Theoretical Papers

60/2006

Title: Extraction and visualization of potential pharmacophore points using support vector machines. Application to ligand-based virtual screening for COX-2 inhibitors.

Authors: Franke, L.; Byvatov, E.; Werz, O.; Steinhalber, D.; Schneider, P.; Schneider*, Gisbert.

Institut für Organische Chemie und Chemische Biologie and Institut für Pharmazeutische Chemie, Johann Wolfgang Goethe-Universität

Marie-Curie-Strasse 9, D-60439 Frankfurt, Germany.

E-mail: gisbert.schneider@modlab.de; Tel: 49-69-798-29821; Fax: 49-69-798-29826.

Source: J. Med. Chem. 2005, 48(22), 6997 – 7004.

Compounds:

- a) 172 Structurally diverse compounds
- b) Celecoxib, diclofenac of type I, rofenoxib, SC-558, NAP.

Biological material:

- a) Cyclooxygenase 2 (COX-2);
- b) Thrombin.

Data taken from the literature:

Crystal structures

- atomic coordinates of ligand/COX-2 complexes were taken from the Protein Data Bank (pdb codes are given);

Compound sets

- (two subsets of the COBRA v2.11 collection of pharmacologically active reference compounds used for SVM training: 188 thrombin and 94 COX-2 inhibitors, these subsets were used as a reference for ranking circa 2.7 million substances that are commercially available from different vendors).

Data determined:

IC_{50} [concentration of the test substance (μM) required for 50% inhibition of COX-2 measured using ELISA technique an celecoxib, diclofenac, and rofenoxib as positive controls].

Computational methods:

Molecular modeling

- each compound was represented by a 3PP fingerprint using the fingerprint generator in the software suite MOE v2004.05, an individual 3PP feature was a triangle, all possible triangles with their vertexes located at the atom centers of a molecule were considered, presence or absence of a certain triangle defined the “on” (i.e., bit is set) or “off” state of the corresponding bit in the fingerprint, the vertex could be either donor (D), acceptor (A), polar (P), donor and planar (D), acceptor and planar (A), hydrophobic (H), and hydrophobic and planar (H) as defined by the rule-based atom-typer implemented in MOE which follows the PAT-TY atom-type definition, SVM was used to constructs a surface in the n-dimensional space that separated active from inactive compounds, where 3PPs were employed to describe a molecule];

SVM [Support Vector Machine, a novel learning machine, based on the idea of mapping the data into a higher-dimensional feature space via a non-linear mapping and then to perform linear regression in this space, where C is a constant parameter depending on the noise of present in the data, and γ is a constant parameter of the kernel controlling the amplitude of the Gaussian function, the SVM constructs a surface in the n dimensional space that separates active from inactive compounds].

Results: SVMs were trained to predict COX-2 and thrombin inhibitors. Trained classifiers were obtained using sets of known COX-2 and thrombin inhibitors as “positive examples” and a large collection of screening compounds as “negative examples”. The ligands were encoded by topological pharmacophore-point triangles. In a retrospective virtual screening, 50 – 90% of the known active inhibitors were listed within the first 0.1% portion of the ranked database. Validation of the constructed classifiers was performed by developing a method for feature extraction and visualization using SVM, where potential pharmacophore points were weighted according to their importance for COX-2 and thrombin inhibition. Known thrombin and COX-2 pharmacophore points were correctly recognized by the machine learning system. The extracted potential pharmacophore patterns coincided with known binding models of thrombin and COX-2 inhibitors. Fig. 1 shows the potential pharmacophore points reflecting their relative contribution (weight) to the SVM classifier.

In a virtual screening procedure, the number of potential COX-2 inhibitors were predicted and tested. A benzimidazole derivative exhibited significant inhibitory activity with an IC_{50} of 0.2μM, which outperformed Celecoxib in the assay. The study demonstrated that the SVM machine-learning method can be used in virtual screening and the results can be interpreted yielding a set of rules for designing novel molecules. The proposed method complements the suite of modeling techniques that have been employed for designing selective COX-2 inhibitors previously. The SVM method offers particular advantages over other machine-learning approaches: (i) the SVM class/nonclass boundary is constructed as the maximum margin classifier, i.e., it does not represent an arbitrary solution; (ii) it relies only on the so-called “support-vectors”; i.e., those molecules that define the classifier function.

(B. B.)

Title: POEM. Parameter optimization using ensemble methods. Application to target specific scoring functions.

Authors: Antes*, I.; Merkwirth, C.; Lengauer, T.
RMSD (root mean square error) of the position of the models. Selecting the best sets for the parameters is often a difficult task, depending on the complexity of the model. In this paper a novel parameter optimization method, POEM, is proposed for this task, which employs ensemble methods.

The novel method was applied to the optimization of target specific scoring functions in molecular docking. As test applications the FlexX and ScreenScore scoring functions were fitted to the kinase and ATPase protein classes. Starting from random parameters promising results were obtained. Fig. 1 shows the relative mean square errors for five model types and four data sets (data set and scoring function, number of data points, number of complexes, linear, quadratic, k-NN, ANN, ANN ensemble): ATPasc, 256, 16, 0.75, 0.36, 0.35, 0.31, 0.23; ATPflexx, 256, 16, 0.85, 0.46 0.54 0.56 0.33; Kin.scc, 366, 15, 0.75 0.46 0.57 0.41 0.29; Kin.flexx, 366, 15, 0.71 0.41 0.39 0.39 0.23; average, –, –, 0.77 0.42 0.46 0.27. Fig. 1 shows the correlation plots for five model types for the ATP.scc data set. The data set consists of randomly generated parameter settings within the boundary region. The plot shows the 10-fold cross-validated outputs of the model versus the actual value of the loss function. Model types used for this comparison are linear regression (ridge regression) (A), quadratic polynomial regression (B), k-nearest-neighbors regression (C), single neural networks (D), and ensembles of neural networks (E).

Fig. 1

The parameter sets generated showed superior performance within the scoring functions functional limits compared to the original values. The POEM procedure converged rapidly, in the present case within 200–300 cycles, and led to robust minima of the loss function/parameter landscape, thus rendering the approach appropriate for optimizations of rugged landscapes avoiding of getting stuck in local minima. (B. B.)

Title: LigScore. A novel scoring function for predicting binding affinities.

Authors: Krammer*, A.; Kirchhoff, P. D.; Jiang, X.; Venkatachalam, C. M.; Waldman, M.

Accelrys Inc.

10188 Telesis Court, Suite 100, San Diego, CA 92121, USA.

E-mail: akrammer@accelrys.com

Source: J. Mol. Graphics Modell. 2005, 23(5), 395 – 407.

Compounds: Ca 300 protein-ligand complexes were taken from the Brookhaven Protein Data Bank.

Data set [atomic coordinates of ca 300 protein-ligand complexes].

Biological material: Ca 300 proteins.

Results: The computational docking, binding, and folding is often described by empirical models that are fitted to the physical properties of the process by adjustable parameters. Proper choice of these parameters is critical for the quality of the models. Selecting the best sets for the parameters is often
Protein Data Bank (pdb codes are given) with $K_i$ binding constant values; 122 complexes with
<2.5 Å resolution were retained.  

**Computational methods:**

Molecular modeling [missing residues and side chains of the protein-ligand complexes that were not part of the active site were added and crystallographic waters were retained, hydrogen atoms were added using either InsightII or Cerius2, in order to remove any unfavorable steric clashes and to optimize hydrogen-bonding patterns, the hydrogen atoms were subsequently minimized utilizing either the CFF force field parameter set or the DREIDING force field parameters in conjunction with Gasteiger charges, two coordinate sets of protein-ligand complexes were obtained (one for each force field) that were then used for training the LigScore1 scoring function, two more coordinate sets for each protein-ligand complex were derived, against which the LigScore2 scoring function was trained, the LigScore scoring functions have been developed by employing the genetic function approximation (theory is given), the van der Waals energy of protein-ligand interactions were computed via a Lennard-Jones 9 – 6 potential using van der Waals radii and energy parameters of either the CFF or the DREIDING force field];

(Genetic Algorithm)

**Data calculated:**

Descriptors [descriptors in Eq. 1: $C_{pos_{tot}}$ is the total surface area of the ligand involved in attractive polar interactions with the protein, and $TotPol^2$ is equal to $C_{pos_{tot}} + C_{neg_{tot}}$, where $C_{neg_{tot}}$ is the total surface area of the ligand involved in repulsive polar interactions with the protein, and $\sum$ indicates the summation of all interactions between the ligand and protein atoms; descriptors in Eq. 2: $\text{SolvPlty}_{\text{lig}} = (Bury_{\text{tot}_{\text{lig}} – Bury_{\text{lip}_{\text{lig}}})^2$ $\text{SolvPlty}_{\text{prot}} = (C_{pos_{tot}} + C_{neg_{tot}} + Bury_{\text{lip}_{\text{lig}} – Clip_{\text{lig}}})^2$];

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

**Results:** Two novel empirical scoring functions, LigScore1 and LigScore2, has been developed to accurately predict the binding affinity between ligand molecules and their protein receptors. The LigScore functions consist of three distinct terms that describe the van der Waals interaction, the polar attraction between the ligand and protein, and the desolvation penalty attributed to the binding of the polar ligand atoms to the protein and vice versa. The functional form of LigScore1 is simple and contains only three descriptors (Eq. 1).

$$pK_i = - \beta_1 \cdot E_{vdW} + \beta_2 \cdot C_{pos_{tot}} - \beta_3 \cdot TotPol^2 + C$$  

Utilizing a regression approach on a data set of 118 protein-ligand complexes we have obtained a linear equation, LigScore2, using these three descriptors (Eq. 2).

$$\text{LigScore} = 0.60 \cdot (\text{COSMOfrag})^2 + 2.5$$

It was concluded that the new empirical scoring functions shows high predictive accuracy of ligand-receptor-binding affinities over a wide range of protein classes as well as pKi values, while consisting of only three physicochemical descriptors that can be readily related to fundamental molecular interactions.

(B. B.)

63/2006

**Title:** COSMOfrag. A novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry.

**Authors:** Hornig*, M.; Klamt, A.

COSMOlogic GmbH and Co.

 KG, Burscheider Str. 515, D-51381 Leverkusen, Germany.

E-mail: hornig@cosmolologic.de; Tel.: 49-2171-731-683.

**Source:** J. Chem. Inf. Model. 2005, 45(5), 1169 – 1177.

**Compounds:** 3040 Compounds.

**Data taken from the literature:**

Datasets (5 compound sets for COSMOfrag calculations: 107 pesticides, water solubility; 50 pesticides, soil sorption; 147 BOSS water solubility; 2570 PHYSPRO, logP; 170 Abraham, intestinal absorption).

**Computational methods:**

COSMO-RS [Continuum Solvation Model for Real Solvents] method applied for the accurate prediction of thermodynamic, environmental, or physiologi- cal properties, the molecular information is gathered in the so-called ε profiles (ε/Å²), the
COSMO-RS calculations are restricted to small-to-medium sized compound sets, because of the CPU time demand of the calculations (theory is given).

**COSMO-frag**

[... a faster version of the COSMO-RS algorithm, because it replaces the costly quantum chemistry step with a selection of suitable fragments from a database of 40,000 DFT/COSMO pre-calculated molecules, the COSMOfrag approach replaces the single profile with a composition of partial profiles, selected by the use of extensive similarity searching algorithms (theory is given).]

**Data calculated:**

**Properties** (the following properties were predicted using the COSMOfrag methodology: water solubility, soil sorption, logP, intestinal absorption).

**Chemical descriptors:**

- logP (logarithm of the partition coefficient in 1-octanol/water).

**Results:** A novel tool for high-throughput ADME property prediction and similarity screening based on quantum chemistry, COSMOfrag, is presented. The COSMO-RS method is capable of the accurate prediction of thermodynamic, environmental, or physiological properties. The modified COSMOfrag method utilizes a selection of suitable fragments from a database of, presently, 40,000 DFT/COSMO pre-calculated molecules and replaces the single profile with a composition of partial profiles, selected by the use of extensive similarity searching algorithms. The enhanced performance of COSMOfrag has been demonstrated using five representative datasets. Performance statistics were the following [dataset, n, property, CPU time (s), average CPU time per compound (s)]: pesticides, 107, water solubility, 70, 0.63; pesticides, 53, soil sorption, 40, 0.75; BOSS, 150, water solubility, 150, 1.0; PHYSPROP, 2570, logP, 1750, 0.59; Abraham, 170, intestinal absorption, 90, 0.53. The loss of accuracy by using COSMOfrag versus full COSMO-RS calculations has been found to be small (in the range of 0.05 log units). The increased speed of COSMOfrag makes viable COSMO-RS property calculations because it replaces the costly quantum chemistry approach with a selection of suitable fragments.

**PLS** (Partial Least Squares projections to latent structures analysis) was performed using the Unscrambler statistical software.

**Data calculated:**

**MEP** [Molecular Electrostatic Potential (kcal/mol) was calculated based on a classical point charge model, MEP for each molecule was obtained by moving a unit positive point charge across the van der Waals surface, and it was calculated at various points on this surface];

**Statistical parameters** [n (number of molecules); Nopt (number of significant PLS components); SEC, RMSEC (standard error of calibration, and root mean square error of calibration, respectively); SEPbs, RMSEPb (standard error of prediction after bootstrapping, and root mean square error of prediction after bootstrapping, respectively); bias (systemic difference between predicted and observed values); rpred (predictive correlation coefficient)].

**Results:** An approach of combining MEP surface properties (autocorrelation vectors) with the conventional PLS analysis has been employed for the prediction of the human A₃ receptor antagonist activities using 358 structurally diverse human A₃ receptor antagonists representing the largest mo-

---

**Title:** Autocorrelation of molecular electrostatic potential surface properties combined with partial least squares analysis as new strategy for the prediction of the activity of human A₃ adenosine receptor antagonists

**Authors:** Moro*, S.; Bacilieri, M.; Cacciari, B.; Spalluto, G.

Molecular Modeling Section, Dipartimento di Scienze Farmaceutiche, Università di Padova
via Marzolo 5, I-35131 Padova, Italy.
E-mail: stefano.moro@unipd.it; Tel.: 39-049-827-5704; Fax: 39-049-827-5366.

**Source:** J. Med. Chem. 2005, 48(18), 5698 – 5704.

**Compounds:** 358 Compounds

**Biological material:** Human adenosine A₃ receptor.

**Data taken from the literature:**

Dataset (a molecular library of 358 compounds including all 21 important chemical classes of human A₃ antagonists currently discovered).

---

**Computational methods:**

**Molecular modeling** [the 3D models of all 358 A₃ antagonists were obtained by using the 3D structure generator Corina, implemented in the ADRIANA QSAR program suite, conformer generation and best conformer selection have been carried out using standard parameters of Corina, the autocorrelation MEP/PLS model was derived using the autocorrelation vectors as molecular descriptors presented as an intrinsic descriptor of the distribution of an atomic property along the molecular graph, each component of the autocorrelation vector is calculated using the formula B(d) = ∑ p_i p_j, where B is the autocorrelation coefficient referring to atom pairs i, j, p is the atomic property, and d is the 1–j topological distance, thus a new 3D descriptor has been introduced that was based on the autocorrelation of properties at distinct points on the molecular surface];

**PLS (Partial Least Squares projections to latent structures analysis) was performed using the Unscrambler statistical software.**
molecular collection used to generate a general human A2 antago-

nist structure-activity relationship model. A robust quantitative

tive model has been obtained as described by the following statistical parameters (n, N_qfit, r, r_cv, r_no, slope, offset, SEC, RMSEC, SEP, RMSEP, bias): 358, 6, 0.82, 0.81, 0.68, –0.61, 0.77, 0.76, 0.79, 0.79, 5.37 e–8. It was suggested that the proposed MEP/PLS approach can be considered as an alternative hit identification tool in virtual screening applications.

(B. B.)

65/2006

Title: G-protein-coupled receptor affinity prediction based on the use of a profiling dataset. QSAR design, synthesis, and experimental validation.

Authors: Rolland*, C.; Gozalbes, R.; Nicolae, E.; Pau, M.-F.; Coussy, L.; Barbosa, F.; Horvath, D.; Revah, F.

Cerep, 128 rue Danton F-92500 Rueil-Malmaison, France.
E-mail: C.Rolland@cerep.fr; Tel.: 33-1-55-94-84-55; Fax: 33-1-55-94-84-10.
Source: J. Med. Chem. 2005, 48(21), 6563–6574.
Compounds: 1939 Compounds.

Biological material: 40 Different G-protein-coupled receptors (GPCRs), trans-membrane proteins that play a critical role in signal transduction.

Data taken from the literature:
Dataset (1939 diverse compounds with associated IC50 values measured using 40 different GPCRs).

Data determined:
ACTexp [experimental value for the GPCR average activity of the validation set compounds (pIC50, µM)].

Computational methods:
PNB [Predictive Neighborhood Behavior modeling based on the “neighborhood behavior” principle (similar structures→ similar properties), and PNB models tend to extrapolate the property of a novel compound as an average of properties of reference molecules that are shown to be structurally similar, according to a well-defined computed similarity score, in addition to the calculated property, the model also returns, for each compound, two confidence thresholds controlling the relative trust in the PNB model prediction: a “density criterion” expressing how well the current compound is surrounded by relevant neighbors, as a function of their dissimilarity to the candidate compound, and a “homogeneity criterion” measuring the (weighted) variance of the property within the set of selected neighbors];
Synergy model [the calibration of a synergy model consists of finding, with regard to the confidence indices of the PNB prediction, an optimal balance of weights for the linear vs the PNB prediction, such that the returned “synergy” estimations (weighted averages of the two independent linear and PNB predictions, respectively) are as close as possible to the experimental values];

QSAR (the final models are “synergy” approaches, combining the predictions of two conceptually independent parent approaches, based on an equation and respectively on neighborhood behavior, to return a more accurate prediction as a weighted average of the estimates provided by each parent).

Results: GPCR affinity prediction was based on the use of a profiling dataset. A QSAR model accounting for “average” GPCR binding was built from a large set of experimental standardized binding data of 1939 compounds tested over different GPCRs, and applied to the design of a library of “GPCR-predicted” compounds. The model was able to capture critical structural features for GPCR binding. 360 of these compounds were randomly selected and tested in 21 GPCR binding assays. Positive compounds were defined by their ability to inhibit the binding of reference compounds at 10µM by more than 70%. A 5.5-fold enrichment in positives was achieved when comparing the “GPCR-predicted” compounds with 600 randomly selected compounds predicted as “non-GPCR” from a general compound set. The model was capable of predicting the strongest binders, since enrichment was greater for higher cutoffs. Significant enrichment was also obtained for peptidic GPCRs and receptors not used to develop the QSAR model. Fig. 1 shows the experimental (ACTexp) versus predicted (ACTpred) values for the GPCR average activity of the validation set, where full circles mark the position of the synergy model predictions: when both linear and PNB predictions are available, the corresponding marker is spiked. By contrast, hollow circles mark the predicted values for the molecules failing to be predicted by the PNB approach (the case when the synergy predictions equal the linear model-based prediction). For cases, both linear and PNB values exist, the dotted bars span the range between the predicted values of the linear (plus sign) and PNB models (x sign), respectively.

The prediction ability of the model was evaluated by synthesizing and testing a collection of GPCR-predicted compounds on a large series of GPCRs from different families. It was suggested that the model is useful for the design of ligand binding with newly identified GPCRs, including orphan ones. The approach should allow identification of new privileged scaffolds for each of these families and understanding how
specific structural modulations provide pharmacological selectivity and specificity.

(B. B.)

Pharmacology

Title: QSAR analyses of 3-(4-benzylpiperidin-1-yl)-N-phenylpropylamine derivatives as potent CCR5 antagonists.

Authors: Roy*, Kunal; Leonard, J. T.

Drug Theoretics & Cheminformatics Lab, Division of Medicinal & Pharmaceutical Chemistry, Department of Pharmaceutical Technology Jadavpur University

Kolkata 700 032, India.

E-mail: kunalroy_in@yahoo.com; Tel.: 91-33-2867-0786.

Source: J. Chem. Inf. Model. 2005, 45(5), 1352 – 1368.

E-mail: kunalroy_in@yahoo.com; Tel.: 91-33-2867-0786.

Kolkata 700 032, India.

Compounds:
a) 93 Compounds of type I, where R1, R2, and R3 are diverse substituents;
b) [125I]RANTES;

Biological material: Chinese hamster ovary (CHO) cells expressing human CCR5 receptor.

Data taken from the literature:

Ks [Michaelis inhibition constant (mM)] representing the affinity of the substrate to displace [125I]RANTES at the human CCR5 receptor.

Computational methods:

MLR (Multivariate Linear Regression analysis);

PCRA (Principal Component Regression Analysis, using forward selection method);

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

MSA (Molecular Shape Analysis);

MFA (Molecular Field Analysis);

RSA (Receptor Surface Analysis);

GFA (Genetic Function Approximation for the generation of multiple QSAR models by evolving random initial models using a genetic algorithm);

Cross- (leave-one-out, leave-15%-out, and validation leave-25%-out cross-validation, LOO, L15O, L25O, respectively).

Data calculated:

JursFPSA_2 (fractional charged partial surface area);

JursRNCS (relative negative charge surface area);

JursRNCG (relative negative charge);

Shadow_XZfrac (area of the molecular shadow in the XZ plane);

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

SPRESS (standard deviation of cross-validated predictions);

RMSE (Root Mean Square Error);

SDEP (Standard Deviation Error in Prediction);

LOF (Friedman’s Lack Of Fit measure);

q2 (cross-validated correlation coefficient);

R2 pred (predictive correlation coefficient).

Chemical descriptors:

\(\alpha_{\text{Hammett}}, \beta_{\text{Hammett}}\) (Hammett’s constants, characterizing the electron-withdrawing power of the substituent in meta- and para-position, respectively);

\(\pi\) (Hansch-Fujita’s substituent constant characterizing hydrophobicity);

L, B1, B2, [STERIMOL steric parameters (Å)];

MR (molar refractivity);

Chiralcenters (indicator variable 1 for the presence of chiral centers).

Results: CCR5 receptor binding affinity data of a set congeners of type I was subjected to QSAR study using the Hansch analysis. Hansch indicator variables encoding different group contributions and different physicochemical variables such as \(\alpha, \beta, \text{MR and STERIMOL}\) parameters of phenyl ring substituents of the compounds were used as predictor variables for the binding affinity. 3D-QSAR analyses of the same data set using MSA, RSA, and MFA techniques were also performed. The best model with reasonable statistical quality was derived from the MSA (Eq. (1)) method, R2, F value (LOO), PRESS, R2 pred): PCA, 0.697, 17.5, 0.585, 28.5, 0.592; PCRA, 0.951, 146.6, 0.898, 7.0, –; MSA, 0.722, 21.9, 0.650, 22.4, 0.774; RSA, 0.613, 0.582, 19.4, 0.535, 29.2, –1.652; MFA, 0.774, 0.719, –, 0.660, 21.8, 0.327.

\(pK_s = 0.011(\pm 0.006) V_m + 45.828(\pm 19.206) JursRNCG – 0.438(\pm 0.192) JursRNCS + 6.277(\pm 2.833) JursFNSA_1 + 1.243(\pm 0.774) JursFPSA_2 – 1.362(\pm 0.338) Chiralcenters – 3.220(\pm 2.979) Shadow_XZfrac – 8.202(\pm 4.052)\)

\(n = 67 \ell = 0.850 R^2 = 0.722 \text{RMSE} = 0.515 F = 21.9\)

\(q^2 = 0.650 \text{PRESS} = 22.4 \text{SPRESS} = 0.616\)

MSA revealed the importance of the RNCG, i.e., substituents with a high RNCG value have more binding affinity than the unsubstituted piperidine and phenyl (R1 position) congeners. The RNCS was found to be detrimental (e.g., \(R^2 = 3.4\text{-Cl}\)) for the activity. An increase in the length of the molecule in the Z dimension (Lz) is useful (e.g., \(R_3 = 3.4\text{-Cl}\)) for the activity. An increase in the length of the molecule in the Z dimension (Lz) is useful (e.g., \(R_3 = 3.4\text{-Cl}\)) for the activity. An increase in the length of the molecule in the Z dimension (Lz) is useful (e.g., \(R_3 = 3.4\text{-Cl}\)) for the activity.
oped models were also subjected to a randomization test (99% confidence level). The MSA derived models had excellent statistical qualities both for the training as well as test sets and outperformed the RSA and MFA modeling results.

67/2006

Title: Effect of selection of molecular descriptors on the prediction of blood-brain barrier penetrating and nonpenetrating agents by statistical learning methods.

Authors: Li, H.; Yap, C. W.; Ung, C. Y.; Xue, Y.; Cao, Z. W.; Chen*, Y. Z.

Bioinformatics and Drug Design Group, Department of Computational Science, National University of Singapore Blk SOC1, Level 7, 3 Science Drive 2, Singapore 117543, Singapore.

E-mail: cscyyz@nus.edu.sg; Tel.: 65-6874-6877; Fax: 65-6774-6756.

Source: J. Chem. Inf. Model. 2005, 45(5), 1376 – 1384.

Compounds: 415 Compounds.

Biological material: Central nervous system (CNS).

Data taken from the literature: Dataset (415 compounds with known BB ratios); ratio BB of the steady-state concentrations of a drug in the brain and blood taken from Micromedex, American Hospital Formulary Service, and literature sources.

Computational methods:
Molecular modeling (the 3D structure of each of the compounds was generated using ChemDraw and DS ViewerPro v5.0, which were converted into 3D structures using CONCORD and fully optimized using the semiempirical AM1 method);
Statistical learning methods (Bayesian neural network, BNN; principal component analysis, PCA; neural network, NN; linear regression, LR; linear discriminate analysis, LDA; C4.5 decision tree, C4.5 DT, probabilistic neural network, PNN; k-nearest neighbors, k-NN; support vector machine, SVM);

Data calculated:
Descriptors (R\text{mxy} molecular rugosity geometrical; S_{hpl} hydrophilic region geometrical; S_{hpb} hydrophobic

pIC_{50} = -0.117(±0.032) logP^2 + 0.832(±0.22) logP + 1.096(±0.73) BEHv5 − 1.172(±1.91)
\text{n} = 19 \; r = 0.965 \; s = 0.145 \; F = 67.0 \logP_{opt} = 3.56

pIC_{50} = 0.638(±0.16) logP − 0.868(±0.21) log(2β + 1) + 1.048(±0.66) BEHv5 − 1.356(±1.72)
\text{n} = 19 \; r = 0.974 \; s = 0.130 \; F = 63.9 \logP_{opt} = 3.04 \logβ = −2.594

It was suggested that the results of the study provide the basis for the design of new congeners with higher activity.

68/2006

Title: 2-[(Carboxymethyl)sulfanyl]-4-oxo-4-arylbutanoic acids selectively suppressed proliferation of neoplastic human HeLa cells. A SAR/QSAR study.

Authors: Drakulić*, B. J.; Juranić, Z. D.; Stanojković, T. P.; Juranić, I. O.

Institute for Chemistry, Technology and Metallurgy Njegoševa 12, Belgrade, Serbia.

E-mail: bdrakulit@helix.chem.bg.ac.yu; Tel/Fax: 38111 3281867.

Source: J. Med. Chem. 2005, 48(17), 5600 – 5603.

Compounds: 20 Compounds of type I (CSABs), where R = H, 3,4-di-Me, 2,4-di-Me, 2,5-di-Me, 4-Me, 4-Et, 4-i-Pr, 4-n-Bu, 4-n-dodecyl, 2-Cl-4-Me, 4-Cl, 4-Br, 2,3,4-tri-MeO, 2,5-di-i-Pr, 2,4-di-i-Pr, 2,4-di-tert-Bu, 2,4,6-tetra-Pr.

Biological material: Target cells: (i) human cervix carcinoma HeLa cells; (ii) human peripheral mononuclear blood cells (PBMC); and (iii) PBMC + phytohemagglutinin (PHA) cells.

Data determined:
IC_{50} [concentration of the test substance (µmol/L)] required for 50% decrease of the target cell survival.

Computational methods:
BCUT (the BCUT metrics of Pearlman are whole molecule atom nonadjacent atoms, 29 standard 3D H-suppressed BCUT and a description of the nominal bond-type for adjacent and nonadjacent atoms).

Data calculated:
\text{logP} (logarithm of the partition coefficient in 1-octanol/water was calculated employing Crippen's fragmentation method);
BEHv5 (BCUT descriptor defined as the highest eigenvalue n.5 of Burden matrix/weighted by atomic van der Waals volumes calculated using the Dragon v3.0 program).

Results: A set of 20 alkyl-, halo-, and methoxy-aryl-substituted 2-[(carboxymethyl)sulfanyl]-4-oxo-4-arylbutanoic acids were synthesized and biologically evaluated. The new compounds, called CSAB, inhibited proliferation of human cervix carcinoma HeLa cells, in vitro in a concentration range of 0.644 to 29.48 µM/L. Two compounds displayed antiproliferative activity at submicromolar concentrations, and five compounds, called CSAB, inhibited proliferation of human cervix carcinoma HeLa cells, in vitro in a concentration range of 0.644 to 29.48 µM/L.

It was suggested that the results of the study provide the basis for the design of new congeners with higher activity.

(B. B.)
region geometrical; \( C_{\text{p}y} \) capacity factor geometrical; \( H_{\text{p}ya} \) hydrophobic interplay moment; \( H_{\text{p}ya} \) amphiphilic moment; \( \text{dis}1, \text{dis}2, \text{dis}3 \), length vectors; \( S_{\text{unc}} \) sum of solvent accessible surface areas of negatively charged charges; \( \gamma_{\text{CT}} \) valence molecular connectivity \( \gamma \) index for cycle of five atoms; \( \gamma_{\text{C6}} \), valence molecular connectivity \( \gamma \) index for cycle of six atoms; \( \epsilon_f \), hydrogen bond donor acidity (covalent hbda); \( m \), molecular dipole moment; \( \mu_{\text{cp}} \) chemical potential; \( \gamma_{\text{elec}} \) electron negativity index; \( Q_{\text{H}1} \), most positive charge on H atoms; \( Q_{\text{O}1} \), most positive charge on O atoms; \( Q_{\text{H}1} \), most negative charge on H atoms; \( S_{\text{H}1} \), sum of estate indices of halogen atoms; \( S(i) \), atom type E-state indices of diverse atoms.

Results: BBB-nonpenetration is desirable for non-CNS-targeting drugs to minimize potential CNS-related side effects. In this study the effect of selection of molecular descriptors on the prediction of blood-brain barrier-penetrating effects. In this study the effect of selection of molecular descriptors important for distinguishing between BBB-Penetrating agents demonstrating that RFE sub-

Molecular descriptors can improve both the BBB+ and BBB- accuracy of the statistical learning methods tested, the SVM appeared to give a slightly higher prediction accuracy than the other methods for both BBB+ and BBB- agents. It was found that the effect of proper selection of molecular descriptors can improve both the BBB- and the overall accuracies of statistical learning methods. Molecular descriptors were selected by using RFE. The method was tested using 415 BBB+ and BBB- agents demonstrating that RFE substantially improved both the BBB- and the overall accuracy for all of the methods studied. Differences in the values of descriptors important for distinguishing between BBB-Penetrating (BBB+) and -nonpenetrating (BBB-) agents have been modeled by six statistical learning methods yielding outstanding accuracies of 75–92% and 60–80%, respectively. The majority of these predictions gave a substantially lower accuracy, and thus overall accuracy, than the BBB+ accuracy. Of the six statistical learning methods tested, the SVM appeared to give a slightly higher prediction accuracy than the other methods for both BBB+ and BBB- agents. It was suggested that the prediction accuracy of statistical learning methods may be further improved by consideration of factors such as hydrogen bonding, active transport, and the relationship with pharmacodynamic properties.

(b. b.)

69/2006

Title: Use of surface charges from DFT calculations to predict intestinal absorption.

Authors: Jones*, Ron; Connolly, P. C.; Klamt, A.; Diederhofen, M.

Strategic Technologies, Pharmacy Division, GlaxoSmithKline Research and Development

South Eden Park Road, Beckenham, Kent BR3 3BS, England.

E-mail: ron.6.jones@gsk.com; Tel.: 44-20-8639-6447; Fax: 44-1920-88-3873.

Source: J. Chem. Inf. Model. 2005, 45(5), 1337 – 1342.

Compounds:

1. Lamivudine of type I
2. a) 241 Structurally diverse compounds.

Biological material: Human intestinal organs.

Data taken from the literature:

Dataset [169 drugs with associated reliable human intestinal absorption data (%Abs)].

Computational methods:

Molecular modeling of the compounds were generated using CORINA, low-energy conformations were obtained using the MMFF force field, geometry optimization was performed using the semiempirical AM1 method implemented in MOPAC 2000 setting the van der Waals radii to 1.17 times of the default values;

COSMO (Conductive Shielding MOdel, COSMO-RS combines an electrostatic theory of locally interacting molecular surface descriptors (which are available from QM calculations) with a statistical thermodynamics methodology; the QM-COSMO calculations provide a discrete surface around a molecule embedded in a virtual conductor (theory is given);

COSMO-

%Abs (method for the a priori prediction of percentage intestinal absorption of drugs based on \( \alpha \)-moments as molecular descriptors, which are derived from quantum chemical density functional calculations combined with the continuum solvation model, COSMO);

DFT (Density Functional Theory).

Data calculated:

\( \alpha \)-Moment (using the MOPAC geometry as input, a single point density functional calculation was carried out to produce a COSMO file holding the surface polarization charges employing Turbomole and Gaussian, for Turbomole the SVP basis set, the BP functional, and the RIFDT method were used, the \( \alpha \)-moments were calculated using the COSMOtherm software);

\( K_{\text{abs}} \) [partition coefficients calculated as \( K_{\text{abs}} = \%\text{Abs}/(100 – \%\text{Abs}) \)]

RMSE [root mean square error (%)].

Results: A model for prediction of \( \%\text{Abs} \) of neutral molecules has been developed based upon surface charges of the molecule calculated by DFT. The surface charges were decomposed into \( \alpha \) moments which are correlated to a partition coefficient representing transfer of the molecule between water and the epithelial membrane. A predictive model was built and tested using a data set of 241 drugs yielding RMSE = 13% on a training set of 38 compounds as well as on
a test set of 107 drugs with high quality experimental %Abs data. Fig. 1 shows the plot of the experimental versus calculated, respectively, %Abs values calculated by COSMO (training set denoted by open squares: n = 38, RMSE = 12.5; high quality test set denoted by closed squares: n = 107, RMSE = 12.8).

The predictive quality of COSMO-%Abs was comparable to previous models. Property maps of the molecule, visualizing the atoms that enhance or hinder absorption, were produced to aid drug design. Fig. 2 shows the surface of lamivudine coded by ki, where regions surface regions with a positive effect on absorption are denoted as (A), regions of strong negative influence are indicated by (B) and (C), and regions of negligible influence are denoted with (D).

Rapid calculation of o-profiles via the COSMOfrag method only slightly decreases the accuracy and may offer an approach to rapid screening.

3D QSAR

Title: Modeling ligand-receptor interaction for some MHC Class II HLA-DR4 peptide mimetic inhibitors using several molecular docking and 3D QSAR techniques.

Authors: Wei, H.-Y.; Tsai, K.-C.; Lin*, T.-H.

Institute of Molecular Medicine & Department of Life Science, National Tsing Hua University Hsinchu 30013, Taiwan.

E-mail: thlin@life.nthu.edu.tw; Fax: 886-3-571-5934;

Source: J. Chem. Inf. Model. 2005, 45(5), 1343–1351.

Compounds: 30 HLA-DR4 peptidomimetic inhibitors, e.g., Ac-(Cha)RAMASL-NH2 (Inh.1), Ac-(L-Nba)RAMASL-NH2, or Ac-(L-m-F-Phe)RAMASL-NH2.

Biological material:

a) HLA-DR4 receptor, a class II MHC molecule
b) MHC class II peptides, cell surface proteins, that perform an essential function in immunological detection using cells.

Data taken from the literature:

IC50 [concentration of the peptidomimetic inhibitor test substance required for 50% inhibition of HLA-DR4 (details not given)];

MRP [Mean Relative Potency of a peptidomimetic inhibitor expressed as a ratio of IC50 against that of Ac-(Cha)RAMASL-NH2];

Crystal (atomic coordinates of the compounds were determined by X-ray diffraction techniques).

Computational methods:

Molecular modeling (the ligand-receptor interaction between some peptidomimetic inhibitors and a class II MHC peptide presenting molecule, the HLA-DR4 receptor, was modeled using 3D-QSAR methods such as the CoMFA, CoMSIA, and the Catalyst program);

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, hydrophobic, and hydrogen bond donor and acceptor) were evaluated using a common probe atom with 1 Å radius, +1.0 charge];

Catalyst [computer program for automatic generation of pharmacophore models for a training set of molecules specifying the relative alignments and active conformations of the ligands consistent with the binding to a common receptor site, the pharmacophore model (hypothesis) consists of a collection of features necessary for the biological activity of the ligands arranged in 3D space];

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), scoring was performed using the Xscore program];

GLIDE (docking program comparing the distances from a grid point to the receptor surface to distances from the ligand center to the ligand surface, the resultant refined poses are kept, and then minimized with a distance-dependent dielectric constant, and the conjugate gradient algorithm, the final poses are scored with Glide-Score with an inclusion of an energy score).

Data calculated:

\[ r_s = 1 - \frac{6 \cdot \Sigma d_i^2}{n(n - 1)} \], 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 (Å) of the position of the corresponding atoms of two superimposed molecular structures].

Results: Ligand-receptor interactions for some MHC Class II HLA-DR4 peptide mimetic inhibitors were modeled using several molecular docking and 3D QSAR techniques. The conformations of the peptidomimetic inhibitors investigated were defined by docking them into the known structure of HLA-DR4 receptor employing the GOLD, GLIDE Rigidly, GLIDE Flexible, and Xscore programs. The goodness of a docking result for docking a series of peptidomimetic inhibitors into the HLA-DR4 receptor was judged by comparing the Spearman’s rank correlation coefficient computed between each docking result and the activity data taken from the literature. The best CoMFA and CoMSIA models were obtained by using the aligned structures of the best docking result. The structural features selected by a stepwise CoMSIA were transferred into the Catalyst program for an automatic generation of conformations and constructing pharmacophore models for the peptidomimetic inhibitors studied. Most inhibitors were accurately predicted by the best pharmacophore model, the Catalyst Hypo1 hypothesis. The Hypo1 hypothesis was mapped onto the corresponding structures of the inhibitors. Fig. 1 shows the mapping of the Hypo1 hypothesis onto the structure of the peptidomimetic inhibitor Ac-(Cha)OMASL-NH2, where the pharmacophore features indicated are: (A) for hydrophobic, (B) for hydrogen-bond donor, and (C) for positive ionizable feature.

CoMSIA investigated all the single or possible combinations of structural features involved in the binding process and identified only two to be statistically important for pharmacophore construction. A Q² value of 0.71 was obtained for structures aligned by the GOLD docking while that obtained for structures aligned by some correspondence points was only 0.57 for the same H+S field indexes employed in the CoMSIA studies. The results indicated that the field alignment methods utilized by GOLD are superior to the point alignment method used by the CoMSIA model. The only binding pocket that remained unaccounted by the best CoMSIA model was p6 which prefers binding with some hydroxy-alkyl groups as defined by the M13 phage display libraries by others. This feature is included as D or a hydrogen-bond donor in the pharmacophore construction process by the Catalyst program.

(B. B.)

Fig. 1

CoMSIA investigated all the single or possible combinations of structural features involved in the binding process and identified only two to be statistically important for pharmacophore construction. A Q² value of 0.71 was obtained for structures aligned by the GOLD docking while that obtained for structures aligned by some correspondence points was only 0.57 for the same H+S field indexes employed in the CoMSIA studies. The results indicated that the field alignment methods utilized by GOLD are superior to the point alignment method used by the CoMSIA model. The only binding pocket that remained unaccounted by the best CoMSIA model was p6 which prefers binding with some hydroxy-alkyl groups as defined by the M13 phage display libraries by others. This feature is included as D or a hydrogen-bond donor in the pharmacophore construction process by the Catalyst program.

(B. B.)

CoMSIA investigated all the single or possible combinations of structural features involved in the binding process and identified only two to be statistically important for pharmacophore construction. A Q² value of 0.71 was obtained for structures aligned by the GOLD docking while that obtained for structures aligned by some correspondence points was only 0.57 for the same H+S field indexes employed in the CoMSIA studies. The results indicated that the field alignment methods utilized by GOLD are superior to the point alignment method used by the CoMSIA model. The only binding pocket that remained unaccounted by the best CoMSIA model was p6 which prefers binding with some hydroxy-alkyl groups as defined by the M13 phage display libraries by others. This feature is included as D or a hydrogen-bond donor in the pharmacophore construction process by the Catalyst program.

(B. B.)
Inh% [inhibition rate (%) of the SARS coronavirus 3C-like proteinase at 250 μmol/L inhibitor concentration].

**Computational methods:**

- **Molecular modeling** [a flexible 3D homology model of SARS 3C-like proteinase was built based on the structure of transmissible-gastroenteritis-virus coronavirus 3C-like proteinase by using the Modeller v6.0 program, virtual docking screening was performed using DOCK v4.0 running on a Linux Cluster Platform containing 128 CPUs, a pharmacophore model was developed employing the POCKET module of LigBuilder, consensus scoring was performed using the Score and Xcscore Programs, and finally “Drug-like” filters were employed to further improve the hit rate];

- **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:**

- **Score, Xscore** [docking score values (kJ/mol) for evaluation of the DOCK hits].

**Results:** The SARS coronavirus 3C-like proteinase is an important drug design target for the development of drugs against severe acute respiratory syndrome (SARS). In this study a “flexible” 3D model has been built for the proteinase by homology modeling and multicanonical MD method and used for virtual screening of molecule libraries. The test compounds were docked into the protease homology model, followed by the employment of a pharmacophore model, consensus scoring, and “drug-like” filters in order to improve the hit rate and the success rate of the virtual docking hit list. Refinement of the enzyme with MD and using a “flexible model” for docking screening were found to be important aspects of the design. Fig. 1 shows the binding mode of C3930 with SARS coronavirus 3C-like proteinase.

- **Fig. 1**

40 Compounds were purchased and their affinities were tested against the proteinase. Three compounds, including calmidazolium, were found to inhibit the enzyme with an apparent K, from 61 to 178μM. Fig. 2 shows the plot of the docking/score energies using homology model structure and SARS coronavirus 3C-like proteinase crystal structures.

**It was suggested that these active inhibitors and their binding modes provide useful information for understanding the receptor-ligand interaction and might help further selective drug design against SARS and other coronavirus viruses. (B. B.)**

72/2006

**Title:** Comparative protein modeling of methionine S-adenosyltransferase (MAT) enzyme from Mycobacterium tuberculosis. A potential target for antituberculosis drug discovery.

**Authors:** Khedkar, S. A.; Malde, A. K.; Coutinho*, E. C. Department of Pharmaceutical Chemistry, Bombay College of Pharmacy

- **Kalina, Santacruz (E), Mumbai 400098, India.**
- **E-mail:** evans-im@eth.net; **Tel.:** 91-22-2667-0871; **Fax:** 91-22-2667-0816.

**Source:** J. Mol. Graphics Modell. 2005, 23(4), 355 – 366.

**Compounds:** S-adenosylmethionine (SAM).

**Biological material:** Methionine S-adenosyltransferase (MAT) enzyme, a methyl donor essential for mycolipid biosynthesis, from Mycobacterium tuberculosis bacterium (Mtb).

**Data taken from the literature:**

- **Crystal structure** [atomic coordinates of E. coli MAT and rat MAT in complex with the inhibitor L-cis-AMB were taken from the Brookhaven Protein Data Bank (pdb codes: 1MXA, 1QM4, respectively)];

- **Sequence** [the amino acid sequence of Mtb-MAT was obtained from the protein database from NCBI (GenBank accession no. P77899; gi: 3915763)].

**Computational methods:**

- **Molecular modeling** [all calculations and molecular modeling of Mtb-MAT were carried out on a Silicon Graphics O2 R4400 using the INSIGHTII molecular modeling package, the HOMOLOGY program was used for the comparative protein modeling, the PSI-BLAST algorithm was used to identify homologous structures for the Mtb-MAT sequence by searching the structural database of protein sequences in the PDB, the MAT crystal structures of E. coli and rat from the PDB were selected as template structures for homology modeling of the Mtb-MAT enzyme, sequence alignment, loop and end region modeling, as well as side chain and splice repair was followed by simulated annealing and restrained energy minimizations of the Mtb-MAT model, the homodimer of Mtb-MAT was generated with the Pdbset module of CCP4].
(Windows V5.0 using the Biomt transformations, the hydrogens were added at pH 7.4, the homodimer was analyzed with the ProStat module, the accuracy and validity of the model was tested with Profiles-3D and graphically portrayed the properly folded and misfolded region(s) in the protein structure by performing an Eisenberg analysis of the model, the homology model of human MAT (hMAT) was generated using the same protocol as described for Mtb-MAT using only the rat MAT (1QM4) crystal structure as the reference protein).

**Data calculated:**
MEP [Molecular Electrostatic Potential (kcal/mol)] was calculated using MOLCAD implemented in Sybyl v6.7.

**Results:** MAT is an important target for antituberculosis drug discovery. In this study a homology model of MAT has been constructed using the X-ray structures of E. coli MAT and rat MAT as templates. The resulting model showed correct stereochemistry as assessed by the Profiles-3D score (Ramachandran plot and good 3D structure compatibility). All structurally and functionally important residues in the active site of Mtb-MAT have been identified that are also present in the E. coli and rat MAT crystal structures and the reported point mutation data. The homology model conserves well the crucial topological and active site features of the MAT family of proteins. There are some differences in the active sites of Mtb and human MAT, which can be exploited to design specific and selective inhibitors of Mtb-MAT. Fig. 1 shows the active site residues of Mtb-MAT (bold) and hMAT after the superposition onto the rMAT crystal structure (1QM4). The residues of rMAT have not been shown as it appears that the Mtb-MAT crystal structure as the reference protein]

These differences appear when one compares the molecular electrostatic potential of mycobacterial and human MAT. The differences in the MEP maps of Mtb and human MAT indicated that selective and specific Mtb-MAT inhibitors can be designed using the homology model.

**Computational methods:**

- **Molecular modeling** [in silico screening was performed using AutoDock v3.0, the top 10 configurations according to δG binding energy were energy minimized using the CHARMM force field employing the InsightII/CHARMM programs, a total of more than 500 compounds in the in-house database were processed by this method, the 3D structure of the P2Y12 receptor was constructed on the basis of the coordinates and the conformation of the 3D model of the P2Y1 receptor by both Fourier transform analysis and homology modeling using the bovine rhodopsin structure as a template, the transmembrane regions (TM) of the P2Y12 receptor were determined by alignment with those of the P2Y receptors and bovine rhodopsin, additionally, high temperature molecular dynamics (MD) simulation was performed for energy minimization using the InsightII/CHARMM parameter];
- **AUTO-DOCK** (automated docking program for finding potential docking sites of whole ligands with user-specified dihedral angle flexibility on rigid protein binding sites of known structure by using a Metropolis Monte Carlo algorithm of simulat-
ed annealing for positional and conformational searching in combination with a rapid energy evaluation through precalculated grids of molecular affinity potentials allowing the inclusion of van der Waals, electrostatic, and hydrogen bonding interactions; evolutionary relationship was analyzed based on the similarity of the primary sequence in the alignment of proteins, the alignments were calculated by ClustalX, and uncertain portions, such as gaps, were deleted using BioEdit.

**Data calculated:**
\[ \delta G \] [AutoDock binding energy (kcal/mol)].

**Results:** Endogenous ligands acting on a human P2Y12 receptor, were searched by in silico screening using AutoDock against an in house molecule library containing more than 500 animal metabolites. Fig. 1 shows the 3D model of PRPP ligand-bound human P2Y12 receptor. The side chains of the important residues in proximity to the docked PRPP are highlighted and labeled.

In addition to the known P2Y12 ligands, such as 2MeSADP and ADP, the screening yielded a selection of cysteinyleukotrienes (CysLTs) and PRPP, with high free energy changes. These ligands were subjected to an in vitro Ca\(^{2+}\) assay using the CHO cells stably expressing P2Y\(_{12}\)-G16 a fusion proteins. It was found that CysLTE4 and PRPP behaved at the P2Y\(_{12}\) receptor as agonists with the EC\(_{50}\) values of 1.3 and 7.8 nM, respectively. Moreover, the phylogenetic relationship of the P2Y, P2Y-like, and CysLT receptors was analyzed based on sequence alignment followed by evolutionary analyses. The phylogenetic analyses revealed that the P2Y\(_{12}\), P2Y\(_{13}\), P2Y\(_{14}\), GPR87, CysLT-1, and CysLT-2 receptors formed a P2Y-related receptor subfamily with common sequence motifs in the transmembrane regions.

**Phylogenetic analysis**

**Data set:** (33 HIV IN inhibitory DKAs with associated IC\(_{50}\) values).

**Computational methods:** Molecular (the algorithm HypoGen implemented in the Catalyst package was used to derive automated SAR pharmacophore hypotheses from a data set of \(\beta\)-diketo-derivatives); Catalyst [computer program for automatic generation of pharmacophore models for a training set of molecules specifying the relative alignments and active conformations of the ligands consistent with the binding to a common receptor site, the pharmacophore model (hypothesis) consists of a collection of features necessary for the biological activity of the ligands arranged in 3D space, the common ones being hydrogen bond acceptor, hydrogen bond donor, and hydrophobic features, the Catalyst/HypoGen module takes activity data into account and uses active and both inactive compounds in an attempt to identify hypotheses that are common among the active compounds but not among the inactives).

**Results:** Pharmacophore-based design of HIV IN inhibitors was performed employing the Catalyst/HypoGen module using a set of 33 \(\beta\)-diketo-acid HIV IN inhibitors (17 training set and 16 test set compounds) reported in the literature. The derived models showed quantitative predictive ability in terms of activity. The best statistical hypothesis consisted of four features (one hydrophobic aromatic region, two hydrogen-bond acceptors, and one hydrogen-bond donor) with correlation coefficient of 0.96. Fig. 1 shows the top scoring HypoGen pharmacophore Hypo1 mapped onto the most active compound (L-870,810) in the training set, where HBD, HBA, and HyAr denote hydrogen bond donor, hydrogen bond acceptor, and hydrophobic aromatic region features of the pharmacophore.

**Fig. 1**

The resulting pharmacophore model guided the rational design of benzylindoles as novel IN inhibitors containing a benzylindole skeleton. The synthesized molecules were found...
to be potent HIV IN inhibitors acting by blocking the strand transfer process. The microwave-assisted synthesis and biological evaluation of the new compounds is described. It was suggested that the best HypoGen pharmacophore can be used as a 3D query for the search of novel potential IN inhibitors in large 3D molecule libraries.

(B. B.)

75/2006

Title: Structural basis of acetylcholinesterase inhibition by triterpenoidal alkaloids.

Authors: Khalid, A.; Azim, M. K.; Parveen, S.; Atta-ur-Rahman; Choudhary*, M. I.

Dr. Panjwani Center for Molecular Medicine and Drug Research, International Center for Chemical Sciences, University of Karachi

Karachi 75270, Pakistan.

E-mail: hej@cyber.net.pk; Fax: 92-21-924-319-091.

Source: Biochem. Biophys. Res. Commun. 2005, 331(4), 1528 – 1532.

Compounds:
a) 2 Triterpenoidal alkaloids of type I: buxamine-B, where \( R_1 = \text{H} \); \( R_2 = \text{CH}_3 \); buxamine-C, where \( R_1 = \text{CH}_3 \);
b) 4 AChE inhibitors: galanthamine, endrophonium, decemethonium, BW284C51.

Biological material: Torpedo californica acetylcholinesterase (type VI-S) (AChE, EC 3.1.7).

Data taken from the literature:
Crystal structure [atomic coordinates of the aromatic gorge of AChE taken from the Protein Data Bank (pdb code: 1ACL)].

Data determined:
Ki [Michaelis inhibition constant (\( \mu \text{M} \)) representing the affinity of the substrate to T. californica AChE].

Computational methods:
Molecular modeling [the 3D structures of buxamine-B and -C were constructed using the Sybyl program, energy minimization was carried out using the Tripos force field, the docking studies were carried out using FlexX docking software, Buxamine-B and -C models were docked in the aromatic gorge of AChE (PDB: 1ACL), the docking results were analyzed by LIGPLOT and WebLab ViewerPro];
FlexX (program for automatic protein-ligand docking based on incremental construction while the receptor is kept rigid);
LPC (program for analysis of protein-ligand contacts based upon an approach known as surface complementarity).

Results: Structural basis of AChE inhibition by triterpenoidal alkaloids of type I has been performed by enzyme kinetics and FlexX molecular docking experiments. Buxamine-C has been found to be 20-fold potent than buxamine-B (Ki = 5.5\( \mu \text{M} \) and 110\( \mu \text{M} \), respectively). The results of the docking experiments suggested that the cyclopentanophenanthrene skeleton of these inhibitors precisely fits into the aromatic gorge of AChE (pdb code: 1ACL). It was suggested that the C-3 and C-20 amino groups of both alkaloids mimic the well-known bis-quaternary ammonium type inhibitors such as decamethonium and interact with the Trp84 and Trp279 residues of the enzyme, respectively. Fig. 1 shows the superposition of both natural inhibitors and decamethonium docked into the aromatic gorge of AChE (pdb code: 1ACL).

The C-3 amino group in buxamine-C appears to adopt a good position at the bottom of the aromatic gorge exerting a minimum of destabilizing contacts, which seems to be crucial for the inhibitory activity of such inhibitors. These observations are supported by significant difference between complementarity values calculated for docked buxamine-B (0.3) and buxamine-C (0.5) by the LPC software.

(B. B.)

76/2006

Title: Modeling aided design of potent glycogen phosphorylase inhibitors.

Authors: Deng*, Qiaolin; Lu, Z.; Bohn, J.; Ellsworth, K. P.; Myers, R. W.; Geissler, W. M.; Harris, G.; Willoughby, C. A.; Chapman, K.; McKeever, B.; Mosley, R.

Department of Molecular Systems, Merck Research Laboratories, Merck & Co. Inc.

P. O.Box 2000, Rahway, NJ 07065, USA.

E-mail: qiaolin_deng@merck.com; Tel.: 1-732-594-0618.

Fax: 1-732-594-4224.

Source: J. Mol. Graphics Modell. 2005, 23(5), 457 – 464.

Compounds:
a) 14 Compounds of type I, where \( R = \text{H}, \text{NO}_2, \text{Cl}, \text{OMe}, \text{CF}_3, \text{Et}, \text{Me} \) (both naphthyl and phenyl derivatives);
b) 3 Known GP inhibitors: Bayer diacid compound (W 1807) of type II, caffein, and CP320626.
Biological material: Glycogen phosphorylase (GP) is a key enzyme in the regulation of glycogen metabolism, catalyzing the breakdown of glycogen to glucose-1-phosphate; among the GP isozymes, human liver GP (HLGP) is the preferred target.

Data taken from the literature: Crystal structure [atomic coordinates of rabbit muscle glycogen phosphorylase complexed with a W1807, caffeine at the inhibitor site, and CP320626 at the dimer dimer interface site (1C50) were taken from the Brookhaven Protein Data Bank (pdb codes: 3AMV, AGFZ, 1C50, respectively)].

Computational methods: Molecular modeling conformers of type I with R = NO₂ (Ia) were generated by distance geometry approach, the conformer set was energy minimized with a distance-dependent dielectric of 2r with the MMFFs force field calculations and superposed onto known inhibitors in complex with GP (3AMV), caffeine at the inhibitor site (1GFZ), and CP320626 at the dimer interface site (1C50), the AMP allosteric site was identified as the most likely binding site for compound Ia through the SQ overlay, a docking study was carried out using ICM software yielding 100 initial docking poses within the AMP allosteric site, each of the complexes obtained was energy optimized using the MMFFs force field, the total energy of the complex, the individual energies of the ligand and the enzyme, and the interaction energy between the ligand and the enzyme were calculated, the binding pocket was characterized by generating a grid for FLOG as a series of iso-energetic surfaces to characterize the binding pocket by its polar (hydrogen bond donor and acceptor) and hydrophobic nature as well as its van der Waal limits].

Data calculated: E_{int} [interaction energy between the ligand and the enzyme as well as the total energy of the complex, and the individual energies of the ligand and the enzyme, respectively, were calculated (kcal/mol)].

Results: A novel type of potent glycogen phosphorylase inhibitors were designed using molecular modeling methodologies based on a phenyl diacid lead, W1807. In the absence of suitable competitive binding assays, compound W1807 was predicted to bind at the AMP allosteric site. The allosteric site was located by superposition onto known inhibitors which bind at different sites in the enzyme and analyses of the surrounding protein environment associated with these distinct sites. Possible docking modes of W1807 at the AMP allosteric site were further modeled using the complex crystal structure 3AMV. W1807 was predicted to interact with positively charged arginines at the AMP allosteric site in the docking model. A reasonable docking model was computationally determined and subsequently confirmed by X-ray crystallography. Fig. 1 shows the docking model of W1807 inside the AMP allosteric site. The heavy atoms of W1807 are shown in ball and stick, the GP structure is shown in solid ribbon with helices, hydrogen atoms have been omitted for clarity.

Characterization of the binding pocket by FLOG revealed a large empty hydrophobic region near the central phenyl ring suitable for accommodating compounds with larger hydrophobic groups and thus improve binding. Based on the results a fused ring analogues were designed with increased hydrophobic bulk in this unfilled region to improve binding. As predicted, this exercise resulted in a new series of GP inhibitors with significantly improved potency. The new ligands showed significantly improved potency. (B. B.)

77/2006

Title: Carbonic anhydrase inhibitors. Stacking with Phe131 determines active site binding region of inhibitors as exemplified by the X-ray crystal structure of a membrane-impermeant antitumor sulfonamide complexed with isozyme II.

Authors: Menchise, V.; De Simone*, G.; Alterio, V.; Di Fiore, A.; Pedone, C.; Scozzafava, A.; Supuran, C. T. Istituto di Biostrutture e Bioimmagini-CNR via Mezzocannone 16, I-80134 Naples, Italy. E-mail: gmg@chemistry.unina.it; Tel.: 39-081-253-4579; Fax: 39-081-253-6642.

Source: J. Med. Chem. 2005, 48(18), 5721–5727.

Compounds: 10 Structurally diverse sulfonamides, e.g., type I.
Biological material: 2 Human carbonic anhydrases (CAs, EC 4.2.1.1): hCA II, hCA IX.

Data taken from the literature:
Crystal structure [atomic coordinates of CA II were taken from the Brookhaven Protein Data Bank (pdb code: 1CA2)].

Data determined:
Crystal structure (atomic coordinates of the hCA II/type I complex were determined by X-ray diffraction techniques).

Computational methods:
Molecular modeling (the structure of the hCA II/type I complex was analyzed by difference Fourier techniques, the model was refined using the CNS program to crystallographic R-factor and R-free values of 0.204 and 0.251, respectively, model building and map inspections were performed using the program O, the stereochemical quality of the model was assessed by Procheck, the most favored and additionally allowed regions of the Ramachandran plot contained 98.6% of the nonglycine residues).

Data calculated:
rmse [root mean square deviation (Å) of the position of the corresponding atoms of two superimposed molecular structures].

Results: Examination of the crystallographic structure of the complex revealed the following features: the phenylethyl moiety of the ligand fills the active site, engaging in van der Waals interactions with side chains of Gln192, Val121, Phe131, Leu198, and Thr200. The 2,4,6-trimethylpyridinium moiety is located at van der Waals distance from the aliphatic chain of Ile91 es-

Biological material: Tyrosinase, a key enzyme catalyzing the first two steps of melanin biosynthesis, it is a copper protein widely distributed in nature.

Data taken from the literature:
IC_{50} [concentration of the test substance (μM) required for 50% inhibition of tyrosinase].

Computational methods: QSAR modeling [the main steps for the application of this method in QSAR were the following: (i) draw the molecular pseudographs for each molecule of the data set, using the software drawing mode; (ii) use appropriate atom weights in order to establishing a strong offset face-to-face stacking with Phe131. Other X-ray crystal structures of hCA II-sulfonamide/sulfamate adducts led to the proposal of two principal binding modes: some inhibitors bind with their tail within the hydrophobic half of the CA II active site, defined by amino acid residues Phe131, Val135, Leu198, Pro202, Leu204. A second group of derivatives, including the positively charged sulfona-
mide investigated here, produces binding with their tail in a different region of the active site, pointing toward the hydrophilic half of it and making an offset face-to-face stacking with Phe131. It appears that this last interaction orients the inhibitor toward the hydrophilic part of the active site, whereas impossibility to participate in this stacking leads to the binding of inhibitors within the hydrophobic half. It was sug-
gested that these findings are relevant for the design of better inhibitors targeting isozymes II, IX, and XII, some of these being overexpressed in hypoxic tumors showing bad disease prognosis.
differentiate the molecular atoms, where each atomic nucleus was characterized with the following parameters: atomic mass (M), atomic polarizability (P), atomic Mulliken electronegativity (K), van der Waals atomic volume (V), and the atomic electronegativity in Pauling scale (G); TOMOCOMD-CARDD [Topological MOlecular COMputer Design-Computer Aided “Rational” Drug Design] that has been developed to generate molecular fingerprints on the basis of the application of discrete mathematics and linear algebra theory to chemistry, in this sense, atom, atom-type, and total quadratic and linear molecular fingerprints have been defined in analogy to the quadratic and linear mathematical maps (theoretical framework and algorithm is given); LDA [Linear Discriminant Analysis implemented in STATISTICA v6.0; statistical parameters: the Wilks’s, parameter (U-statistic), square Mahalanobis distance (D²), and Fisher ratio (F) for the training set; linear canonical regression coefficient (R can) and chi-squared (χ²) as the measure of the statistical quality of the developed models].

Data calculated:
ΔP% (posterior classification probability classifying a compound as active or inactive by the discriminant equation).

Results: The search for novel tyrosinase inhibitors requires the study of the role of tyrosinase in hyperpigmentation and melanogenesis disorders. In this study novel tyrosinase inhibitors were identified by using TOMOCOMD-CARDD descriptors and pattern recognition techniques for finding functions that discriminate between potent tyrosinase inhibitor compounds and inactive ones. Significant LDA models were developed and globally good classifications of 93.51% and 92.46% were achieved for the best models using non-stochastic and stochastic linear indices respectively, in the training set. For external prediction sets accurate classifications of 91.67% and 89.44% were obtained. Fig. 1 shows the plot of the ΔP% calculated by one of the developed equations using stochastic linear indices for each compound in the training and test sets (the compounds 1–183 and 184–246 are active (tyrosinase inhibitors) in training and test sets, respectively; chemicals 247–541 and 542–658 are inactive (non-inhibitors of tyrosinase) in both training and test sets, respectively.

The new method increases the speed of the discovery of new lead-like compounds, as a suitable alternative to the screening and in vitro assay. This was proven experimentally through the isolation and characterization of six new cycloartane compounds isolated from herbal plants followed by tyrosinase inhibitory assay. A good concordance was shown between the theoretical and experimental results. It was suggested that the results obtained provides a QSAR tool that can be used in the identification of new tyrosinase inhibitor compounds.

(B. B.)

79/2006

Title: Theoretical and experimental design of atypical kinase inhibitors. Application to p38 MAP kinase.

Authors: McClure*, K. F.; Abramov, Y. A.; Laird, E. R.; Barberia, J. T.; Cai, W.; Carty, T. J.; Cortina, S. R.; Danley, D. E.; Dipesa, A. J.; Donahue, K. M.; Dombroski, M. A.; Elliott, N. C.; Gabel, C. A.; Han, S.; Hynes, T. R.; LeMotte, P. K.; Mansour, M. N.; Marr, E. S.; Letavic, M. A.; Pandit, J.; Ripin, D. P.; Sweeney, F. J.; Tan, D.; Tao, Y.

Pfizer Global Research and Development, Groton Laboratories Eastern Point Road, Groton CT 06340, USA.

E-mail: kim.f.mcclure@pfizer.com; Tel.: 1-860-441-8223; Fax: 1-860-715-4610.

Source: J. Med. Chem. 2005, 48(18), 5728–5737.

Compounds:
a) 5 Bioisostere structures used for molecular modeling: 1-methyl-1H-benzotriazole, 5, 3-methyl benzo[d]isoxazole, 3-methyl[1,2,4]triazolo[4,3-a]pyridine, pyridine, 3-dimethyl-1,3-dihydro benzimidazol-2-one-pyridine
b) Compounds of type I, type II, type III, type IV.

QSAR Comb. Sci. 25, 2006, No. 3, 265–292  www.qcs.wiley-vch.de © 2006 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim 281
Biological material: Stress-activated protein kinase, p38 MAP kinase p38R, a component of the complex cytokine-signaling pathway.

Data taken from the literature:
Crystal structure: atomic coordinates of p38α/SB-203580 complex were taken from the Brookhaven Protein Data Bank (pdb code: 1IAN).

Data determined:
IC$_{50}$ [concentration of the test substance (nM) required for 50% inhibition of p38α]; Crystal structure (atomic coordinates of type II complexed by p38α were determined by X-ray diffraction techniques).

Computational methods:
Molecular modeling: the geometries of the five fragments were fully optimized in the gas phase at the B3LYP/6-31G(d) level of theory, the molecular dipole moments, isotropic polarizabilities, and atomic charges were evaluated by single-point calculations at the B3LYP/6-311G(d,p) level, the ESP atomic charges were calculated according to the CHELPG algorithm, the polarized continuum model (PCM) was adopted at the HF/6-31+G(d) level of theory for the geometry optimizations in water.

Data calculated:
Descriptors [solvation free energy, ΔGsolv, and absolute chemical hardness, $\eta$; descriptors were calculated using single-point calculations using the Gaussian98/03 program suites, the molecular electrostatic potentials (MEPs) were calculated at the B3LYP/6–311G(d,p) level of theory and plotted using Spartan’02 software, the octanol-water partition coefficients, logP, were predicted using ACD/labs v6.00 software package, the van der Waals molecular volume, V, and water-accessible hydrophobic surface area, ASA-H, descriptors were evaluated by the MOE software, only atoms with absolute values for the ESP atomic charges of less than 0.2 e were considered in calculating the ASA-H, the molecular geometries optimized in solution were adopted for these calculations].

Results: Theoretical and experimental design of atypical kinase inhibitors has been performed and applied to the p38 MAP kinase. More specifically, 5 mimics of the benzimidazole nucleus found in inhibitors of p38 kinase were proposed, and their theoretical potential as bioisosteres were investigated. A set of calculated descriptors relevant to the anticipated binding interaction for the bioisostere fragments were calculated. The design considerations and synthesis of p38 inhibitors based on these H-bond acceptor fragments are described in detail. Fig. 1 shows the cartoon representation of p38α/SB-203580 complex 1IAN of a benzimidazole in the active site of p38α.

Comparative evaluation of the pyridine-, benzimidazole-, benzotriazole-, and triazolopyridine-fragment containing inhibitors shows the triazoles type II and IV to be significantly more potent experimentally than the benzimidazole structure of type II after which they were modeled. An X-ray crystal structure of type IV bound to the active site showed that the triazole group serves as the H-bond acceptor but unexpectedly as a dual acceptor, inducing movement of the crossover connection of p38R. Fig. 2 shows the X-ray structure of type IV bound to p38α.

As a result of the study, significantly more potent compounds were identified with superior calculated physicochemical properties than the benzimidazole group that they were designed to replace. The computed descriptors for the hydrophobic and d-d interaction capacities were the most useful in ranking potency.

(B. B.)

80/2006

Title: Antibiotic binding to monozinc Cpha β-lactamase from Aeromonas hydrophila. Quantum mechanical/molecular mechanical and density functional theory studies.

Authors: Xu, D.; Zhou, Y.; Xie, D.; Guo*, H.
Department of Chemistry, University of New Mexico Albuquerque NM 87131, USA.
Compounds: Biapenem of type I, an antibiotic, shown in Fig. 1;

Biological material: Monozinc Cpha α-lactamase enzyme from Aeromonas hydrophila.

Data taken from the literature:

Crystal structure (atomic coordinates of biapenem were determined by X-ray diffraction techniques);

Crystal structure of the lactamase enzyme-reaction intermediate of the biapenem hydrolysis complex were taken from the Brookhaven Protein Data Bank (pdb code: 1X8I).

Computational methods:

Molecular modeling [the docked lactamate/biapenem model was constructed by modifying the enzyme-intermediate complex (code: 1X8I), after removing the intermediate the optimized biapenem structure was docked in the active site, the resulting complex was solvated in a 25 Å radius of pre-equilibrated TIP3P water sphere centered at the zinc ion and optimized according to the QM/MM and DFT protocols, the quantum region in the QM/MM simulations, which included the Zn(II) ion and its ligands, the antibiotic molecule, the catalytic water, and an active-site histidine residue, was treated using the SCC-DFTB model, 500 ps molecular dynamics (MD) simulations of both the apo enzyme and the substrate-enzyme complex were performed employing the SHAKE algorithm, all DFT calculations were performed using Gaussian 03];

QM/MM, DFT [Quantum Mechanical/Molecular Mechanical, and Density Functional Theory (DFT) methods, respectively];

SCC-DFTB (Self-Consistent Charge Density Functional Tight Binding);

SHAKE (the SHAKE algorithm used to constrain all bond lengths of the molecules to their equilibration values).

Results: The active-site dynamics of the binding of biapenem to monozinc Cpha α-lactamase has been studied employing QM/MM and DFT methods. The quantum region in the QM/MM simulations was treated using the SCC-DFTB model. Biapenem was docked at the active site corresponding to a recent X-ray structure of an enzyme-intermediate complex (code: 1X8I). Fig. 1 shows the important bonding interactions at the active site of the CphA-biapenem complex, where the biapenem is drawn with bold lines.

The modeling experiments revealed that the substrate is connected in a metal binding through its 3-carboxylate oxygen, and anchored by several hydrogen bonds between the substrate and active-site residues, particularly those made possible by the conformational change of Asn233. An active-site water was found in a pocket near the lactam carbonyl carbon of the substrate hydrogen bonded primarily with the carboxylate side chain of the metal-binding Asp120. It was proposed that the Asp120 residue, like His118, may serve as a general base to activate the catalytic water. The energetic stability of the metal-ligand bonds and the hydrogen-bond network was checked by MD simulations of both the apo enzyme and the substrate-enzyme complex. The structure and dynamics of the substrate-enzyme complex model provided valuable insights into the mode of catalysis in these enzymes that play a central role in the bacterial resistance to α-lactam antibiotics. The theoretical simulations complementing experimental results provided important details of the process and yielded quantities that would be hard to measure accurately. It was suggested that a molecular level modeling of the binding and catalysis of the metallo-α-lactamases is essential in understanding the origin of bacterial resistance and in designing novel and effective inhibitors. (B. B.)

81/2006

Title: Angiotensin II pseudopeptides containing 1,3,5-trisubstituted benzene scaffolds with high AT2 receptor affinity.

Authors: Georgsson, J.; Sköld, C.; Plouffe, B.; Lindeberg, G.; Botros, M.; Larhed, M.; Nyberg, F.; Gallo-Payet, N.; Gogoll, A.; Karlén, A.; Hallberg*, Anders.

Department of Medicinal Chemistry, Division of Organic Pharmaceutical Chemistry, Uppsala University P. O. Box 574, SE-751 23 Uppsala, Sweden.

E-mail: anders.hallberg@orgfarm.uu.se; Tel.: 46-18-471-4284; Fax: 46-18-471-4474.

Source: J. Med. Chem. 2005, 48(21), 6620 – 6631.

Compounds: 5 Pseudopeptide analogues of angiotensin II (Ang II of type I, Asp-Arg-Val-Tyr-Ile-His-Pro-Phe), where the Val-Tyr-Ile, Val-Tyr, or Tyr-Ile portions are replaced. [125I]Ang II.

Biological material: Angiotensin receptors: AT1 in the rat liver membranes, and AT2 in the pig uterus myometrium.

Data determined:

Kc1 [Michaelis inhibition constant (nM) representing the affinity of the substrate to displace [125I]Ang II at the AT1 or AT2 receptor].
Computational methods:
Molecular modeling [conformational search of the Ang II and its analogues was conducted using the systematic unbound multiple minimum (SUMM) search method in the BatchMin program, the OPLS- AA force field and the general Born solvent accessible (GB/SA) surface area method for water as implemented in MacroModel v8.5, were used in the calculations, pharmacophore model of these agonists were derived using DISCOtech implemented in the SYBYL v6.9 package];

DISCO [Distance COmparison used for finding superimpositions of compounds that contain at least one conformation of each molecule and user specified numbers of point types, if a clique meets the user-specified set-up criteria (e.g. a minimum number of site points matched, distance tolerances, etc.) a pharmacophore is defined, the pharmacophore points are hydrogen bond donor atoms (DL), hydrogen bond acceptor atoms (AL), charged atom centres (CHG), centres of hydrophobic rings (HY), in addition Donor Site (DS) and Acceptor Site (AS) exist in the biological receptor];

DISCO-tech [an enhanced, faster version of DISCO, given a set of molecules that are related by their ability to bind to a protein receptor, DISCOtech identifies features that could be elements in a pharmacophore model, DISCOtech operates in distance space and can perform clique detection to generate pharmacophore hypotheses on up to 300 conformers per molecule).

Data calculated:  
rmssd [root mean square deviation (Å) of the position of the corresponding atoms of two superimposed molecular structures].

Results: Two 1,3,5-trisubstituted aromatic scaffolds intended to serve as \( \gamma \)-turn mimetics have been synthesized and incorporated in five pseudopeptide analogues of Ang II. Molecular modeling supported the supposition that 1,3,5-trisubstituted aromatic scaffolds may serve as relevant \( \gamma \)-turn mimetics. Fig. 1 shows the image of models of scaffolds superimposed onto an inverse \( \gamma \)-turn found in a crystallized protein.

Thus, a tentative model has been proposed for the activation of the AT\(_2\) receptor by angiotensin II analogues. (B. B.)

Title: A proposed molecular basis for the selective resveratrol inhibition of the PGHS-1 peroxidase activity.

Authors: Kümmerle, A. E.; da Silva, G. M. S.; Sant’Anna, C. M. R.; Barreiro, E. J.; Fraga*, C. A. M.

Laboratório de Avaliação e Síntese de Substâncias Bioativas (LASSBio), Faculdade de Farmácia, Universidade Federal do Rio de Janeiro (UFRJ)
Rio de Janeiro, P. O.Box 68023, RJ 21944-971, Brazil.
E-mail: cmfraga@pharma.ufrj.br; Tel.: 55-21-2562-6503; Fax: 55-21-2562-6644.

Source: Bioorg. Med. Chem. 2005, 13(21), 5981 – 5985.

Compounds: Resveratrol of type I.

Biological material: Prostaglandin endoperoxide H\(_2\) synthase isomorphs (PGHS-1 and PGHS-2), a homodimeric bifunctional enzyme that catalyzes the first two steps in the biosynthesis of prostaglandins from arachidonic acid.

Data taken from the literature: Crystal (crystal coordinates of PGHS-1 in complex with the inhibitor flurbiprofen (pdb code: 1EQH) and PGHS-2 in complex with SC-558 (pdb code: 1CX2) have been taken from the Brookhaven Protein Data Bank].
**Computational methods:**

Molecular modeling (sketching, geometry optimization and conformational search of resveratrol was performed BioMedCAche v5.0 software, the minimum energy conformation was obtained first with the MM2 method followed by application of semi-empirical AM1 minimization using two dihedral angle search by step, the charges of resveratrol and heme group were computed using the Tripos force field with the Gasteiger-Hückel method, docking of resveratrol was carried with FlexX in Sybyl v7.0 into the peroxidase site of the crystallographic PGSH-1 and PGSH-2 structures, scaling function was used to penalize deviations of the docked conformer from the ideal geometry);

FlexX (program for automatic protein-ligand docking based on incremental construction while the receptor is kept rigid).

**Data calculated:**

$\Delta G_{\text{bind}}$ [binding energy (kcal/mol) estimated as the sum of free energy contributions from hydrogen bonding, ion-pair interactions, hydrophobic and $\pi$-stacking interactions of aromatic groups, and lipophilic interactions].

**Results:** Resveratrol was docked into the PGSH-1 and PGSH-2 receptor in order to study the molecular basis for the selective inhibition of the PGHS-1 peroxidase activity by resveratrol. The docking model predicted a slightly less favorable $\Delta G_{\text{bind}}$ ($-17.9 \text{ kcal/mol}$) value of the resveratrol to the PGHS-2 peroxidase site than its corresponding binding to the PGHS-1 ($-20.4 \text{ kcal/mol}$) site. Fig. 1 and Fig. 2 show the molecular docking results of resveratrol with the PGHS-1 and PGHS-2 peroxidase sites, respectively, where only the main amino acid residues within 3 Å around the inhibitor are shown for clarity. In both complexes resveratrol is attached to the peroxidase site of the corresponding PGHS isoform, describing the molecular surface contour of the ligand, and dashed lines represent the hydrogen bonds formed between resveratrol and the corresponding PGHS isoform.

The formation of hydrogen bonds among the hydroxyl groups of the resveratrol phenyl rings, the backbone of Fe-heme and the carbonyl group of Leu294 inside the PGHS-1 peroxidase site, as well as the absence of His214 in the backbone of PGHS-1 were the crucial features required to keep the aromatic rings of the natural product parallel to the Fe-heme group and transverse to the peroxidase access channel. This positioning represented a large steric hindrance at this site and led to selective inhibition.

(B. B.)

83/2006

**Title:** Structural model of the Plasmodium CDK, Pfmrk, a novel target for malaria therapeutics.

**Authors:** Peng, Y.; Keenan, S. M.; Welsh*, W. J.

Department of Pharmacology, University of Medicine & Dentistry of New Jersey-Robert Wood Johnson Medical School (UMDNJ-RWJMS) and the Informatics Institute of UMDNJ.

Piscataway NJ 08854, USA.

E-mail: welshwj@umdnj.edu; Tel.: 1-732-235-3234; Fax: 1-732-235-3475.

**Source:** J. Mol. Graphics Modell. 2005, 24(1), 72 – 80.

**Compounds:** 10 Oxindole-based compounds of type I, where $R_1 = \text{diverse structural moieties; } R_2 = \text{H,OMe,Br,I}; e.g., GW5074 where } R_1 = \text{C-Ph-3,5-diBr-4-OH}; R_2 = \text{I}.

**Biological material:** Malaria falciparum kinase (Pfmrk) that is homologous to the mammalian cyclin-dependent kinases (CDKs) (Pfmrk shows 46% sequence identity with CDK7).

**Data taken from the literature:**

Protein sequences [all sequence data were obtained from the National Center for Biotechnology Information (accession numbers of Pfmrk and human CDK2 (hCDK2): AAD55782 and P24941, respectively, multiple sequence alignments for a series of CDKs, as well as pairwise alignment between hCDK2 and Pfmrk, were conducted using the ClustalW1.8 routine with default parameters];

Crystal structure [atomic coordinates of hCDK2 were taken from the Brookhaven Protein Data Bank (pdb code: 1QMZ)].

**Data determined:**

IC$_{50}$ [concentration of the test substance (µM) required for 50% inhibition of Pfmrk (details not given)].

**Computational methods:**

Molecular [all calculations were conducted on Silicon Graphics Octane R12000 machines, ligand structures were built with Sybyl 6.8 and optimized with the MMFF94 force field and atomic charges, the Pfmrk structural model was con-
structed in five steps, the active crystal structure of hCDK2 (pdb ID = 1QMZ) was chosen as the template for the homology modeling of Pfmrk, ab initio parameterization involving the EM and MD refinement procedures of the Pfmrk-ATP structural model required parameterization of the Mg²⁺ ion within its hexacoordinate environment: atomic charges, van der Waals radius and force constant for nonbonded interactions, the RHF (spin-restricted Hartree-Fock) Hamiltonian with the 6-31G* basis set implemented in the Gaussian98 package was employed to calculate the electrostatic potential (ESP) for the Mg²⁺ coordination complex, refinement of the Pfmrk structural model was conducted by EM and MD calculations using AMBER7 supplemented by the Mg²⁺-related parameters calculated, geometry of the energy minimized structures was analyzed with PROCHECK, a set of Pfmrk inhibitors of type I were docked to the structural model of Pfmrk using GOLD and the G Score function: [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)].

Results: Recent genome sequencing and molecular cloning projects have identified several enzymes from P. falciparum that may represent novel drug targets, including CDKs which are expected to play a crucial role in parasitic growth. In this study a 3D structural model of Pfmrk, a putative human CDK activating kinase (CAK) homolog in P. falciparum has been constructed. Significant features of this structural model were (i) parameterization of the Mg²⁺ hexacoordination system using ab initio quantum chemical calculations to accurately represent the ATP-kinase interaction; and (ii) comparison between the GOLD docking scores and measured binding affinities for a series of oxindole-based Pfmrk inhibitors of known activity. The refined structural model was evaluated by the docking of a series of known Pfmrk inhibitors. A reasonable correlation (r = 0.82) has been calculated between the G Scores and inhibition constants (pIC₅₀) validated the accuracy of the developed Pfmrk structural model.

Fig. 1 shows the plot of the G Score versus pIC₅₀ values for the oxindole-based Pfmrk inhibitors.

Fig. 2 shows the binding mode of GW5074 in the ATP-binding pocket of Pfmrk, where the docked GW5074 is rendered in ball-and-stick mode, the protein residues are rendered in stick mode, and the hydrogen bonds are highlighted by dotted lines.

Analysis of the ligand-Pfmrk binding interactions revealed specific residues (Met66, Met75, Met91, Met94, and Phe143) within the Pfmrk binding pocket that may play an important role in inhibitor binding affinity and selectivity. It was suggested that the Pfmrk structural model developed, together with information obtained from analysis of ligand-receptor interactions, should help the rational design of novel, potent and selective Pfmrk inhibitors as antimalarial agents. Several novel compounds were identified that exhibit in vitro inhibitory activity against Pfmrk. The chemical scaffolds of these inhibitors were distinct from those associated with known CDK inhibitors.

(B. B.)

84/2006

Title: Novel properties of malarial S-adenosylmethionine decarboxylase as revealed by structural modelling.

Authors: Wells, G. A.; Birkholtz, L.-M.; Joubert, F.; Walter, R. D.; Louw*, Abraham, I.

Department of Biochemistry, School of Biological Sciences, Faculty of Natural and Agricultural Sciences, University of Pretoria

Pretoria 0002, South Africa.

E-mail: braam.louw@bioagric.up.ac.za; Tel.: 27-12-420-2480; Fax: 27-12-362-5302.

Source: J. Mol. Graphics Modell. 2006, 24(4), 307 – 318.

Compounds: a) S-adenosylmethionine b) Putrescine, a polyamine produced by ODC by decarboxylation from ornithine.

Biological material: Ornithine decarboxylase (ODC, EC 4.1.1.17) and S-adenosylmethionine decarboxylase (AdoMetDC, EC 4.1.1.50) domains in Plasmodium falciparum (PfAdoMetDCare) represented by a single bifunctional protein.

Data taken from the literature:
Crystal [atomic coordinates of the AdoMetDC crystal structure structures of the potato enzyme (2.3 Å, PDB]
entry 1MHM) and the human enzyme irreversibly complexed with the substrate methyl-ester (2.7 Å, PDB entry 1I7B) were taken from the Protein Data Bank;

Protein sequences [28 protein sequences were taken from the Swissprot database].

Computational methods:
Molecular modeling [protein alignment was carried out using Clustalx 1.81, the AdoMetDC domain was modeled using the human and potato X-ray crystal structures as templates, initial models were built with Modeller 6v2 using high refinement to generate 100 models, hydrogens were added for van der Waals and electrostatic interactions with a distance dependent dielectric constant of 4.0, results of modeling were analyzed using Procheck, Ligplot 4.1.1, and Pymol, the DSSP algorithm was used to assign secondary structure as implemented in the DSSPCMBI program, ligand docking was performed using the SA-docking module of InsightII].

Data calculated:
rmse [root mean square deviation (Å) of the position of the corresponding atoms of two superimposed molecular structures].

Results: In P. falciparum, the two main regulatory activities of polyamine biosynthesis, ODC and AdoMetDC, occur in a single bifunctional protein. In this study, novel features of the P. falciparum AdoMetDC domain was modeled using the human template. The AdoMetDC domain was modeled without the two largest inserts, to yield and rmse of 1.85 Å from the human template. Fig. 1 shows the important active site residues in the P. falciparum AdoMetDC, with the ligand AdoMetDC, where Glu72 and Glu438 are acidic residues, Phe5 and Phe415 are aromatic residues, and Cys87 is a polar residue. Hydrogen bonds are highlighted using dashed lines.

Contact with the rest of the bifunctional complex was predicted to occur on one face of the PfAdoMetDC domain. In the active site four substitutions were identified compared to the human template. It was hypothesized that one of these substitutions may be responsible for the lack of inhibition by Tris, compared to mammalian AdoMetDC. The model also gave an explanation for the lack of putrescine stimulation in PfAdoMetDC compared to mammalian AdoMetDC (in some organisms the activity and/or proteolytic processing of AdoMetDC is stimulated by putrescine). It was shown that a network of residues that connects the putrescine-binding site with the active site in human AdoMetDC was conserved in the malarial and plant cognates. Internal basic residues were found to assume the role of putrescine, based on the model and site-directed mutagenesis: the presence of Arg11 was crucial for normal activity, while disrupting Lys15 and Lys215 each cause 50% inhibition of AdoMetDC activity. It was suggested that these novel features of malarial AdoMetDC can guide further experimental investigations that may lead to the discovery of parasite-specific inhibitors.

(B. B.)

85/2006

Title: Design, synthesis and biological activity of acyl substituted 3-amino-5-methyl-1,4,5,7-tetrahydropyrazolo[3,4-b]-pyridin-6-ones as potential hypnotic drugs.

Authors: Falco*, J. L.; Lloveras, M.; Buira, I.; Teixidó, J.; Borrell, J. I.; Méndez, E.; Terencio, J.; Palomer, A.; Guglietta, A.

Medicinal Chemistry Department, R&D Center, Ferrer Internacional S. A.

Juan de Sada 32, E-08028 Barcelona, Spain.

E-mail: jlfalco-research@ferrergrupo.com; Tel.: 34-93-509-3252; Fax: 34-93-411-2764.

Source: Eur. J. Med. Chem. 2005, 40(11), 1179 – 1187.

Compounds:
a) 20 Compounds of type I, where R₁ = Me, (2-furyl)CO, (2-thienyl)CO; R₂ = Me, Phe, p-Me,C₆H₄, p-Cl,C₆H₄, Ph(CH₂)₂, PhNH, p-Me,C₆H₄NH, p-Cl,C₆H₄NH
b) 3 Non-benzodiazepinic hypnotic drugs: zolpidem, zaleplon, alpidem, indiplon
c) [³H]lumazenil.

Biological material:
a) α₁-GABA receptors from male Sprague-Dawley rat cerebellum
b) Caco2 cells.

Data determined:
%Bind [specific binding (%) of the test substance at α₁-GABA receptors in rat cerebellum membranes calculated as %Bind = (X – N/T – N) × 100, where X = amount of ligand for every concentration of compound; T = total binding, maximum amount bound to the radiolabeled ligand; N = non-specific binding, amount of radiolabeled ligand bound in a non-specific way irrespective of the receptor used].
Computational methods:

VolSurf [computer program that automatically converts 3D molecular fields into simpler molecular descriptors related to absorption, distribution, metabolism, excretion (ADME) properties, which are easy to understand and to interpret]; (Lipinski’s rule set postulating that poor oral absorption is probable for a compound that have any two of the following properties: molecular weight above 500 amu, ClogP above 5, and having more than 5 hydrogen bonds donors, or 10 hydrogen bond acceptors present);

“rule of 5” [computer program for automatic generation of pharmacophore models for a training set of molecules specifying the relative alignments and active conformations of the ligands consistent with the binding to a common receptor site, the pharmacophore model (hypothesis) consists of a collection of features necessary for the biological activity of the ligands arranged in 3D space, the common ones being hydrogen bond acceptor, hydrogen bond donor, and hydrophobic features].

Catalyst [computer program for automatic generation of pharmacophore models for a training set of compounds of type I as potential hypnotic drugs were designed and synthesized based on the scaffold shown in Fig. 1.]

Data calculated:

ADME [ADME properties such as Caco2 (Gastro-Intestinal Barrier) and BBB (Blood-Brain Barrier)];

CLOGP (1-octanol/water partition coefficients of the compounds were estimated using the CLOGP algorithm).

Chemical descriptors:

N.O, (number of H bond acceptors and H bond donors, respectively);

NH,OH [molecular weight (g/mol)].

Results: Among the known non-benzodiazepinic hypnotic drugs acting on the γ1 subunit of the GABA-A receptor, zolpidem zaleplon and indiplon show high affinity and selectivity. Following a design methodology including Catalyst pharmacophoric requirements and ADME-predicted properties, a set of compounds of type I as potential hypnotic drugs were designed and synthesized based on the scaffold shown in Fig. 1.

Figure 1

The acyl substituted 3-amino-4,5-dihydro-1H-pyrazolo[3,4-b]pyridin-6(7H)-ones were completely devoid of any inhibitory activity (%Bind < 20%), while Zolpidem showed an inhibitory activity of 99.4% (at a concentration of 1.0·10^{-5} M). This lack of activity could be attributed to two possible reasons: (i) in the pharmacophoric model shown in Fig. 1 Zolpidem shows a hydrophobic interaction of an aryl substituent present in N2 of the pyrazole ring; such interaction, which is not possible for the molecules of type I, could explain the lack of inhibitory activity; (ii) a literature search has revealed that Zaleplon is metabolized by oxidation of the pyrimidine ring to afford an inactive metabolite that presents a lactam carbonyl group at the same relative position of the pyridone carbonyl group of the molecules of type I. It was concluded that the presence of such carbonyl group, the distinctive feature of the applied synthetic methodology, could be responsible of the loss of activity of the novel compounds.

(B. B.)

Multivariate Analysis

86/2006

Title: Structure-based approaches to improve selectivity.

CDK2-GSK3β binding site analysis.

Authors: Vulpetti*, Anna; Crivori, P.; Cameron, A.; Bertrand, J.; Brasca, M. G.; D’Alessio, R.; Pevarello, P.

Nerviano Medical Sciences

Viale Pasteur 10, I-20014 Nerviano (MI), Italy.

E-mail: anna.vulpetti@nervianoms.com; Tel. 39-0331-581-544; Fax 39-0331-581-366.

Source: J. Chem. Inf. Model. 2005, 45(5), 1282 – 1290.

Compounds: 10 Protein ligands: pyrrolopyrazole, thiazole, benzodipryrazole, pyrazole, 3-anilino-4-arylmaleimide, alsterpaullone, staurosporine, benzodipryrazole, AMP-PNP, indirubin-3’-monoxime.

Biological material: 2 Protein kinases: CDK2/cyclin A and GSK3β having a well-conserved ATP binding pocket.

Data determined:

Crystal structure [atomic coordinates of CDK2/cyclin and GSK3β with different inhibitors were taken from the Brookhaven Protein Data Bank (pdb codes are given)];

Crystal structure [atomic coordinates of CDK2/cyclin and GSK3β with different inhibitors were determined by X-ray diffraction techniques].

Computational methods:

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);

GRIND [GRid-INdependent Descriptors were calculated, analyzed, and interpreted using the program ALMOND v3.3, GRINDs are calculated from molecular interaction fields (MIFs) computed on the basis of the GRID force field using different probes, one obtains a set of positions defining a virtual receptor site (VRS), the procedure for obtaining GRIND involves three steps: (i) computing a set of MIF; (ii) filtering the MIF to extract the most relevant nodes;
and (iii) encoding the filtered MIF into the GRIND variables];

GOLPE (Generating Optimal Linear PLS Estimations program using D-optimal design to pre-select non-redundant variables and fractional factorial design to run PLS analyses with different variable combinations);

Probes (OH2 water, DRY hydrophobic probe, C3 sp3 methyl probe, N1 Neutral flat NH, e.g., amide, N1: sp3 N with lone pair, N1 + sp3 amine NH cation, NM3 trimethylammonium cation, OH phenol or carboxy OH, O- sp2 phenolate oxygen, O sp2 carbonyl oxygen);

CPCA (Consensus Principal Component Analysis).

Results: An evaluation and comparison of two different approaches, GRID/CPCA and GRIND/CPCA, that can be used for visualizing the structural differences between related proteins is presented. Ten crystal structures of CDK2/cyclin A and GSK3β complexed with different inhibitors were compared to identify regions that could be potential sites for achieving selectivity for CDK2 versus GSK3β. The study pointed out marked differences in the bottoms of the CDK2-GSK3β ATP binding pockets that guided the optimization to obtain a selective benzodipyrazole CDK2 inhibitor. It had been reported that a bromine substituent at position 6 of indirubin increases the selectivity for GSK3β over CDK2. The superposition of the crystal structure of 6-bromo-indirubin-3′-oxime bound into the GSK3β and the benzo-pyrazole ligand bound into the CDK2 crystal structure is shown in Fig. 1. The attachment points of the bromine and the gem-dimethyl almost overlap. The two moieties occupy the same area of the ATP pocket (which correspond to the NM3 region, previously described).

The gain in selectivity could be associated with two main differences in the ATP pockets of the enzymes. Phe80 of CDK2, the so-called gatekeeper residue that had been often exploited for the design of kinase selective ligands, was replaced by a leucine in GSK3β, and Ala144 was replaced by a cysteine. As a result of these mutations, the CDK2 had a less elongated and less flat buried region at the back of the ATP pocket. The selectivity regions identified using the GRIND/CPCA approach were in good agreement with those obtained using the well-known GRID/CPCA method. The analysis performed using the GRIND descriptors confirmed the presence of favorable hydrophobic areas near the Phe80 of CDK2 that can be fitted optimally by adding two hydrophobic moieties above and below the plane identified by the benzodipyrazole scaffold. Fig. 2 shows the DRY-DRY GRIND variable able to discriminate CDK2-GSK3β proteins. The nodes A and B indicate the position of the favorable hydrophobic areas at a distance of about 3 Å from each other.

![Fig. 1](image1)

Both methods provided insights on structural features that might be useful for improving selectivity against both GSK3β and CDK2 kinases. In particular, the GRID/CPCA approach was instrumental in rationally designing a novel, potent, and selective CDK2 benzodipirazole inhibitor.

(B. B.)

87/2006

Title: MTD-PLS. A PLS variant of the minimal topologic difference method. 3. Mapping interactions between estradiol derivatives and the alpha estrogenic receptor.

Authors: Kurunczi*, L.; Seclaman, E.; Oprea, T. I.; Crisan, L.; Simon, Z.

University of Medicine and Pharmacy “V. Babes” 2 Eftimie Murgu, RO-300041 Timisoara, Romania. E-mail: dick@acad-icht.tmu.edu.ro

Source: J. Chem. Inf. Model. 2005, 45(5), 1275–1281.

Compounds: a) Estradiol and 60 estradiol derivatives (51 training set and 10 test set compounds);

b) 6,7-[3H]estradiol.

Biological material: Uterus estrogen receptor alpha (ERα) from rat (Ratus norvegicus).

Data taken from the literature:

Crystal [atomic coordinates of human ERα LBD complexed with 17α-estradiol were taken from the Brookhaven Protein Data Bank (pdb code: 1ere)].

Data determined:

RBA [relative binding affinities defined as RBA = 100[E]s/[C]o, where [E]s is the unlabeled estradiol concentration that reduces the 6,7-[3H]estradiol saturation of the receptor to 50% and [C]o is the ligand concentration producing the same effect].

Computational methods:

Molecular modeling [the hypermolecule was derived starting with the 3D structure of 17α-estradiol, as modeled with HyperChem v5.11 Pro, conformational analysis was performed only for side chains, atomic partial charges were derived employing the semiempirical AM1 method in HyperChem, fragmental van der Waals volumes were computed using the procedure described by Ohal, fragmental hydrophobicities from Rekker were adapted by introducing reasonable approximations to split some group contributions into atomic ones, the X-ray structure of the human ERR LBD complexed with 17α-es-
tradiol (pdb code: 1ere), CHIME and NCBF (Noncovalent Bond Finder) were used for structure visualization, key ligand receptor interactions, were investigated by superimposing the steroid skeleton of each ligand in the set in the MTD-PLS-selected conformation, to the 17α-estradiol structure from 1ere inside the protein, and then, we the 17α-estradiol was deleted, thus the binding position proposed by MTD-PLS for almost each ligand was investigated without any assumptions regarding protein flexibility; MTD [Minimal Steric (Topological) Difference is a measure of the misfit of the ligand molecule within the receptor site, a hypermolecule is calculated by superposition of the ligands, the vertices of a hypermolecule are obtained by an optimization procedure, the vertices are attributed to three regions: receptor cavity (εi = -1), receptor walls (εi = +1) and exterior (εi = 0), the MTD value for the ligand L is equal to the number of occupied wall vertices, plus the number of unoccupied cavity vertices (Eq. 1),

\[ MTD_L = s + \sum_{j} m_{ij} x_{ij} \]

where \( j \) = vertex enumeration, \( s \) = total number of cavity vertices, \( x_{ij} = 1 \) if molecule \( i \) occupies vertices \( j \) and \( x_{ij} = 0 \), if not;

MTD-PLS (Partial Least Squares projections to latent structures analysis performed SIMCA P v9.0); LOO (Leave-One-Out cross-validation).

Data calculated: [fragmental descriptors used in this work:
- \(-\text{OH}, -\text{O}, -\text{NH}, -\text{S}, -\text{CH}, -\text{F}, -\text{Cl}, -\text{Br},
- \text{I}, \text{N} \text{(nitro),} -\text{NH}_2, -\text{CH}_2, -\text{CH}_3, \text{O} \text{(nitro)},
- \text{C}<, -\text{O}, =\text{O} \text{(keto),}=\text{O} \text{(ester),} =\text{CH}, =\text{C}, =\text{N}-\text{]};

VIP (Variable Influence on Projection); q² (cross-validated correlation coefficient).

Results: MTD-PLS, a PLS variant of the minimal topologic difference method has been used for mapping interactions between estradiol derivatives and the ERα. A set of 45 estrogen agonist derivatives with associated RBA values was used. Fig. 1 shows the hypermolecule construction and vertex numbering, where the common rigid skeleton is used in superposition, and only the differing vertices are numbered.

Using this method spatially assigned analysis of fragment properties can provide receptor site maps, within the limits of the training sets. A steric misfit was identified for the steroidal position 2; Advantageous hydrophobic and van der Waals interactions (enhanced by high polarizability) were found for the 17R-CHdCH-X fragment. The MTD-PLS mapping results were confirmed by the experimentally derived estradiol-estrogen receptor binding site contacts based on X-ray crystallography data. The results suggest that the proposed MTD-PLS method can yield useful insights for interactions with receptors of unknown 3D structure and, generally, for the steric rigidity of receptor sites.

(B. B.)

Data Base Search, Virtual Screening

88/2006

Title: Enhancing the effectiveness of similarity-based virtual screening using nearest-neighbor information.

Authors: Hert, J.; Willett*, P.; Wilton, D. J.; Acklin, P.; Arizona, K.; Jacoby, E.; Schuffenhauer, A.

Krebs Institute for Biomolecular Research and Department of Information Studies, University of Sheffield, Western Bank Sheffield S10 2TN, England.

E-mail: p.willett@sheffield.ac.uk; Tel.: 44-114-222-2633; Fax: 44-114-278-0300.

Source: J. Med. Chem. 2005, 48(22), 7049 – 7054.

Compounds: 102 514 Compounds.

Data taken from the literature:

Datasets (11 sets of bioactive compounds MDL Drug Data Report database including 8294 actives:
- SHT3 antagonists, SHT1A agonists, SHT reuptake inhibitors, D2 antagonists, renin inhibitors,
- angiotensin II AT1 antagonists, thrombin inhibitors, substance P antagonists, HIV protease inhibitors,
- cyclooxygenase inhibitors, protein kinase C inhibitors).

Computational methods:

SIM (similarity-based virtual screening based on similarity search, e.g., using Tanimoto coefficient or fingerprint-based similarity searching, the searches are evaluated by the recall at 5%, i.e., the percentage of the database molecules belonging to the same activity class as the reference structure that is retrieved in the top 5% of a ranking of the database);

TurboSim (similarity-based virtual screening involving the nearest neighbors of the reference structure);

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 samples belongs].

Data calculated:

\[ T_C \] (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: A simple way of enhancing the effectiveness of similarity-based virtual screening by using information about the nearest neighbors (NNs) of the initial target structure in a similarity search is reported. In this study the underlying hypothesis was tested that fusing the outputs of similarity
searches based on a single bioactive reference structure and on its nearest neighbors (of unknown activity) (TurboSim) is more effective (in terms of numbers of high-ranked active structures) than a similarity search involving just the reference structure. The study demonstrated that TurboSim outperforms Sim in simulated virtual screening searches of the MDL Drug Data Report database (activity class, actives, Sim, number of NNs in TurboSim using 5, 10, 20, 50, 100 NNs): 5HT3 antagonists, 752, 31.7, 34.8, 36.8, 38.6, 42.1, 44.0; 5HT1A agonists, 827, 26.3, 28.1, 29.6, 31.8, 34.5, 36.2; 5HT reuptake inhibitors, 359, 21.6, 23.4, 24.0, 23.8, 24.3, 24.1; D2 antagonists, 395, 25.1, 25.8, 26.9, 27.5, 29.1, 30.3; renin inhibitors, 1130, 90.4, 91.2, 92.1, 93.1, 94.3, 94.7; angiotensin II AT1 antagonists, 943, 77.4, 80.8, 83.5, 86.7, 90.2, 92.0; thrombin inhibitors, 803, 44.5, 45.6, 47.1, 48.3, 51.0, 50.7; substance P antagonists, 1246, 28.6, 30.5, 31.7, 32.2, 33.3, 34.1; HIV protease inhibitors, 750, 51.6, 51.9, 52.6, 53.3, 54.5, 55.2; cyclooxygenase inhibitors 636, 13.7, 14.6, 15.0, 15.3, 15.1, 14.4; protein kinase C inhibitors, 453, 21.0, 21.1, 21.1, 20.9, 20.6; average over all classes, 754, 39.2, 40.7, 41.9, 42.9, 44.5, 45.1. Fig. 1 shows the average probability of a compound being active as a function of its rank. The inset shows the left-hand part of the plot (right at the top of the ranking) in greater detail.

It was suggested that the proposed turbo similarity searching approach provides a simple way to enhance the effectiveness of simulated virtual screening searches. The approach requires no modifications to existing similarity software other than the ability to fuse the outputs of the multiple searches to give a single combined ranking of the database structures. In this study 2D fingerprints and the Tanimoto coefficient were used, but any other type of similarity measure can be used that satisfies the similar property principle.

(B. B.)

89/2006

**Title:** Multiple-ligand-based virtual screening. Methods and applications of the MTree approach.

**Authors:** Hessler, G.; Zimmermann, M.; Matter, H.; Evers, A.; Naumann, T.; Lengauer, T.; Rarey*, M.

Center for Bioinformatics, University of Hamburg

Bundesstr. 43, D-20146 Hamburg, Germany.

E-mail: rarey@zbh.uni-hamburg.de; Tel.: 49-40-428-387-351; Fax: 49-40-428-387-351.

**Source:** J. Med. Chem. 2005, 48(21), 6575 – 6584.

**Compounds:** 47 691 Compounds.

**Biological material:**

a) Angiotensin converting enzyme (ACE), a type membrane-anchored dipeptidyl carboxypeptidase that is essential for blood pressure regulation and electrolyte homeostasis

b) α1 adrenergic receptors, that belong to the family of G-protein coupled receptors (GPCRs).

**Data taken from the literature:**

**Dataset** [the data set for model generation and virtual screening was extracted as a drug-like subset from the World Drug Index (WDI), the filtered candidate database contained 47 691 compounds, comprising 331 α1a and 108 ACE inhibitors].

**Computational methods:**

**MTree** [all molecules were described by the established feature tree descriptor, which is derived from a topological molecular graph, a consistent topological molecular alignment based on chemically reasonable matching of corresponding functional groups was performed employing a new pairwise alignment algorithm (theory is given), the MTree models generated were used for retrospective virtual screening in compound libraries];

**Catalyst** [computer program for automatic generation of pharmacophore models for a training set of molecules specifying the relative alignments and active conformations of the ligands consistent with the binding to a common receptor site, the pharmacophore model (hypothesis) consists of a collection of features necessary for the biological activity of the ligands arranged in 3D-space, the common ones being hydrogen bond acceptor, hydrogen bond donor, and hydrophobic features; the virtual screening was performed using the pharmacophores derived by using Catalyst].

**Results:** A novel approach for ligand-based virtual screening is presented by combining query molecules into a multiple feature tree model called MTree. The proposed approach capable of dealing with a set of known actives simultaneously.
Retrospective virtual screening with MTree models generated for angiotensin-converting enzyme and the α1a receptor on a large candidate database. The resulting models were in very good agreement with the available X-ray structural information and known pharmacophores demonstrating the quality and possibility for chemical interpretation in these models. Fig. 1 shows the derived ACE MTree model, where the bold hexagons indicated significant correspondence with potency.

The screening yielded enrichment factors up to 71 for the first 1% of the screened database. MTree models developed outperformed database searches using single feature trees in terms of hit rates and quality. The MTree screening procedures additionally identified alternative molecular scaffolds not included in any of the query molecules. Moreover, the new methodology identified relevant molecular features, that are known to be important for affinity to the target. The methodology allows for introducing weights of individual nodes of the MTree model, giving differing importance to these features. Such weights could be derived on the basis of experimental affinity data.

(B. B.)

Fig. 1 shows the derived ACE MTree model, where the bold hexagons indicated significant correspondence with potency.

Fig. 2 shows the derived α1A MTree model, where the bold hexagons indicated significant correspondence with potency.

Fig. 2