JOURNAL OF combinatorial CHEMISTRY



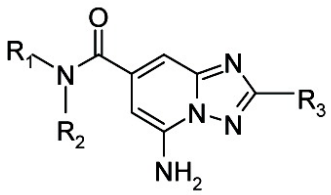
Article

# Ligand-Based Combinatorial Design of Selective Purinergic Receptor (A) Antagonists Using Self-Organizing Maps

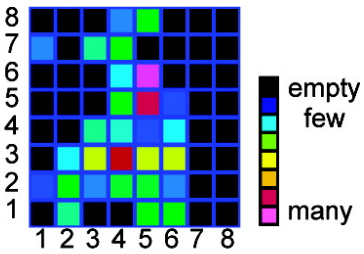
Gisbert Schneider, and Matthias Nettekoven

*J. Comb. Chem.*, **2003**, 5 (3), 233-237• DOI: 10.1021/cc020092j • Publication Date (Web): 27 February 2003






Triazolopyridines **1**




Self-organizing map of Combi-products

## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML



Journal of Combinatorial Chemistry is published by the American Chemical Society.
1155 Sixteenth Street N.W., Washington, DC 20036

# Ligand-Based Combinatorial Design of Selective Purinergic Receptor ($A_{2A}$) Antagonists Using Self-Organizing Maps

Gisbert Schneider† and Matthias Nettekoven*

*F. Hoffmann-La Roche AG, Pharmaceuticals Division, Lead Generation, Grenzacherstrasse 124, CH-4070 Basel, Switzerland*

*Received October 15, 2002*

A virtual screening procedure based on a topological pharmacophore similarity metric and self-organizing maps (SOM) was developed and applied to optimizing combinatorial products functioning as $P_1$ purinergic receptor antagonists. The target was the human $A_{2A}$ receptor. A SOM was developed using a set of biologically tested molecules to establish a preliminary structure−activity relationship. A combinatorial library design was performed by projecting virtually assembled new molecules onto the SOM. A small focused library of 17 selected combinatorial products was synthesized and tested. On average, the designed structures yielded a 3-fold smaller binding constant (∼33 vs ∼100 nM) and 3.5-fold higher selectivity (50 vs 14) than the initial library. The most selective compound obtained revealed a 121-fold relative selectivity for $A_{2A}$ with $K_i$ ($A_{2A}$) = 2.4 nM, and $K_i$ ($A_1$) = 292 nM. This result demonstrates that it was possible to design a small, activity-enriched focused library with an improved property profile using the SOM virtual screening approach. The strategy might be particularly useful in projects in which structure-based design cannot be applied because of a lack of receptor structure information, for example, in the many projects aiming at finding new GPCR modulators.

## Introduction

Novel selective purinergic receptor ($P_1$) antagonists might provide a basis for the treatment of neurodegenerative diseases.[1] The $P_1$ receptor family belongs to the class of G-protein couples receptors (GPCR) comprising the subtypes $A_1$, $A_{2A}$, $A_{2B}$, and $A_3$ that are activated by adenosine as their endogenous ligand.[2,3] Adenosine acts as a neuromodulator, possessing global importance in the modulation of molecular mechanisms underlying many aspects of physiological brain function by mediating central inhibitory effects. An increase in neurotransmitter release follows traumas, such as hypoxia, ischemia, and seizures. These neurotransmitters are ultimately responsible for neural degeneration and neural death, which causes brain damage or death of the individual. The $A_{2A}$ receptor (409−412 amino acids) was cloned from various species and has emerged as a main target within the $P_1$ receptor family.[4] It is preferentially distributed in dopamine-rich brain regions and coexpressed with the $D_2$ receptor, with which it interacts in an antagonistic manner.[5] $A_{2A}$ receptor antagonists inhibit the motor depressant effects of dopamine antagonists, such as haloperidol, which makes them of particular interest for treatment of neurodegenerative disorders, such as Parkinson's disease.[6]

Currently, the application of virtual screening and molecular design approaches relying on a detailed three-dimensional model of the receptor binding pocket, for example, automated molecular docking, is hampered by the fact that for most GPCR, such a model is unavailable.[7] Therefore, we followed an entirely ligand-based approach. The starting point for the project was a set of 153 combinatorial products derived from scaffold structure **1** (Scheme 1) with known $K_i$ values for the two adenosine receptor subtypes $A_{2A}$ and $A_1$.[8,9] The idea was to rely on this information for the development of a preliminary structure−activity relationship (SAR) model, which could be used to virtually synthesize novel combinatorial products with improved activity and selectivity for the $A_{2A}$ subtype. For identification of best-suited building blocks $R_1$, $R_2$, and $R_3$ decorating scaffold **1** (Scheme 1), a new-fangled virtual screening procedure was followed, which is based on artificial neural networks (self-organizing map, SOM)[10,11] and topological pharmacophore similarity (CATS method):[12]

1. Encoding the set of tested compounds by the CATS topological pharmacophore descriptor,
2. Training of a SOM for feature mapping (SAR model),
3. Identification of "seed" compounds from the SOM,
4. Variation of the seed by virtual library enumeration,
5. Projection of the virtual library onto the SOM obtained by step 2,
6. Selection of candidates for synthesis and testing (focused library design),
7. Chemical synthesis and determination of in vitro activity ($K_i$), and
8. Go to step 1 or terminate.

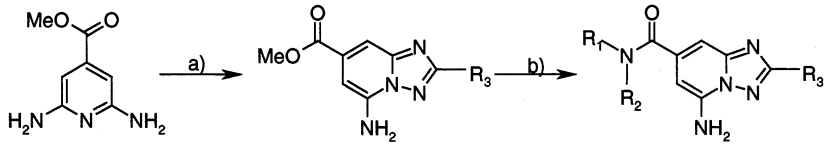
With the exception of step 7 of this scheme, optimization takes place in silico. Synthesis and testing of the designed

* Tel: +41-61 6886227. Fax: +41-61 6886459. E-mail: matthias.nettekoven@roche.com.

† Present address: Institut für Organische Chemie und Chemische Biologie, Johann Wolfgang Goethe-Universität, Marie-Curie-Strasse 11, D-60439 Frankfurt, Germany. Tel: +49-69 79829821. Fax: +49-69 79829826. E-mail: gisbert.schneider@modlab.de.

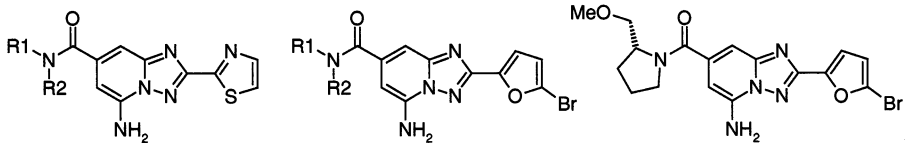
10.1021/cc020092j CCC: $25.00 
Published on Web 02/27/2003

**Scheme 1.** Synthesis of 5-Amino-1-aryl-[1,2,4]triazolo[1,5-a]pyridine-7-carboxylic Acid Amide Derivatives **1**[a]




[a] Reagents and conditions: (a) (i) *O*-mesitylenesulfonylhydroxylamine, (ii) $R_3CHO$, (iii) KOH/MeOH, $O_2$; b) $AlMe_3$, $R_1R_2NH$, dioxane, 90 °C.

**Scheme 2.** Seed Structures Identified for Virtual Synthesis (**4, 5**) and the Most $A_{2A}$-Selective Product Obtained (**6**) Yielding 121-Fold Selectivity versus $A_1$[a]




[a] [$K_i$ ($A_{2A}$) = 2.4 nM, $K_i$ ($A_1$) = 292 nM].

molecules is of great importance, because the results help to adjust the virtual screening process and provide the necessary information enabling selection of molecular building blocks with a high possibility of revealing the desired activity profile. To see whether the procedure can actually be of use in the hit-to-lead optimization, we performed a single optimization round, resulting in an average 3-fold increase of both bioactivity and selectivity compared to the initial combinatorial library.

## Materials and Methods

**(1) Library Synthesis.** Triazolopyridine derivatives have previously demonstrated their ability to potently and selectively bind to the human $A_{2A}$ receptor. Preliminary results were obtained from 5-amino-7-aryl-[1,2,4]-triazolo-[1,5-a]-pyridines.[8] To further investigate the potential of close analogues, 5-amino-1-aryl-[1,2,4]triazolo[1,5-a]pyridine-7-carboxylic acid amide derivatives **1** were synthesized with an analogous protocol (Scheme 1).[9]

The required 2,6-bis-amino-pyridine-4-carboxylic acid methyl ester **2** can be synthesized according to literature procedures,[13] and the subsequent conversion to the 5-amino-1-aryl-[1,2,4]triazolo[1,5-a]pyridine-7-carboxylic acid methyl ester derivatives **3** follows a three step/one pot procedure commencing with the regioselective N-amination of the pyridine nitrogen with *O*-mesitylenesulfonylhydroxylamine. The intermediate can be condensed with a wide variety of aldehydes (substituted benzaldehydes, furfurals, pyridine carboxaldehydes, and thiophene carboxaldehydes), and subsequent cyclization and oxidation under basic conditions yielded triazolo-pyridine methyl ester derivatives **3** in yields from 18 to 51%. Conversion to the desired triazolopyridine carboxylic acid amide derivatives **3** was conveniently performed by heating a mixture of ester **3** with the respective amine, which was premixed with $AlMe_3$, in a solvent-like dioxane for several hours. The purification of the library members was routinely performed by preparative reversed-phase HPLC.

**(2) Biological Testing.** The assessment of in vitro binding data for human $A_{2A}$ and $A_1$ was performed in analogy to procedures described in the literature.[3]

**(3) Virtual Screening. Building Block Retrieval.** Secondary amines for virtual coupling to scaffolds **4** and **5** (Scheme 2) were selected from the Available Chemicals Directory, ACD (Molecular Design Limited, San Leandro, CA). The maximum molecular weight cutoff was 200; amides were excluded. The initial hit list of 487 molecules was further reduced manually to a set of 96 chemically tractable building blocks that were not included in the initial set of 153 products derived from scaffold **1**.

**Molecule Encoding.** For SAR modeling by self-organizing networks, all molecules were represented by the CATS pharmacophore descriptor, which is based on a topological correlation of generalized atom types.[12] The definition of lipophilic, positive and negative charge centers, and hydrogen-bond donors and acceptors followed the LUDI implementation.[14] For five atom types, there are 15 possible pairs. Distance between pairs of atoms was defined as the shortest topological path connecting the two nodes in the molecular graph. Distances up to 10 bonds were considered in the present study, resulting in a 150-dimensional vector description of a molecule giving a relative frequency of atom type pairs over bonds in the molecular graph. Raw counts were scaled by the number of non-hydrogen atoms present in a molecule. This molecular representation is rotation- and translation-invariant, and thus, the problem of pairwise structural alignment was avoided. The CATS descriptor has been shown to be suited for database similarity searching and de novo design in the absence of three-dimensional conformer models.[7,12,15] The idea of the topological pharmacophore space reflects Carhart's concept of atom-pair descriptors.[16]

**Self-Organizing Feature Map (SOM).** For graphical display and feature mapping, the molecule distribution in the 150-dimensional pharmacophore space were projected onto a plane spanned by (8 × 8) 64 neurons (clusters) of a SOM. Kohonen's algorithm was applied to perform the mapping procedure,[10] as implemented in the NEUROMAP package.[11] As a result of this nonlinear mapping procedure, each neuron represents a cluster of molecules having certain features in common, that is, the molecules belonging to its "receptive field" are more similar to each other than to any

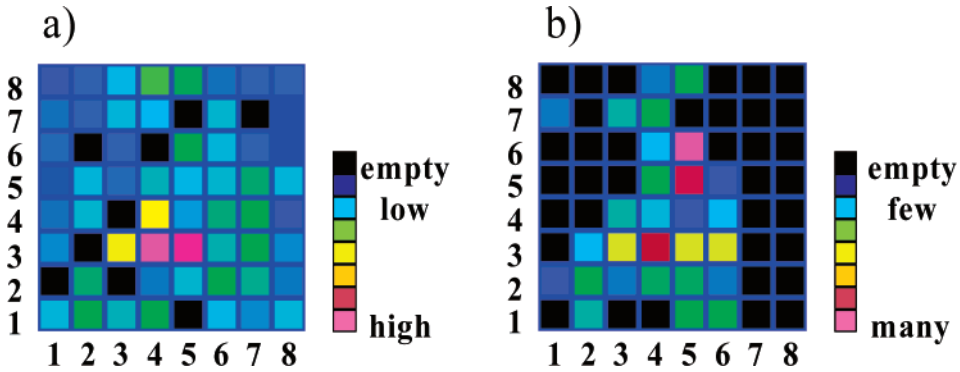



**Figure 1.** Self-organizing maps showing the distribution of selectivity values [$K_i$ ($A_1$)/$K_i$ ($A_{2A}$)] of the initial 153-member library (a), and the 192 virtual combinatorial products generated (b). Seed structure **4** (cf. Scheme 2) was selected from neuron (5/3); seed structure **5** (cf. Scheme 2), from neuron (4/3).

other molecule cluster in the data set. SOM development is comparable to Voronoi tesselation of the high-dimensional data space, reflecting a vector quantization process. For focused library design, virtual combinatorial products were projected onto the SOM, and the molecules falling into the receptive fields of selected neurons (in our project, the neurons (3/3), (4/3), and (5/3) shown in Figure 2 became members of the focused library.

## Results and Discussion

The first step of our virtual screening procedure was molecule encoding. Using the CATS topological pharmacophore descriptor, all available and tested 153 molecules were converted to a 150-dimensional vector representation. Then, a planar (8 × 8) SOM was developed with these data, that is, the distribution of molecules in the 150-dimensional space was projected onto the plane by the SOM algorithm. The projection is topology-preserving, which means that molecules that are close to each other in the two-dimensional projection are also in close vicinity in high-dimensional space. Each of the 64 neurons forms a "receptive field" containing molecules having certain pharmacophore features in common. The averaged selectivity values of the molecules, $K_i$ ($A_1$)/$K_i$ ($A_{2A}$), belonging to each receptive field are represented in a color code in Figure 1a.

The adjacent neurons (3/3), (4/3), (4/4), and (5/3) contain the most selective $A_{2A}$ antagonists, as indicated by yellow and red coloring (Figure 1a). Because of the observation that a small area of pharmacophore space seems to represent characteristic molecular features responsible for relative $A_{2A}$ selectivity, we decided to define neurons (3/3), (4/3), and (5/3) as the "target area" for molecular design. The idea was to virtually generate new compounds using "seed structures" as combinatorial templates that are representatives of the respective fields of these neurons. Seed structure **4** was selected from neuron (5/3); seed structure **5**, from neuron (4/3) (Scheme 2). The thiazole ring of seed **4** and the bromofuran moiety of seed **5** in position $R_3$ of scaffold **1** were the only two substituents found among the set of molecules clustered in neurons (3/3), (4/3), and (5/3). We concluded that these two building blocks might be responsible for the relatively high selectivity of the compounds for the $A_{2A}$ receptor.

Having fixed position $R_3$ in scaffold **1**, the next step was the combinatorial optimization of the secondary amine representing $R_1$ and $R_2$. A set of 96 secondary amines that were not contained in the initial set of 153 products derived from scaffold **1** was retrieved from the ACD. Then, seeds **4** and **5** were assembled with the 96 secondary amines, and each virtual product was encoded by the topological pharmacophore descriptor and projected onto the SOM. The distribution of the compounds is shown in Figure 1b. As expected, the majority are located close to the target area. From the 192 virtual products, 9 fell into the respective field of neuron (3/3), 7 molecules were assigned to neuron (4/3), and 6 compounds fell in neuron (5/3). Of the 22 virtual product molecules, 14 contained the seed structure **5**, and 8 virtual products contained seed structure **4**. From the 22 virtual product molecules, 5 had to be eliminated because of unavailability or structural features considered to cause adverse side reactions under the general reaction protocol of the respective amines. These remaining 17 compounds formed a small focused library that was subsequently synthesized and tested for their activity in $A_{2A}$ and $A_1$ binding assays. Table 1 summarizes the results.

On average, the small library of 17 structures displayed a 3-fold lower binding constant (~33 vs ~100 nM) and 3.5-fold higher selectivity (50 vs 14) than the initial library. The most selective compound, **6**, of the product library has a 121-fold relative selectivity for $A_{2A}$ with $K_i$ ($A_{2A}$) = 2.4 nM, and $K_i$ ($A_1$) = 292 nM. This result demonstrates that it was possible to design a small, activity-enriched focused library with an improved property profile using our virtual screening approach (Table 2). The strategy might be particularly useful for projects in which structure-based design cannot be applied because of a lack of receptor structure information, for example, in the many projects aiming at finding new GPCR modulators.

The idea of using SOMs for molecular feature mapping and virtual screening is not new and has already been applied in a number of successful projects by us and other research groups.[17−19] One particular appeal of the method presented here is the fact that both SAR modeling and focused library design can be performed in one step by nonlinear mapping. Traditional SAR models usually aim at identifying individual parameters that have influence on molecular activity. The SOM approach inherently considers all possible features in

**Table 1.** In Vitro Biological Activity of 17 Triazolopyridine Carboxylic Acid Amides **1** versus Human $A_{2A}$ and Their $A_1$ Selectivity

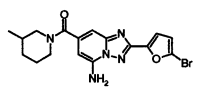
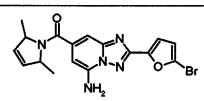
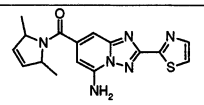
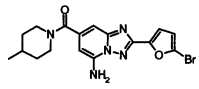
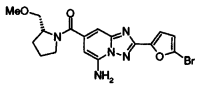
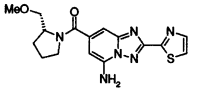
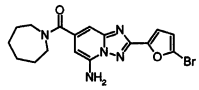
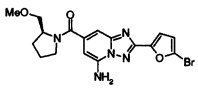
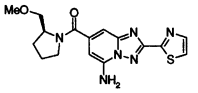
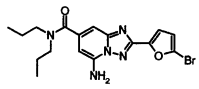
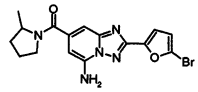
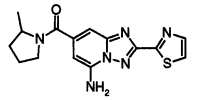
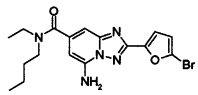
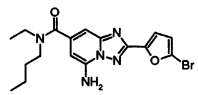
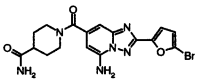
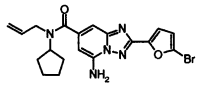
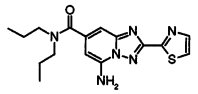
| Neuron (3/3) | | | Neuron (4/3) | | | Neuron (5/3) | | |
|---|---|---|---|---|---|---|---|---|
| | $K_i$ $A_{2A}$ (nM) | Sel. vs. $A_1$ | | $K_i$ $A_{2A}$ (nM) | Sel. vs. $A_1$ | | $K_i$ $A_{2A}$ (nM) | Sel. vs. $A_1$ |
|  | 2.9 | 45 |  | 3.9 | 65 |  | 104 | 19 |
|  | 6.3 | 29 |  | 2.4 | 121 |  | 51 | 66 |
|  | 1.7 | 32 |  | 11.7 | 53 |  | 84 | 53 |
|  | 3 | 65 |  | 4.4 | 65 |  | 32 | 92 |
|  | 3.2 | 40 |  | 89 | 17 | | | |
|  | 75 | 5 | | | | | | |
|  | 1.6 | 44 | | | | | | |
|  | 87 | 31 | | | | | | |

**Table 2.** Average Activity and Selectivity Values of the Initial Compound Library ($N = 153$) and the Designed and Tested Combinatorial Products ($N = 17$)[a]

| | activity ⟨Ki⟩/nM | | |
|---|---|---|---|
| | $A_{2A}$ | $A_1$ | ⟨selectivity⟩ for $A_{2A}$ |
| initial library | 102 (101) | 860 (1154) | 14 (19) |
| focused library | 50 (93) | 974 (1264) | 33 (23) |

[a] Standard deviations in brackets.

parallel, that is, in our case, the 150 elements of the topological pharmacophore descriptor. The SOM uses an adaptive weighing scheme (neuron vectors or centroids of the receptive fields) to account for the individual contributions to the model. Selecting molecules from adjacent neuron clusters to form a focused library is comparable to similarity searching. Advantages we see in the SOM method are (i) the possibility of visually selecting clusters, thereby avoiding strict similarity threshold values; (ii) similarity searching is performed by considering multiple molecules as pharmacophore "seeds";[20,21] and (iii) visual comparison of compound distributions, by which one can get an overview of the library focusing process. The degree of library "diversity", that is, here coverage of topological pharmacophore space, can be influenced by selecting compounds from more or less distant SOM neurons. The method is fast, easily implemented, and applicable to a large variety of medicinal chemistry projects. It was designed to assist the medicinal chemist during the hit-to-lead phase of drug discovery.

Certainly, this method for focused combinatorial library design has limitations regarding the resolution power and is, thus, probably confined to crude activity and selectivity optimization. The activity profile of scaffold **1** could not be further improved during subsequent cycles using the available set of 96 secondary amines selected from ACD (data not shown). For fine-tuning of structures during lead optimization, additional quantitative SAR models are needed, and a significantly augmented set of building blocks for combinatorial exploration of chemical space will be essential. Although the method seems to be restricted to hit-profiling rather than lead optimization, it was demonstrated that our SOM-based virtual screening procedure can be used to rapidly and efficiently construct focused combinatorial libraries and identify products with enhanced properties and significant biological activity.

**Acknowledgment.** The authors are grateful to Drs. A. Alanine, A. Flohr, R. Jakob-Roetne, R. Norcross, and C. Riemer for helpful discussions and for support during the preparation of this manuscript. Technical assistance is gratefully acknowledged from C. Kuratli. This research was supported by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Frankfurt.

## References and Notes


(1) Dunwiddie, T. V.; Masino, S. A. *Annu. Rev. Neurosci.* **2001**, *24*, 31−55.
(2) Klinger, M.; Freissmuth, M.; Nanoff, C. *Cell. Sign* **2002**, *14*, 99−108.
(3) Klotz, K.-N.; Hessling, J.; Hegler, J.; Owman, C.; Kull, B.; Fredholm, B. B.; Lohse, M. J. *Naunyn-Schmied. Arch. Pharmacol*. **1998**, *357*, 1−9.
(4) Muller, C. E.; Scior, T. *Pharm. Acta Helv.* **1993**, *68*, 77−111.
(5) Johansson, B.; Georgiev, V.; Fredholm, B. B. *Neuroscience* **1997**, *80*, 1187−1207.
(6) (a) Svenningsson, P.; Le Moine, C.; Fisone, G.; Fredholm, B. B. *Prog. Neurobiol.* **1999**, *59*, 355−396. (b) Ongini, E.; Monopoli, A.; Cacciari, B.; Baraldi, P. G. *Farmaco* **2001**, *56*, 87−90.
(7) Schneider, G.; Böhm, H.-J. *Drug Discovery Today* **2002**, *7*, 64−70.
(8) Nettekoven, M. *Synlett* **2001**, *12*, 1917−1920.
(9) Brodbeck, B.; Püllmann, B.; Schmitt, S.; Nettekoven, M. *Tetrahedron Lett*. **2003**, in press.
(10) Kohonen, T. *Self-Organization and Associative Memory*; Springer; Heidelberg, 1984.
(11) Schneider, G.; Wrede, P. *Prog. Biophys. Mol. Biol.* **1998**, *70*, 175−222.
(12) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. *Angew. Chem., Int. Ed. Engl*. **1999**, *38*, 2894−2896.
(13) Von Beck, E. R.; Meyer, H. *Monatshefte Chem. verw. Teile and. Wiss*. **1915**, *36*, 731−749.
(14) Böhm, H.-J. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61−78.
(15) Schneider, G.; Clement-Chomienne, O.; Hilfiger, L.; Schneider, P.; Kirsch S.; Böhm, H.-J.; Neidhart, W. *Angew. Chem., Int. Ed. Engl.* **2000**, *39*, 4130−4133.
(16) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64−73.
(17) Sadowski, J. *Curr. Opin. Chem. Biol.* **2000**, *4*, 280−282.
(18) Polanski, J.; Jarzembek, K.; Gasteiger, J. *Comb. Chem. High Throughput Screening* **2000**, *3*, 481−495.
(19) Schneider, G.; So, S.-S. *Adaptive Systems in Drug Design*; Landes Bioscience: Austin, 2002.
(20) Beno, B. R.; Mason, J. S. *Drug Discovery Today* **2001**, *6*, 251−258.
(21) Tropsha, A.; Zheng, W. *Comb. Chem. High Throughput Screening* **2002**, *5*, 111−123.


CC020092J