SYBYL QSAR WITH COMFA

Create & Visualize Structure-Activity Relationships That Accurately Predict Ligand Affinity

Quantitative structure-activity relationships (QSARs) relate a molecule's chemical properties or biological activity to its structure in order to design products with increased effectiveness.

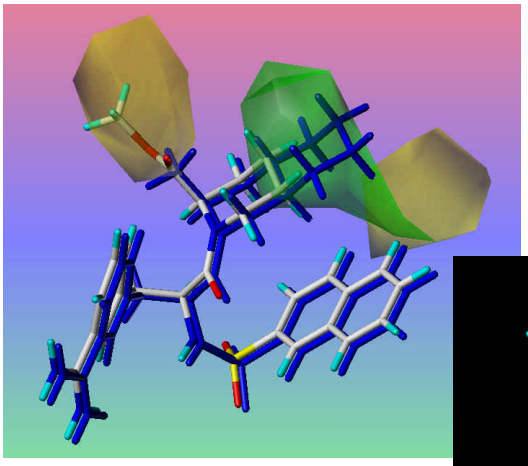
QSAR with CoMFA® provides tools to build statistical and graphical models of activity from molecular structure, and uses these models to make accurate predictions for the activity of untested compounds.

QSAR with CoMFA organizes structures and their associated data into Molecular Spreadsheets, calculates molecular descriptors, and performs sophisticated statistical analyses that reveal patterns in structure-activity data.

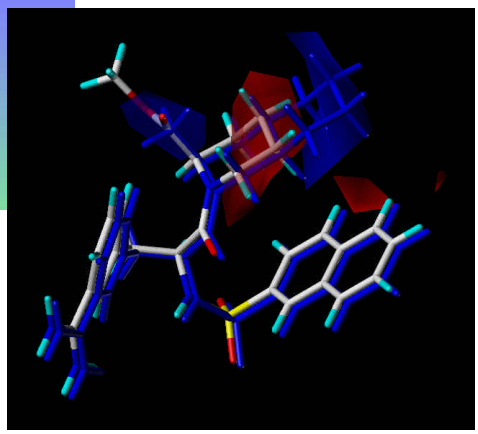
QSAR with CoMFA is fully integrated with SYBYL® to enable visualization and analysis of structure-activity relationships.

Applications

- Develop quantitative structure-activity relationships
- Predict the properties and activities of untested molecules
- Compare different QSAR models statistically and visually
- Optimize the properties of a lead compound
- Validate models of receptor binding sites
- Generate hypotheses about the characteristics of a receptor binding site
- Prioritize compounds for synthesis or screening
- Determine key structural requirements for high affinity receptor ligands



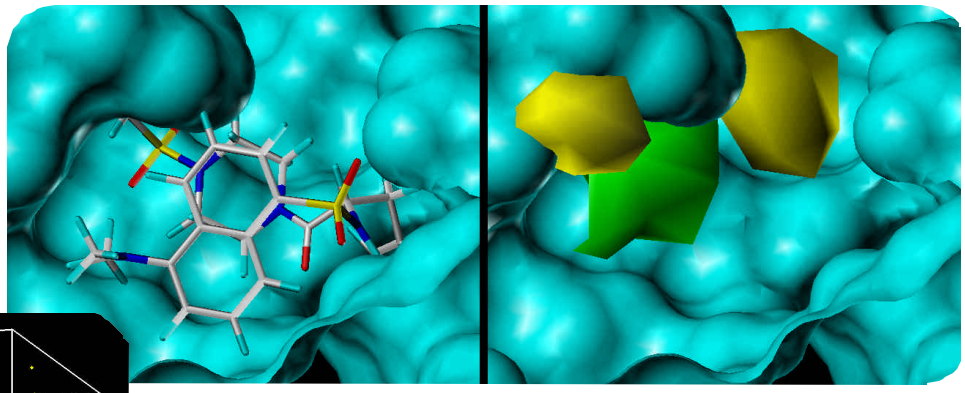
Contour plots from a CoMSIA analysis of thrombin inhibitors. (Left) Regions of favorable steric interactions are shown in green; sterically unfavorable regions are shown in yellow.

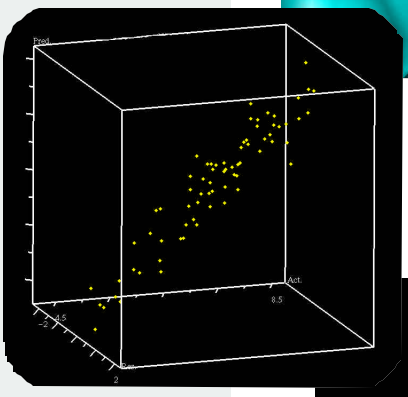


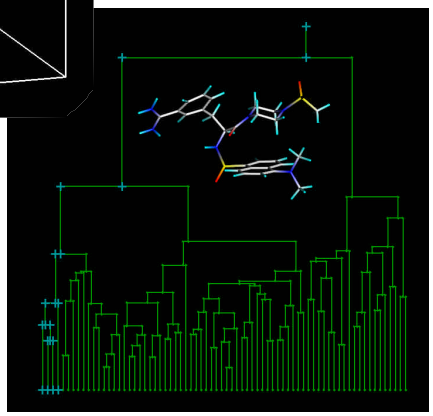
(Right) Blue contours indicate regions where hydrophobic interactions enhance binding; red contours show regions where hydro-phobic properties decrease affinity. These contours were used to design a novel inhibitor[1], displayed in blue, predicted to have ~100x greater affinity. The piperidine ring of the original inhibitor was enlarged to a decaline system in order to occupy regions that favor both steric bulk and hydrophobic groups. The methyl ester was changed to a methyl group to reduce unfavorable steric interactions while still occupying a region favorable for hydrophobic interactions.

Tripos

Create & Visualize QSAR Models







A CoMSIA analysis of thrombin inhibitors. (Above) A MOLCAD solvent-accessible surface of the thrombin binding site. On the left, an inhibitor is positioned in the site based on X-ray coordinates. On the right, CoMSIA contours show regions predicted to prefer steric bulk (green) and other regions (yellow) where steric interactions are unfavorable. The sterically favorable contour lies within the pocket, while the sterically unfavorable contours intersect the surface, confirming the CoMSIA analysis.

(Far Left) A scatter plot showing the predicted versus actual binding affinity for a training set of 72 thrombin inhibitors. The residual for each compound is plotted in the third dimension. (Left) Hierarchical clustering of the thrombin inhibitors in the training set based on CoMSIA descriptors. Clusters can be selected interactively from the dendrogram, and compounds in each cluster can be viewed.

Features

- Multiple statistical methods for generating predictive models, including PCA, PLS, and SIMCA
- Crossvalidation of models for confidence in predictive ability, including SAMPLS
- Hierarchical clustering of compounds based on properties
- Storage of project details and analyses
- Summary statistics
- Interactive graphs that display property distribution, predicted vs actual activity, and residuals

Statistical tools in QSAR with CoMFA include Principal Component Analysis[5] (PCA or Factor Analysis) for uncovering relationships between descriptors, Partial Least Squares[6] (PLS) regression for analyzing continuous response data (IC_{50}, etc.), and Soft Independent Modeling of Class Analogy[7] (SIMCA) for analyzing data that is categorical rather than continuous (i.e., active versus inactive). A hierarchical clustering tool groups compounds into classes having similar properties.

Bootstrapping and crossvalidation techniques are provided to test a model's predictive power, diagnose chance correlation, and insure model robustness. An interface to SAMPLS[8] enables very fast crossvalidation analyses.

Data and results of statistical analyses can be displayed as scatter plots, distributions, or histograms. Graphs, structures, and spreadsheet interact with each other to facilitate exploration of the data. By selecting a row in a spreadsheet or a point in a graph, the corresponding areas in the other displays will be highlighted as well.

The results of CoMFA or CoMSIA analyses are displayed as color-coded contours around molecules, allowing visual identification of regions responsible for favorable or unfavorable interactions with the receptor.

QSAR with CoMFA stores project details, making it possible to regraph, compare analyses, and predict the properties of new compounds. All data can be easily reevaluated if the underlying molecular structures are modified.

Calculate Descriptors, Organize and Store Data

Features

- Extensive set of built-in 2D & 3D descriptors and property calculators, including EVA
- Automatic calculation of CoMFA and CoMSIA molecular fields
- Calculate or import custom descriptors
- Interfaces to external programs for calculating descriptors
- Supports multiple conformers for each molecule
- Automated alignment of structures for 3D QSAR analyses
- SYBYL Molecular Spreadsheet™ for organizing and managing structures, descriptors, and properties

An extensive set of physicochemical descriptors – structural, conformational, geometric, electronic, and thermodynamic – are built into QSAR with CoMFA.

Interfaces to separately licensed programs provide access to other descriptors such as logP and molar refractivity (ClogP/CMR); HOMO or LUMO values (AMPAC); and specialized 2D fingerprints (HQSAR).

Molecular structures, descriptors, and properties are organized and managed within SYBYL's Molecular Spreadsheet. Custom descriptors can be readily imported. Using SYBYL Programming Language, descriptors unique to a project can be calculated and automatically entered into the Spreadsheet.

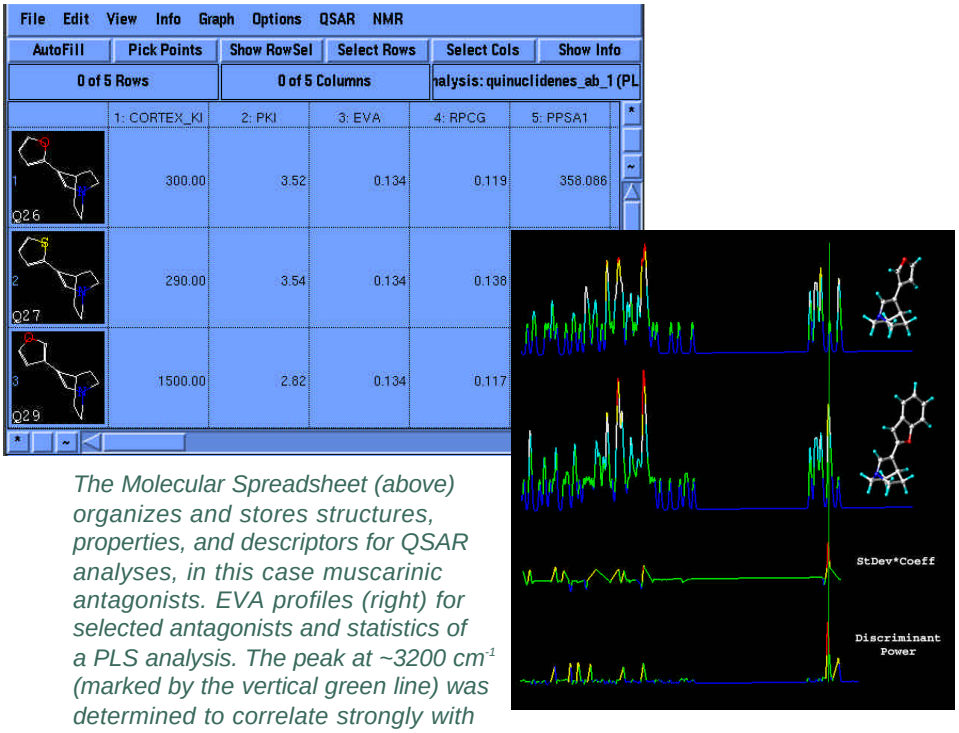
Three-dimensional QSAR methods such as Comparative Molecular Field Analysis[2] (CoMFA) require a set of aligned molecules. QSAR with CoMFA includes methods for automatically aligning molecules. Field Fit optimizes the alignment of molecules to a previously calculated steric or electrostatic field. Alternatively, molecules in a database can be aligned to a template molecule based on a common substructure.

Once a set of molecules is aligned, CoMFA calculates the steric and electrostatic interaction energy of a probe atom with each molecule at points on a grid surrounding the molecules. CoMFA descriptors can be used alone or in conjunction with other descriptors.

Comparative Molecular Shape Indices Analysis[3] (CoMSIA) is similar to CoMFA, but uses a Gaussian function rather than Coulombic and Lennard-Jones potentials to assess steric, electrostatic, hydrophobic, and hydrogen bond donor/acceptor fields.

If the correct conformation of a molecule is not known, multiple conformers can be stored in the Molecular Spreadsheet. This allows alternative conformers to be considered in a CoMFA or CoMSIA analysis.

Eigenvalue (EVA) descriptors[4] are vectors based on eigenvalues corresponding to a molecule's vibrational modes. Like CoMFA and CoMSIA, EVA incorporates 3D information. However, it is not sensitive to molecular alignment and is only slightly sensitive to molecular conformation.



The Molecular Spreadsheet (above) organizes and stores structures, properties, and descriptors for QSAR analyses, in this case muscarinic antagonists. EVA profiles (right) for selected antagonists and statistics of a PLS analysis. The peak at ~3200 cm^{-1} (marked by the vertical green line) was determined to correlate strongly with activity and is prominent in the profile of the benzofuran derivative (second from top), but is sharply attenuated in the less active 3-furanyl analog.

Validation

Tripos' patented CoMFA has been used successfully in hundreds of published QSAR studies.[9]

Acknowledgements

CoMSIA was developed by Professor Gerhard Klebe of the University of Marburg. EVA was developed at Shell Research Ltd.

Hardware and Software Requirements

QSAR with CoMFA requires a separate license in addition to a license for SYBYL/Base. SYBYL and QSAR with CoMFA run on Silicon Graphics R4000 and higher platforms operating under IRIX 6.5 and higher.

Complementary Software

Descriptor Calculation

AMPAC™ (SemiChem, Inc.) for calculating structural and electronic properties of molecules using semiempirical quantum mechanical methods

ClogP/CMR (BioByte Corporation) for calculating partition coefficients and molar refractivity as molecular descriptors

HQSAR™ for calculating novel 2D fingerprints (holograms) as descriptors

Pharmacophore Elucidation and Molecular Alignment

DISCO™ for pharmacophore elucidation from sets of precomputed conformers

FlexS™ for flexible, automatic alignment of molecules

GASP™ for generating pharmacophore hypotheses from a set of ligands with complete conformational flexibility

RECEPTOR™ for refining pharmacophore geometries and molecular alignments

QSAR Tools

Advanced CoMFA® for additional molecular fields, more clustering tools, Region Focusing to increase field resolution, and Progressive Scrambling to validate models

Distill™ for clustering and aligning compounds based on their common substructures

De Novo Design

LeapFrog® for *de novo* ligand design or optimization of a lead compound starting from a CoMFA model

References

1. M. Bohm, J. Sturzebecher, and G. Klebe *Three-Dimensional Quantitative Structure-Activity Relationship Analyses Using Comparative Molecular Field Analysis and Comparative Molecular Similarity Indices Analysis To Elucidate Selectivity Differences of Inhibitors Binding to Trypsin, Thrombin, and Factor Xa* J. Med. Chem. 42:458-477(1999).

2. R.D. Cramer III, D.E. Patterson, and J.D. Bunce *Comparative Molecular Field Analysis (CoMFA) 1. Effect of Shape on Binding of Steroids to Carrier Proteins* J. Am. Chem. Soc. 110:5959 (1988).

3. G. Klebe *Comparative Molecular Similarity Indices: CoMSIA* In 3D QSAR in Drug Design, H. Kubinyi, G. Folkers, and Y.C. Martin (Eds.), Kluwer Academic Publishers, Great Britain 3:87 (1998).

4. A.M. Ferguson, T. Heritage, P. Jonathon, et al. *EVA: A new theoretically based molecular descriptor for use in QSAR/QSPR analysis* J. Comp.-Aided Mol. Des. 11:143 (1997).

5. R.D. Cramer III *BC(DEF) Parameters 1. The Intrinsic Dimensionality of lntramolecular Interactions in the Liquid State* J. Am. Chem. Soc. 102:1837 (1980).

6. R.D. Cramer III, J.D. Bunce, D.E. Patterson, and I.E. Frank *Cross-Validation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies* Quant. Struct-Act. Relat. 7:18-25 (1988).

7. S. Wold and M. Sjostrom *SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy* In Chemometrics: Theory and Application, B.R. Kowalski, (Ed.), American Chemical Society, Washington (1977).

8. B.L. Bush & R.B. Nachbar *Sample-distance Partial Least Squares: PLS optimized for many variables, with application to CoMFA* J. Comp.-Aided Mol. Des. 7:587 (1993).

9. K.H. Kim, G. Greco, and E. Novellino *A Critical Review of Recent CoMFA Applications* In 3D QSAR in Drug Design, H. Kubinyi, G. Folkers, and Y.C. Martin (Eds), Kluwer Academic Publishers, Great Britain 3:257 (1998).

TR 090 03/01DS2001