7.2369	0.3161
15	GLACHQLCA	6.380	6.0594	0.3206	60	LLLCLIFLL	7.585	7.1406	0.4444
16	LIGNESFAL	6.415	7.0559	-0.6409	61	SLDDYNHLV	7.585	7.1764	0.4086
17	ALAKAAAAV	6.419	6.4857	-0.0667	62	VLIQRNPQL	7.644	6.9473	0.6967
18	LLAVGATKV	6.477	6.5115	-0.0344	63	SLYADSPSV	7.658	7.7106	-0.0526
19	ALAKAAAAAL	6.511	6.2262	0.2848	64	ILSQVPFSV	7.699	7.6472	0.0518
20	WILRGTSFV	6.556	6.9084	-0.3524	65	IMDQVPFSV	7.719	8.0305	-0.3115
21	IISCTCPTV	6.580	6.6649	-0.0849	66	QLFEDNYAL	7.764	7.4713	0.2927
22	FLGGTPVCL	6.623	6.8756	-0.2526	67	ALMDKSLHV	7.770	7.5250	0.2450
23	ALIHHTHL	6.623	6.7908	-0.1677	68	YAILPVS	7.796	7.6075	0.1885
24	NLSWLSLDV	6.639	6.0466	0.5924	69	FVWLHYYSV	7.824	8.1149	-0.2909
25	YMIMVKCWM	6.663	6.6427	0.02035	70	MLGTHTMEV	7.845	7.3180	0.5270
26	VLQAGFLL	6.682	7.0412	-0.3592	71	LLFGYPVYV	7.886	8.0253	-0.1393
27	GTLGIVCPI	6.714	6.5233	0.1907	72	ILKEPVHGV	7.921	7.5915	0.3295
28	VILGVLLLI	6.785	7.4728	-0.6878	73	YLMGPVTV	7.932	7.9139	0.0181
29	VTWHRHYHLL	6.793	6.5597	0.2333	74	WLDQVPFSV	7.939	7.9514	-0.0124
30	PLLPIFFCL	6.796	7.5217	-0.7257	75	KTWGQYWQV	7.955	7.6934	0.2616
31	TLGIVCPIC	6.815	5.9499	0.8651	76	ALMPYACI	8.000	7.4383	0.5617
32	CLTSTVQLV	6.832	7.1061	-0.2741	77	YLAPGPVTA	8.032	7.6408	0.3912
33	ILLCLIFL	6.845	6.7815	0.0635	78	YLYPGPVTV	8.051	8.3112	-0.2602
34	FAFRDLCIV	6.886	6.6689	0.2171	79	LLMGTLGIV	8.097	7.6769	0.4201
35	FLEPGPVTA	6.898	7.4940	-0.5960	80	YLWPGPVTV	8.125	8.0916	0.0334
36	ALAKAAAAA	6.947	6.8081	0.1389	81	FLLTRILTI	8.149	7.8796	0.2694
37	LMAVVLASL	6.954	7.4908	-0.5368	82	GLLGWSPQA	8.237	8.2184	0.0185
38	YVITQHWL	6.983	6.3410	0.6420	83	ILYQVPFSV	8.310	8.7197	-0.4097
39	LLCLIFLLV	6.996	7.5015	-0.5055	84	GILTVILGV	8.347	7.8414	0.5056
40	ITAQVPFSV	7.020	6.6685	0.3515	85	NMVPFFPPV	8.398	8.0854	0.3126
41	YLEPGPVTL	7.058	7.1483	-0.0903	86	ILDQVPFSV	8.481	7.6904	0.7906
42	YTDQVPFSV	7.066	7.0742	-0.0082	87	YLFPGPVTA	8.495	8.3473	0.1477
43	NLYVSLLLL	7.114	6.9769	0.1371	88	YLDQVPFSV	8.638	8.1326	0.5054
44	ILHNGAYSL	7.127	7.3493	-0.2223	89	ILFQVPFSV	8.699	8.4335	0.2655
45	SIISAVVGI	7.159	7.3048	-0.1458	90	ILWQVPFSV	8.770	8.5002	0.2698

Statistical Indices:  
R = 0.887132

R<sup>2</sup> = 0.787003

RES = 0.366873

SEE = 0.038672

表 3 二层次 QSAR 预测模型性质参数的系数和序列位置参数的系数

Table 3 Prediction coefficients of two-level QSAR model position

No.	Property	Coefficient { $a_i$ }	No.	Position (Residue)	Coefficient { $b_j$ }
1	Lip	-0.02445	1	R1	2.53268
2	Hyd	0.19258	2	R2	8.36712
3	S <sup>L</sup>	-0.00212	3	R3	3.06856
4	S <sup>H</sup>	0.00348	4	R4	-4.89559
5	P <sub><math>\alpha</math></sub>	0.15367	5	R5	3.12686
6	P <sub><math>\beta</math></sub>	0.07823	6	R6	2.45367
7	P <sub><math>c</math></sub>	0.19764	7	R7	1.24669
8	Vol	0.00366	8	R8	-3.50416
-	-	-	9	R9	-3.79249

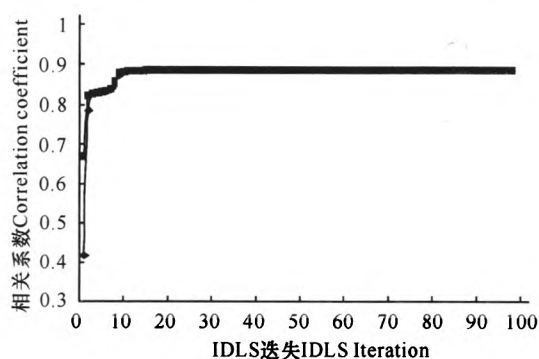


图 3 计算的与实验的活性相关系数随 IDLS 迭代的变化

Fig. 3 The correlation coefficients in IDLS iterations

◆: 从 { $a_i^{(n)}$ } 得到的相关系数; ■: 从 { $b_j^{(n)}$ } 得到的相关系数。  
◆:  $R_a$  iteration for { $a_i$ }, ■:  $R_b$  iteration for { $b_j$ }.

表 4 测试集中的 40 个 MHC-I 短肽和测试计算的结果

Table 4 Prediction results of 40 MHC-I peptides

No.	Sequence	Expt pIC <sub>50</sub>	Pred pIC <sub>50</sub>	pIC <sub>50</sub> Diff	No.	Sequence	Expt pIC <sub>50</sub>	Pred pIC <sub>50</sub>	pIC <sub>50</sub> Diff
1	LLGCAANWI	5.301	5.1708	0.1302	21	ITFQVPFSV	7.179	7.3750	-0.1960
2	SAANDPIFV	5.342	4.8592	0.4828	22	FTDQVPFSV	7.212	6.8379	0.3741
3	TTAEAAAGI	5.380	5.4678	-0.0878	23	RLMKQDFSV	7.342	7.5681	-0.2261
4	LTVILGVLL	5.580	5.3216	0.2584	24	KLHLYSHPI	7.352	6.6450	0.7070
5	HLLVGSSGL	5.792	6.4811	-0.6891	25	ITMQVPFSV	7.398	7.2641	0.1340
6	GIGILTVIL	6.000	5.7321	0.2679	26	KIFGLAFL	7.478	6.7818	0.6962
7	TVILGVLL	6.072	5.4662	0.6058	27	ALVGLFVLL	7.585	7.3852	0.1998
8	WTDQVPFSV	6.145	6.8930	-0.7480	28	YLSGPVTV	7.642	7.2387	0.4033
9	AIKAAAAV	6.176	6.4480	-0.2720	29	GLYSTVPV	7.699	7.6303	0.0687
10	ILTVILGVL	6.419	7.0160	-0.5970	30	YLYPGVTA	7.772	8.6335	-0.8615
11	AVAKAAAAV	6.495	5.9131	0.5819	31	YLAPGPVTV	7.818	7.3184	0.4996
12	ILDEAYVMA	6.623	7.4445	-0.8215	32	VVLGVVFGI	7.845	7.4509	0.3941
13	LLWFHISCL	6.682	6.3594	0.3226	33	MMWYWGPSL	7.921	7.4007	0.5203
14	TLDSQVMSL	6.793	7.2566	-0.4636	34	ILAQVPFSV	7.939	7.7270	0.2120
15	HLYQGCQVV	6.832	7.6799	-0.8479	35	FLLSLGIHL	8.053	8.1578	-0.1048
16	QLFHLCLII	6.886	7.6475	-0.7615	36	ILMQVPFSV	8.125	8.3225	-0.1975
17	ITDQVPFSV	6.947	6.6320	0.3150	37	YLFPGPVTV	8.237	8.0249	0.2121
18	ALCRWGLLL	7.000	7.2766	-0.2766	38	YLMGPVTA	8.367	8.2363	0.1307
19	NLGNLNSI	7.119	7.0974	0.02160	39	YLWGPVTA	8.495	8.4140	0.0810
20	HLYSHPIIL	7.131	7.5663	-0.4353	40	FLDQVPFSV	8.658	7.8964	0.7616
Statistical Indices:		$R^2 = 0.753202$		$RES = 0.469728$		$SEE = 0.074271$			
$R = 0.867872$									

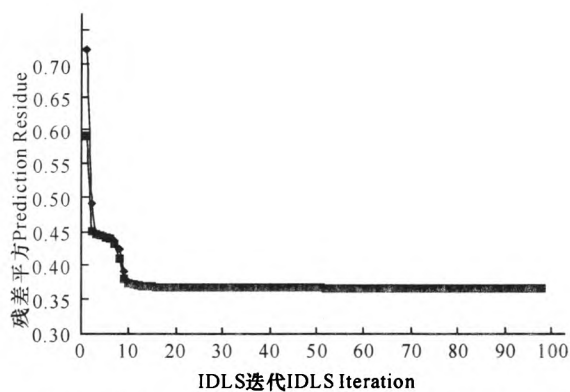


图 4 计算与实验的活性残差平方和随 IDLS 迭代的变化

Fig. 4 The summation of squared differences in the IDLS iterations

◆: 从 { $a_i^{(n)}$ } 得到的残差平方和; ■: 从 { $b_j^{(n)}$ } 得到的残差平方和。  
◆:  $Q_a$  iteration for { $a_i$ }, ■:  $Q_b$  iteration for { $b_j$ }.

### 3 结论

基于分子片段的二层次 QSAR 预测网络的最大优点是把普通 QSAR 使用的分子整体性质参数改为分子片段的性质参数, 准确地描述了药物分子取代基的细微变化造成的分子生物活性的改变。迭代的双最小二乘 (IDLS) 法把普通 QSAR 的单方向的算法改为双方向、有反馈和学习功能的算法, 因而大大提高了 QSAR 模型的预测能力。在本文的算例中, 最初的相关系数 ( $R^{(1)} = 0.4167$ ) 实际上是使用分子整体

性质的普通 QSAR 的预测能力的度量,而迭代收敛后的相关系数( $R^{(98)} = 0.8817$ )是双层次、双方向的 QSAR 预测网络和 IDLS 算法的改进效果。从信息论的角度看,二层次的预测网络比单层次的预测模型有更大的信息量。二层次的 QSAR 预测网络可以为药物分子设计提供分子性质和片段结构两方面的信息。

#### 致谢

感谢杜奇石教授(美国戈登生命科学研究所资深研究员)的帮助和指导!

#### 参考文献:

- [1] Kubinyi H. QSAR: In Hansch analysis and related approaches; Chap4[M]. VCH; Weinheim, 1993.
- [2] Lowis D R. In HQSAR: A new highly prediction QSAR technique[J]. Tripos Technical Notes, 1997, 1(5): 1-17.
- [3] Heritage T W, Lowis D R. Molecular hologram QSAR. Ch 14. In Rational drug design[P]. ACS symposium series 719, ACS, Washington DC. 212(1999).
- [4] Hansch C, Leo A. In exploring QSAR: Fundamentals and applications in chemistry and biology. ACS Professional Reference Book [M]. ACS, Washington DC. (1995).
- [5] Tripos International. SYBYL-X 1.3[CP]. 1699 South Hanley Rd, St Louis, MO, 63144, USA.
- [6] Chemical computing group Inc. Molecular Operating Environment (MOE), 2013. 08[CP]. 1010 Sherbooke St. West, Suite # 910, Montreal, QC, Canada, H3A 2R7, 2013.
- [7] Du Q S, Huang R B, Wei Y T, et al. Fragment-based quantitative structure-activity relationship (FB-QSAR) for fragment-based drug design[J]. J Comput Chem, 2009, 30(2): 295-304.
- [8] Du Q S, Huang R B, Wei Y T, et al. Multiple field three dimensional quantitative structure-activity relationship (MF-3D-QSAR)[J]. J Comput Chem, 2008, 29(2): 211-219.
- [9] Imai J, Hasegawa H, Maruya M, et al. Exogenous antigens are processed through the endoplasmic reticulum-associated degradation (ERAD) in cross-presentation by dendritic cells[J]. Int Immunol, 2005, 17(1): 45-53.
- [10] Koopmann J O, Albring J, Hüter E, et al. Export of antigenic peptides from the endoplasmic reticulum intersects with retrograde protein translocation through the Sec61p channel [J]. Immunity, 2000, 13 (1): 117 - 127.

(责任编辑:尹 闯)

## 13 项标准入围广西 2011 年度重要技术标准项目

新闻时间:2014-1-16

日前,广西壮族自治区政府办公厅公布广西 2011 年度重要技术标准项目,综合近年来经济社会效益等因素考虑,3 项国家标准、2 项行业标准和 8 项地方标准入围最终榜单。

据了解,广西年度重要技术标准项目审核周期一般为两年,专家组结合项目成效、经济社会效益等因素做出评定。在广西 2011 年度重要标准项目榜单当中,3 项国家标准分别为缆索用环氧涂层钢(GB/T 25835—2010)、工程机械子午线轮胎无损检验方法 X 射线法(GB/T 26276—2010)、交直流电表校验仪校准规范(JJF 1284—2011),2 项行业标准为液化石油气发动机技术条件(QC/T 693—2011)、肉桂丰产栽培技术规程(LY/T 1945—2011),8 项地方标准为桑蚕主要病害防治技术规程(DB45/T 731—2011)、制糖企业能源计量器具配备和管理要求(DB45/T 737—2011)、大型游乐设施日常维护保养规范(DB45/T 726—2011)、猪脑心肌炎病毒的检测反转录聚合酶链反应法(DB45/T 749—2011)、人面果绿化苗木培育技术规程(DB45/T 770—2011)、黄藤种苗质量要求(DB45/T 709—2010)、阳朔民居旅馆服务质量要求与等级划分(DB45/T 765—2011)、里当鸡养殖技术规程(DB45/T 734—2011)。

据悉,2011 年度,广西企事业单位、高等院校和科研院所以及标准化工作者和科技人员充分发挥能动性和创造性,以实施技术标准发展战略为抓手,围绕千亿元产业和战略性新兴产业及优势特色产业发展需要,主导制定了一批具有前沿性、关键性和共性技术的国家标准、行业标准和地方标准,显著提高了标准对经济社会发展的引领支撑作用和贡献率。

(摘自人民网)