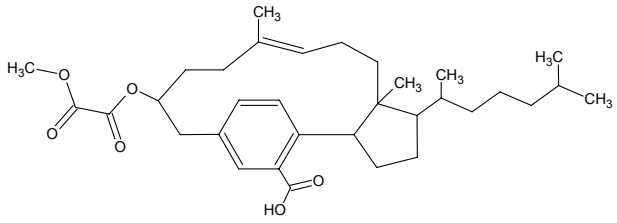
Inverse-QSAR for Inhibitors of Phosphate Cdc25B

Derick Weis and Donald Visco
Tennessee Technological University
Box 5013
Cookeville, TN 38505
{dcweis21, dvisco}@tntech.edu

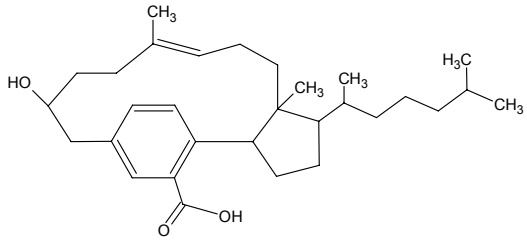
Mike Brown, Shawn Martin, and Jean-Loup Faulon
Sandia National Laboratories
P. O. Box 5800
Albuquerque, NM 87185-0310
{wmbrown, smartin, and jfaulon}@sandia.gov

Quantitative structure-activity relationships (QSARs) provide a description of the correlation between the structure of a molecule and a specific molecular property of interest. QSARs can be employed to refine the search for molecules matching a desired property in an existing database, but ideally one would like to examine potential compounds outside the database through solving the inverse-QSAR (I-QSAR) problem. Signature is a powerful molecular descriptor [1] with previous success in solving the I-QSAR problem. [2,3] Focused libraries of compounds with desired predicted values are created from which a high-quality lead compound can be developed. In this study, we explore the I-QSAR technique to develop inhibitors of phosphate cdc25B.

Cdc25 phosphates play an important role in the growth and development of eukaryotic cells. In humans, there are three types: cdc25A, B, and C. Cdc25B is often over-expressed in various forms of cancer including lung, breast, and prostate. [4] Blocking the cdc25B with chemical inhibitors has the potential to interfere with tumor cells. A database of 28 known cdc25B inhibitors with IC_{50} data was obtained to create a QSAR. Examples of compounds from this database are provided below in Figure 1.



IC_{50} = 11 μM



IC_{50} = 7.9 μM

Figure 1: Sample training set compounds from the Cdc25B database.

The Signature molecular descriptor is a type of topological index that encodes all atoms a pre-defined height h away from the root atom. The unique height 1 atomic Signatures for the cdc25B dataset are given in Table 1. As an example, the first atomic Signature, x1, encodes a carbon atom double bonded to another carbon atom, and two hydrogen atoms. The molecular Signature for a compound is the summation of each atomic Signature multiplied by the occurrence of that atomic Signature in the given compound.

Table1: Height 1 atomic Signature database.

x1	[C](=[C][H][H])	x13	[C]([C][H][H][O])
x2	[C]([C]=[C][H])	x14	[C]([C][O]=[O])
x3	[C]([C][C]=[C])	x15	[C]([C]p[C]p[C])
x4	[C]([C][C]=[O])	x16	[C]([H][H][H][O])
x5	[C]([C][C][C][C])	x17	[C](p[C]p[C][H])
x6	[C]([C][C][C][H])	x18	[C](p[C]p[C][O])
x7	[C]([C][C][C][O])	x19	[H]([C])
x8	[C]([C][C][H][H])	x20	[H]([O])
x9	[C]([C][C][H][O])	x21	[N]([C][C][C])
x10	[C]([C][H]=[O])	x22	[O](=[C])
x11	[C]([C][H][H][H])	x23	[O]([C][C])
x12	[C]([C][H][H][N])	x24	[O]([C][H])

The key feature of Signature is its ability to setup the I-QSAR problem by developing a set of Diophantine (integer coefficients and solutions) constraint equations [3]. These equations are divided in two groups: consistency equations and graphicality equation. Equations 1-8 are referred to as the consistency equations, and are created from the height 1 atomic Signatures. The purpose of the consistency equations is to account for the fact that a bond in one atomic Signature must appear in another atomic Signature, but in the reverse order. For example, consider equation 1 which has two variables (x12 and x21). For a compound to be consistent, there must an equal number of carbon to nitrogen bonds as nitrogen to carbon bonds. For x12, there is a single carbon to nitrogen bond, but for x21 there are three nitrogen to carbon bonds. Thus, three times the occurrence of x21, minus the occurrence of x12 must be zero for consistency.

(1) -x12+3*x21=0

(2) -x20+x24=0

(3) MOD(2*x15+2*x17+2*x18 , 2)=0

(4) MOD(x1+x2+x3 , 2)=0

(5) -x4-x10-x14+x22=0

(6) -x7-x9-x13-x14-x16-x18+2*x23+x24=0

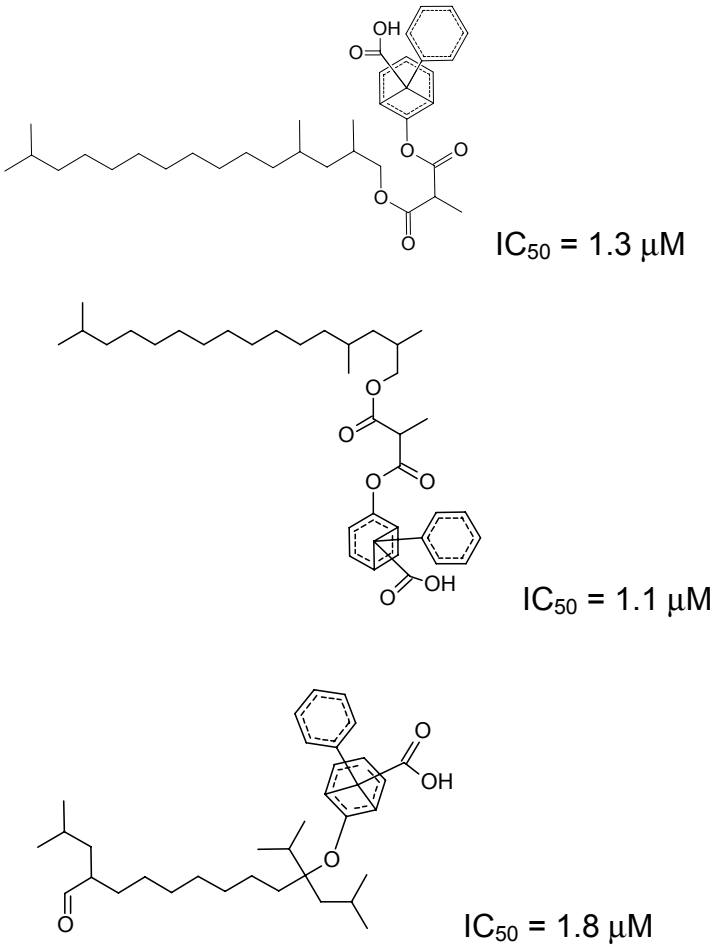
(7) -2*x1-x2-x6-2*x8-x9-x10-3*x11-2*x12-2*x13-3*x16-x17+x19=0

(8) MOD(x2+2*x3+2*x4+4*x5+3*x6+3*x7+2*x8+2*x9+x10+x11
+x12+x13+x14+x15 , 2)=0

The second type of constraint equation is known as the graphicality equation. It is developed from the height 0 atomic Signatures, and is a *necessary* condition for a connected graph. [3] The conditions for satisfying the graphicality equation are that the sum of vertex (root atom) degrees must be even, and the number of vertices of an odd degree must be even. The graphicality equation for the cdc25B dataset is provided in equation 9.

(9) MOD(x1+x2+x3+x4+2*x5+2*x6+2*x7+2*x8+2*x9+x10+2*x11
+2*x12+2*x13+x14+x15+2*x16+x17+x18-x19-x20+x21-x22 , 2)=0

Any solution which satisfies all the constraint equations is evaluated for fitness by means of a QSAR generated on the training set. Therefore, those compounds having a predicted IC_{50} value lower than the strongest inhibitor in the training set comprise a focused database. Note that solving the problem in this manner generates new compounds without merely modifying substituent groups around a scaffold. Accordingly, we report some preliminary candidates in Figure 2 (below). Note that all candidates will be evaluated for energetic stability.



IC_{50} = 1.3 μM

IC_{50} = 1.1 μM

IC_{50} = 1.8 μM

Figure 2: Compounds generated by solving the I-QSAR problem are promising candidates for strong inhibitors of Cdc25B.

A previous project applied Signature to hydrofluoroethers (HFEs) for a polyurethane foam blowing agent replacement. [3] Solving the inverse problem resulted in seven high-quality candidates outside of the original HFE training set. The generic structure for most of the seven compounds had already been proposed as foam blowing agents by 3M in a patent filing. [5] The HFE constraint equations were straightforward to solve due to the saturated bonds, and size (12-24 atoms) of compounds in the training set. As the size and number of bond types for training set compounds increase, the number of atomic Signatures increase. Thus, the number and complexity of constraint equations increase as well. The cdc25B constraint equations were more difficult to solve because the training set compounds were larger (77 to 100 atoms) with aromatic and double bonds present. Accordingly, we developed modified algorithms to handle the solution of the Diophantine equations that were generated. Additional steps in the I-QSAR algorithm were developed to apply the Lipinski Rule of Fives, which is a set of guidelines for pharmaceuticals [6]. Ultimately, we will arrive at a focused database of compounds ready for addition evaluation (stability, toxicity, etc.)

[1] J. F. Faulon, D. P. Visco, Jr. and R. S. Pophale, "The Signature Molecular Descriptor. 1. Extended Valence Sequences vs. Topological Indices in QSAR and QSPR studies", *J. Chem. Inf. Comput. Sci.,* **43**, 707 – 720 (2003).

[2] C. Churchwell, M. D. Rintoul, S. Martin, D. P. Visco, Jr., A. Kotu, R. S. Larson, L.O. Sillerud, D. C. Brown and J. L. Faulon , "The Signature Molecular Descriptor. 3. Inverse Quantitative Structure-Activity Relationship of ICAM-1 Inhibitory Peptides", *J Molecular Graphics and Modelling,* **22**, 263 – 273 (2004).

[3] D. Weis, J. L. Faulon, R. LeBone, D. Visco, "The Signature Molecular Descriptor. 5. The Design of Hydrofluoroether Foam Blowing Agents Using Inverse-QSAR", *Ind. Eng. Chem. Res*, **44**, 8883-8891 (2005).

[4] S. Baurle, T. Blume, J. Gunther, D. Henschel, R. Hillig, M. Husemann, A. Mengel, C. Parchmann, E. Schmid, and W. Skuballa, "Design and synthesis of macrocyclic inhibitors of phosphatase Cdc25B", Bioorg. Med. Chem. Lett., **14**, 1673-1677 (2004).

[5] G. Moore, R. Flynn, M. Guerra, J. Owens, Omega-Hydrofluoroalkyl Ethers, precursor carboxylic acids and derivatives thereof, and their preparation and application. US 6204299 B1, March 20, 2001.

[6] C. Lipinski, ACD/LogP – Rule of 5, http://www.acdlabs.com/products/phys_chem_lab/logp/ruleof5.html. 2005.