

# Molecular modeling for the BioPrint<sup>®</sup> pharmaco-informatics platform

Vincent Berenz, Fabien Tillier, Frédérique Barbosa, Boryeu Mao, Dragos Horvath, Nicolas Froloff  
Cerep, 19 avenue du Quebec, 91951 Courtaboeuf Cedex, France - tel +33 (0)1 60 92 60 00 - www.cerep.com



BioPrint is a database of a large and homogeneous set of in-house generated experimental data<sup>1,2</sup>: More than 2,500 compounds (marketed drugs, compounds which failed in clinical trials, as well as reference compounds) have been profiled under standardized conditions in more than 180 well-characterized *in vitro* assays (receptors, enzymes, ion channels, cellular function, *in vitro* ADME). This database also contains *in vivo* effects of the compounds (therapeutic indications, mechanisms of action, pharmacokinetics, and adverse drug reactions in man).

This database is combined with proprietary data management software. These Java/Oracle-based applications, the "BioPrint tools", use ChemAxon APIs and Cerep proprietary algorithms to make possible the establishment of relationships between chemical structures and their *in vitro* pharmacology and ADME properties and the determination of *in vitro* pharmacology and/or ADME patterns that correlate with specific biological activities or clinical effects.

## the BioPrint tools

<b>search tools</b>	retrieval of compounds by name, ID (partial or complete), activities on target and/or structure. Display of the results using customizable heat-maps with SDF and Excel file exportabilities.
<b>adverse drug reactions (ADR) tools</b>	searching compound ADRs and ADR-assay associations. These associations are statistically assessed using chi-square significance test between each pharmacology assay and ADR.
<b>profile search tools</b>	retrieval of compounds having similar pharmacological and/or ADME profiles

## ■ QSAR models builder

This interface allows the generation of predictive QSAR models. It is an independent Cerep software designed to be used with any data-set<sup>3,4</sup>.



### Cerep 3D structural descriptors

Cerep descriptors are based on 6 pharmacophoric features: hydrophobicity, aromaticity, hydrogen-bond donor, hydrogen-bond acceptor, positive charge, negative charge.

**Fuzzy bipolar pharmacophoric autocorrelograms (FBPA):** a 3D fingerprint of a molecule built from the distribution of pairwise distances between pharmacophoric features based on multiconformational models.

**Pharmacophore type areas (PTA)** representing molecular areas corresponding to each type of pharmacophoric feature and Generic Polar Surface Area (PSA).

**Field overlap (FOI - EFO)** representing spatial distribution of pharmacophoric features, volume integrals of the pairwise products of local field intensities associated with each of the possible combinations of the pharmacophore types. Refined pharmacophoric features assignment for EFO calculation.

Pharmacophoric features are detected using ChemAxon Screen and 3D conformers are generated using ChemAxon 3D clean. Descriptors are calculated using Cerep proprietary algorithms.

### Project configuration

- Selection of structural descriptors to be used (BioPrint activity values may be used as additional descriptors). Sigmoid and Gaussian transformations also available. (see Cerep 3D structural descriptors)

- Selection of BioPrint structures using activity criterion. Split of the compounds set into a training set and a test set randomly chosen but maintaining similar distribution of activity values.

- Selection of the activity target value: %inhibition, pIC<sub>50</sub> or physico-chemical. Possible specification of activity value range for low/medium/high activity.

### QSAR models generation

At project completion, several models are generated, using different strategies

- Linear QSAR modeling: optimal descriptors selected by genetic algorithm, multi-linear regression.

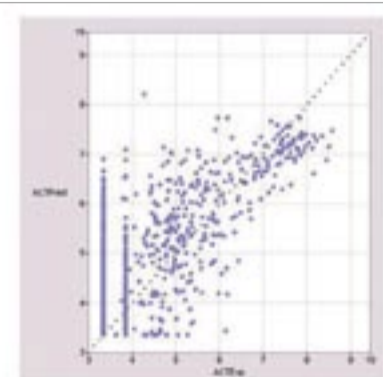
- Neural network modeling: descriptors from linear regression, three-layer fully interconnected neural net.

- Neighborhood behavior modeling: based on similar compounds, use of specific parameters for the similarity metric to predict a specific property, confidence scores on predictions: density and homogeneity.

- Synergy models combining predictions from all these methods.

### Model selection

From these models, one will be selected for future prediction based on various statistical indicators that are available for quality evaluation. For each of the learning and validation sets, a graph displays the comparison of the predicted activity versus the experimental activity.



ID	93026
Experimental Property	8.780
Predicted	8.563
ZDF2_IPC	10.844
ZS03_MONH	10.581
ZS03_AR_HB07	10.248
ZS01_MONH	9.791
ZS03_AR_HB07	9.754
ZS01_SHB	9.665
ZDF1_SH2	9.027
AR_AR3	8.932
ZDF2_IP_HB04	8.462
ZS03_AR_HB04	8.414
ZS03_IP_NC9	8.326
ZDF2_IP_PC8	8.314

**Theofylline HCl**  
drug class - drug  
bioassay number: 7000569-052  
id: 50882574

This graph shows the comparison of experimental activities with predicted activities for the validation set. When the user clicks on a dot, the corresponding structure is displayed along with its descriptors values.

## ■ Full BioPrint profile prediction

General neighborhood behavior (GNB) modeling is an interactive BioPrint tool which estimates the entire activity profile of a compound using the experimentally determined profiles of its nearest neighbors. When predicting a profile, the following steps are performed:

### Step 1: structure selection

The candidate structure is either drawn or loaded from a file. Its pharmacophoric flags are then detected and its FBPA descriptors are calculated. The number of neighbors to base the profile prediction upon is user specified (default: 10).

Pharmacophoric flags	
Hydrogen bond acceptor	14
Aromatic	15 16 17 18 19 20 21 22 23 24 25 26
Positive charge	10
Hydrophob	1 2 3 4 5 6 7 8 9 11 12 13 27
Hydrogen bond donor	10
Negative charge	no atoms

### Step 2: neighbor retrieval & ComPharm superposition

A first-pass screening uses the FBPA descriptors for similarity selection of the best matching BioPrint structures<sup>5,6</sup>. ComPharm overlay scores are calculated on these selected neighbors.

The **ComPharm algorithm**<sup>7</sup> seeks an optimal superposition of pharmacophorically similar functional groups of two chemical structures. Scores are derived from this optimal alignment:

- Overlay score: proportion of pharmacophoric groups of the reference molecule that are not in the candidate molecule.
- Dissimilarity score: both proportions of groups of the reference that are not in the candidate and vice versa.

### Step 3: user specified neighbors selection

Overlays can be visually inspected and either kept or discarded for GNB profile prediction.

selection	STRUCTURE	NAME	Overlay score	Dissimilarity score	BioPrint Number	Drug class
<input type="checkbox"/>		EA2P	0.348	0.312	7000516-55	Non-drug
<input type="checkbox"/>		ZS03_HB07	0.446	0.414	7000521-46	Non-drug
<input type="checkbox"/>		Neomycin	0.459	0.459	7000605-33	Drug

### Step 4: predicted profile visualisation

General neighborhood profile (GNB) profile prediction is a weighted average over the BioPrint experimental profiles of the user-approved neighbors. Results are displayed using heatmap and bar charts display.

**Heat-map display:** the predicted profile is displayed in a table: each column corresponds to an assay. The first line corresponds to the %inhibition predicted values, the second to the pIC<sub>50</sub> values. Furthermore, strong hits are displayed in red cells, medium hits in orange cells and low hits in yellow cells.

**Bar chart display:** the complete profile is displayed in a bar chart, each bar corresponding to a predicted assay value. The primary values chart presents predicted %inhibition values while the secondary values chart presents the predicted pIC<sub>50</sub> values.

## CONCLUSION

- The BioPrint pharmaco-informatics platform contains a dataset of high quality and homogeneous *in vitro* pharmacology and ADME data on a highly diverse collection of drugs and drug-like compounds. This constitutes an ideal training set for benchmarking new molecular descriptors, various QSAR building techniques, as well as *in vitro* profile prediction tools based on the neighborhood behavior concepts in biological space.
- Using Cerep proprietary algorithms and the ChemAxon chemoinformatics toolkit we have developed a collection of user-friendly and web-enabled modeling tools: an independent QSAR application designed to build customizable models and a profile prediction application that uses the general neighborhood behavior (GNB) approach to estimate the entire activity profile of a compound.

## References

- Krejsa, C.M. et al. (2003) *Curr. Op. Drug Disc. & Dev.*, **4**: 470-480
- Froloff, N. et al. (2006) "Construction of a homogeneous and informative *in vitro* profiling database for anticipating the clinical effects of drugs". In: Jacoby, E. (Ed.), "Chemogenomics Knowledge-based Approaches to Drug Discovery", Imperial College Press, London, (in press)
- Gozalbes, R. et al., (2006) "The BioPrint approach for the evaluation of ADMET properties: application to the prediction of cytochrome P450 2D6 inhibition". In: Testa, B., Krämer, S.D., Wunderli-Allenspach, H. & Folkers, G. (Eds.), "Pharmacokinetic profiling in drug research: biological, physicochemical and computational strategies", HCA, Zurich, Wiley-VCH, Weinheim, 395-415
- Rolland, C. et al., (2005) *J. Med. Chem.*, **48**, **21**: 6563-6574
- Horvath, D. et al., (2003) *J. Chem. Inf. & Comp. Sci.*, **43**: 680-690
- Horvath, D. et al., (2003) *J. Chem. Inf. & Comp. Sci.*, **43**: 691-698
- Horvath, D. et al., (2001) "ComPharm - automated comparative analysis of pharmacophoric patterns and derived QSAR approaches, novel tools in high throughput drug discovery: A proof of concept study applied to farnesyl protein transferase inhibitor design". In: Diudea, M. (Ed.), "QSAR/QSAR studies by molecular descriptors", Nova Science Publishers, 395-439