

Computer Aided Drug Design

——Quantitative Structure- Activity Relationships (QSAR)

Qin Xu

<http://cbb.sjtu.edu.cn/~qinxu/CADD.htm>

Structure-Activity Relationships (SAR)

- Rule of 5 (drug like)
 - donor < 5, acceptor > 10, mass < 500Da, LogP < 5
- Rule of 3 (lead like)
 - donor ≤ 3, acceptor ≥ 3, mas < 300Da, LogP ≥ 3, rotational bond ≤ 3, polar SA ≤ 60Å

Molecular
Parameters

log P

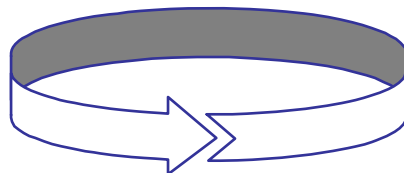
σ

MR

MLP

...

experiences



Biological activity

IC₅₀

K_i

MIC

Permeation

...

Quantitative Structure-Activity Relationships (QSAR)

- **Mathematical Models**

Found a quantitative relationship

Molecular
Parameters

$\log P$

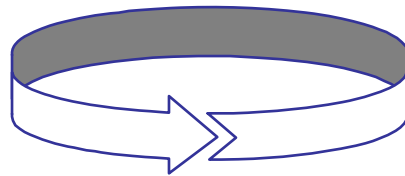
σ

MR

MLP

...

Statistical tool



Biological activity

IC_{50}

K_i

MIC

Permeation

...

Traditional methods - 2D(1964)

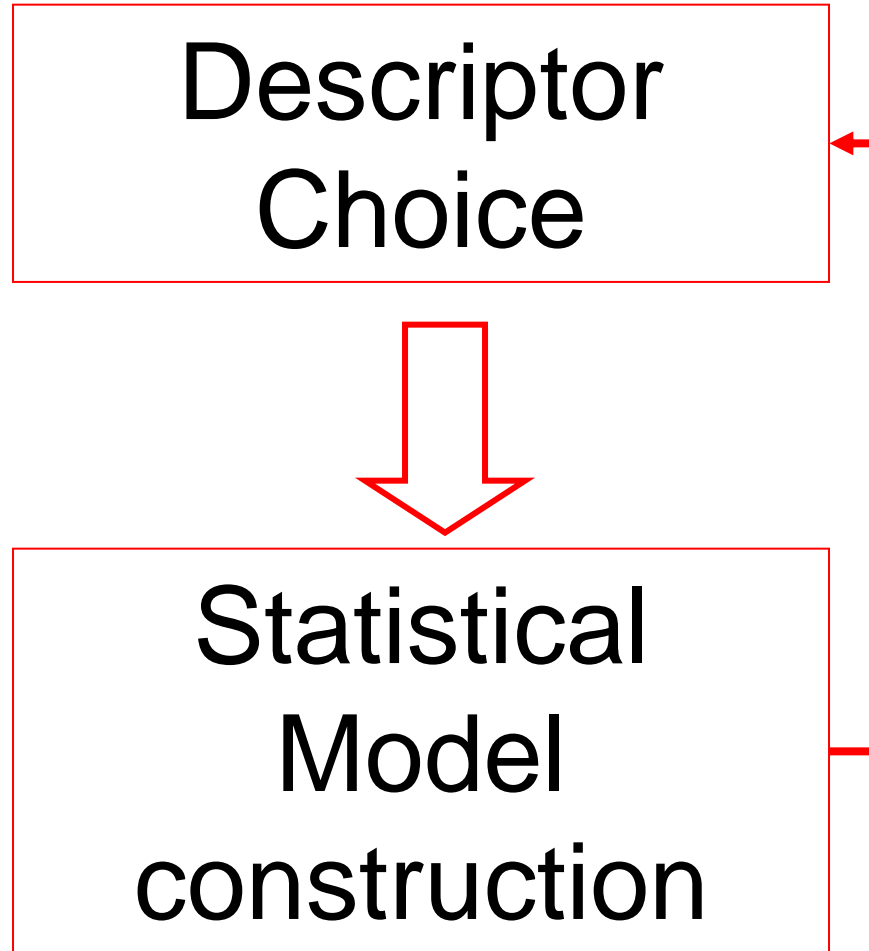
- parameters:
 - Lipophilicity ($\log P$, π)
 - Electronics (σ , ρ , Hammett correlation)
 - Sterics (E_s , molar refractivity)
- Statistical analysis
 - One parameter : Linear regression
 - Several Parameters : Multilinear regression
- Models of linear equations:

$$\text{Log } 1/C = k_1 P_1 + k_2 P_2 + k_3 S + k_4 E_s + k_5$$

3D-QSAR(1988)

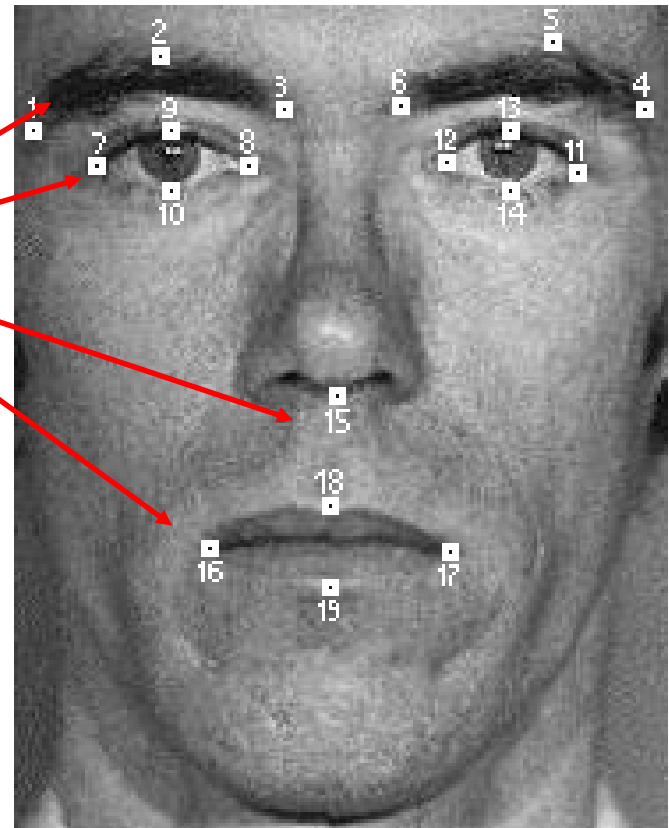
- Parameters
 - Molecular Field calculation (software)
 - ^fField contributions: Steric and electrostatic fields from CoMFA.
Steric, Electrostatic, Hydrophobic, H-bond Donor and Acceptor fields
 - Interpretation of CoMFA results
 - Three dimensional structure
- Describe ligand-receptor interactions
 - In vitro bioactivity

The key of QSAR



Feature Selection

- Face identification
- Key features.
- The same applies to molecules.



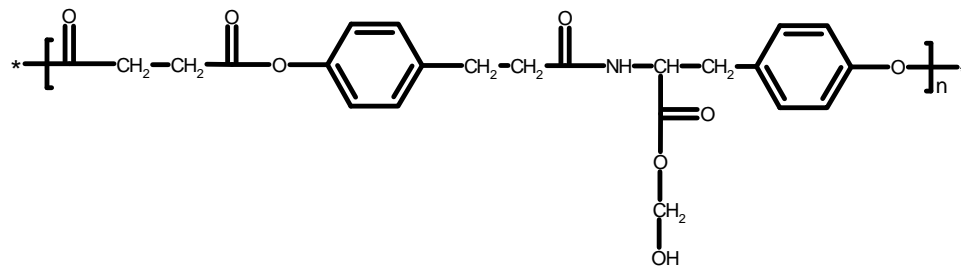
Descriptor Choice



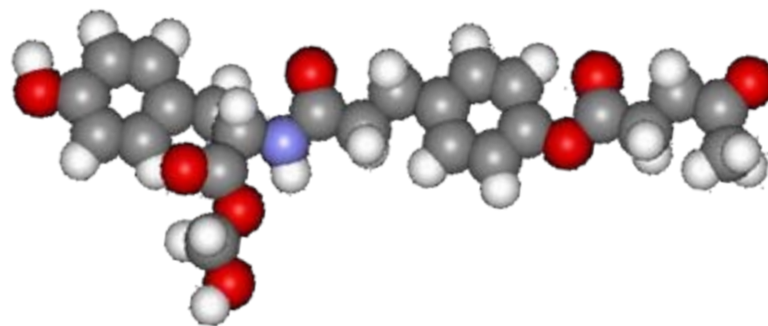
- 乔木、灌木
- 阔叶、针叶
- 落叶、常绿

Types of descriptor for molecules

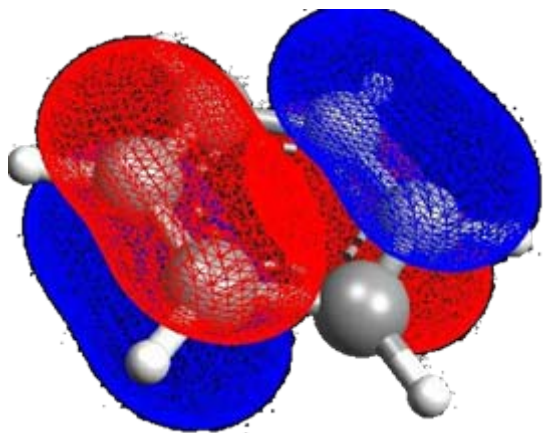
Topology



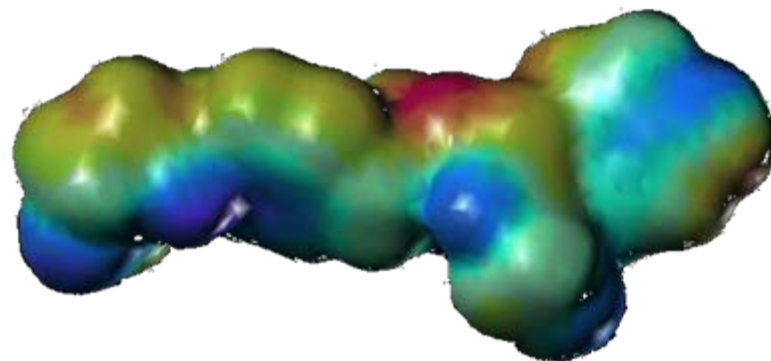
Geometry



Quantum



Electrostatic

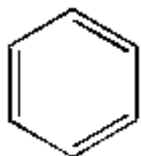


Fingerprint

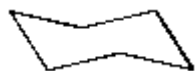
If a universe of features

$U = \{ \text{is-aromatic, has-ring, has-C, has-N, has-O, has-S, has-P, has-halogen} \}$
there are $2^8=256$ possible fingerprints

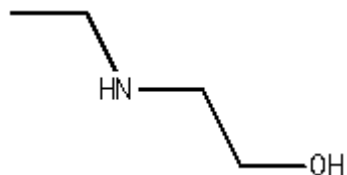
Tanimoto coefficient between fingerprints X and Y is defined to be:
 $\# \text{ features in intersect}(A,B) / \# \text{ features in union}(A,B)$



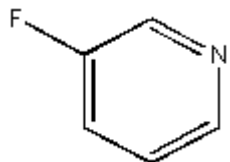
{ is-aromatic, has-ring, has-C }



{ has-ring, has-C }



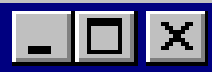
{ has-C, has-N, has-O }



{ is-aromatic, has-ring, has-C, has-N, has-F }



Fingerprint Chooser (Cluster)



Fingerprint:

Field	Description
FP:BIT_MACCS	MACCS Structural Keys (Bit packed)
FP:MACCS	MACCS Structural Keys
FP:TAD	Typed Atom Distances
FP:TAT	Typed Atom Triangles
FP:TGD	Typed Graph Distances
FP:TGT	Typed Graph Triangles



Similarity Metric:

Code	Description
tanimoto	Tanimoto Coefficient
tanimoto-ss	Tanimoto Superset/Subset



OK

Cancel

Search

Database: i:/moe/sample/mol/gmp_inhibitors.mdb

Entries: all selected visible hidden

Query

Fingerprint: MACCS Structural Keys (Bit packed) (FP:BIT_MACCS)
Metric: Tanimoto Coefficient (tanimoto)

Overlap:  85

Atoms: Form Query From Selected Atoms Only

Action

Visibility: Show Hits Hide Hits Nothing

Selection: Select Hits Unselect Hits Nothing

No Query

Set Fingerprint...

Close

QSAR models

C = molar concentration that causes a certain biological effect

values of the regression coefficients

95% confidence intervals of the coefficients and the constant term

$$\text{Log } 1/C = 1,15 (\pm 0,2) \pi - 1,46 (\pm 0,4) \sigma^+ + 7,82 (\pm 0,2) \quad (8)$$

logarithms of reciprocal values are the correct scaling

lipophilicity parameter

electronic parameter

constant term

$$(n = 22; r = 0,945; s = 0,196; F = 78,6; Q^2 = 0.841; s_{\text{PRESS}} = 0.238)$$

number of compounds

correlation coefficient r ; measure for the relative quality of a model

Fisher value; measure for the statistical significance

standard deviation s ; measure for the absolute quality of a model

standard deviation of cross-validation predictions and squared cross-validation correlation coefficient (measures for internal predictivity)

Useful parameters

- **Log P**:(partition coefficient) Hydrophobicity
 $P = [\text{drug}] \text{ in octanol} / [\text{drug}] \text{ in water}$
- **p**:substituent hydrophobicity constant
 $\pi_x = \log P_x - \log P_H$
- **Hammett constant (1940)** σ :
 - Measure e-withdrawing or e-donating effects (compared to benzoic acid & how affected its ionization)
- **Taft's steric factor (Es)** (~1956)
 - an experimental value based on rate constants
- **Molar refractivity (MR)**
 - measure of the volume occupied by an atom or group
 - equation includes the MW, density, and the index of refraction
- **Verloop steric parameter-**
 - computer program uses bond angles, van der Waals radii, bond lengths

Sterimol parameter

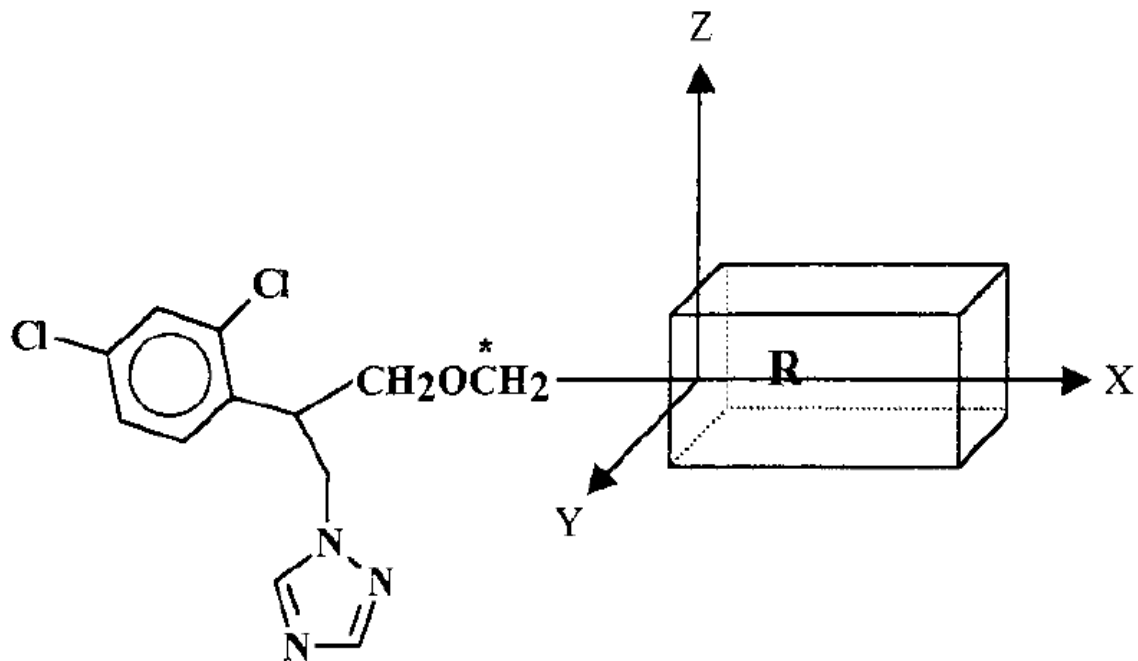


Figure 2. Definition of some of the parameters of Table 3. Geometrical descriptors: $|X|_{\max} = X1$; $|Y|_{\max} = Y1$; $|Y|_{\min} = Y2$; $|Z|_{\max} = Z1$; $|Z|_{\min} = Z2$. Charge: * indicates the carbon used for the charge.

3D-QSAR Parameters

- Traditional CoMFA fields

- Steric fields, Lennard-Jones potential

$$E_j = \sum_{k=1}^{\text{natoms}} \epsilon \cdot \left[\left(\frac{r_{\text{probe}} + r_k}{r_{ij}} \right)^{12} - 2 \cdot \left(\frac{r_{\text{probe}} + r_k}{r_{jk}} \right)^6 \right]$$

- Electrostatic fields, Coulomb potential

$$E_j = \sum_{k=1}^{\text{natoms}} \frac{q_{\text{probe}} \cdot q_k}{\epsilon \cdot r_{jk}}$$

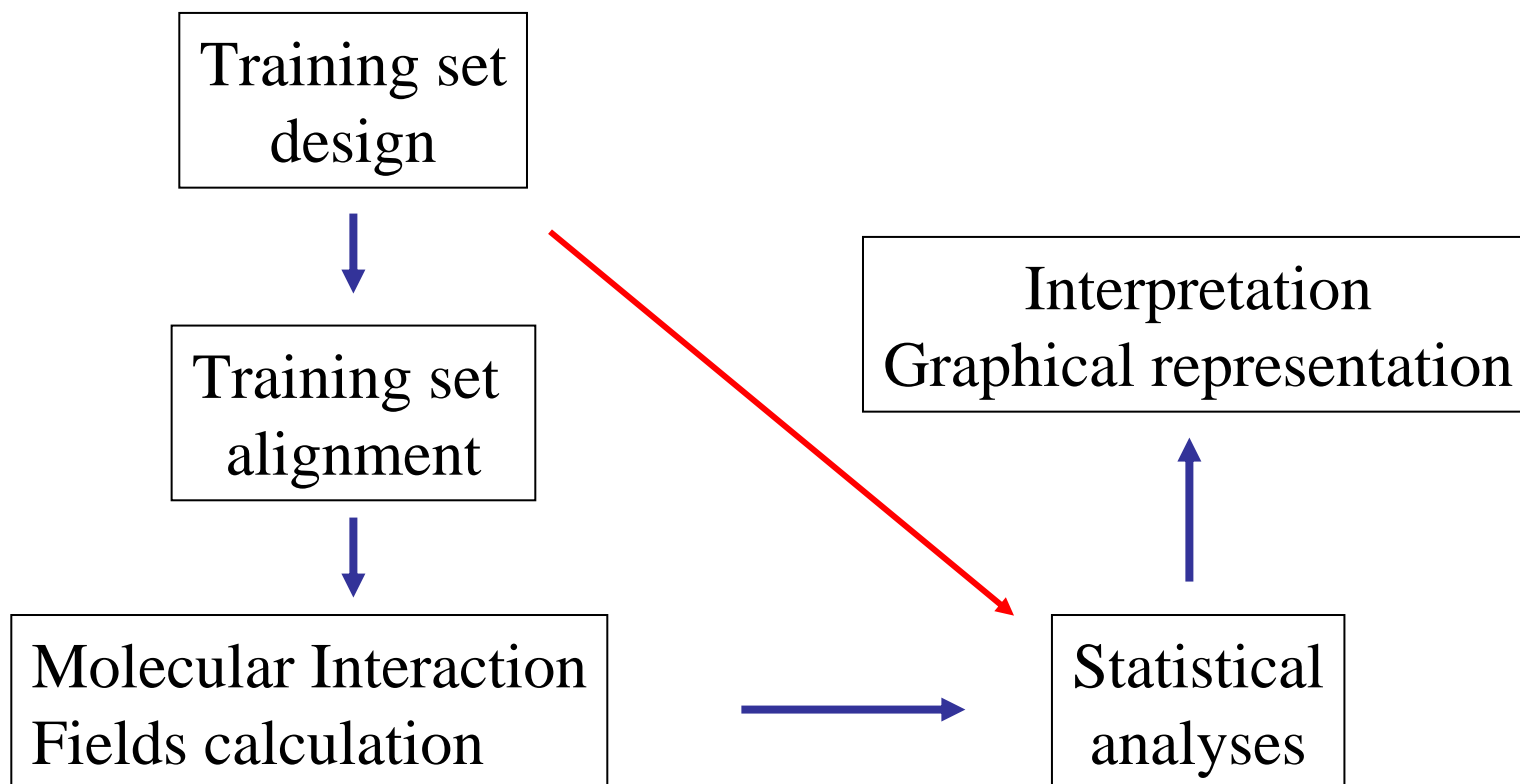
More 3D-QSAR Parameters

- Additional fields in CoMFA
 - Interaction energies with functional groups (GRID software)
 - hydrophobic field (HINT software)
 - Molecular Lipophilicity Field (CLIP software)

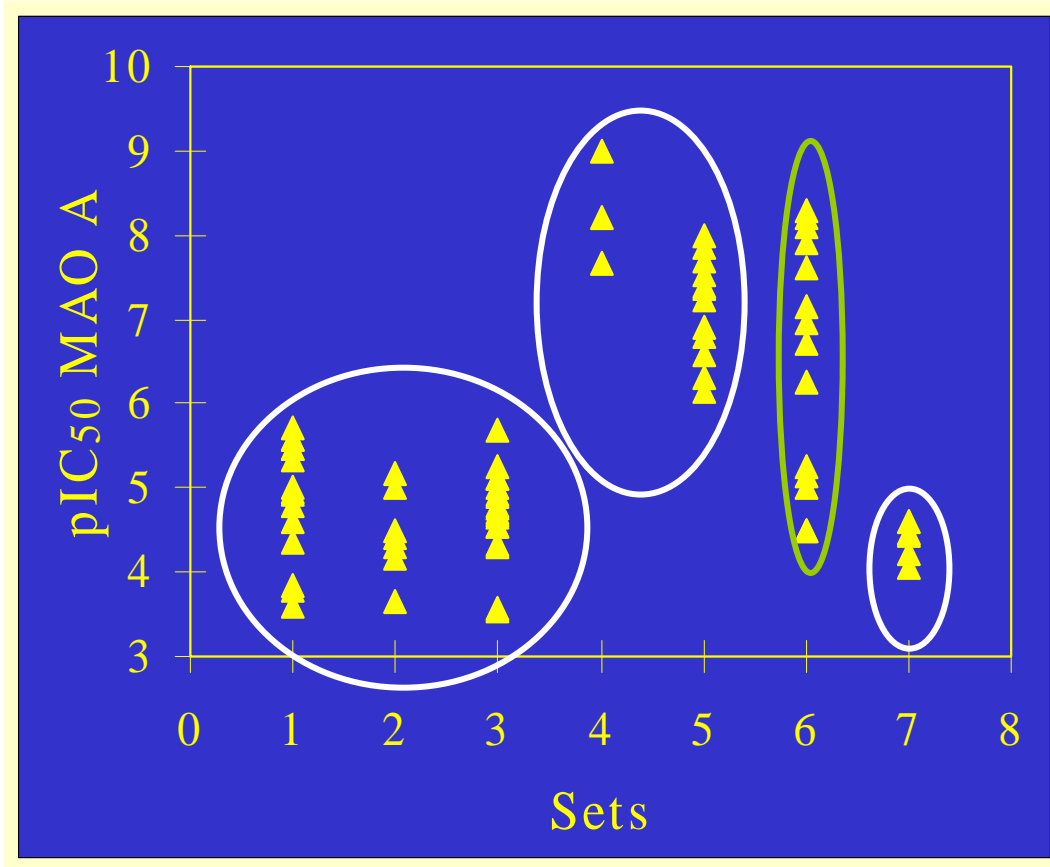
Biological activity parameters

- IC50 (50% inhibiting concentration),
- Ki (inhibitory constant),
- MIC (minimum inhibitory concentration),
- Permeation

Workflow



Training set design



QSAR method

- Hansch method
- Free-Wilson method
- Craig Plots
- Topliss operational schemes

3D-QSAR methods

- **Methods**

- Molecular-shape analysis [1]
- Hypothetical Active Site Lattice (HASL) [2]
- RECEPT Programs [3]
- Crippen's distance geometry [4]
- Voronoi binding sites [5]
- Comparative Molecular Field Analysis (CoMFA) [6]

- [1] Hopfinger A.J. *J. Am. Chem. Soc.* **1980**, 102, 7196-7206
- [2] Doweiko A.M. *J. Med. Chem.* **1988**, 31, 1396-1406
- [3] Kato Y.; Itai A.; Iitake Y. *Tetrahedron* **1987**, 43, 5229-5236
- [4] Ghose A.K.; Crippen G.M. *J. Med. Chem.* **1985**, 28, 333-346
- [5] Boulu L.D.; Crippen G.M. *J. Comp. Chem.* **1989**, 10, 673-682
- [6] Cramer R.D. III; Patterson D.E.; Bunce J.D. *J. Am. Chem. Soc.* **1988**, 110, 5959-5967

Statistical models

- Linear (quantitative model)
 - LR (linear regression)
 - MLR (multi-linear regression)
 - PCR (principle component regression)
 - PLS (partial least square)
 - SVR (support vector regression)
- Nonlinear (qualitative model)
 - SVM (support vector machine)
 - Bayesian statistics

Evaluation of QSAR models

拟合效果

- 相关系数**R**,

$$R = \sqrt{1 - \frac{\sum (y_{calc} - y_{exp})^2}{\sum (y_{calc} - y_{mean})^2}}$$

- 均方根偏差**RMSE**

$$RMSE = \sqrt{1 - \frac{\sum (y_{calc} - y_{exp})^2}{n - k - 1}}$$

- **Fisher**检验值**F**

$$F = \sqrt{\frac{R^2 (n - k - 1)}{k(1 - R)^2}}$$

y_{calc} 为计算活性; y_{exp} 为实验活性; y_{mean}

为计算活性平均值; n 为样本个数; k 为变量个数

Descriptor Choice

编号	描述子名称	物理意义	类别
1	Pmi	主惯性矩	空间参数
2	E_vdw	范德华构成的势能	电子参数
3	PmiY	以Y做主惯性矩	空间参数
4	PmiZ	以Z做主惯性矩	空间参数
5	Zagreb	Zagreb指数	拓扑参数
6	PEOE- VSA+1	在 $[0.05, 0.1)$ 之间 q_i 的 v_i 之和	空间参数

描述子及其所代表的物理量及类型

Descriptor Combinations	RMSE	R	F 值
1	0.40642	0.37643	2.698
1+2	0.26828	0.79120	11.671
1+2+3	0.23969	0.70148	12.369
1+2+3+4	0.21909	0.86641	13.366
1+2+3+4+5	0.16442	0.92729	22.814
1+2+3+4+5+6	0.09823	0.97461	30.763

根据QSAR模型的拟合能力评价指标确定描述子的组合

Evaluation

- 交叉验证

- 相关系数

$$R_{cv} = \sqrt{1 - \frac{\sum (y_{pred} - y_{exp})^2}{\sum (y_{calc} - y_{mean})^2}}$$

- 预测误差平方

$$PRESS = \sum (y_{pred} - y_{exp})^2$$

- 预测集的预测

$R_{cv} \rightarrow 1$ or $PRESS \rightarrow 0$

Better prediction

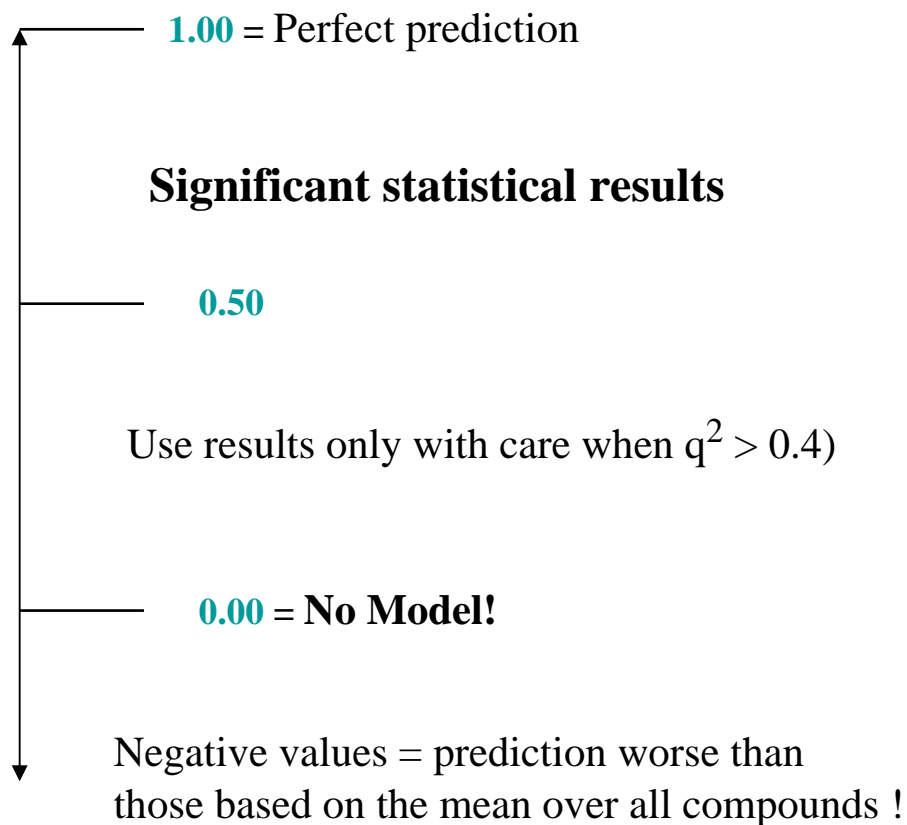
Statistical value for 3D-QSAR

- For the step of crossvalidation:
 - Cross-validated correlation coefficient, q^2 .
 - Optimal number of components, N .

- For the final model:
 - Squared of correlation coefficient, r^2 .
 - Standard error of estimate, s .
 - Residuals.
 - F values.

Crossvalidated r^2_{cv} (q^2)

$$q^2 = 1.0 - \frac{\sum (y_{predicted} - y_{experimental})^2}{\sum (y_{predicted} - \bar{y})^2}$$



A example by MOE

- Compound database
 - mytrps.mdb
- Statistical Model
 - PLS

Descriptor Choice

QuaSAR-Descriptor

Database File: / /mytrps.mdb Selected Entries Only

Molecule Field: molecule

Auto Select: Database Fields Selected Database Fields Clear Selection

Descriptors Selected: 32

CODE	CLASS	DESCRIPTION
PEOE_VSA+0	2D	Total positive 0 vdw surface area
PEOE_VSA+1	2D	Total positive 1 vdw surface area
PEOE_VSA+2	2D	Total positive 2 vdw surface area
PEOE_VSA+3	2D	Total positive 3 vdw surface area
PEOE_VSA+4	2D	Total positive 4 vdw surface area
PEOE_VSA+5	2D	Total positive 5 vdw surface area
PEOE_VSA+6	2D	Total positive 6 vdw surface area
PEOE_VSA-0	2D	Total negative 0 vdw surface area
PEOE_VSA-1	2D	Total negative 1 vdw surface area
PEOE_VSA-2	2D	Total negative 2 vdw surface area
PEOE_VSA-3	2D	Total negative 3 vdw surface area
PEOE_VSA-4	2D	Total negative 4 vdw surface area
PEOE_VSA-5	2D	Total negative 5 vdw surface area
PEOE_VSA-6	2D	Total negative 6 vdw surface area
PEOE_VSA_FHYD	2D	Fractional hydrophobic vdw surface area
PEOE_VSA_FNEG	2D	Fractional negative vdw surface area
PEOE_VSA_FPNEG	2D	Fractional polar negative vdw surface area
PEOE_VSA_FPOL	2D	Fractional polar vdw surface area
PEOE_VSA_FPOS	2D	Fractional positive vdw surface area
PEOE_VSA_FPPOS	2D	Fractional polar positive vdw surface area

Class: All 2D 3D x3D

Filter: pео

OK Cancel

Database Viewer: ...md+drugdesign/计算机辅助药物设计/第四章 定量构效...

	molecule	index	pki-trypsin	PEOE_VSA+0	PEOE_VSA+1	PEOE_VSA-0
1	↖	1	6.7700	166.6551	58.8359	12
2	↖	14	7.1310	111.4545	61.3827	20
3	↖	8	6.2010	52.2182	63.1013	0
4	↖	56	4.8540	133.1898	23.5355	3
5	↖	22	5.9590	87.9260	115.8015	14
6	↖	20	6.2840	35.1567	117.9567	0
7	↖	39	6.1020	213.2828	43.9829	0
8	↖	66	4.5380	105.5288	26.6814	30
9	↖	70	3.9280	25.0511	61.3827	20
10	↖	62	4.4950	139.9789	43.9107	0
11	↖	6	6.7700	159.9957	107.2708	0
12	↖	19	6.6580	99.9298	73.1477	22
13	↖	7	6.2010	121.7980	31.0351	8
14	↖	82	6.0090	209.4729	43.9829	12
15	↖	42	6.9210	113.1221	103.0054	0
16	↖	35	6.0040	151.0610	56.7791	0
17	↖	48	6.6580	35.1567	160.0401	0
18	↖	69	4.5380	35.1567	73.1477	0
19	↖	54	6.0710	177.2480	88.7919	0
20	↖	80	6.1490	86.5924	112.1223	0
21	↖	63	4.6020	71.1915	73.1477	0
22	↖	71	4.5090	37.3652	122.8972	0
23	↖	25	6.1610	190.0442	43.9829	0
24	↖	36	5.9210	90.5787	115.8015	3
25	↖	5	6.1190	200.9421	43.9107	12

87 entries 0 selected all visible 36 fields 0 selected all visible

DBV|Compute|Descriptor

87 molecules altogether

Subsets

Popup|Sort up

Diverse Subset

Database: / /mytrps.mdb

Selected Entries Only

Molecule: molecule ▾

Output Limit: 56

Cluster Field: (none) ▾

Method: Descriptors | Fingerprint | Conformation

Options:

index	PEOE_VSA+6	SlogP_VSA0
pki-trypsin	PEOE_VSA-0	SlogP_VSA1
PEOE_VSA+0	PEOE_VSA-1	SlogP_VSA2
PEOE_VSA+1	PEOE_VSA-2	SlogP_VSA3
PEOE_VSA+2	PEOE_VSA-3	SlogP_VSA4
PEOE_VSA+3	PEOE_VSA-4	SlogP_VSA5
PEOE_VSA+4	PEOE_VSA-5	SlogP_VSA6
PEOE_VSA+5	PEOE_VSA-6	SlogP_VSA7

Normalize Descriptors to Unit Variance.

OK Cancel

	SMR_VSA3	SMR_VSA4	SMR_VSA5	SMR_VSA6	SMR_VSA7	SDIVPRIO
	40.7999	195.3002	81.3450	110.7670		1
	32.8972	250.6802	55.8411	39.9668		2
	38.0431	250.6802	44.4659	39.9668		3
	32.8972	127.1780	51.7825	133.3041		4
	38.0431	212.9434	107.5301	6.6407		5
	38.0431	322.0856	10.0734	6.6407		41
	32.8972	162.4643	49.7246	239.4940		30
	65.7944	89.4411	81.3450	38.3319		6
	38.0431	212.9434	15.7610	6.6407		13
	32.8972	127.1780	7.5867	206.1680		88
	38.0431	212.9434	86.1750	72.4351		26
	38.0431	250.6802	44.4659	26.7860		9
	32.8972	146.0464	67.3870	39.9668		15
	32.8972	124.7275	81.3450	244.9287		11
	38.0431	357.3358	49.7246	6.6407		12
	32.8972	162.4643	52.2113	142.0027		21
	38.0431	212.9434	81.3450	6.6407		31
	38.0431	212.9434	15.7610	6.6407		17
	32.8972	271.6065	52.2113	206.5968		7
	32.8972	160.0139	86.3184	39.9668		36
	38.0431	212.9434	7.5867	39.5379		27
	40.7999	283.5162	51.4149	6.6407		29
	32.8972	162.4643	44.4659	239.9229		34
	38.0431	320.8242	15.3322	42.0246		24
25	28.3388	32.8972	127.1780	81.3450	211.6027	16

87 entries, 0 selected, all visible. 36 fields, 0 selected, all visible.

DBV|Compute|Diverse Subset

Training set and Test set

The screenshot shows a software window with a menu bar (File, Entry, Field, Compute, Display, Window, Help) and a 'Cancel' button. Below the menu is a table with columns: molecule, index, pki-trypsin, PEOE_VSA+0, PEOE_VSA+1, and PEOE. Rows 41-44 are visible. Two dialog boxes are overlaid on the table. The top dialog, 'Save Database As', has a 'Filename:' field with a 'Browse...' button, 'Open in Viewer:' options (None, New, Current), and checkboxes for 'Selected Fields Only' and 'Selected Entries Only'. The bottom dialog is a file browser showing a list of files in the 'Files' pane, including binary-test.mdb, binary-train.mdb, binary.mdb, merged_bbb.mdb, merged_sareport.mdb, my_bb.mdb, mytrps-test.mdb, mytrps-train-refined1.mdb, mytrps-train-refined2.mdb, mytrps-train-refined3.mdb, mytrps-train.mdb, mytrps.mdb, sareport.mdb, and tangjian.mdb. The status bar at the bottom reads: '87 entries, 56 selected, all visible. 36 fields, 1 selected, all visible.'

	molecule	index	pki-trypsin	PEOE_VSA+0	PEOE_VSA+1	PEOE
41	↖	20	6.2840	35.1567	117.9567	0
42	↖	59	5.1020	118.5069	77.4130	3
43	↖	29	5.3470	35.1567	116.1977	6

Save Database As

Filename: Browse...

Open in Viewer: None New Current

Selected Fields Only

Selected Entries Only

OK Cancel

.mdb

Directories

Files

- binary-test.mdb
- binary-train.mdb
- binary.mdb
- merged_bbb.mdb
- merged_sareport.mdb
- my_bb.mdb
- mytrps-test.mdb
- mytrps-train-refined1.mdb
- mytrps-train-refined2.mdb
- mytrps-train-refined3.mdb
- mytrps-train.mdb
- mytrps.mdb
- sareport.mdb
- tangjian.mdb

OK Cancel

87 entries, 56 selected, all visible. 36 fields, 1 selected, all visible.

Statistical model

Database: / /mytrps.mdb Selected Entries Only

Status: No Model

Activity Field: pki-trypsin Method: PLS

Weight Field: (NONE)

Options: Component Limit: 0 Condition Limit: 1e+006

Binary Threshold: 0 Smooth: 0.25

Model:

Descriptor
SlogP_VSA9
SMR_VSA0
SMR_VSA1
SMR_VSA2
SMR_VSA3
SMR_VSA4
SMR_VSA5
SMR_VSA6
SMR_VSA7
\$DIVPRIO

Sort by Normalized Sort by Name



Database: /mytrps-train.mdb Selected Entries Only

Status: PLS Model: RMSE=0.289242 R2=0.890215

Activity Field: pki-trypsin Method: PLS

Weight Field: (NONE)

Options: Component Limit: 0 Condition Limit: 1e+006

Binary Threshold: 0 Smooth: 0.25

Model:

Descriptor	Coefficient	Normalized
index		
pki-trypsin		
PEOE_VSA+0	-0.06167	-3.61906
PEOE_VSA+1	-0.06600	-2.63682
PEOE_VSA+2	-0.01319	-0.10532
PEOE_VSA+3	-0.17390	-1.37540
PEOE_VSA+4	-0.08432	-0.64509
PEOE_VSA+5	-0.08280	-0.62412
PEOE_VSA+6	-0.07364	-0.49471
PEOE_VSA-0	-0.06894	-2.21951

Sort by Normalized Sort by Name

DBV|Compute|QuaSAR-Model

Validation

QuaSAR-Model

Database: /mytrps-train.mdb Selected Entries Only

Status: Waiting for Validation Options...

Activity Field: pki-trypsin Method: PLS

Weight Field: (NONE)

Options: Component Limit: 0 Condition Limit: 1e+006

Binary Threshold: 0 Smooth: 0.25

Descriptor	Coefficient	Normalized
index		
pki-trypsin		
PEOE_VSA+0	-0.06167	-3.61906
PEOE_VSA+1	-0.06600	-2.63682
PEOE_VSA+2	-0.01319	-0.10532
PEOE_VSA+3	-0.17390	-1.37540
PEOE_VSA+4	-0.08432	-0.64509
PEOE_VSA+5	-0.08280	-0.62412
PEOE_VSA+6	-0.07364	-0.49471
PEOE_VSA-0	-0.06894	-2.21951

Sort by Normalized Sort by Name

QuaSAR-Model

Database: /mytrps-train.mdb Selected Entries Only

Status: RMSE=0.289242 R2=0.890218 XRMSE=1.79447 XR2=0.033869

Activity Field: pki-trypsin Method: PLS

Weight Field: (NONE)

Options: Component Limit: 0 Condition Limit: 1e+006

Binary Threshold: 0 Smooth: 0.25

Descriptor	Coefficient	Normalized
index		
pki-trypsin		
PEOE_VSA+0	-0.06167	-3.61906
PEOE_VSA+1	-0.06600	-2.63682
PEOE_VSA+2	-0.01319	-0.10532
PEOE_VSA+3	-0.17390	-1.37540
PEOE_VSA+4	-0.08432	-0.64509
PEOE_VSA+5	-0.08280	-0.62412
PEOE_VSA+6	-0.07364	-0.49471
PEOE_VSA-0	-0.06894	-2.21951

Sort by Normalized Sort by Name

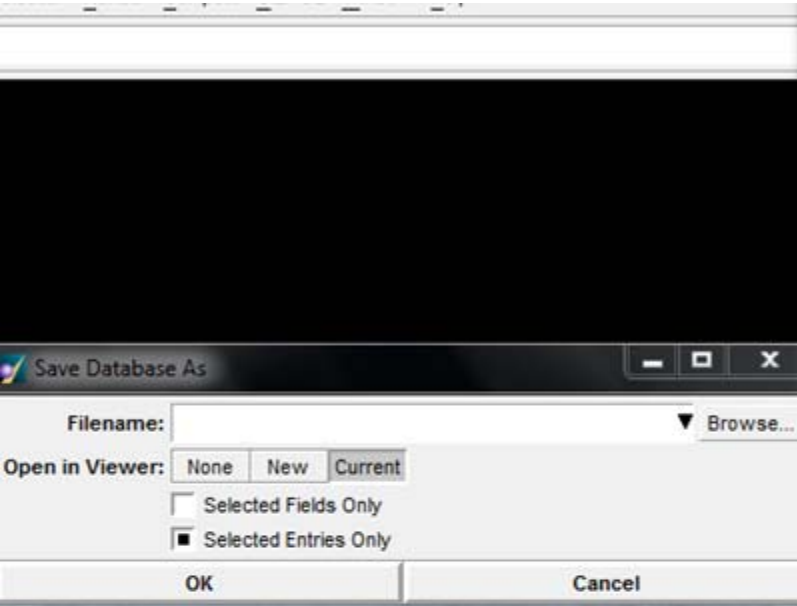
QuaSAR Fit Validation

Database: /mytrps-train.mdb

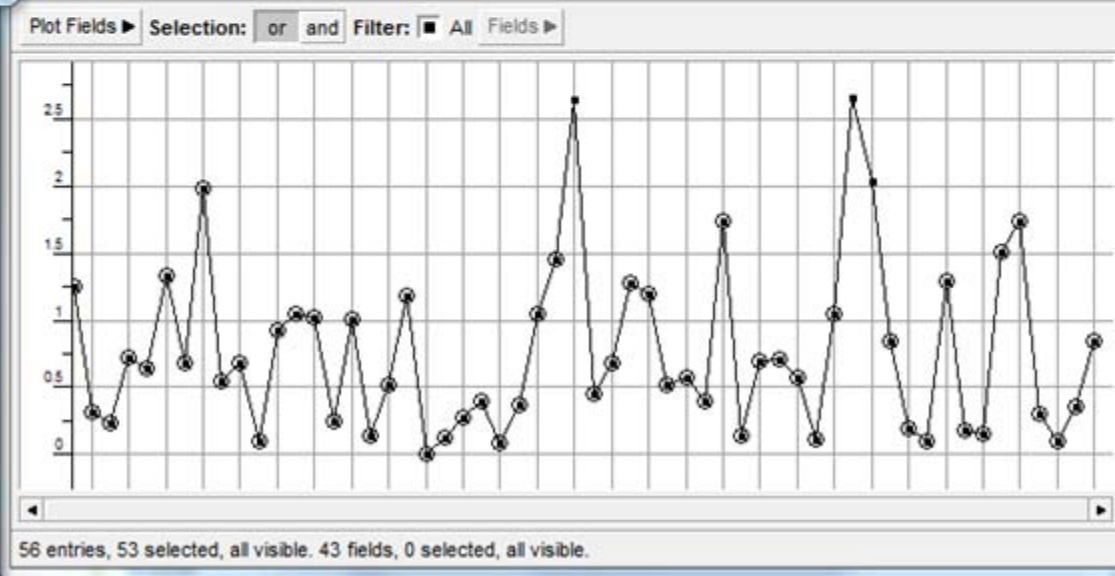
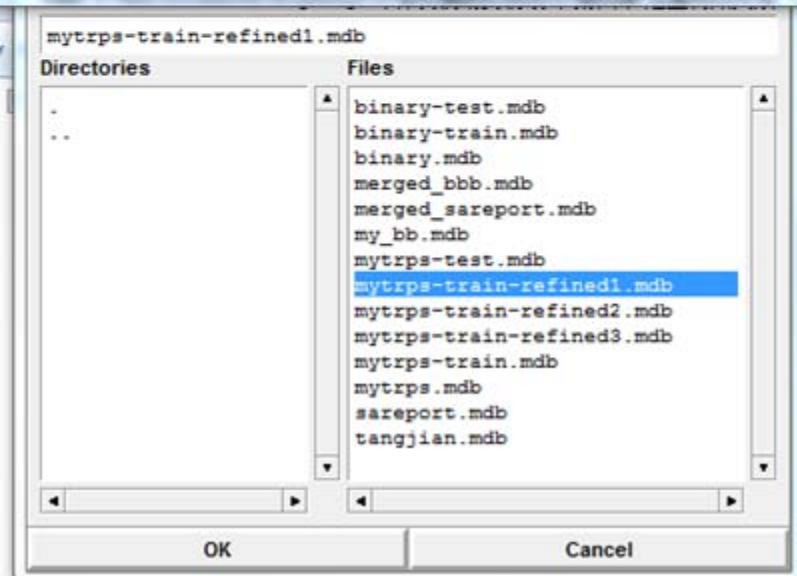
Model Validation: Write Predictions
 Write Prediction Residuals
 Write Prediction Z-Scores

Cross Validation: Write Predictions
 Write Prediction Residuals
 Write Prediction Z-Scores

Outliers?



	activity	\$SPRED	\$RES	\$Z-SCORE	\$XPRED	\$XRES	\$XZ-SCORE
36	0	5.3499	0.5041	1.7428	4.9237	0.9303	3.3602
37	0	5.7046	0.0404	0.1398	5.6793	0.0657	0.2252
38	0	5.4763	0.2017	0.6973	5.2809	0.3971	1.3723
39	0	5.4518	0.2062	0.7129	5.2971	0.3609	1.2465
40	0	5.4526	0.1674	0.5789	5.2764	0.3436	1.1847
41	0	5.5698	0.0322	0.1114	5.5519	0.0501	0.1716
42	0	5.7473	-0.3033	1.0486	10.3032	-4.8592	20.1099
43	0	6.1673	-0.7693	2.6597	6.7053	-1.3073	5.0545
44	0	4.7569	0.5901	2.0402	3.3839	1.9631	7.7526
45	0	5.3480	-0.2460	0.8505	5.4828	-0.3808	1.3180
46	0	4.7980	0.0560	0.1936	-1.8944	6.7484	24.1155



Fit model II → Validation II

The image displays the QuaSAR software interface, showing the model fit and validation options. The main window is titled "QuaSAR-Model" and shows the following settings:

- Database: /mytrps-train-refined1.mdb
- Status: Waiting for Validation Options...
- Activity Field: pki-trypsin
- Method: PLS
- Weight Field: (NONE)
- Options: Component Limit: 0, Condition Limit: 1e+006, Binary Threshold: 0, Smooth: 0.25
- Model: Descriptor Coefficient Normalized

The model coefficients are listed below:

Descriptor	Coefficient	Normalized
index		
pki-trypsin		
PEOE_VSA+0	-0.02264	-1.28749
PEOE_VSA+1	-0.03021	-1.19084
PEOE_VSA+2	-0.05629	-0.44958
PEOE_VSA+3	-0.10136	-0.79753
PEOE_VSA+4	+0.02838	+0.21530
PEOE_VSA+6	-0.01320	-0.09692
PEOE_VSA+6	-0.03520	-0.22798
PEOE_VSA-0	-0.01880	-0.58696

The "QuaSAR Fit Validation" dialog box is open, showing the following options:

- Database: /mytrps-train-refined1.mdb
- Model Validation: Write Predictions, Write Prediction Residuals, Write Prediction Z-Scores
- Cross Validation: Write Predictions, Write Prediction Residuals, Write Prediction Z-Scores

The "Database viewer" window shows a table with the following columns: \$RES, \$Z-SCORE, \$XPRED, \$XRES, and \$XZ-SCORE. The table contains 25 rows of data.

		\$RES	\$Z-SCORE	\$XPRED	\$XRES	\$XZ-SCORE
1	76	-0.0276	0.1452	7.2490	-0.4790	2.4833
2	93	-0.0083	0.0437	7.8261	-0.6951	3.6346
3	93	0.0017	0.0089	5.4241	0.7769	4.0589
4	18	0.0422	0.2223	-0.3584	5.2124	28.9333
5	96	-0.1306	0.6885	7.2087	-1.2497	6.8235
6	01	0.1239	0.6531	6.0738	0.2102	1.1050
7	60	0.3560	1.8765	5.6117	0.4903	2.6855
8	11	-0.0331	0.1744	6.4133	-1.8753	9.9545
9	45	0.0235	0.1237	2.2456	1.6824	8.8764
10	95	0.1055	0.5564	4.3354	0.1596	0.8372
11	81	-0.3181	1.6767	7.6699	-0.8999	5.0965
12	72	0.0808	0.4258	4.8314	1.8266	9.9288
13	31	0.0979	0.5160	3.9998	2.2012	12.2028
14	15	-0.1025	0.5403	6.1910	-0.1820	0.9548
15	09	0.1401	0.7384	6.7005	0.2205	1.1608
16	30	-0.5290	2.7883	7.2717	-1.2677	8.2197
17	71	0.1209	0.6375	5.7693	0.8887	4.7767
18	80	-0.0800	0.4217	7.6494	-3.1114	17.4220
19	15	-0.2125	1.1200	6.5175	-0.3685	1.9648
20	32	-0.2912	1.5351	4.9902	-0.3882	2.0898
21	11	0.0079	0.0419	4.3195	0.1895	0.9896
22	57	-0.2747	1.4481	6.5833	-0.4223	2.2750
23	04	-0.0794	0.4183	6.1002	-0.1792	0.9390
24	72	0.1718	0.9054	5.7739	0.3451	1.8305
25	81	-0.1141	0.5012	6.5771	-0.1871	0.8550

53 entries, 0 selected, all visible. 42 fields, 0 selected, all visible.

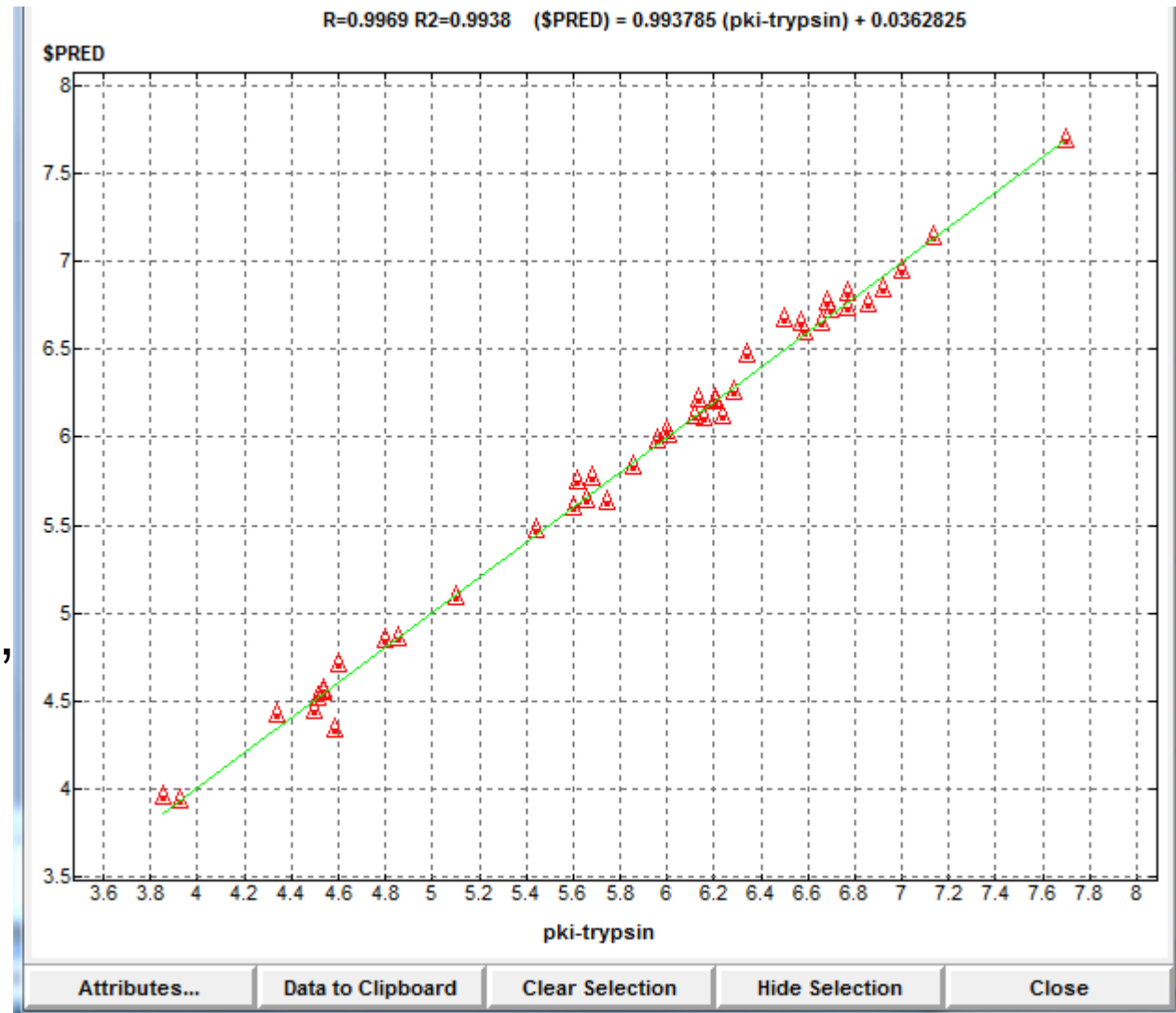
After refinement

The screenshot shows the QuaSAR-Model software interface. The window title is "QuaSAR-Model". The "Database:" field contains "/mytrps-train-refined1.mdb" and there is a checkbox for "Selected Entries Only". The "Status:" field displays "RMSE=0.18971 R2=0.954685 XRMSE=1.18992 XR2=0.306414", with the latter two values highlighted in a red box. The "Activity Field:" is set to "pki-trypsin" and the "Method:" is "PLS". The "Weight Field:" is "(NONE)". The "Options:" section includes "Component Limit:" (0), "Condition Limit:" (1e+006), "Binary Threshold:" (0), and "Smooth:" (0.25). The "Model:" section contains a table with three columns: "Descriptor", "Coefficient", and "Normalized". The table lists several descriptors, with the last three rows highlighted in blue.

Descriptor	Coefficient	Normalized
\$DIVPRIO		
\$PRED		
\$RES		
\$XPRED		
\$XRES		
\$XZ-SCORE		
\$Z-SCORE		
PEOE_VSA+0	-0.02264	-1.28749
PEOE_VSA+1	-0.03021	-1.19084
PEOE_VSA+2	-0.05629	-0.44958

At the bottom of the model list, there are checkboxes for "Sort by Normalized" (unchecked) and "Sort by Name" (checked), and a "Select Database Fields" button. The bottom of the window features five buttons: "Fit", "Report", "Validate", "Save", and "Close".

Training dataset III



Again, delete outliers with Z-SCORE>1.5,

→44 entries

RMSE=0.0739945,

R2=0.993785,

XRMSE=0.48619,

XR2=0.770392

→trypsinpred.fit

DBV|Compute|Analysis|Correlation plot

Model Evaluation

DBV|Compute|Model-Evaluate

Model-Evaluate

Database: / /mytrps-test.mdb Browse...

Model File: / /trypsinpred.fit Browse...

Field: trypsin-pred

Entries: All Selected Visible

Use Descriptors in Database if Available

OK Cancel

Molecular Database Calculator

Database: sign/ / /mytrps-test.mdb

Use Selected Entries Only

<< < Del > >>

Available Fields:

- {pki-trypsin} - {trypsinpred}
- SMR_VSA7
- \$DIVPRIO
- trypsin-pred
- trypsin-res
- trypsinpred

pow	log	sqr	abs	sin	cos	tan	ceil
exp	log10	sqrt	cbrr	asin	acos	atan	floor
PI	e	index	selected	RAND	()	USER
<	>	max	min	7	8	9	+
<=	>=	mean	std_dev	4	5	6	-
==	<>	fquart	tquart	1	2	3	*
and	or	median	sum	0	.	+/-	/

USER Function:

Destination Field: trypsinres

Destination Type: float

Evaluate Clear Close

File Entry Field Compute Display Window Help Cancel

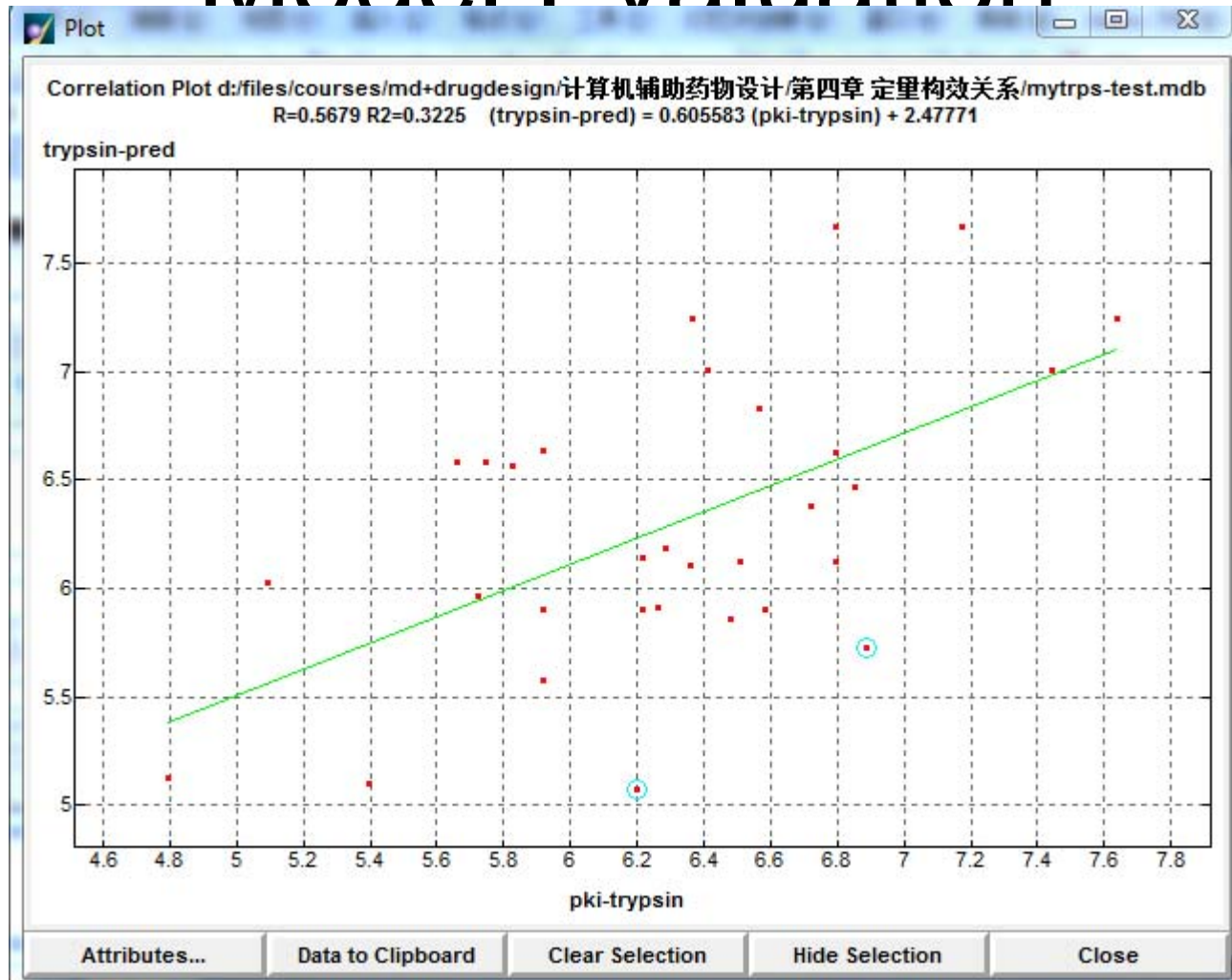
	6	SMR_VSA7	\$DIVPRIO	trypsin-pred	trypsin-res	trypsinpred
1	50	144.9507	88	6.6244	0.1716	6.6244
2	59	39.9668	88	6.4688	0.3852	6.4688
3	13	42.0246	88	7.0056	0.4384	7.0056
4	22	6.6407	88	5.7245	1.1615	5.7245
5	59	6.6407	88	5.9092	0.3508	5.9092
6	59	73.2928	88	6.1868	0.0972	6.1868
7	59	6.6407	88	5.8592	0.6218	5.8592
8	17	75.3506	88	6.5786	-0.9207	6.5786
9	63	6.6407	88	6.5644	-0.7404	6.5644
10	63	39.9668	88	5.1040	0.2940	5.1040
11	13	42.0246	88	7.0056	-0.5966	7.0056

Plot Fields Selection: or and Filter: All Fields

31 entries, 2 selected, all visible. 39 fields, 2 selected, all visible.

DBV|Compute|Calculator

Model Evaluation



Course Outline

- Introduction and Case Study
- Drug Targets
 - Sequence analysis
 - Protein structure prediction
 - Molecular simulation
- Molecular Docking
- Drug Design
 - QSAR
 - Pharmacophore
 - De novo Drug Design
 - Combinatorial library