Title: Connecting traditional QSAR and molecular simulations of papain hydrolysis—importance of charge transfer.

Authors: Lepp, Z.; Chuman*, H.

Department of Pharmaceutical Sciences, The University of Tokushima Shomachi
Tokushima 770–8505, Japan.
E-mail: hchuman@ph.tokushima-u.ac.jp; Tel./fax: 81-88-633-9508.

Source: Bioorg. Med. Chem. 2005, 13(9), 3093–3105.

Compounds: 23 N-benzoylglycine derivatives if type I, where

\[ R = \text{H, NH}_2, \text{F, CH}_3, \text{OCH}_3, \text{Cl, CN, SO}_2\text{NH}_2, \text{CONH}_2, \text{COCH}_3, \text{NHCOCH}_3, \text{CF}_3, \text{C}_2\text{H}_5. \]

Biological material: Papain, a cystein protease.

Data taken from the literature:

- [Michaelis-Menten constant for the papain hydrolysis of the N-benzoylglycine esters (dimension not given)];
- [the crystallographic structure of the cystein protease papain was obtained from the Brookhaven Protein Data Bank (PDB code 1KHQ), this crystal structure contains a covalently bonded diazomethylketone inhibitor connected to the sulfur atom of residue cystein°, which was removed before building complexes].

Computational methods:

- Molecular modeling [the protein-ligand complexes used for molecular dynamics (MD) simulations were generated by docking the ligand into the active site by the FlexX, Cscore calculations were used for ranking and the docked conformers the active site was defined by a radius of 6.5 Å around the inhibitor of the crystallographic structure, N benzoylglycine ligands were built by PC Spartan, the coordinates of hydrogens were optimized by molecular mechanics minimization using the Tripos force field and Sybyl v6.91, MD simulations were used to obtain optimal conformation for complexes, the Amber ff99 and GAFF force fields were used for describing the protein and ligands, respectively, the atomic charges for ligands were calculated by carrying out single point calculation on the HF/6 31G* level using Gaussian98 and were fitted using the standard RESP method in Amber v7, the protein was reported to be positively charged and neutralized by teap by adding five chlorine anions, a TIP3 water cap of 20 Å radius around the Cys° residue was added];

\[ pK_m = 3.976 (\pm 1.881) F + 0.168 (\pm 0.093) Z + 0.884 (\pm 0.218) \tau + 3.104 \]

\[ n = 23.3 \pm 0.734 s = 0.303 F = 7.4 \]

POC, PCA (Principal Coordinates Analysis and Principal Component Analysis, respectively, were performed using the program MVSP v3.12);

FlexX (program for automatic protein-ligand docking based on incremental construction without manual intervention implemented in Sybyl v6.91);

MLR (Multivariate Linear Regression analysis was performed using XLstat pro7.5).

Data calculated:

- Geometry descriptors [generation of geometrical descriptors for regression analyses (Fig. 1): D1-D3: dihedral (torsion) angles between atoms C1–C3–O1–C7, C3–O1 C7–C8, and O1–C7–C8–N1, respectively; Z—van der Waals distance from O1 to the end of the molecule through the axis of O1–C3];

\[ C2, O1 \] (charge on atoms C2 and O1);

\[ WS_{O1} \] (average watershell around atom O1);

\[ WdW_{Vol} \] (van der Waals volume (Å³));

\[ Z \] (Z distance as depicted in Fig. 1).

Chemical descriptors:

- p (Hansch-Fujita’s substituent constant characterizing hydrophobicity);
- o (Hammett’s constant characterizing the electron-withdrawing power of the substituent);
- F, R (Swain-Lupton’s electronic parameters, characterizing the field and resonance effects, respectively).

Results: MD simulations and full structure LocalSCF semi-empirical quantum mechanics calculations of receptor-ligand complexes were carried out to investigate structure-activity relationship for the papain hydrolysis of a set of 23 N-benzoylglycine esters of type I and various atomic charge and structural parameters of complexes as descriptors. Traditional QSAR descriptors were reinterpreted using detailed structural information. A moderately significant linear regression equations were calculated (Eqs. 1, 2) and it was shown that the pattern of charge distribution on the ester group was different if charges of free or complex ligands were analyzed. The charges of two significant atoms, namely the O1, which is at the reaction center and the C2, were used which are the closest independent non-substituted atom to the substituent, water-shell and torsional angle values as descriptors.
pK_m = 0.242 ± 0.056 C2 – 0.316 ± 0.062 O1 
- 0.186 ± 0.059 WS_O1 – 0.218 ± 0.062 D2 
+ 4.289 ± 0.053 n = 23 r = 0.833 s = 0.253 F = 10.2 (2)

The effects of complexation on the electronic structure of ligands were also studied by multivariate analyses (PCA and PCO) and a trend was found for the change of inter-ligand correlation of atomic charges by complexation. The correlation among these charges after complex formation was different for the case of molecules substituted at para- or at diameta positions. It was suggested that the results can help to understand how traditional QSAR descriptors, e.g., F or o, interact with other electronic effects during complex formation.

(B.B.)

121/2005

**Title:** Prediction of octanol-air partition coefficients of semivolatile organic compounds based on molecular connectivity index.

**Authors:** Zhao, H.; Zhang, Q.; Chen, J.; Xue, X.; Liang*, Xinmiao. Dalian Institute of Chemical Physics, Chinese Academy of Sciences Dalian 116011 People's Republic of China. E-mail: liangxm@dicp.ac.cn; Tel.: 86-411-8369-9015; Fax: 86-411-8360-6605.

**Source:** Chemosphere 2005, 59(10), 1421–1426.

**Compounds:** 10 polychlorinated dibenzo-p-dioxins and dibenzofurans (PCDD/Fs), 24 polychlorinated naphthalenes (PCNs), 4 polycyclic aromatic hydrocarbons (PAHs), 6 chlorobenzenes (CBs), and 13 polybrominated diphenyl ethers (PBDEs).

**Data taken from the literature:** logKOA (logarithm of the 1-octanol/air partition coefficient).

**Computational methods:** MLR (stepwise Multivariate Linear Regression analysis).

**Data calculated:** \( \gamma \) (nOth order, mOth type molecular connectivity indices proposed by Kier and Hall).

**Results:** Octanol-air partition coefficients of semivolatile organic compounds (PCNs, CBs, PAHs, PCDD/Fs and PBDEs) have been modeled based on molecular connectivity indices. The following statistically significant linear regression equations were calculated using stepwise MLR.

\[
\begin{align*}
\log K_{OA} & = 4.066 \gamma_6 - 2.376 \gamma_p + 1.387 \\
n &= 24 r = 0.991 s = 0.127 F = 598 \\
\log K_{OA} & = 1.354 \gamma_4 - 1.902 \\
n &= 4 r = 0.999 s = 0.061 F = 827 \\
\log K_{OA} & = 1.305 \gamma_p + 3.062 \\
n &= 10 r = 0.990 s = 0.173 F = 394 \\
\log K_{OA} & = 1.488 \gamma_6 - 1.217 \\
n &= 6 r = 0.996 s = 0.082 F = 558 \\
\log K_{OA} & = 2.25 \gamma_p - 0.86 \gamma_p - 4.879 \\
n &= 13 r = 0.963 s = 0.275 F = 64
\end{align*}
\]

Fig. 1 shows the plot of the experimental versus predicted value of logKOA values of PCNs.

Higher accuracy has been obtained in comparison with the models based on theoretical molecular descriptors. The correlation coefficients are greater than 0.99 except that for PBDEs, and the standard deviation is less than 1.83 log units, which is less than error measured by Harner et al. In addition, the computation of molecular connectivity index used in this paper is very simple and the acquisition of data is very easy. Therefore, this method is a viable alternative to predict the octanol-air partition coefficients from molecular structures.

(B.B.)

**Classical QSAR: Pharmacology**

122/2005

**Title:** Topological models for prediction of anti-HIV activity of acylthiocarbamates.

**Authors:** Bajaj, S.; Sambi, S. S.; Madan*, A. K. Faculty of Pharmaceutical Sciences, MD University Rohtak 124001, India. E-mail: madan_ak@yahoo.com; Tel.: 91-1262-212-111; 91-1262-272-535.

**Source:** Bioorg. Med. Chem. 2005, 13(9), 3263–3268.

**Compounds:**

- a) 33 Compounds of type I, where R1 = H, CH3; R2 = C6H5, CH2C6H5, C6H5OCH2, 2-furyl; R3 = C6H5, 4-F-C6H4; R4 = diverse acyl groups;
- b) 27 Compounds of type II. Where R = H, CH3, CH2CH3, Cl, OCH3, CF3, COCH3, F, Br, NO2, I, N(C2H5)2, OCH2CH3; Acyl = benzoyl, 2-thienoyl, 4-Cl-benzoyl, furoyl, 2-Cl-nicotinoyl.

**Biological material:**

- a) Human immunodeficiency virus type-1 (HIV-1);
- b) HIV-1 reverse transcriptase (HIV RT).

**Data taken from the literature:**

Activity [anti-HIV activity (details not given)].
Computational methods:

Model calculation (EC\textsubscript{50} values of the test compounds were estimated using an house program system and correlated with anti-HIV activity taken from the literature).

Data calculated:

EC\textsubscript{50} \textit{(effective concentration of the test substance (\textmu M) required for 50% inhibition of HIV RT was estimated using an house program system):}

W (Wiener index);

\(\gamma\) (first order molecular connectivity index);

\(\gamma^c\) (augmented eccentricty connectivity index calculated as \(\gamma^c = 2^{-s} (Mi/Ei), \text{where} M_i \text{ is the of all degrees of all vertices } (U_i), \text{ adjacent to vertex} i; \ E_i \text{ is the eccentricity, and} \ n \text{ is the number of vertices in graph} \ G, \text{ for a molecular graph (G),} U_1, U_2, \ldots, U_n \text{ are vertices and the number of first neighbors of a vertex} U_i \text{ is this vertex}).

Results: HIV RT inhibitory activity of 61 acylthiocarbamates of types I and II has been modeled using topological descriptors (W, \(\gamma\), and \(\gamma^c\)). Resulting data were analyzed and predictive models were developed after identification of the active ranges. In the next step biological activity was assigned to each of the compounds using these models which were compared with the reported anti-HIV activity. Very high accuracy of prediction ranging from 95% to 98% was achieved using the developed topological models (n, r, s, and F as well as the analytical form of the model based upon the augmented eccentric topological index is not given).

\[
EC_{50} = -0.0016 \ W + 5.42
\]

(1)

\[
EC_{50} = 3.9331 (\gamma)^2 - 107.89 \gamma + 740.15
\]

(2)

Comparison of the three QSAR models reveals the following data (descriptor, % classification, accuracy of prediction): W, 72.13, 97.72; \(\gamma\), 55.73, 97.06; \(\gamma^c\), 100, 95.08. Analysis of the structure-activity relationships of the compounds in the active range revealed the following features: activity is shown only by the O-2-(phthalimidoethyl) for the acylthiocarbamates of type II. All compounds of type I having other groups are inactive. Compounds having halogen or nitro group at other positions are inactive. Amongst five-member heterocyclic rings, activity is better with furfuryl group. The developed topological models possess a useful potential for designing lead structures for development of further potent anti-HIV agents.

123/2005

Title: Geometry, topology, and atom-weights assembly descriptors to predicting A1 adenosine receptors agonists.

Authors: González*, M. P.; Terán, C.; Tejeira, M.; Besada, P.

Service Unit, Experimental Sugar Cane Station ‘Villa Clara-Cienfuegos’

Ranchuelo, Villa Clara C. P. 53100, Cuba.

E-mail: mpgonzalez76@yahoo.es; Tel.: 53-42-281-473; Fax: 53-42-281-130.

Source: Bioorg. Med. Chem. 2005, 15(10), 2641–2645.

Compounds:

a) 32 Adenosine analogues of type I, where \(R^1 = H, 4-NO_2-\text{Ph}-CO, 5\text{-Cl-pyridin-2-yl-NH-CO-}: \ R^2 = \text{diverse substituents;} \ R^1 = H, I, n-C_3H_7-C=C, Ph-C=C, Ph(CH_2)OH-C=C, n-C_3H_7-C=C, Ph-C=C, Ph(CH_2)C=C;

b) [\(^3H\)]R-PIA.

Biological material: Adenosine receptor (A\(_1\)) in rat membranes.

Data taken from the literature: \(K_i\) (Michaelis inhibition constant (nM) representing the affinity of the substrate to displace [\(^3H\)]R PIA at the A1 receptor in rat brain membranes).

Computational methods:

GET-\textit{AWAY} (GEometry, Topology, and Atom-Weights As-
sem\textit{b}ly descriptors used for QSAR modeling).

Data calculated:

Descriptors [Topological, Galvez Topological Charges in-
dexes, Randic Molecular Profiles, Geometrical, WHIM, calculated using DRAGON software, and GETAWAY descriptors: CIC1, complementary information content (neighborhood symmetry of 1-order); SP12, shape profile no.12; SP03, shape profile no.03; SP04, shape profile no.04; SP12 shape profile no.12; FDI, folding degree index; H8v, H autocorrelation of lag 8/weighted by atomic van der Waals volumes; REIG, first eigenvalue of the R matrix; R2u+, R maximal autocorrelation of lag 2/un-weighted; R7u+, R maximal autocorrelation of lag 7/unweighted; R5v, R autocorrelation of lag 5/weighted by atomic van der Waals volumes; R1v+, R maximal autocorrelation of lag 1/ weighted by atomic van der Waals volumes; T(N.S), sum of topological distances between N.S; pIPCO9, molecular multiple path count of order 09; MPC09, molecular path count of order; VRA1, Randic-type eigenvector-based index from adjacency matrix; MSD, mean square distance index (Balaban); DP01, molecular profile no.01; SP07, Shape profile no.07; SP13, shape profile no.13; JGI5, mean topological charge index of order 5; GGH10, topological charge index of order 10; GG19, topological charge index of order 9; GGI8, topological charge index of order 8; GGI3, topological charge index of order 3; GGI2 topological charge index of order 2; W3D, 3D-Wiener index; AGDD, average geometric distance de-
gree; DDl, D/D index; ADDD, average distance/distance degree; MAXDP, maximal electro topological positive variation; Gu, G total symmetry index/unweighted; E2 s, second component accessibility directional WHIM index/weighted by atomic electro topological states; L2 s, second component size directional WHIM index/weighted by atomic electro topological states; G3 m, third component symmetry directional WHIM index/weighted by atomic masses; P2 m, second component shape directional WHIM index/weighted by atomic masses; E3u, third component accessibility directional WHIM index/unweighted;

\( q^2 \) (cross-validated correlation coefficient).

Results: The GETAWAY descriptors have been used for modeling the quantitative structure-activity relationships of 32 A1 adenosine receptors agonists of type I. A regression model (Eq. 1) has been developed accounting for more than 77% of the variance in the experimental activity (K).

\[
\log K_i = -8.143 H_{88} - 45.062 R_{2u} + 10.686 R_{2u}^2 + 91.678 R_{7u} - 11.937 R_{5v} + 29.425 R_{1v} + 23.886 \quad (1)
\]

Five further regression models employing topological, Galvez topological charges, Randic molecular profiles, geometrical, and WHIM descriptors, failed to give satisfactory models \( R^2 = 0.70 \) for this property with the same number of variables in the equation (descriptor type, r, s, F, q). \( q^2 \): topological, 0.837, 0.439, 9.8, 0.491, 0.573; Galvez topological charges indexes, 0.816, 0.464, 8.3, 0.427, 0.608; Randic molecular profiles, 0.800, 0.482, 7.4, 0.403, 0.620; geometrical, 0.789, 0.493, 6.9, 0.366, 0.640; WHIM, 0.879, 0.383, 0.481, 7.5, 0.371, 0.637. It was proposed that the GETAWAY approach should be used as an efficient alternative to DRAGON descriptors, in order to develop new analogues of the A1 adenosine receptor agonists.

(B. B.)

1242005

Title: Molecular modelling of human microsomal epoxide hydrolase (EH) by homology with a fungal (Aspergillus niger) EH crystal structure of 1.8 Å resolution. Structure-activity relationships in epoxides inhibiting EH activity.

Authors: Lewis*, D. F. V.; Lake, B. G.; Bird, M. G.

School of Biomedical and Molecular Sciences, University of Surrey

Guildford, Surrey, GU2 7XH, England.

E-mail: d.lewis@surrey.ac.uk; Tel.: 44-1483-686-477; Fax: 44-1483-300-803.

Source: Toxicol. in Vitro 2005, 19(4), 517–522.

Compounds:
a) 9 Compounds of type I, where R = CCl3, Ph, t-Bu, CF3, Bn, ClICH, FCH2, CH2, CH3;
b) Styrene oxide (EH substrate).

R-CH=CH2

\[
\text{O}^\circ
\]

Biological material: Human microsomal epoxide hydrolase (EH).

Data taken from the literature:

I% \[\text{% inhibition of EH (details not given)}\];

logP \[\log_{10}(\text{partition coefficient in 1-octanol/water})\];

Crystal structure \[\text{atomic coordinates of several mammalian EHs, including rabbit, pig, rat, mouse, as well as fungal (Aspergillus niger) EH)}\]

Computational methods:

Molecular modeling \[\text{molecular modelling procedures were carried out using the Sybyl software suite as implemented on a Silicon Graphics Indigo2 R10000 workstation, molecular orbital calculations were carried out employing the semiempirical AM1 method, human EH was created via homology modeling based on the crystallographic structures of several mammalian EHs, including rabbit, pig, rat, mouse as well as fungal EH)}\];

Data calculated:

Descriptors \[\text{generation of electronic parameters such as dipole moment (\( \mu \)), orbital energies (\( E_{\text{HOMO}} \), \( E_{\text{LUMO}} \)) and partial atomic charges (e.g. net charge on the epoxide oxygen atom, QO), electrostatic potential minima (V_mn) were calculated using the semiempirical AM1 method);}\n
logP \[\log_{10}(\text{the partition coefficient in 1-octanol/water was estimated using via CLogP software)}\];

Results: Molecular modelling of human microsomal EH by homology with A. niger EH crystal structure has been carried out to 1.8 Å resolution. In the homology model the active site lies in a well-defined, essentially hydrophobic, pocket within the enzyme structure. Two tyrosine residues, that are conserved in all known mammalian EH sequences, form hydrogen bonds with the epoxide oxygen atom on the known EH substrate, styrene oxide. A small hydrophobic cleft is present in the active site region, where the phenyl group of styrene oxide can bind, but this cleft appears to be of limited size such that the bulky side-chains of an incoming epoxide molecule may obstruct binding. The inhibitory activity of epoxide hydrolase substrates results from an association between a relatively low QO value for hydrogen bonding to the active site tyrosines, and a fairly high lipophilicity in the form of logP which is related to a combination of molecular size of the substituent and its overall hydrophobic character. This feature set optimizes affinity for the essentially hydrophobic active site region lined by complementary amino acid residues such as leucine, methionine and tryptophan. Consequently, logP/QO provided a relevant descriptor for the explanation of potency differences with the set of epoxides, as shown by Eqs. 1–4.

\[
\log I% = 0.119 (\pm 0.008) \log P/QO + 0.925 \quad (1)
\]

\[ n = 6 \quad T = 0.9906 \quad s = 0.0440 \quad F = 209.8 \]

\[
I% = 9.670 (\pm 1.507) \log P/QO - 8.006 \quad (2)
\]

\[ n = 9 \quad r = 0.9245 \quad s = 13.556 \quad F = 41.2 \]

\[
\log I% = 1.083 (\pm 0.156) \log P + 0.823 (\pm 0.315) \mu - 1.789 \quad (3)
\]

\[ n = 6 \quad r = 0.9427 \quad s = 0.3217 \quad F = 24.0 \]
\[
I% = 25.819 \pm 7.072 \log P - 12.350 \pm 4.660 E_L + 22.932 \quad (4)
\]

The generated QSAR models may be used to predict EH substrate and/or inhibitor potencies of untested compounds.

(B.B.)

125/2005

**Title:** QSAR and classification models of a novel series of COX-2 selective inhibitors. 1,5-diarylimidazoles based on support vector machines.

**Authors:** Liu, H. X.; Zhang*, R. S.; Yao, X. J.; Liu, M. C.; Hu, Z. D.; Fan, B. T.

Department of Chemistry, Lanzhou University
Lanzhou 730000, China.
E-mail: ruison@public.lz.gs.cn; Tel.: 86-931-891-2578; Fax: 86-931-891-2582.

**Source:** J. Comput.-Aided Mol. Design 2004, 18(6), 389–399.

**Compounds:** 15 Compounds of type I, 14 compounds of type II, 33 compounds of type III, 12 compounds of type IV, 11 compounds of type V, where R¹ and R² are diverse substituents.

**Biological material:** Human cell lines expressing cyclooxygenase 2 (COX-2).

**Data taken from the literature:**

| IC₅₀ | [concentration of the test substance (µM) required for 50% inhibition of COX-2]. |

**Results:** QSAR and classification models have been developed for a novel set of COX-2 selective inhibitors of types I–V employing SVM methodology. Each compound was described by CODESSA descriptors. The heuristic method was used to search the descriptor space and select the descriptors relevant to inhibitory activity. The approach yielded a seven-descriptor model based on SVMs with root mean square errors of 0.107 and 0.136µM for training and prediction sets, respectively.

\[
pIC_{50} = 0.2785 \text{RPCS} - 818.8700 \text{HACA-2/T} + 0.1239 \text{HACA-1};
\]

\[
+ 10.3010 \text{ASIC0} + 0.7853 \text{NBR} - 0.0075 \text{PNSA-1};
\]

\[
+ 1.9507 \text{ABON} + 0.6826 \quad (1)
\]

n = 68 r = 0.845 s = 0.306 F = 16.06

**Computational methods:**

- **Molecular modeling**
  - Molecular structures were pre-optimized using the MM + molecular mechanics force field implemented in HyperChem, full optimization was performed using the semiempirical PM3 method implemented in MOPAC employing the Polak-Ribiere algorithm.

- **SVM**
  - Support Vector Machine approach for classification and regression problems based on the “structural risk minimization” principle which defines the trade-off between the approximation quality of a given dataset and the complexity of the approximating function as opposed to the empirical risk minimization concept, that concentrates on the approximation quality for the dataset (theory is given), all calculation programs implementing SVM were written in R-file using Libsvm based on R script, the kNN algorithm was also performed by R software, all scripts were run on a Pentium IV PC with 256 MB RAM.

- **CODESSA**
  - (Comprehensive Descriptors for Structural and Statistical Analysis, a chemical multipurpose QSAR/QSPR statistical analysis and prediction program package for the calculation of constitutional, topological, geometrical, electrostatic, quantum mechanical, and thermodynamic descriptors solely from the structure of compounds and searching for the best multiple linear relationships between the calculated descriptors and experimental property data, CODESSA v2.61 employs the “The Heuristic Method” and “The Best Multilinear Regression Method”);

- **kNN**
  - [k-Nearest Neighbors, a pattern recognition method, where the distance (usually Euclidean) between the pattern vector of an unknown and each of the pattern vector of the training set is first computed, the k nearest samples to the unknown are selected and it is classified in the group to which the majority of the k (k = 3) samples belongs].

**Data calculated:**

- **Descriptors**
  - [selected CODESSA descriptors: RPCS relative positive charged SA (RPCS); HACA-2/TMSA (HACA2/T); HACA-1 (HACA-1); average structural information content (order 0) (ASIC0); number of benzene rings (NBR); PNSA-1 partial negative surface area (PNSA 1); average bond order of a N atom (ABON)];

- **rmse**
  - [root mean square deviation (µM) between the experimental and predicted IC₅₀ values].

Q SAR Comb. Sci. 2005, 24 Abstr. 125 © 2005 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim 667
Fig. 1 shows the plot of the predicted versus experimental pIC$_{50}$ values calculated using the developed regression equation, where squares and circles represent training set and test set compounds, respectively.

The best classification results were derived using SVMs. The accuracy for training and test sets was 91.2% and 88.2%, respectively. The proposed method represents a novel and effective method for drug design and screening.

B. B.

126/2005

Title: Benzene sulfonamide analogs of fluoroquinolones. Antimicrobial activity and QSAR studies.

Authors: Nieto, M. J.; Alovero, F. L.; Manzo, R. H.; Mazzieri*, M. R.

Departamento de Farmacia, Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, Ciudad Universitaria 5000 Córdoba, Argentina.

E-mail: mrmazzie@dqo.fcq.unc.edu.ar; Tel.: 54-351-433-4127.

Source: Eur. J. Med. Chem. 2005, 40(4), 361–369.

Compounds: 10 Benzene sulfonamide analogs of fluoroquinolones (BSFQs) of type I, where R = H, NH$_2$, NHCOCH$_3$, N(CH$_3$)$_2$, CH$_3$, OCH$_3$, Cl, NO$_2$.

Biological material: 2 Bacterium species: Streptococcus aureus ATCC 29213 and Escherichia coli ATCC 25922.

Data taken from the literature:

log$k_w$ [logarithm of the capacity factor, extrapolated to pure water, measured by reversed-phase liquid chromatography (RPLC)];

$R_m$ [measure of lipophilicity expressed as $R_m = \log(1/R_0 - 1)$, where $R_0$ is the retention factor obtained by reversed-phase thin-layer chromatography (RP-TLC)].

The accuracy for training and test sets was 91.2% and 88.2%, respectively. The proposed method represents a novel and effective method for drug design and screening.

Computational methods:

CLOGP (1-octanol/water partition coefficients of the compounds were estimated using the Biobyte Software CLOGP v1.0.0).

Chemical descriptors:

- $\pi$ (Hansch-Fujita’s substituent constant characterizing hydrophobicity);
- $\sigma$ (Hammett’s constant characterizing the electron-donating/withdrawing power of the substituent);
- $F, R$ (Swain-Lupton’s electronic parameters, characterizing the field and resonance effects, respectively);
- $E_s$ (Taft’s constant, characterizing the steric effects of the substituent);
- L, B$_1$, B$_5$ [STERIMOL steric length and width parameters (Å)];
- I [indicator variable 1 for ciprofloxacin derivatives (CIP) and 0 for norfloxacin derivatives (NOR)].

Results: A QSAR study of novel antimicrobial BSFQs obtained by the derivatization of N4-piperazinyl atom of ciprofloxacin (CIP) has been performed. The behavior of the new BSFQ series was similar to the studied previously norfloxacin (NOR) analogs allowing the QSAR analysis of a complete set of BSFQs. Hansch analysis of the activity data showed a linear correlation of the activity with electronic and steric parameters. Small electron-donor groups increased the in vitro activity against Gram-positive bacteria. Hydrophobic properties played a relatively minor role in modulating MIC values, while lipophilic parameters obtained from HPLC turned to be more reliable descriptors. In this study the amino- and the methyleno derivatives were the most active analogs. Single parameter regression equations for pMIC were statistically poor. The best multilinear regression equations were the followings (Eqs. 1–5):

\[
pMIC = -1.36 \pm 0.14 \quad \sigma_g - 1.02 \pm 0.13 \quad B_1 + 0.50 \pm 0.10 \quad I + 7.15 \pm 0.19 \quad (1) \\
n = 16 \quad r = 0.96 \quad s = 0.20 \quad F = 46.5
\]

\[
pMIC = -1.48 \pm 0.32 \quad R - 1.15 \pm 0.12 \quad B_1 + 0.58 \pm 0.17 \quad I + 7.05 \pm 0.33 \quad (2) \\
n = 16 \quad r = 0.87 \quad s = 0.34 \quad F = 12.64
\]

\[
pMIC = -1.48 \pm 0.25 \quad \sigma_g - 1.21 \pm 0.27 \quad B_1 + 0.17 \pm 0.18 \quad \log k_w + 6.26 \pm 0.27 \quad (3) \\
n = 16 \quad r = 0.88 \quad s = 0.33 \quad F = 13.5
\]

\[
pMIC = -1.48 \pm 0.25 \quad \sigma_g - 1.20 \pm 0.27 \quad B_1 + 0.19 \pm 0.22 \quad \text{CLOGP} + 7.62 \pm 0.40 \quad (4) \\
n = 16 \quad r = 0.88 \quad s = 0.34 \quad F = 13.4
\]

\[
pMIC = -0.81 \pm 0.31 \quad \sigma_g - 0.17 \pm 0.18 \quad \pi + 0.52 \pm 0.24 \quad I + 5.69 \pm 0.18 \quad (5) \\
n = 16 \quad r = 0.74 \quad s = 0.47 \quad F = 4.71
\]

Fig. 1 shows the plot of the observed versus predicted, respectively, pMIC values calculated using Eq. 1.
The results confirmed again previous biological findings about the presence of new interactions with target topoisomerases.

(B. B.)

Fig. 1

The descriptors were ranked according to their relevance to the model displaying 22 MOE descriptors at the top of the list. The results produced by the CR model was statistically as good as the PLS model, owing to the cross-validation method employed (RMSEC, RMSEP, RMSD). The nonlinear methods did not performed better than the linear ones. The good statistical quality of the linear PLS and CR render them the method of choice to be used for predictions when it is difficult or impossible to make experimental measurements e.g., for virtual screening, combinatorial library design, and lead optimization.

(B. B.)

Fig. 1
Compositions:

a) Hydroxypropyl-β-cyclodextrin (HP-β-CD);
b) 25 Structurally diverse drugs: alphaxalone, betamethasone, carmofur, cholesterol, citronellol, diazepam, ethisterone, furosemide, hydrocorisone, ibuprofen, indomethacin, ketoprofen, limonene, lorazepam, methotrexate, naproxen, oxaazeamp, phenytoin, piroxicam, prednisolone, progesterone, retinol, spironolactone, testosterone, 17-methyl-testosterone;
c) 16 Structurally diverse drugs (test set): cimetidine, eprisodine, propranolol, thiazolobenzimidazole, acyclovir, acetazolamide, taxol, iraconazole.

Data taken from the literature:

\[ \log(S_0)/C_0 = 0.182 \text{ CMR} - 0.150 \text{ ClogP} - 0.00683 \text{ PSA} - 0.0844 \delta_{\text{tot}} + 3.766 \]

\[ n = 25 \quad r = 0.81 \quad s \text{ not given} \quad F = 19.2 \quad q^2 = 0.711 \quad (1) \]

\[ \log(S_0)/C_0 = -0.0058 \text{ MW} - 0.0122 \text{ MW} - 0.179 \text{ ClogP} - 0.00547 \text{ TPSA} + 1.827 \]

\[ n = 25 \quad r = 0.873 \quad s \text{ not given} \quad F = 0.605 \quad (2) \]

External model validation was carried out employing a test set of six compounds that were representative of the training set used. Out of the six compounds in the testing set, only one compound, namely Zolpidem, had a residual of 1.09 log unit, whereas the remaining five compounds were predicted with a residual of < 1 log unit. Fig. 1 shows the plot of the predicted versus observed \( \log(S_0)/C_0 \) values, where bars represent the standard error of prediction.

These equations can allow formulation scientists to rapidly estimate the potential of HP-β-CD in increasing solubility of poorly water-soluble drugs at the early stage of drug development. The role of the selected descriptors in model description selected could be reasonably rationalized. As for the effect of hydrogen bonding and cohesive forces on the solubility of drug-CD complexed species, it seems very complicated and difficult to understand.

(B. B.)

129/2005

Title: QSAR treatment of the soil sorption coefficients of organic pollutants.

Authors: Kahn, I.; Fara, D.; Karelson, M.; Maran*, Uko; Andersson, P. L.

Institute of Chemical Physics, Department of Chemistry, University of Tartu
2 Jakobi Str., Tartu, EE-51014, Estonia.
E-mail: uko@chem.ut.ee

Source: J. Chem. Inf. Model. 2005, 45(1), 94 – 105.

Compounds: 344 Organic pollutants including 14 chemical classes: acetanilides, alcohols, amides, anilines, carbamates, dinitroanilines, esters, carbamates, dinitroanilines, esters, nitrobenzenes, organic acids, phenols and benzonitriles, phenylureas, phosphates, triazines, and triazoles.

Data taken from the literature:

K_{OC} (soil sorption coefficient defined as the ratio between the concentration of a chemical sorbed by the soil and dissolved in the soil water normalized to soil organic carbon).

Computational methods:

Molecular modeling [molecular structures were drawn and optimized using the Merck Molecular Force Field (MMFF) as implemented in the MacroModel v7.0 program, conformational analysis was carried out employing the Monte Carlo Multiple Minimum (MCMCM) method, energy minimization was performed using the semiempirical AM1 method implemented in MOPAC v6.0];

CODES-SA (Comprehensive Descriptors for Structural and Statistical Analysis, a chemical multipurpose
QSAR/QSPR statistical analysis and prediction program package for the calculation of descriptors solely from the structure of compounds and searching for the best multiple linear relationships between the calculated descriptors and experimental property data, CODESSA employs the “The Heuristic Method” and “The Best Multilinear Regression Method, BMLR”;

PLS (Partial Least Squares projections to latent structures analysis);
PLS (Partial Least Squares projections to latent structures analysis);
PCA (Principal Component Analysis);
LOO (Leave-One-Out cross-validation).

Data calculated:

Descriptors (constitutional, topological, geometrical, electrostatic, quantum mechanical, and thermodynamic descriptors were calculated using CODESSA);
WHIM (Weighted Holistic Invariant Molecular indices to rotations and translations contain information about the 3D structure in terms of size, shape, symmetry, and atom distribution derived from Cartesian coordinates);
\( \eta^\text{max} \) (absolute hardness, AM1);
\( \eta^\text{PNSA-1} \) (partial negative surface area, AM1);
\( q^2 \) (cross-validated correlation coefficient).

Results: General and class-specific QSPR models have been developed for soil sorption (logKoc) of 344 organic pollutants (0 < logKoc < 4.94) using a large diverse set of theoretical molecular descriptors based only on molecular structure. Two general QSAR models were obtained. The two-parameter Model_1 (logP and \( \eta^\text{max} \)) was derived for a structurally representative set of 68 chemicals showing \( R^2 = 0.76 \) and \( s = 0.41 \) parameters. Fig.1 shows the plot of the calculated versus experimental logKoc values by Model_2 for 344 pollutants, where three compounds denoted by lines (near circles) had the largest residuals.

Model_1 was validated using the test set comprising the remaining 276 pollutants (\( R^2 = 0.70, s = 0.45 \)). An additional validation of both models was carried out employing an independent set of 48 pollutants. Both Model_1 and Model_2 predicted the logKoc values at the level of experimental precision. The theoretical molecular descriptors used in the QSPR models yielded further insight into the mechanisms of soil sorption. Analysis of the distribution of the residuals of the logKoc values calculated by both general models indicated the need and possible advantages of modeling soil sorption for smaller data sets of individual classes of organic pollutants. Accordingly, QSPR models were also developed for the 14 individual chemical classes. The descriptors used in these models were examined and related to the possible interaction mechanisms in soil sorption. The major molecular properties relevant to soil sorption were hydrophobicity, size and shape of the compounds and charge distribution. A larger size and bulkier shape favor non-specific interactions with the soil constituents and the humic matrix. The charge distribution describes non-specific polar and specific interactions.

(B. B.)

130/2005

Title: QSPR using MOLGEN-QSPR. The challenge of fluoroalkane boiling points.

Authors: Rücker*, Christoph; Meringer, M.; Kerber, A. Department of Mathematics, Universität Bayreuth POB 101251, Bayreuth D-95440 Germany.
E-mail: christoph.ruecker@uni-bayreuth.de

Source: J. Chem. Inf. Model. 2005, 45(1), 74 – 80.

Compounds: Eighty-three 1 – 4 fluoro alkanes.

Data taken from the literature:
BP [normal boiling point ( °C) of 83 fluoro alkanes].

Computational methods:

MOLGEN (software combining descriptor generation with calculations of many molecular descriptors with data processing employing various multivariate and statistical methods);
GENQSAR Statistical Analysis, a chemical multipurpose QSAR/QSPR statistical analysis and prediction program package for the calculation of constitutional, topological, geometrical, electrostatic, quantum mechanical, and thermodynamic descriptors solely from the structure of compounds and searching for the best multiple linear relationships between the calculated descriptors and experimental property data, CODESSA employs the “The Heuristic Method” and “The Best Multilinear Regression Method”;

MLR (Multivariate Linear Regression analysis);
LOO (Leave-One-Out cross-validation).

Data calculated:

Descriptors [all substructures of one to four bounds in the fluoroalkanes considered: C–C, C–F, C–C–C, F–C–F, C–C–C–F, C–C–C–C, F–C–C–C, F–C–C–C, F–C–C–C–F, C–C–C–C–F, C–C–F(C)–C, C–C–C–F, C–C–F(C)–C, C–C–F(C)–C, C–C–F(C)–C, C–C–C–C–F, F–C–C–F(C)–C, C–F–C–C–F(C)–C]; [further descriptors: bip = n(CHF) + n(CH₂F), bip = bip + n(C–CH₂F–C); xsF = Nᵢ – n(CH₂F, F) – Nᵢ, n(CH₂F, F) = Nᵢ, n(CH₃, F), unless this number is negative, in which case xsF is set to 0; n(CF₃)³; n(CH₃)², n(CF₃, CH₃); (rel(Np))² = (Np/number of all atoms)², \( F_{\text{inv}} \) = \( (Np/(2Nc+2))² \).
Results: MOLGEN-QSPR, has been used for QSAR modeling of the boiling points of lower fluoroalkanes. The derived models are based exclusively on simple descriptors derived from molecular structure. The following multilinear MOLGEN-QSPR models were calculated for the set of fluoroalkanes (Eqs. 1–3).

\[
\begin{align*}
BP &= 83.2226X_u^{0.85} - 25.1841 \gamma^{0.92} - 12.2045 \delta_F - 5.21304 S(\text{ssssC}) - 34.247 n(F-C-F) - 9.2969 \text{molecular volume (cm}^3\text{/mol)} - 41.3515 \text{dipole moment (debye)} - 25.1841 \\nS &= 10 n_{(F-	ext{C-F})} - 1.09815 \text{heat of formation (kcal/mol)} \\
T &= 85.9418 \text{ retention time (min)} \\
R &= 81.9815 \text{ retention time (min)}
\end{align*}
\]

The models described the BPs of nearly twice as many fluoroalkanes to a higher precision than did and previous attempts by the same statistical method and the same number of descriptors. The predictive capability of the models were validated by LOO cross-validation.

131/2005

Title: Theoretical analysis of the retention behavior of alcohols in gas chromatography.

Authors: Song*, Yuanzhi; Zhou, J.; Zì, S.; Xie, J.; Ye, Y. Department of Chemistry, Huaiyin Teachers College, Jiangsu Province Key Laboratory for Chemistry of Low-Dimensional Materials.

Huaian 223300, People’s Republic of China. E-mail: songyuanzhi@sina.com; Tel.: 86-517-3766-342; Fax: 86-517-3766-342.

Source: Bioorg. Med. Chem. 2005, 13(9), 3169 – 3173.

Compounds: 1-Pentanol, 3-methyl-1-pentanol, 2-butanol, 2-pentanol, 1-butanol, 2-methyl-1-propanol, 1-propanol, ethanol, methanol, ethyl-acetate.

Data determined: t [retention time (min)] was measured using SP-2305 gas chromatograph coupled Superox 20M-diglycerol polarity stationary phase).

Computational methods: Molecular structures were drawn in ChemOffice 2004, geometry optimization was performed using the semiempirical AM1 method implemented in MOPAC final geometry was obtained with the Hartree-Fock method at 6-31G(d) level;

MLR (Multivariate Linear Regression analysis).

Data calculated: E_LUMO [energy (hartree) of the lowest unoccupied and E_HOMO highest occupied molecular orbitals, respectively]; Q ( Mulliken negative atom charges with hydrogens summed into heavy in a molecule calculated at the ab initio 6-31G(d) level);

HOF [heat of formation (kcal/mol)]; D [ dipole moment (debye)]; V [ molecular volume (cm^3/mol)]; R [ molecular radius (Å)].

Results: Quantitative structure-retention relationship (QSR) analysis of the gas chromatographic retention times of alcohols have been developed. The QSPR model has been developed using MLR and a set of molecular descriptors of the alcohols calculated using AM1 and ab initio methods. Statistically highly significant liner regression equations were obtained for the retention times.

\[
\begin{align*}
t &= 0.1284 \text{HOF} + 0.4395 \text{V} + 41.0189 \text{Q} + 116.2407 \text{E}_{\text{LUMO}} - 21.8435 (1) \\
n &= 43 r = 0.9941 s = 4.6242 F = 508 \text{ r}_c = 0.9826 s_c = 5.652 \\
t &= 0.1698 \text{HOF} + 0.4624 \text{V} - 3.7628 \text{D} + 146.5715 \text{E}_{\text{LUMO}} - 295734 (2) \\
n &= 10 r = 0.8454 s = 3.6476 F = 6.64 \\
t &= 0.1068 \text{HOF} + 27.0203 \text{R} + 47.2433 \text{Q} + 103.085 \text{E}_{\text{LUMO}} - 85.9418 (3) \\
n &= 10 r = 0.8955 s = 4.0773 F = 5.06 \\
t &= 0.2549 \text{HOF} + 28.0636 \text{R} - 40.9711 \text{D} + 145.6966 \text{E}_{\text{LUMO}} - 100.6382 (4) \\
n &= 10 r = 0.8944 s = 4.0969 F = 5.00
\end{align*}
\]
The retention mechanism of alcohols of separation operating in the gas chromatogram was discussed. The most important molecular descriptors governing retention were Q or V and LUMO. The developed QSAPR models allow the estimation of retention times for similar compounds in cases where retention values are not readily available.

(B.B.)

Classical QSAR: Toxicology and Environmental Sci.

132/2005

Title: Description of the electronic structure of organic chemicals using semiempirical and ab initio methods for development of toxicological QSARs.

Authors: Netzeva, T. I.; Aptula, A. O.; Benfenati, E.; Cronin*, M. T. D.; Gini, G.; Lessigiaris, I.; Maran, U.; Vracko, M.; Schürmann, G.

School of Pharmacy and Chemistry, Liverpool John Moores University

Byrom Street, Liverpool, Merseyside L3 3AF England.

E-mail: m.t.cronin@livjm.ac.uk

Source: J. Chem. Inf. Model. 2005, 45(1), 106–114.

Compounds: 568 Structurally diverse organic compounds.

Biological material: Fish (fathead minnow, Pimephales promelas).

Data taken from the literature:

LC50 [concentration of the compound (mg/L) required to kill 50% of P. promelas in a 96 h test].

Computational methods:

Molecular modeling (initial 3D structures and subsequent geometry optimization was performed employing CORINA 2001 software and Sybyl v6.8 modeling environment);

MLR (Multivariate Linear Regression analysis);

PLS (Partial Least Squares projections to latent structures analysis performed using SIMCA-P v9.0);

LOO (Leave-One-Out cross-validation).

Data calculated:

Descriptors (a total of 162 descriptors were calculated using the AM1 method and 121 descriptors using the B3LYP/6-31G** method; 4 descriptors based on molecular orbital energies, 31 descriptors based on charge distribution, 75 descriptors based on molecular surface area and net atomic charges, 41 descriptors based on molecular orbital wavefunctions and energies, 11 geometrical descriptors);

logP (logarithm of the partition coefficient in 1-octanol/water calculated using PALLAS v3.0);

RMSE (root mean square error of the fit);

r2 (cross-validated correlation coefficient performed using MINITAB v13.1);

q2 (cross-validated correlation coefficient for the PLS model).

Results: The electronic descriptors of the compounds in a large and chemically diverse organic compound set was calculated using semiempirical and ab initio methods for the development of toxicological quantitative structure-activity relationships models employing MLR and PLS analyses. The use of logP in a single parameter regression equation yielded the following regression equation (8 outliers were omitted) (Eq. 1).

\[
\text{pLC}_{50} = -0.700 (\pm 0.022) \log P - 0.720 (\pm 0.058)
\]

\(n = 560\) \(r = -0.806\) \(s = 0.803\) \(F = 1034\) \(r^2_{cv} = 0.647\)

Fig. 1 shows the plot between pLC60 versus logP, where the outliers were denoted by empty circles.

Statistically similar and significant models were obtained using AM1 and B3LYP calculated descriptors and in addition calculated logP values. The quality of the models derived using the two sets of quantum chemical descriptors depended mainly on the type of descriptors employed. It was found that for modeling large data sets \(n = 568\) irrespective of the mechanism of toxic action, the use of precise but computer time demanding ab initio descriptors did not offer considerable advantage over the semiempirical ones. For the best MLR models the following statistical parameters were obtained (model, descriptor calculation method, number of parameters, \(r^2\), \(r^2_{cv}\), s, F): AM1, 2, 0.663, 0.658, 0.805, 555; AM1, 3, 0.676, 0.671, 0.790, 392; AM1, 4, 0.698, 0.692, 0.763, 325; B3LYP, 2, 0.667, 0.663, 0.800, 565; B3LYP, 3, 0.685, 0.679, 0.778, 409; B3LYP, 4, 0.693, 0.686, 0.770, 317. For the best PLS models the following statistical parameters were obtained (descriptor calculation method, n, number of parameters, number of components, \(r^2\), \(q^2\), RMSE): AM1, 562, 13, 2, 0.726, 0.719, 0.721; B3LYP, 562, 13, 3, 0.724, 0.711, 0.723. With B3LYP the charged partial surface area (CPSA) descriptors were generally more significant, while with AM1 the preference was given to size descriptors and delocalizability indices.

133/2005

Title: Use of structure descriptors to discriminate between modes of toxic action of phenols.

Authors: Spycher, S.; Pellegrini, E.; Gasteiger*, J.

Computer-Chemie-Centrum and Institute of Organic Chemistry, University of Erlangen-Nürnberg

Nägelpbachstrasse 25, D-91052 Erlangen, Germany.

E-mail: Gasteiger@chemie.uni-erlangen.de; Tel.: 49-9131-85-26570; Fax: 49-9131-85-26566.

Source: J. Chem. Inf. Model. 2005, 45(1), 200–208.

Compounds: 220 Substituted phenols.

Biological material: Tetrahymena pyriformis.
Data taken from the literature:

**Compounds**

[220 phenols taken from the open NCI database data set with four associated Modes of Toxic Action (MOA): (i) uncouplers of oxidative phosphorylation; (ii) precursors to soft electrophiles (proelectrophiles); (iii) soft electrophiles; (iv) polar narcotic agents].

**Computational methods:**

- **CPG NN** [Counter-PropaGation Neural Networks for classification studies, an extension of the Kohonen Self-Organizing Maps (SOM), where one or more output layers are added the Kohonen input layer];
- **Multinom** (MULTINOMial logistic regression);
- **LR** [Logistic Regression, a type of regression analysis, where the dependent variable \(Y\) is a dummy variable (coded 0, 1), where the Maximum Likelihood Estimation (MLE) method is used in order to derive the regression coefficients];
- **CV** (k-fold cross validation method);
- **PCA** (Principal Component Analysis).

**Data calculated:**

- **Descriptors** (the PETRA program package employs various physicochemical methods for the calculation of atomic physicochemical properties of organic compounds, e.g., total, \(\sigma\) and \(\pi\)-charge distribution, atomic polarizability, lone-pair \(\sigma\)- and \(\pi\) electronegativities, and molecular properties such as hat of formation, molecular polarizability, etc.);

\(T_i\)

(Hostelling score calculated by a method utilized for the determination of the prediction space covered by the developed models based on the Hostelling’s \(T^2\) statistic including PCA).

**Results:** Structure descriptors were employed to discriminate between modes of toxic action of phenols utilizing two classification models based on a data set of 220 phenols with four associated MOAs. CPG NN and multinom were used as classification methods. The combination of topological autocorrelation of empirical \(\pi\)-charge and \(\sigma\)-electronenegativity and of surface autocorrelation of hydrogen-bonding potential yielded a 21-dimensional model that significantly discriminated between the four MOAs. Fig. 1 shows the PCA-score plot of the training set with 21 descriptors, where the ellipse defines the 95% confidence region (full circles, full triangles, x signs, and open triangles denote polar narcotics, uncouplers, proelectrophiles, and soft electrophiles, respectively.

The calculation of such descriptors was very fast which made them ideal for screening of databases. The predictive power of the approach was found to be equal to previously published quantum mechanical based descriptors. The overall predictive power of the approach was estimated to be 92% using 5-fold CV. In the next step, a simple score value for the distance to the training data was used to determine the prediction space of the model and employed in a study of the phenols included in the NCI database. The use of a prediction space metric was proved to be an essential requisite for the screening of this very diverse database. The prediction space covered by the developed model was rather limited due either to the limited diversity and size of the training set or to the high dimensionality of the descriptors employed.

(B.B.)

134/2005

**Title:** QSAR model of the phototoxicity of polycyclic aromatic hydrocarbons.

**Authors:** Ribeiro, F. A. L.; Ferreira*, Marcia, M. C.

Laboratório de Quimiometria Teórica e Aplicada, Instituto de Química, Universidade Estadual de Campinas 13081-970 Campinas, SP Brazil.

E-mail: marcia@iqm.unicamp.br; Tel.: 55-193788-3102; Fax: 55-19-3788-3023.

**Source:** Theochem 2005, 719(1–3), 191–200.

**Compounds:** 67 Polycyclic aromatic hydrocarbons (PAHs) containing 2–7 rings and 5–6 carbon atoms e.g., pentaphene of type I, benzo[b]chrysene of type II and diben[a,j]anthracene of type III.

**Biological material:** Daphnia magna.

**Data taken from the literature:**

- Phototoxicity [phototoxicity of the PAHs (details not given)];
- \(LT_{50}\) [phototoxicity of PAHs for Daphnia magna, expressed as the median adjusted lethal time (ALT) normalized to a constant concentration to eliminate (dimension not given)].

**Computational methods:**

- **Molecular** (the geometry of all molecules was optimized and the frontier orbital energies and the GAP value were obtained from semi-empirical AM1 molecular orbital calculations, the QSAR mod-
el for phototoxicity was constructed with the PLS method on mean-centered data and validated by LOO cross-validation, the calculations were carried out by employing the programs implemented in Pirouette v2.02 and PLS Toolbox for MATLAB; Partial Least Squares projections to latent structures analysis; Cluster Analysis.

**Data calculated:**

\[ E_{\text{LUMO}} \] (energy (eV) of the lowest unoccupied and highest occupied molecular orbitals, respectively, calculated using Spartan software for Unix); 
\[ E_{\text{HOMO}} \] (descriptor calculated as \( E_{\text{HOMO}} - E_{\text{LUMO}} \) (eV)); GAP (descriptor as \( SQR(\text{PRESS}/n) \), where \( n \) is the number of compounds in the training set); LOO (Leave-One-Out cross-validation); PRESS (sum of the squared deviation between the predicted and measured binding affinities for every molecule); q’ (cross-validated correlation coefficient).

**Results:** Upon sunlight exposure, PAHs undergo rapid structural modification generally via oxidation reactions. The modified products are in many cases more toxic than their parent compounds. The toxicity is due to the \( \pi \)-orbital system of PAHs, which strongly absorbs in the UV and visible regions of the solar spectrum. In this study a QSAR study of 67 PAHs has been performed and a prediction rule for the phototoxicity of these compounds was proposed. EHOMO, ELUMO and GAP were used as descriptors. The relationships between these molecular descriptors and the photo-induced toxicity were found to be non-linear, and Gaussian type functions were used to linearize them. Statistically significant PLS models were calculated and the models were validated by predicting the phototoxicity for a set of molecules not used in model development. Pentaphene of type I, benzo[ \( b \) ]chrysene of type II, and dibenz[a,j]anthracene were among the compounds potentially phototoxic as predicted by the model. A new GAP range (7.2 \( \pm \) 0.7 eV) was proposed for the classification of phototoxic compounds, and a larger cutoff was suggested for the normalized lethal time as \( \log(1/\text{ALT}) \leq -2.95 \). An unscaled regression equation was obtained from the PLS model (Eq. 1) by using a routine for unscaling the regression vector obtained from autoscaled data.

\[
\text{Phototoxicity} = 0.4737 E_{\text{HOMO}} - 0.3692 E_{\text{LUMO}} + 1.1081 \text{GAP} - 4.3355
\]
\[ n = 13 r = 0.92 s \text{ not given F not given } q^2 = 0.79 \] (1)

The photo-induced toxicity for 53 PAHs was estimated from the QSAR study. Based on the results, a new scale for toxic compounds was proposed and the predicted values of phototoxicity allowed the classification of these molecules into toxic or non-toxic.

(B. B.)

**Title:** Quantitative structure-activity relationships of nitroaromatics toxicity to the algae (Scenedesmus obliquus).

**Authors:** Yan, X.-F.; Xiao*, H.-M.; Gong, X.-D.; Ju, X.-H.

Department of Chemistry, Nanjing University of Science and Technology Nanjing 210094, PR China.

E-mail: xiao@mail.njust.edu.cn; Tel.: 86-25-8431-5947; Fax: 86-25-8443-1622.

**Source:** Chemosphere 2005, 59(4), 467 – 471.

**Compounds:** 25 Nitroaromatic compounds: nitrobenzene, 4-nitrotoluene, 2-nitroaniline, 3-nitroaniline, 4-nitroaniline, 2-nitroisocoumarin, 3-nitroanisole, 4-nitroanisole, 2,4-nitrophenol, 3-nitrophenol, 4-nitrophenol, 2,4-dinitrotoluene, 2,6-dinitrotoluene, 1,2-dinitrobenzene, 1,3-dinitrobenzene, 1,4-dinitrobenzene, 2,4-dinitroaniline, 2,4-dinitrophenol, 2-CN, 3-CN, 4-CN, 3-CNBr, 4-CNBr, 2,5-DCNB, 3,4-DCNB.

**Biological material:** Algae (Scenedesmus obliquus).

**Data taken from the literature:**

\[ EC_{50} \] (concentration of the test substance (mol/L) required for causing 50% acute toxic effect against S. obliquus); \[ \log P \] (logarithm of the partition coefficient in 1-octanol/water).

**Computational methods:** Molecular [the compounds were fully optimized using the B3LYP method and 6-311G** basis set, vibrational analysis indicated that all structures are stable and correspond to the minimum point on the potential energy surface, quantum chemical calculations were carried out by the Gaussian98 program running on PC]; MLR (Multivariate Linear Regression analysis was performed employing the SPSS statistical program package).

**Data calculated:**

\[ E_{\text{LUMO}} \] (energy (eV) of the lowest unoccupied and highest occupied molecular orbitals, respectively); \( Q_{\text{NO}_2} \) (charge of the nitro group).

**Results:** Quantitative structure-activity relationships of the toxicity of a set of 25 nitroaromatics toxicity against algae (S. obliquus) has been performed using MLR. The descriptors \( \log P, E_{\text{LUMO}} \), and \( Q_{\text{NO}_2} \) were used as independent variables to develop the QSAR models. For 18 mononitro derivatives, the hydrophobicity parameter \( \log P \) explained the toxic activity successfully (Eq. 1).

\[
pEC_{50} = 0.668 \log P + 2.310 \]
\[ n = 18 r = 0.9044 s = 0.158 F = 71.9 \] (1)

The use of \( E_{\text{LUMO}} \) and \( Q_{\text{NO}_2} \) yielded Eq. 2 for all the compounds (Eq. 2).

\[
pEC_{50} = 4.634 Q_{\text{NO}_2} - 26.468 E_{\text{LUMO}} + 2.924 \]
\[ n = 25 r = 0.85 s = 0.291 F = 28.5 \] (2)

Omitting three outliers gave the statistically highly significant Eq. 3.

\[
pEC_{50} = 6.481 Q_{\text{NO}_2} - 25.053 E_{\text{LUMO}} + 3.746 \]
\[ n = 22 r = 0.926 s = 0.206 F = 56.9 \] (3)

Eqs. 1 and 3 indicate that there are two different mechanisms for the mono-and dinitro aromatic compounds. The toxicity of mononitro aromatics to the algae was found to be proportional to their hydrophobicity. However, for dinitro aromatic compounds, due to the electrophilic nature of the parent compounds, there is probably a mechanism of stepwise reduction of \(- Q_{\text{NO}_2} \). It was suggested that Eq. 3 can be used
to predict the toxicity of nitro aromatic compounds against algae studied here.
(B. B.)

136/2005

Title: Toward an optimal procedure for PC-ANN model building. Prediction of the carcinogenic activity of a large set of drugs.

Authors: Hemmateenejad*, B.; Safarpour, M. A.; Miri, R.; Nesari, N.

Medicinal & Natural Products, Chemistry Research Center, Shiraz University of Medical Sciences
Shiraz, Iran.
E-mail: hemmatb@sums.ac.ir; Tel.: 98-711-3872; Fax: 98-711-233-2225.

Source: J. Chem. Inf. Model. 2005, 45(1), 190–199.

Compounds: 735 Drug molecules.

Data taken from the literature:
DFcarc [carcinogenic activity expressed in the DF scale of 735 compounds retrieved from the Merck index based on the annual report of carcinogenesis (details not given)].

Computational methods:
Molecular modeling (molecular structures were built using HyperChem v7, no geometry optimization was performed); PCA (Principal Component Analysis); GA (Genetic Algorithm, a stochastic optimization method that mimics the process of evolution by manipulating a collection of data structures called chromosomes); ANN (feed-forward, back propagation Artificial Neural Networks); Combined procedures [eigenvalue ranking, correlation ranking, and genetic algorithm (ER-PC-ANN, CR-PC-ANN, PC-GA-ANN, respectively)]; CV (Leave-One-Out cross-validation).

Data calculated:
Descriptors (a total number of 1350 theoretical descriptors are calculated for each molecule using DRAGON software: 47 constitutional descriptors, 255 topological indices, 21 molecular walk counts, 64 Burden eigenvalues, 21 charge topological indices, 218 3D and geometrical descriptors, 99 WHIM descriptors, 150 radial distribution function descriptors, 241 functional group and atom-centered groups, 6 empirical and chemical descriptors);

\[ r^2_{cv} \] (cross-validated correlation coefficient);
\[ \text{RMSE}_{cv} \] (root mean square error of the cross-validated model).

Results: The performances of the three novel QSAR algorithms, ER-PC-ANN, CR-PC-ANN, and PC-GA-ANN were compared by application of these models to the prediction of the carcinogenic activity of a diverse set of 735 drugs, each described by a total number of 1350 theoretical descriptors. The data matrix (735 x 1350 dimension) was subjected to PCA. PCA explained 95% of the variances in the matrix by the first 137 principal components (PC’s). ER, CR, and GA were employed to select the best PC’s for PC-ANN modeling. In the ER-PC-ANN approach, the PC’s were stepwise entered into the ANN based on their decreasing eigenvalue. In the CR-PC-ANN, the ANN was first employed to model the nonlinear relationship between each one of the PC’s and the carcinogen activities separately. In the next step, the PC’s were ranked based on their decreasing correlating ability and entered to the input layer of the ANN one after another. Finally, GA was used to find the best set of PC’s. Both external and cross-validation methods were used to validate the predictive performances of the derived models (model, \[ r^2_{cv} \], \[ \text{RMSE}_{cv} \]: ER-PC-ANN, 0.714, 1.89; CR-PC-ANN, 0.911, 0.95; PC-GA-ANN, 0.922, 0.88). Fig. 1 shows the plot of calculated versus experimental activity for the test set compounds using the PC-GA-ANN model.

It was found that the PC-GA-ANN and CR-PC-ANN procedures outperformed the EV-PC-ANN procedure. The results revealed that the results yielded by the PC-GA-ANN algorithm were better than those produced by CR-PC-ANN. However, the difference was not significant.
(B. B.)

Classical QSAR: ADME

137/2005

Title: A new approach for the tissue-blood partition coefficients of neutral and ionized compounds.

Author: Zhang, H. B.

Chemistry Department, Beijing Normal University 100875 Beijing, Peoples R China
E-mail: hbzhang@bnu.edu.cn; Tel.: 86-10-5880-5194; Fax: 86-10-5880-0567.

Source: J. Chem. Inf. Model. 2005, 45(1), 121 – 127.

Compounds: 265 Structurally diverse compounds.

Biological material: Kidney, brain, muscle, lung, liver, heart, and fat.

Data taken from the literature:
PC (tissue-blood partition coefficient of 36 organic chemicals for human fat, kidney, muscle, lung, and heart, and of 10 compounds for rabbit fat, kidney, muscle, lung, and heart).

Computational methods:
MLR (Multivariate Nonlinear Regression analysis was performed using the standard regression analysis program GFA BASIC v4.38);
LOO (Leave-One-Out cross-validation).

Data calculated:
Descriptors (molecular descriptors were calculated using the HYBOT/HYBOT-PLUS-98 program pack-
age: molecular polarizability, $\alpha$; maximum positive charge, $q^+_{\text{max}}$; sum of all positive partial atomic charges for all acceptor substructures in the molecule, $\Sigma_{\text{Ca}}$; sum of the H-bond factor values for all donor atoms in a molecule, $\Sigma_{\text{Cd}}$; maximum H-bond acceptor descriptor in a molecule, $C_{\text{am}}$; $f_{\text{ui}}$ [fraction of unionized and ionized basic compounds is calculated using Eq. 1.]

$$f_{\text{ui}} = 1/(1 + 10^{pK_a - 7.4}), f_{\text{iu}} = 1 - f_{\text{ui}} \quad (1)$$

PRESS (sum of the squared deviation between the predicted and measured binding affinities for every molecule).

**Results:** Recently the author has developed a nonlinear model for the tissue-blood partition coefficients of neutral compounds. In this study a new approach is presented for the tissue-blood partition coefficients of ionized compounds partitioning into kidney, brain, muscle, lung, liver, heart, and fat. In this paper a nonlinear model equation based on tissue composition (a content of lipids, proteins, and water) for the tissue-blood partition coefficients of compounds was further developed to account for the neutral and ionized forms of the compounds (the complex equations of the nonlinear model are given). Using the developed model a nonlinear regression analysis for neutral and ionized compounds partitioning into kidney, brain, muscle, lung, liver, heart, and fat yielded equations with high predictive power for the training set ($n = 201$, $r = 0.905$, $s = 0.291$, $q = 0.890$) and test set compounds ($n = 64$, $r = 0.906$, $s = 0.247$). Fig. 1 shows the plot of the calculated versus experimental PC$_t$ values for 201 data points in the training set.

Fig. 2 shows the plot of the calculated versus experimental PC$_t$ values for 64 data points in the test set.

The results demonstrated that the equilibrium distribution of a compound in a several tissues is essentially the equilibrium distribution of the compound in tissue (chemical) compositions. It was also shown that neutral and ionic forms of a compound as well as in different tissue (chemical) compositions have a different mechanism of action in vivo. The nonlinear model equation may be considered as the expressive form of the Hansch equation in nonlinear spaces (or multi-phase system).

(B. B.)

---

**Title:** CoMFA and docking studies on glycogen phosphorylase inhibitors as antidiabetic agents.

**Authors:** Prathipati, P.; Pandey, G.; Saxena*, A. K.

**Medicinal Chemistry Division, Central Drug Research Institute Chatter Manzil Palace, Lucknow-226 001, India.**

E-mail: anilsak@hotmail.com; Tel.: 91-0522-221-24-11/4268; Fax: 91-0522-222-3405.

**Source:** J. Chem. Inf. Model. 2005, 45(1), 136–145.

**Compounds:**

a) 13 Compounds of type I, where $R^1 = H, F, Cl, Br$; $R^2 = \text{Ph, Cy}$; $R^3 = \text{CONMe}_2, \text{CONHMe, CONH}_2, \text{COOMe, COOH, CH}_2\text{OH}$;

b) 12 Compounds of type II, where $R^1 = H, \text{F, Cl, Br, OMe}$; $R^2 = H, \text{F;} R^3 = \text{CONMe}_2, \text{CONHMe, COOMe, COOH, CO(1-piperidin-4-ol)}$.

**Biological material:** Glycogen phosphorylase (GP$_a$).

**Data taken from the literature:**

Crystal [atomic coordinates of GP$_a$ were taken from structure the Brookhaven Protein Data Bank (pdb code: 1lwo)].

**Data determined:**

IC$_{50}$ [concentration of the test substance (dimension not given) required for 50% inhibition of GP$_a$].

**Computational methods:**

LUDI (program implemented in InsightII to determine possible binding geometries for a ligand that interacts with hydrogen bonding and hydrophobic sites of the receptor using statistical data from small-molecule crystal structures);

ComFA [Comparative Molecular Field Analysis of the molecules was carried out represented by their steric and electrostatic fields sampled at the intersections of a three-dimensional lattice (2 Å grid increment) using an sp$^3$ carbon atom probe with a charge of +1, H-bonding fields, indicator fields, and parabolic fields were also used, and all regression analyses were done using PLS algorithms in SYBYL v6.9];

---

3D QSAR

**138/2005**

**Title:** CoMFA and docking studies on glycogen phosphorylase inhibitors as antidiabetic agents.

**Authors:** Prathipati, P.; Pandey, G.; Saxena*, A. K.

**Medicinal Chemistry Division, Central Drug Research Institute Chatter Manzil Palace, Lucknow-226 001, India.**

E-mail: anilsak@hotmail.com; Tel.: 91-0522-221-24-11/4268; Fax: 91-0522-222-3405.

**Source:** J. Chem. Inf. Model. 2005, 45(1), 136–145.

**Compounds:**

a) 13 Compounds of type I, where $R^1 = H, F, Cl, Br$; $R^2 = \text{Ph, Cy}$; $R^3 = \text{CONMe}_2, \text{CONHMe, CONH}_2, \text{COOMe, COOH, CH}_2\text{OH}$;

b) 12 Compounds of type II, where $R^1 = H, \text{F, Cl, Br, OMe}$; $R^2 = H, \text{F;} R^3 = \text{CONMe}_2, \text{CONHMe, COOMe, COOH, CO(1-piperidin-4-ol)}$.

**Biological material:** Glycogen phosphorylase (GP$_a$).

**Data taken from the literature:**

Crystal [atomic coordinates of GP$_a$ were taken from structure the Brookhaven Protein Data Bank (pdb code: 1lwo)].

**Data determined:**

IC$_{50}$ [concentration of the test substance (dimension not given) required for 50% inhibition of GP$_a$].

**Computational methods:**

LUDI (program implemented in InsightII to determine possible binding geometries for a ligand that interacts with hydrogen bonding and hydrophobic sites of the receptor using statistical data from small-molecule crystal structures);

ComFA [Comparative Molecular Field Analysis of the molecules was carried out represented by their steric and electrostatic fields sampled at the intersections of a three-dimensional lattice (2 Å grid increment) using an sp$^3$ carbon atom probe with a charge of +1, H-bonding fields, indicator fields, and parabolic fields were also used, and all regression analyses were done using PLS algorithms in SYBYL v6.9];
CoMSIA | Comparative Molecular Similarity Indices
Analysis of the molecules was carried out as an alternative approach to CoMFA based on similarity indices calculated at the intersections of a three dimensional lattice, the five physicochemical properties for CoMSIA (steric, electrostatic, hydro-phobic, and hydrogen bond donor and acceptor) were evaluated using a common probe atom with 1 Å radius, ±1.0 charge, and hydrophobicity and hydrogen bond property values of +1.0, the value of an attenuation factor α was 0.3 for the Gaussian-type distance dependence:

LOO (Leave-One-Out cross-validation);
MLR (Multivariate Linear Regression analysis).

Data calculated:
CFH (the common features generated by the Common Feature Hypothesis approach: A hydrogen-bond acceptor; R, aromatic ring; H, hydrophobic);
Cont (steric contact scores);
Lipo Score (Ludi score function value);
No.rot.-bonds (number of rotatable bonds);
Sum (sum of all scores);
SPRESS (standard deviation of cross-validated predictions);
\( r^{2}_\text{pred} \) (predictive correlation coefficient);
\( q^{2} \) (cross-validated correlation coefficient).

Chemical descriptors:
No.HB (number of hydrogen bonds).

Results: Glycogen phosphorylase (GP) is an attractive target for the design of inhibitors that may prevent glyco- genolysis at high glucose levels in type II diabetes. The carboxamides represent one of the major classes of GP inhibitors other than glucose derivatives. In this study CoMFA methodology and docking of ligands into GP, was performed in order to elucidate the essential structural and physicochemical requirements responsible for binding to the GP enzyme and to develop predictive models for designing indole-2-carboxamide derivatives. The Common Feature Hypothesis approach generated 10 hypotheses containing the RHDA, RHDD, HHDA, or HHDD features. All 25 molecules belonging to the training and test sets mapped every feature of all hypotheses. Fig. 1 shows the structure of the template molecule of type II with \( R_{1} = \text{Cl}, R_{2} = \text{F} \), and \( R_{2} = \text{CO(1-piperidin-4-ol)} \), onto which all the other 24 molecules were superimposed using the RMS fitting procedure.

Rigid body docking of the inhibitors into GP, using Ludi yield the following regression models (Eqs. 1,2).

\[
\begin{align*}
pIC_{50} &= 0.085(\pm 0.063) \text{No.HB} - 0.005(\pm 0.002) \text{Lipo Score} - 0.035(\pm 0.012) \text{Cont} + 2.662(\pm 1.848) \text{CONSTANT} \\
n = 25 \quad r = 0.563 \quad s = 0.564 \quad F = 3.247 \quad (P = 0.03)
\end{align*}
\]

\[
\begin{align*}
pIC_{50} &= 0.133(\pm 0.049) \text{No.HB} - 0.007(\pm 0.002) \text{Lipo Score} - 0.043(\pm 0.009) \text{Cont} + 4.244(\pm 1.504) \text{CONSTANT} \\
n = 23 \quad r = 0.760 \quad s = 0.427 \quad F = 8.554 \quad (P = 0.001)
\end{align*}
\]

A statistically significant CoMFA model has been developed using pharmacophoric alignments and hydrogen-bonding fields demonstrated high predictive ability for the training set (\( r^{2} = 0.98, q^{2} = 0.68 \)) and the test set (\( r_{\text{pred}}^{2} = 0.85 \)) compounds. PLS coefficient contour maps from CoMFA have been mapped onto the GP model showing a high level of compatibility. The manual docking and scoring results by LUDI indicated the positive contribution of hydrogen bonding groups and the negative contribution of lipophilic and steric contacts.

(B. B.)

139/2005

Title: A CoMFA study of COX-2 inhibitors with receptor based alignment.

Authors: Datar, P. A.; Coutinho*, Evans C.

Department of Pharmaceutical Chemistry, Bombay College of Pharmacy, Kalina, Santacruz (E) Mumbai 400 098, India.

E-mail: evans-im@eth.net; Tel.: 91-22-2667-0871; Fax: 91-22-2667-0816.

Source: J. Mol. Graphics Modell. 2004, 23(3), 239–251.

Compounds:
a) 57 COX-2 inhibitors comprising the cyclopentene-, spiroleptene-, benzene-, pyrazole- pyrrole-, imidazole-, pyrimidine-, isoxazole- thiadiazole-, thiadiazole-, and oxadiazole class of compounds, e.g., 5 cyclopentene class of compounds of type I, where \( R^{1} = \text{diverse aromatic moieties, } R^{2} = \text{NH}_{2}, \text{CH}_{3} \).
b) SC-558 a selective COX-2 inhibitor.

Biological material: Cyclooxygenase-2 (COX-2).

Data taken from the literature:

\[
\begin{align*}
\text{IC}_{50} & \quad \text{[concentration of the test substance (M) required for 50\% inhibition of COX-2 (details not given)]} \\
\text{Crystal structure} & \quad \text{[atomic coordinates of the X-ray crystal structure of SC-558 complexed with the COX-2 enzyme were taken from the Brookhaven Protein Data Bank (PDB code: 1CX2)].}
\end{align*}
\]

Computational methods:

Molecular modeling [structures were built employing the BUILD-ER module in INSIGHT II v.98, the structures were optimized with the CVFF force field, two different methods were employed for generation of conformation and alignment, in the first method, the X-ray crystal structure of SC-558 was used as a template for superimposition...]
(Model 1), a second mode of alignment was carried out using docking strategies with the DOCK v4.0 program (Model 2), the ligand structures were refined using the BFGS method, the molecules aligned in these manner were subjected to CoMFA analysis; DOCK [program for finding potential docking sites on proteins of known structure by starting with solvent accessible surface, and filling cavities with overlapping spheres to make binding pockets, ligands of known structure (e.g., found by searching a database) are then automatically docked into this site];

AFFINITY (program based on both Monte Carlo and simulated annealing strategies, to generate simultaneously the conformations and alignment needed for the CoMFA study);

ComFA [Comparative Molecular Field Analysis of the molecules was carried out represented by their steric and electrostatic fields sampled at the intersections of a three-dimensional lattice (2 Å grid increment) using an sp3 carbon atom probe with a charge of +1, all regression analyses were done using PLS algorithms in SYBYL v6.7];

PLS (Partial Least Squares projections to latent structures analysis);

LOO (Leave-One-Out cross-validation).

Data calculated:

SDEP (Standard Deviation Error in Prediction);

PRESS (sum of the squared deviation between the predicted and measured binding affinities for every molecule);

$q^2$ (cross-validated correlation coefficient).

Results: A CoMFA study of a diverse training set of 53 COX-2 inhibitors has been performed employing two different alignment methods. The first method of alignment of the molecules was based on the binding information obtained from a crystallographic study, that yielded CoMFA Model 1. The second mode of alignment was generated by docking the inhibitors in the binding pocket using the DOCK v4.0 and AFFINITY suite of programs, that yielded CoMFA Model 2. The Model 2 was slightly better than Model 1 in terms of the statistical parameters $r^2$ and $q^2$ (model, n, $N_{opt}$, $q^2$, PRESS, $r^2$, $s$, F, SDEP, steric, electrostatic molecular fields): Model 1, 33, 6, 0.624, 0.985, 0.971, 0.274, 145.1, 0.455, 0.327, 0.673; Model 2, 34, 6, 0.733, 0.804, 0.989, 0.160, 418.6, 0.509, 0.297, 0.703. Fig. 1 shows the plot of the predicted versus observed pIC$_{50}$ values for the test set calculated with Model 2.

Fig. 2 shows the contour map of the most active molecule in the series superimposed on the topology of the COX-2 enzyme, where in Model 2 the dark and light contours represent regions, where negative electrostatic potential around the ligand enhance and decrease, respectively, activity.

Improvement in the external predictivity of Model 2 suggested that the lowest energy conformation of the ligand does not always satisfactorily describe binding for all molecules. The different binding configurations generated by AFFINITY helped to identify the different arrangements of the ligand in the binding pocket. The docking studies give a meaningful insight into the H-bonding interactions between the inhibitors and residues in the active site of the enzyme, which can be exploited in designing better inhibitors.

(B. B.)

Title: 3D-QSAR and preliminary evaluation of anti-inflammatory activity of series of N-pyrrolylcarboxylic acids.

Authors: Lessigierska, I.; Nankov, A.; Bocheva, A.; Pajeva, I.; Bijev*, A.

University of Chemical Technology and Metallurgy 8 “Kl. Ohridski”, Blvd. 1756, Sofia, Bulgaria E-mail: a.bijev@uctm.edu; Tel.: 359-2-625-4445; Fax: 359-2-685-488.

Source: Farmaco 2005, 60(3), 209–218.

Compounds:

a) 30 Compounds of type I, where $R^1 = H, F, Cl, CH_3, CH_2, CH_3, CF_3, CN, SO_2CH_3, CONH_2, CO_2H, NO_2, OH, OCH_3, OCH_2CH_3, NH_2, N(CH_3)_2, N(CH_3)CH_3, CH_2OH$; $R^2 = CF_3, CHF_2$;

b) SC-558 known COX inhibitor.

Biological material:

a) Male Wistar rats;

b) Cyclooxygenase enzyme (COX-1 and COX-2 subtypes);

c) SC-558 known COX inhibitor.

Data taken from the literature:

IC$_{50}$ [concentration of the test substance (µmol/L) required for 50% inhibition of the COX enzymes in vitro reported by Penning et al.];
Crystal structure (atomic coordinates of the of structure of SC-558 bound to COX-2 were taken from the Brookhaven Protein Data Bank).

Computational methods:

Molecular modeling (modelling calculations were performed using Sybyl v6.6 software, geometry optimizations were performed employing the Tripos force field with the Powell method, the AM1 semi-empirical method implemented in Mopac v6.0 was applied for quantum chemistry calculations, the crystal structure of SC-558 bound to COX-2, obtained from was used as a template structure for building SC-558 derivatives, the following structural features were used for the alignment of the compounds: (i) the centroid of the R substituted phenyl ring, (ii) the carbon atom of the two or three fluoro substituted methyl group and the sulfur atom of phenyl-sulfonamide group);

ComFA [Comparative Molecular Field Analysis of the molecules was carried out represented by their steric and electrostatic fields sampled at the intersections of a three-dimensional lattice (2 Å grid increment) using an sp3 carbon atom probe with a charge of +1, all regression analyses were done using PLS algorithms in SYBYL v6.6];

CoMSIA [Comparative Molecular Similarity Indices Analysis of the molecules was carried out as an alternative approach to CoMFA based on similarity indices calculated at the intersections of a three dimensional lattice, the five physico-chemical properties for CoMSIA (steric, electrostatic, hydro-phobic, and hydrogen bond donor and acceptor) were evaluated using a common probe atom with 1 Å radius, +1.0 charge, and hydrophobicity and hydrogen bond property values of +1, the value of an attenuation factor α was 0.3 for the Gaussian-type distance dependence].

Data calculated:

pKa (negative logarithm of the acidic dissociation constant were calculated with ACD/ChemSketch software);

SEP (Standard Error of Prediction);

q² (cross-validated correlation coefficient).

Results: The study aimed at the development of new potential inhibitors of COX-2. 3D-QSARs of compounds of type I were investigated using CoMFA and CoMSIA methodologies. Statistically significant CoMFA and CoMSIA models were calculated using SC-558 as template structure (model, n, N_mol, molecular field, q², SEP): CoMFA, 14, 3, electrostatic, 0.761, 0.284; CoMFA, 14, 5, steric, 0.667, 0.374; CoMSIA, 14, 4, electrostatic, 0.828, 0.253; CoMSIA, 14, 5, steric, 0.625, 0.397; CoMSIA, 15, 3, electrostatic, 0.837, 0.564; CoMSIA, 15, 5, hydrophobic, 0.902, 0.483. Fig. 1 shows the predicted COX-2 activities of the investigated compounds calculated using the developed CoMSIA model (electrostatic field, q² = 0.837) where the compound predicted to be most active (16) is type I with R¹ = 4-CO2H, R² = CHF₂.

The derived new compounds, with high similarity to the template of already recognized selective COX-2 inhibitors, will be an object of forthcoming evaluation of COX-2 selectivity in vitro. (B. B.)
larity indices calculated at the intersections of a three-dimensional lattice, the five physicochemical properties for CoMSIA (steric, electrostatic, hydrophobic, and hydrogen bond donor and acceptor) were evaluated using a common probe atom with 1 Å radius, +1.0 charge, and hydrophobicity and hydrogen bond property values of +1, the value of an attenuation factor \( \alpha \) was 0.3 for the Gaussian-type distance dependence, the X-ray crystal structures of three inhibitors bound to pcDHFR were used for defining the alignment rule, scaled MNDO ESP-fit partial charges were calculated with MOPAC v6.0 using atomic coordinates obtained by energy minimizing the aligned molecules with the MMFF94S force field and MAXIMIN2 routine in Sybyl, the regression analyses were done using PLS algorithms in SYBYL v6.81; PLS (Partial Least Squares projections to latent structures analysis); CV [Leave-One-Out (LOO), Leave-Ten-Out (LTO) cross-validation, and Leave-Several-Out (LSO)].

**Data calculated:**

- rmsd [root mean square deviation (Å) of the position of the corresponding atoms of two superimposed molecular structures];
- MAE (Mean Absolute Error);
- \( s_{\text{PRESS}} \) (standard deviation of cross-validated predictions);
- \( q^2 \) (cross-validated correlation coefficient).

**Results:** 3D-QSAR CoMSIA modelling has been applied to a set of 406 structurally diverse pcDHFR and rIdHFR inhibitors. A QSAR model containing 6 components was developed for pcDHFR employing LTO cross-validation (\( n = 240, q^2 = 0.65 \)), and a 4-component model was calculated for rIdHFR (\( n = 237, q^2 = 0.63 \)), both including steric, electrostatic and hydrophobic contributions (DHFR, \( r^2 \), s, F, \( q^2 \), \( s_{\text{PRESS}} \), \( N_{\text{map}} \), MAE, steric, electrostatic, hydrophobic fields); pcDHFR, 0.80, 0.52, 157.5, 0.65, 0.69, 0.63, 0.18, 0.43, 0.39; rIdHFR, 0.63, 186.9, 0.63, 0.77, 0.4, 0.75, 0.19, 0.37, 0.19, 0.37, 0.44. Fig. 1 shows the plot of the measured versus predicted pIC\(_50\) values for pcDHFR, where training and test set compounds are shown using filled and empty circles, respectively;

CoMSIA contour maps of the contributions for the significant molecular fields were used to identify important ligand-receptor interactions in 3D. Classification models were also developed predicting selectivity for pcDHFR over rIdHFR using SIMCA methodology, with a selectivity ratio of 2 (IC\(_{50}\) rIdHFR/IC\(_{50}\) pcDHFR) for delimiting classes. A 5-component model including steric and electrostatic molecule field contributions displayed cross-validated and test set classification rates of 0.67 and 0.68 for selective inhibitors, and 0.85 and 0.72 for unselective inhibitors. It was concluded that the predictive power of the CoMSIA and SIMCA classification models, together with the structural insights derived from them, might aid in the design of novel inhibitors used against P. carinii infections in immuno-compromised states.

(B. B.)

---

**Molecular Graphics**

**Title:** Identification of a new scaffold for opioid receptor antagonism based on the 2-amino-1,1-dimethyl-7-hydroxytralin pharmacophore.

**Authors:** Grundt, P.; Williams, I. A.; Lewis, J. W.; Husbands*, S. M.

**Department of Pharmacy and Pharmacology, University of Bath**

Bath, BA2 7AY, England.

E-mail: s.m.husbands@bath.ac.uk; Tel.: 44-1225-383103; Fax: 44-1225-386114.

**Source:** J. Med. Chem. 2004, 47(21), 5069–5075.

**Compounds:**

- a) 9 Compounds of type I, where \( R^1 = \text{Me}, \text{n-Pr}, \text{cyclopropylmethyl (CPM)}, \text{allyl}, \text{cinnamyl}; R^2 = \text{H}, \text{Me}, \text{n-Pr}, \text{allyl}, \text{cinnamyl, CPM}; \)
- b) trans-(3,4)-dimethyl-4-(3-hydroxyphenyl)piperidine of type II;
- c) Naltrexone, natrindole, norBNI;
- d) \([\text{H}]\text{DAMGO (μ)}, [\text{H}]\text{Cl-DPDPE (δ)}, [\text{H}]\text{U69,593 (κ)}.\)

**Biological material:** 3 Opioid receptor subtypes: μ, MOR; δ, DOR; κ, KOR.

**Data determined:**

\( K_i \) [Michaelis inhibition constant (nM) representing the affinity of the substrate to displace \([\text{H}]\text{DAMGO}, [\text{H}]\text{Cl-DPDPE}, \) or \([\text{H}]\text{U69,593}\) from the μ, δ, or κ opioid receptor, respectively].
**Computational methods:**

**Molecular** (structures of type I and IIa were drawn using the Builder option in MOE v2004.030 and minimized using the MMFF94x force field, the Flexible Alignment function was used for the overlay, with the Restraints command ensuring that the phenolic rings remained aligned).

**Results:** The trans-(3,4)-dimethyl-4-(3-hydroxyphenyl)piperidines of type II, a class of opioid antagonists, recently provided selective antagonists for three subtypes opioid receptors (MOR and KOR). Molecular modeling studies have been performed indicating strong structural similarity between the parent of this series of compounds and 2-amino-1,1-dimethyl-7-hydroxytetralin of type I. It has been established that type I represents a new scaffold for opioid receptor antagonism.

In binding and in vitro functional assays, the aminotetralin derivatives showed some overlap in structure-activity relationships with that previously reported for the phenylpiperidine series, providing evidence for a common binding mode for the two series at these type opioid receptors. Fig. 1 shows the overlay of type 4a with type II with R1 = cinnamyl, R2 = H (IIa).

Introduction of a methoxy group in the 3-position of the skeleton increased potency at MOR and KOR receptors, suggesting that this aminotetralin derivative offers an alternative scaffold for the design of further receptor selective opioid ligands. The ligands designed and synthesized displayed comparable opioid binding affinity to their analogues of type II. (B. B.)

**Title:** Molecular-modeling based design, synthesis, and activity of substituted piperidines as γ-secretase inhibitors.

**Authors:** Gundersen*, E.; Fan, K.; Haas, K.; Huryn, D.; Jacobsen, J. S.; Kreft, A.; Martone, R.; Mayer, S.; Sonnenberg-Reines, J; Sun, S.-C.; Zhou, H.

Chemical and Screening Sciences, Wyeth Research CN 8000, Princeton, NJ 08543-8000, USA.
E-mail: gundere@wyeth.com; Tel.: 1-484-865-9202; Fax: 1-484-865-6463.

Source: Bioorg. Med. Chem. 2005, 15(7), 1891 – 1894.

**Compounds:**
a) Lead compound of type I, hit compound of type II;
b) 9 Compounds of type III, where R1 = H, Cl, OCH3; R2 = H, CH3, OCH3.

**Biological material:** γ-Secretase, responsible for β-amyloid protein (Aβ40 and Aβ42) deposit formation and the development of Alzheimer’s disease.

**Data determined:**

EC50 [effective concentration of the test substance (µM) required for 50% inhibition of γ-secretase measured in vitro in an ELISA assay].
Three of the designed analogs of type III showed moderate γ-secretase inhibitory activity (EC₅₀ = 3.5 – 20.9 μM; EC₅₀ of type I ≈ 0.1 μM). The data suggested that the compounds acted via γ-secretase inhibition. The measured γ-secretase inhibitory activity within this small library validated the usefulness of the ROCS search design strategy.

**Title:** Inhibitory effects of 2-substituted-1-naphthol derivatives on cyclooxygenase I and II.

**Authors:** Kongkathip, B.; Sangma, C.; Kirtikara, K.; Luangkamin, S.; Hasitapan, K.; Jongkon, N.; Hannongbua, S.; Kongkathip⁎, N.

Natural Products and Organic Synthesis Research Unit (NPOS), Department of Chemistry, Faculty of Science Kasetsart University, Bangkok 10900, Thailand.

E-mail: fscinpk@ku.ac.th; Tel.: 66-2-9428-900x211; Fax: 66-5793-955.

**Source:** Bioorg. Med. Chem. 2005, 13(6), 2167–2175.

**Compounds:**
- a) 13 α-naphthol compounds of type I carrying diverse substituents in the β position;
- b) 4 Known COX-inhibitors: flurbiprofen, naproxen, aspirin, SC-558 of type II.

**Biological material:**
- a) Two isoforms of the cyclooxygenase enzyme (COX-1 and COX-2);
- b) Immortalized mouse PGHS-1 and PGHS-2 cells;
- c) Vero cells (ATCC CCL-81).

**Data taken from the literature:**
- Crystal structure [atomic coordinates of COX-2 complexed with SC-558 and COX-1 bound to flurbiprofen were taken from the Brookhaven Protein Data Bank (pdb codes: 1eqh and 1cx2, respectively)].

**Data determined:**
- IC₅₀ [concentration of the test substance (μM) required for 50% inhibition of the production of PGHS-1 and PGHS-2 using immortalized mouse PGHS-1 and PGHS-2 cells];
- IC₅₀ [concentration of the test substance (μg/mL) required for eliciting 50% cytotoxic effect measured using colorimetric method].

**Computational methods:**
- Molecular modeling [initial structures of eight naphthol compounds modeling of type I were generated by molecular model-
The results provided a model for the binding of the naphthal derivatives to COX-2 which should be very useful to design more potent COX-2 and selective COX inhibitors. (B. B.)

**Multivariate Analyses**

145/2005

**Title:** Volsurf analysis of carbapenem antibiotics.

**Authors:** Doddareddy, M. R.; Cha, J. H.; Cho, Y. S.; Koh, H. Y.; Yoo, K. H.; Kim, D. J.; Pae*, A. N.

Life Sciences Division, Korea Institute of Science and Technology
P. O. Box 131, Cheongryang, Seoul 130-650, South Korea.
E-mail: anpae@kist.re.kr; Tel.: 82-2–958-5185; Fax: 82-2958-5189.

**Source:** Bioorg. Med. Chem. 2005, 13(10), 3339–3349.

**Compounds:** 85 Carbapenem derivatives of type I, where R is structurally divers substituents.

**Biological material:** 2 Bacterial species: Staphylococcus aureus SG 511 (Gram negative), Escherichia coli 078 (Gram positive).

**Data taken from the literature:** MIC [minimum inhibitory concentration (dimension not given) of the test substance against S. aureus and E. coli].

**Chemical descriptors:** (72 descriptors were calculated using Volsurf v3.0 employing the OH2, O, and DRY probes).

**Results:** Volsurf analysis was applied to a set of 70 carbapenem compounds acting as antibacterials using S. aureus and E. coli representing Gram positive and Gram negative bacteria, respectively. PCA of the data matrix of 70 carbapenem analogs and 72 Volsurf descriptors yielded 5 PCs (PC, explained variance, cumulated value): PC1, 45.62, 45.62; PC2, 15.09, 60.72; PC3, 9.62, 70.33; PC4, 7.44, 77.79; PC5, 6.64, 84.41, respectively. The PC score plots showed clustering of compounds according to the activity. The PC loading plots explained the Volsurf descriptors responsible for the separation and behaviour of the compounds. The PC loading plots explained the Volsurf descriptors responsible for the separation and behaviour of the compounds. All the compounds of the test set were predicted fairly well showing residual values less than one log unit. Fig. 1 shows the PLS coefficient plot for the correlation of Volsurf descriptors for S. aureus.

The MIC activity data of S. aureus (Gram positive) was better explained than E. coli (Gram negative) by the PLS models. Fig. 2 shows the plot of experimental versus calculated activities of S. aureus obtained by PLS analysis, where the triangles show the predictions of the training set and stars show the data for the test set compounds.

It was concluded that the Volsurf approach is highly efficient in predicting the biological activities and pharmacokinetic behaviour of these carbapenem antibiotics. (B. B.)
Title: GRIND-derived pharmacophore model for a series of α-tropanyl derivative ligands of the sigma-2 receptor.

Authors: Cratteri, P.; Romanelli, M. N.; Cruciani, G.; Bonaccini, C.; Melani, F.

Department of Pharmaceutical Sciences, University of Florence
via U. Schiff 6, I-50019 Sesto Fiorentino, Firenze, Italy.
E-mail: paola.gratteri@unifi.it; Fax: 39-055-457-3671.

Source: J. Comput.-Aided Mol. Design 2004, 18(5), 361–374.

Compounds: 48 Compounds of type I, where R1, R2 are diverse substituents, R3 = Me, Bn; X = O–, –S–, –NH–, –N(Me)+, or missing; Y = –C(=O)O–, –CH2O–, –C(=O)NH–.

Biological material: Sigma-2 receptor present in the CNS as well as in various peripheral tissues, and are involved in several physiological effects.

Data taken from the literature: Ki [Michaelis inhibition constant (nM) representing the affinity of the substrate to the α1 and α2 receptor].

Computational methods:
Molecular [all molecular modeling calculations were performed on an Silicon Graphics O2 R10000 workstation. 3D structures of the ligands were generated using fragment libraries and/or the builder module of the InsightII 2000 package, energy minimizations were carried out using the conjugate gradient method with AMBER force field parameterized in vacuum and the Discover module of InsightII 2000, compounds were generated in the non-protonated form, the conformational search was performed using a simulated annealing procedure, followed by cluster analysis yielding subsets of conformers for a specific molecule based on a defined rmsd value, combination of FILO and GA techniques provided sets of chromosomes (one for each molecule of the dataset) for which the R2 value was maximized, PCA and the PLS algorithms were used as implemented in the Almond program (47)];

GRID (program for determining energetically favorable binding sites on a molecule by calculating the electrostatic, hydrogen bond and Lennard-Jones interactions of chemically selective probes with the selected target at each node of an interaction grid based on an empirical energy function);

GA (Genetic Algorithm);

PCA (Principal Component Analysis);

PLS (Partial Least Squares projections to latent structures analysis);

LOO (Leave-One-Out cross-validation).

Data calculated:

GRIND (GRid INDependent descriptors, a novel class of alignment independent 3D descriptors allowing a detailed understanding of the internal geometrical relationships of receptor regions with which the ligands establish non-covalent bonds without requiring alignment of the ligands);

S {sigma-1/sigma-2 selectivity calculated as log[Ki (s2/Ki s1)]};

rmsd [root mean square deviation (Å) of the position of the corresponding atoms of two superimposed molecular structures].

Results: A GRIND-derived pharmacophore model has been developed for a set of α-tropanyl derivative ligands of type I of the sigma-2 receptor using GRIND descriptors. Statistically significant PLS models for were calculated for sigma-2 affinity. Sigma-2 model: r2 = 0.83, q2 = 0.63) and S (sigma-1/ sigma-2 selectivity): r2 = 0.72, q2 = 0.46 were derived using a training set of α-tropanyl derivatives. The models provide pictures of the virtual receptor site (VRS) providing a qualitative pharmacophoric representation of the sigma receptor. The analysis performed using the GRIND descriptors has confirmed the presence of two hydrophobic areas and a H-bond donor moiety in the binding site of the sigma receptor, interacting with the lipophilic groups and the electron-rich center of the molecules. They modeled the internal geometrical relationships within two hydrophobic areas (hydrophobic-1 and -2) and a H-bond donor receptor region with which ligands establish non-covalent bonds. Fig. 1 shows the proposed geometrical relationship and maps of the main interaction areas for the α-receptor.

The obtained PLS model predicts the sigma-2 activity of the α-tropanyl derivatives involved in the study, while results from the selectivity analysis highlight the distance within the two hydrophobic areas as the major sensitive element for sigma selectivity.

(B. B.)
Title: Enrichment of ligands for the serotonin receptor using the Shape Signatures approach.

Authors: Nagarajan, K.; Zauhar, R.; Welsh*, W. J.

University of Medicine and Dentistry of New Jersey, Robert Wood Johnson Medical School (UMDNJ-RWJMS) 675 Hoes Lane, Piscataway, NJ 08854, USA.

E-mail: welshwj@umdnj.edu

Source: J. Chem. Inf. Model. 2005, 45(1), 49-57.

Compounds: Over 10,000 compounds.

Biological material: Serotonin receptors.

Data taken from the literature:

- Crystal structure: [atomic coordinates of 14-3-3-ζ complexed by serotonin N-acetyltransferase were taken from the Brookhaven Protein Data Bank (pdb code: 1IB1)];
- Compounds: [825 agonists and 400 antagonists together with approximately 10,000 randomly chosen compounds retrieved from the National Cancer Institute (NCI) database];
- Databases: [NCI Diversity Set (ca 230,000 compounds in total), ACD-3D (ca 280,000 in total), MDDR 3D (ca 120,000 compounds in total) were used for virtual screening purposes];
- Homology model: (a 3D homology model of SARS 3D-like proteinase based on the structure of transmissible gastroenteritis-virus coronavirus 3C-like proteinase was built using the MODELLER 6.0 program).

Computational methods:

- Molecular: (ligand conformations were generated using CORINA, docking studies were performed using GOLD and DOCK);
- GOLD: [flexible protein-ligand docking program featuring a (i) genetic algorithm methodology for protein docking; (ii) full ligand and partial protein flexibility; (iii) energy functions partly based on conformation and non-bonded contact information from the Cambridge Structural Database (CSD)];
- DOCK: [program for finding potential docking sites on proteins of known structure by starting with solvent accessible surface, and filling cavities with overlapping spheres to make binding pockets, ligands of known structure (e.g., found by searching a database) are then automatically docked into this site].

Data calculated:

- Shape of Signature: (the shape signature is a histogram representation the ray segment lengths obtained from the ray tracing within the SAV of a molecule, a shape signature which contains only the length of information is termed a 1D signature, the 2D-MEP signatures encoding both segment length and electrostatic potential information associated with the point of incidence within the SAV);
- SAV: [triangulated solvent accessible volume of a molecule generated by the SMART algorithm];
- Tc: [Tanimoto coefficient is defined as N(AB)/[N(A)+N(B)−N(AB)], where N(AB) is the number of bits set in common by A and B, N(A) is the total number set by A, and N(B) is the total number set B];

Results: Enrichment of ligands has been performed for the serotonin receptor using the Shape Signatures approach, a new 3-dimensional molecular comparison method adapted here to rank ligands of the serotonin receptors. The approach was exemplified using a variety of test databases including the mixture of agonists and antagonists together with approximately 10,000 randomly chosen compounds from the NCI database. Both 1D and 2D Shape Signature databases were compiled for the enrichment studies, and key parameters for searching and matching the molecules were determined. It was found that the 1D Shape Signature approach is highly efficient in separating agonists from a mixture of molecules which includes compounds randomly selected from the NCI database taken as inactives. The method was also equally effective at separating agonists and antagonists from a pool of active ligands for the serotonin receptor. The influence of conformational variation of the shape signature on enrichment was studied by docking a subset of ligands into the crystal structure of serotonin N-acetyltransferase (code: 1IB1). Enrichment studies using the resulting “docked” conformations yielded only slightly improved results compared with the CORINA-generated conformations. Fig. 1 shows the comparison between CORINA generated and docked conformations, where black circles and gray squares denote 2D and 1D scores.

Fig. 1

Parallel enrichment studies were carried out using 2D shape signatures showing high selectivity with more restricted coverage due to the high specificity of 2D signatures.

(B. B.)

Programs, algorithms

Title: Evaluation and application of multiple scoring functions for a virtual screening experiment.

Authors: Xing*, Li; Hodgkin, E.; Liu, Q.; Sedlock, D.

Tripos, Inc.

1099 S. Hanley Road, St. Louis, MO 63144, USA.

E-mail: li.xing@pfizer.com; Fax: 1-636-247-7607.

Source: J. Comput.-Aided Mol. Design 2004, 18(5), 333–344.

Compounds: a) Dx-9065a (Factor Xa ligand);
b) 549 confirmed inhibitors of Factor Xa mixed with 26,000 compounds from Lead Quest.

**Biological material:** Factor Xa.

**Data taken from the literature:**
- Crystal structure: Atomic coordinates of Dx-9065a complexed by Factor Xa were taken from the Brookhaven Protein Data Bank (pdb code: 1FAX);
- Compounds: (26,000 compounds from Lead Quest, a general screening library from Tripos).

**Computational methods:**
- FlexX (program implemented in Sybyl v6.8 for automatic, flexible ligand-protein docking based on incremental construction without manual intervention);
- Scoring: [5 scoring functions, FlexX, DOCK, GOLD, ChemScore and PMF, available from the CScore module in Sybyl v6.8, were evaluated for their abilities to reproduce crystallographic binding mode of ligands to Factor Xa].

**Data calculated:**
- $r_s$ (Spearman’s rank correlation coefficient): $r_s = 1 - \frac{6 \cdot \sum d_i^2}{n(n^2 - 1)}$, where $d_i$ is the difference between two ranks at the point $i$ and $n$ is the total number of points;
- rmsd (root mean square deviation): $\text{Å}$ of the position of the corresponding atoms of two superimposed molecular structures.

**Results:** In order to identify novel chemical classes of factor Xa inhibitors, five scoring functions were employed to evaluate the docking poses generated by FlexX. The compound collection was composed of 549 confirmed potent factor Xa inhibitors and a subset of the LeadQuest screening compound library. Four scoring functions but PMF successfully reproduced the crystal complex (PDB code: 1FAX). This was unexpected since PMF was parametrized on crystal complexes. Fig. 1 shows the representative poses generated by FlexX in reference to co-crystallized DX-9565a.

After docking and scoring by different methods FlexX exhibited the highest hit rate enrichment in the entire screening process, followed by D-SCORE and ChemScore. Hit rate enrichments by G-SCORE and PMF were comparatively moderate. The hit rate of 80% was achieved by FlexX at an energy cutoff of $-40 \text{ kJ/mol}$, which is about 40-fold over random screening (2.06%). The study suggested that presenting more poses of a single molecule to the scoring functions could deteriorate their enrichment factors. A series of potential factor Xa inhibitors was identified from LeadQuest with a potential capability of replacing the benzamidine moiety, yielding compounds with improved pharmacokinetic properties. Several promising scaffolds with favorable binding scores were identified from LeadQuest. Consensus scoring by pair-wise intersection failed to enrich the hit rate yielded by single scorings (i.e. FlexX). It was cautioned that reported successes of consensus scoring in hit rate enrichment could be artificial because their comparisons were based on a selected subset of single scoring and a reduced subset of double or triple scoring. The findings obtained in this study were based upon a single biological system.

(B.B.)