

# Almond™

RAPIDLY GENERATE, VALIDATE, AND APPLY PHARMACODYNAMIC  
QSAR MODELS WITHOUT ALIGNMENT



Recent studies have demonstrated the importance of evaluating multiple types of descriptors and statistical methods in creating a predictive QSAR model for a set of molecules.<sup>1</sup> Almond<sup>2</sup> offers a new generation of molecular descriptors and a statistical workbench with which to rapidly create and test structure-activity relationship hypotheses.

*"Models obtained with GRid-Independent Descriptors (GRIND) are of comparable statistical quality with respect to models obtained with other 3D-QSAR methodologies, from the point of view of the fitting and predictive power of the models, and are readily interpretable. Moreover, from a chemometric point of view, they are quite robust and less prone to overfitting and other undesirable problems than models obtained with other methods. ... In the field of drug design, GRIND provides a fast and simple way to obtain structure-activity methods, not involving a time-consuming and difficult preliminary step related to compound superimposition."*<sup>2</sup>

Traditional 3D-QSAR methods require a two-step preparation to analyze a set of molecules: 1) identification of the functional groups essential for binding; and 2) superimposition of the molecules based on alignment of these groups. This preparation is the most time consuming aspect of the method, and is a source of user bias. By creating 3D-QSAR models without the need to superimpose molecules, Almond is much faster and more objective yet retains the predictive power of more complicated methods.

Because Almond and its descriptors represent the internal geometrical relationship of pharmacophoric regions, they are useful for predicting pharmacodynamic properties (i.e., specific receptor/ligand interactions). This is in contrast to other descriptors more suited for pharmacokinetic properties such as logP, permeability, solubility, etc.

#### Use novel, alignment-independent descriptors

##### *GRid-INdependent Descriptors (GRIND)*

Starting with a set of 3D structures, Almond employs GRID3 force field to generate Molecular Interaction Fields (MIFs). These fields are created by calculating the interaction energy of one or more probe atoms with each molecule at points on a grid surrounding the molecule.

The information in the MIFs is transformed to generate information-rich descriptors independent of the location of the molecules within the grid. The first step is filtering of the MIFs to remove redundant information. MIF points are retained only if they 1) have low energy values, representing highly favorable interactions; and 2) are as far as possible from each other. These criteria limit MIF points to those representing independent pharmacophoric groups by which the ligand can interact with the receptor.

In the second step, the filtered MIF points are subjected to an auto- and cross-correla-

tion transform. Almond "bins" the distances between pairs of filtered MIF points and computes the product of the interaction energies for each pair. For each distance bin, only the maximum product is stored, along with the spatial position of its two MIF points. The values from auto-correlation of each probe's MIF points and the cross-correlation of different probes' MIF points are represented in correlograms, where the products of the interaction energies are plotted versus the distance separating the points.

In other words, the GRIND features describe compounds in terms of their potential capability to bind and the QSAR model shows what features guide the response.

At the end, a matrix of GRid-INdependent Descriptors (GRIND) is obtained, with each row representing a compound by its associated transformed MIF descriptors. This reduced set of variables represents the geometrical relationships between relevant regions of the (unknown) receptor surface and is independent of the coordinate frame in which the MIF is computed. The whole

process of calculating GRIND takes less than 15 sec/compound on an SGI MIPS R12000 300 MHz or 1.5 sec/compound on a Pentium IV 3 GHz Linux® machine.

##### *Shape Descriptors*

Almond's shape descriptors<sup>4</sup> complement the point-based interaction information provided by GRIND and enable the consideration of shape complementarity in QSAR models. A special shape probe, TIP, is used to extract each molecule's isosurface at 1 kcal/mol from the field of a normal GRID calculation. The local curvature is calculated for all the grid points situated on the isosurface. A transform similar to that used for GRIND is applied: the points are filtered to retain only the most convex; and correlations are calculated. A TIP-TIP autocorrelogram indicates the distance between separated TIP patches. Cross-correlograms between TIP and other probe atoms indicate the distance between specific area of interaction (e.g. a hydrogen-bond acceptor site) and the protruding parts of the molecule.

**Advantages**

- Because no initial alignment is required, Almond's 3D-QSAR analysis can be done in a fraction of the time required by other methodologies.
- GRIND descriptors are alignment-independent, fast to compute, compact to store, and readily interpretable through interactive graphic tools.
- Unlike classic autocorrelation methods, Almond's descriptors allow the original information to be reconstructed, making it possible to interpret both the descriptors and the models generated from them in terms of the molecules in the analysis.
- Almond's risk of generating an overfitted model is less than for other 3D-QSAR methods because the variables-to-compound ratio is much smaller as a result of trimming the number of MIF variables.

**A statistical workbench designed for testing and refining QSAR models**

Almond includes chemometric tools specifically adapted to GRIND: Fractional Factorial Design (FFD) variables selection, Principal Component Analysis (PCA), Partial Least Squares (PLS), and PLS model validation with a variety of cross-validation options. Almond simplifies model refinement through graphical interfaces that enable easy removal of descriptors or molecules either by manual selection or based on statistical design criteria.

A Model Library manager makes it possible to save any model generated by Almond, and to call them up for fully automatic prediction of activity starting directly from the 3D structure of chemicals to be tested.

Almond's statistical workbench makes it a fast, simple to use, and highly flexible research tool for generating and testing hypotheses.

**Visualize descriptors and predictions**

In addition to speed, one of the strengths of the GRIND is their interpretability. Almond's correlograms are ordered representations of the GRIND in which the horizontal axis represents the distance between two grid points and the vertical axis is the product of their associated field energies. These plots are interactive, so that by selecting a point, the corresponding interactions are shown by a line overlaying the molecule.

In a correlogram, the height of a peak expresses the intensity of the field at the two corresponding points. Large interaction regions around the compound are represented by wide peaks; and small interactions are represented by narrow peaks. Correlograms can be displayed individually for every molecule in the set or simultaneously for all molecules. When color-coded by activity, by molecule, or other user-definable rule, a user can visually identify which interactions correlate with activity. Compounds interacting with the same receptor can have some peaks in common, expressing the internal geometrical relationships of receptor

regions with which ligands establish non-covalent bonds.

Almond's statistical models can be examined through a set of 2D and 3D graphs, including: PCA scores; PCA/PLS loadings; predicted vs. experimental activity; residuals vs. experimental activity; and  $r^2$ ,  $q^2$ , SDEC and SDEP as a function of model dimensionality.

**Applications**

The descriptors and statistical tools in Almond make it useful for 3D-QSAR analyses relevant to a wide range of different fields, including:

- Receptor binding site studies<sup>5</sup>
- Quantitative structure-activity relationships<sup>6,7</sup>
- Quantitative structure-metabolism relationships<sup>7,8</sup>
- Designing combinatorial libraries
- Virtual screening
- Lead optimization
- Selectivity studies<sup>9</sup>
- 3D searching of databases<sup>10</sup>
- Pharmacophore identification<sup>11</sup>

**Features**

- Accepts standard molecule file formats as input (mol2 files, SDFfiles, etc.)
- Incorporates GRID for molecular interaction field calculation.
- Can import pre-computed MIFs from a variety of sources.
- No arbitrary limits to the size or to the number of molecules.
- Scaling options for both GRIND and shape descriptors.
- A wide range of probe groups, including those that commonly represent strong non-covalent interactions commonly found in ligand/receptor complexes: hydrophobic (DRY), hydrogen bond acceptors (O), and hydrogen bond donors (N1).
- Extensive interactive visualization capabilities, including molecule, descriptor,

and model displays.

**Complementary Software**

- **VolSurf™** for predicting ADME properties based on pre-calculated models.
- **QSAR with CoMFA®** for constructing predictive structure-activity models from a set of aligned molecules.
- **FlexS™** for shape-based screening in the absence of a receptor structure, and automatic structural alignment of molecules.
- **Legion™/CombiLibMaker™** for building virtual combinatorial libraries in cSLN format.
- **OptDesign™** for designing and editing combinatorial libraries.
- **UNITY®** for rapid, flexible 3D searching to identify lead compounds based on a pharmacophore hypothesis.

**Acknowledgements**

Almond was developed by Prof. Gabriele Cruciani, Dr. M. Pastor, and co-workers at the University of Perugia, Italy.

**Validation**

Almond was tested using two sets of compounds that have been extensively used to validate other 3D-QSAR methods<sup>2</sup>: the glycogen phosphorylase inhibitors used to validate GOLPE; and the corticosteroid-binding globulin receptor ligands used to validate CoMFA. In both cases, Almond yielded a model with fitting and predictive capabilities comparable to those other more time-consuming methods.

**Hardware and Software Requirements**

Almond runs on workstations operating under IRIX® (SGI®) or Linux® (x86). An HTML browser is required to use Almond's online help.

## References

- Kovatcheva, A.; Golbraikh, A.; Oloff, S.; et al. "Combinatorial QSAR of ambergris fragrance compounds." *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 582-95.
- Pastor, M.; Cruciani, G.; McLay, I.; et al. "Grid Independent Descriptors (GRIND). A novel class of alignment-independent three-dimensional molecular descriptors." *J. Med. Chem.* **2000**, *43*, 3233-3243.
- (a) Goodford, P. J. "A computational procedure for determining energetically favourable binding sites on biologically important macromolecules." *J. Med. Chem.* **1985**, *28*, 849-857.  
(b) Carosati, E.; Sciabola, S.; Cruciani, G. "Hydrogen Bonding Interactions of Covalently-Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field." *J. Med. Chem.* **2004**, *47*, 5114-5125.
- Fontaine, F.; Pastor, M.; Sanz, F. "Incorporating molecular shape into the alignment-free GRIND-independent Descriptors (GRIND)." *J. Med. Chem.* **2004**, *47*, 2805-15.
- Gutierrez-de-Teran, H.; Centeno, N. B.; Pastor, M.; Sanz, F. "Novel approaches for modeling of the A1 adenosine receptor and its agonist binding site." *Proteins* **2004**, *54*, (4) 705-15.
- Benedetti, P.; Mannhold, R.; Cruciani, G.; Ottaviani, G. "GRIND/Almond investigations on CysLT1 receptor antagonists of the quinoliny(bridged)aryl type." *Bioorg. Med. Chem.* **2004**, *12*, (13) 3607-17.
- Afzelius, L.; Zamora, I.; Masimirembwa, C. M.; et al. "Conformer- and alignment-independent model for predicting structurally diverse competitive CYP2C9 inhibitors." *J. Med. Chem.* **2004**, *47*, (4) 907-14.
- Crivori, P.; Zamora, I.; Speed, B.; et al. "Model based on GRID-derived descriptors for estimating CYP3A4 enzyme stability of potential drug candidates." *J. Comp.-Aided Mol. Design* **2004**, *18*, (3) 155-66.
- Benedetti, P.; Mannhold, R.; Cruciani, G.; Pastor, M. "GBR compounds and mepyraines as cocaine abuse therapeutics: chemometric studies on selectivity using grid independent descriptors (GRIND)." *J. Med. Chem.* **2002**, *45*, (8) 1577-84.
- Cruciani, G.; Pastor, M.; Mannhold, R. "Suitability of molecular descriptors for database mining. A comparative analysis." *J. Med. Chem.* **2002**, *45*, (13) 2685-94.
- Cruciani, G.; Benedetti, P.; Caltabiano, G.; et al. "Structure-based rationalization of antitumor drugs mechanism of action by a MIF approach." *Eur. J. Med. Chem.* **2004**, *39*, (3) 281-9.



WWW.TRIPOS.COM				CONTACT_US@TRIPOS.COM		
AUSTRALIA +61 (7) 5439 9775	CANADA +1 450 4334500	FRANCE +33 1 69 59 29 49	GERMANY +49 89 45 10 300	JAPAN +81 3 5166 1721	UNITED KINGDOM +44 1 908 650000	UNITED STATES 800-323-2960 +1-314-647-1099
©Tripos, Inc. 2005 All rights reserved. 0075 Printed in USA						