# Application of Predictive QSAR Models to Database Mining: Identification and Experimental Validation of Novel Anticonvulsant Compounds

Min Shen,[†] Cécile Béguin,[‡] Alexander Golbraikh,[†] James P. Stables,[§] Harold Kohn,*[,†] and Alexander Tropsha*[,†]

*Division of Medicinal Chemistry and Natural Products, School of Pharmacy, CB# 7360, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7360, Department of Chemistry, University of Houston, Houston, Texas 77204-5641, and Epilepsy Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, 6001 Executive Boulevard, Suite 2106, Rockville, Maryland 20852*

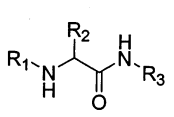
*Received November 20, 2003*

We have developed a drug discovery strategy that employs variable selection quantitative structure-activity relationship (QSAR) models for chemical database mining. The approach starts with the development of rigorously validated QSAR models obtained with the variable selection *k* nearest neighbor (*k*NN) method (or, in principle, with any other robust model-building technique). Model validation is based on several statistical criteria, including the randomization of the target property (*Y*-randomization), independent assessment of the training set model's predictive power using external test sets, and the establishment of the model's applicability domain. All successful models are employed in database mining concurrently; in each case, only variables selected as a result of model building (termed descriptor pharmacophore) are used in chemical similarity searches comparing active compounds of the training set (queries) with those in chemical databases. Specific biological activity (characteristic of the training set compounds) of external database entries found to be within a predefined similarity threshold of the training set molecules is predicted on the basis of the validated QSAR models using the applicability domain criteria. Compounds judged to have high predicted activities by all or the majority of all models are considered as consensus hits. We report on the application of this computational strategy for the first time for the discovery of anticonvulsant agents in the Maybridge and National Cancer Institute (NCI) databases containing ca. 250 000 compounds combined. Forty-eight anticonvulsant agents of the functionalized amino acid (FAA) series were used to build *k*NN variable selection QSAR models. The 10 best models were applied to mining chemical databases, and 22 compounds were selected as consensus hits. Nine compounds were synthesized and tested at the NIH Epilepsy Branch, Rockville, MD using the same biological test that was employed to assess the anticonvulsant activity of the training set compounds; of these nine, four were exact database hits and five were derived from the hits by minor chemical modifications. Seven of these nine compounds were confirmed to be active, indicating an exceptionally high hit rate. The approach described in this report can be used as a general rational drug discovery tool.

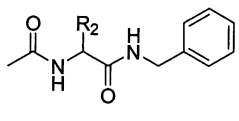
## Introduction

Epilepsy is a chronic disorder, characterized by recurrent unprovoked seizures. In the United States, some 2 million people, including 340 000 children, suffer from epilepsy and its sequelae.[1] Currently the main treatment for epileptic disorder is the long-term and consistent administration of anticonvulsant drugs.[2,3] Unfortunately, current medications are ineffective for more than a third of the patients with epilepsy.[4] Many continue to have seizures, while others experience disturbing side effects (e.g., drowsiness, dizziness, nausea, liver damage).[5] Current therapies have failed to adequately control this disorder, documenting the need for new agents with different mechanisms of action.

Over the years, we have advanced a series of anticonvulsant agents termed functionalized amino acids (**1**,



**1**



**2**

FAA).[6−8] The molecular target(s) of FAA compounds has not been identified. Nearly 250 FAAs have been prepared and evaluated in animal seizure models. Of these, 12 conformed to general structure **2** and provided protection against maximal electroshock (MES) induced seizures,[9,10] at doses comparable to or better than those for phenytoin.[11] The MES test is a proven method for the identification of new drug candidates for partial and generalized seizures. While these studies have provided structural patterns beneficial for FAA activity, it has become increasingly difficult to formulate a useful SAR, as the diversity of molecular structures has increased.

* To whom correspondence should be addressed. Tel. 919-966-2955 (A.T.), 919-966-2680 (H.K.). Fax: 919-966-0204 (A.T.), 919-843-7835 (H.K.). E-mail: alex_tropsha@unc.edu (A.T.), harold_kohn@unc.edu, (H.K.).
[†] University of North Carolina at Chapel Hill.
[‡] University of Houston.
[§] National Institutes of Health.

10.1021/jm030584q CCC: $27.50 
Published on Web 03/20/2004

Molecular modeling can facilitate the understanding of the pharmacological data and permit the development of novel anticonvulsants. Since the structure of macromolecular targets of FAA action remains unknown, ligand-based methods of analysis, such as pharmacophore mapping and quantitative structure−activity relationships (QSAR), represent the most efficient approaches to rationalize available experimental data and improve the design of new FAA.

Several QSAR models for FAA compounds have been developed over the years, including our most recent variable selection *k* nearest neighbor (*k*NN) QSAR models for 48 chemically diverse FAA anticonvulsants (see ref 12 and references therein). These models could be employed for further design and discovery of novel anticonvulsant agents.

One popular computational approach to rational drug discovery is database mining, which relies on the structures of known active molecules as queries.[13] Most often, the queries are derived from three-dimensional (3D) pharmacophores, which entail essential chemical structural features in the particular mutual orientation responsible for the biological activity of a series of compounds.[14] Three-dimensional database mining seeks to identify novel compounds that match the query. If the repository of compound samples as well as relevant biological assays are available, the computational hits can be validated experimentally, ultimately yielding useful lead compounds.[15]

An obvious parallel can be established between the selection of characteristic pharmacophoric features responsible for specific biological activity and the selection of a subset of chemical descriptors implicated in the most statistically significant variable selection QSAR models. This analogy led us earlier to define the latter collection of descriptors as a "descriptor pharmacophore".[16] We proposed that descriptor pharmacophores can be employed in database mining and demonstrated that variable selection QSAR models can be used for both finding molecular structures with the same biological activity as the probe molecules in chemical databases[17] or virtual chemical libraries[18] and predicting the value of the activity. In this paper, we have refined the QSAR-based database-mining approach and employed it for anticonvulsant lead compound discovery.

In a previous publication,[12] we described the development of multiple *k*NN QSAR models for 48 FAA anticonvulsants; the models were extensively validated using several criteria of robustness and accuracy.[19,20] Herein, we describe the application of these models to mining two publicly available chemical databases, i.e., the Maybridge[21] and National Cancer Institute (NCI)[22] collections containing ca. 250 000 compounds combined. Our objective was to identify new lead compounds and experimentally confirm their predicted anticonvulsant activity. The 10 best models with the highest validated predictive power from our previous study[12] have been used for database mining. Twenty-two compounds, which were predicted to have high activities by all or the majority of all models, were chosen as consensus hits. Nine compounds were synthesized and sent to the Anticonvulsant Screening Program (ASP) at the National Institutes of Health (NIH) to test for the anticonvulsant activity; four of these compounds were among the 22 computational hits and five were designed on the basis of two of the hits. Seven out of nine tested compounds were confirmed to be active, indicating an exceptionally high experimental hit rate. The success of this study suggests that our approach, which combines QSAR modeling and database mining, can be explored as a general rational drug discovery tool and applied to a large variety of available datasets of biologically active compounds.

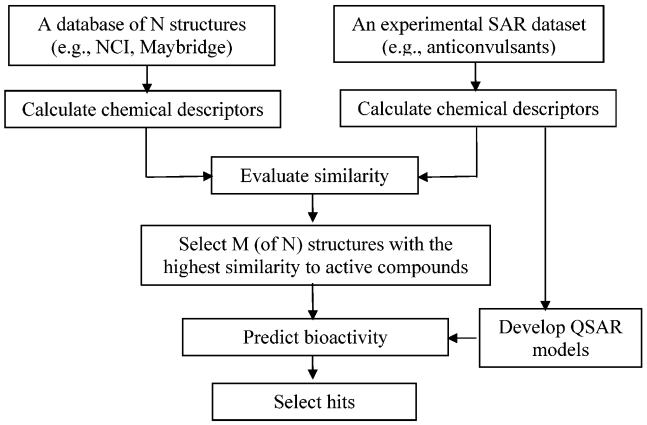



**Figure 1.** Flowchart of database mining that employs predictive QSAR models.

## Computational Details

**Application of QSAR Models to Database Mining.** Our database-mining strategy makes use of validated QSAR models developed for the available series of biologically active compounds. The general flowchart of the database-mining procedure is shown in Figure 1 and includes the following major steps.

1. Develop validated variable selection QSAR models for a dataset of compounds with known structures and activities; define descriptor pharmacophores[16] and applicability domains[19] for all models.

2. Compute chemical descriptors used in QSAR model development for all compounds in the available chemical database(s); in our studies we have used molecular topological indices calculated with the MolConnZ program.[23]

3. Calculate chemical similarity values (we use the Euclidean distance in the descriptor pharmacophore space) between all active probes (i.e., molecules used for QSAR model development) and every structure in the database.

4. Rank all database structures by their similarity to a probe(s), and select *M* structures within certain similarity threshold.

5. Predict biological activity values for these *M* structures based on preconstructed QSAR models using applicability domain.

6. Select structures predicted by all (or a majority of) QSAR models to have high values of biological activity as *computational hits*.

We now discuss individual steps outlined above.

**QSAR Modeling and Descriptor Pharmacophore Identification.** In our previous study of anticonvulsants,[12] we applied two variable selection QSAR algorithms, genetic algorithm (GA) or simulated annealing (SA) partial least squares (GA/SA-PLS)[17,24,25] and *k*NN analyses,[26] to build predictive models for the FAA

dataset. Both of these methods have been shown to produce predictive models for other datasets as well[24−26] and can be easily automatized and adapted to the task of database searching, or virtual screening.[17,18] Both methods typically employ multiple descriptors derived from 2D molecular topology (e.g., molecular connectivity indices or atom pair descriptors), which eliminates the conformational and alignment ambiguities inherent to most 3D QSAR methods. Stochastic optimization algorithms such as GA or SA are used to build robust QSAR models characterized by high values of cross-validated $R^2$ ($q^2$). All models derived in our previous studies[12] were subjected to extensive validation using several important criteria for predictive QSAR modeling. These criteria are discussed extensively in our recent paper;[19] they include the randomization of the target property (*Y-randomization)*, independent assessment of the training set model predictive power using *external* test sets, and the establishment of the model *applicability domain*. All specific details of the calculations can be found in an earlier publication.[12] For this paper, we have used only *k*NN models developed with the MolConnZ descriptors.

As discussed in the Introduction, we have defined descriptor pharmacophores as a subset of descriptor types implicated in successful variable selection QSAR models.[16,17] The similarity searches between known active FAA molecules and compounds in chemical databases were conducted using descriptor pharmacophores found in those variable selection QSAR models that satisfied our rigorous validation criteria.

**Selection of Similarity Probes.** Forty-four compounds out of 48 FAAs with the anticonvulsant activity with $ED_{50}$ less than 100 mg/kg, which is considered promising by the NIH standard, were selected as the similarity probes for database mining. The structures and activity of these 44 compounds were reported in Table 1 of our previous paper;[12] this table is also included in the Supporting Information for reference.

**Molecular Descriptors and Similarity Calculations.** Molecular topological descriptors were calculated with the MolConnZ program[23] for both probes and database molecules. Since absolute scales for the various MolConnZ descriptors can differ by orders of magnitude, range scaling was used to avoid giving descriptors with significantly higher ranges a disproportionably higher weight upon the molecular similarity calculations. Euclidean distance was used as the measure of similarity in the multidimensional descriptor space. The latter distance $d_{ij}$ between any two compounds $i$ and $j$ in $N$-dimensional descriptor space was calculated as

$$d_{ij} = \sqrt{\sum_{n=1}^{N} (X_{in} - X_{jn})^2} \qquad (1)$$

where $X_{in}$ and $X_{jn}$ are the values of $n$th descriptor for compounds $i$ and $j$, respectively, and the summation is over all descriptors. Compounds with the smallest distance (highest similarity) from the active probe were considered as hits and subjected to the prediction of their activity based on QSAR models.

**Applicability Domain of QSAR Models.** Formally, a QSAR model can predict the target property for any compound for which chemical descriptors can be calculated. The training set models are developed in *k*NN QSAR modeling by interpolating activities of the $k$ nearest neighbors of each compound to predict the activity of that compound.[26] This procedure allowed us to derive a special applicability domain (i.e., similarity threshold) specific to each particular *k*NN QSAR model to avoid making predictions for compounds that differ substantially from the training set molecules.[19] In our studies, this threshold, $D_T$ is calculated from the training set models as follows

$$D_T = \bar{y} + Z\sigma \qquad (2)$$

Here, $\bar{y}$ is the average Euclidean distance between each compound and its $k$ nearest neighbors (where $k$ is the parameter optimized in the course of QSAR modeling), $\sigma$ is the standard deviation of these Euclidean distances, and $Z$ is an arbitrary parameter to control the significance level. We set the default value of this parameter $Z$ at 0.5, which formally places the allowed distance threshold at one-half of the standard deviation (assuming a Boltzman distribution of distances between $k$ nearest neighbor compounds in the training set). Thus, if the distance of the external compound from at least one of its nearest neighbors in the training set exceeds this threshold, the prediction is considered unreliable.

**Databases.** Two publicly available databases have been explored: the Maybridge[21] and the NCI[22] databases, including 55 273 and 237 771 chemical structures, respectively. The NCI database was curated; thus, metal-containing compounds as well as compounds with incomplete chemical structures (which could not be processed by MolConnZ) were excluded, leaving 194 736 compounds for database mining. The total collection of compounds subjected to database mining included 250 009 molecules.

## Results and Discussion

**QSAR Modeling of FAA Compounds.** The development of rigorously validated QSAR models for 48 chemically diverse FAA anticonvulsants was reported earlier.[12] Model building included multiple divisions of the original dataset into a training and test set (see ref 27 for details). Multiple QSAR models (760 total) were generated independently for all training sets and validated using the test sets. Generally, we accept models with leave-one-out cross-validation ($q^2$) values for the training set greater than 0.5 and $R^2$ values for predicted vs actual activities of the test set compounds greater than 0.6 as the most significant criteria.[19] Only the 10 best models obtained from multiple *k*NN-QSAR analyses using MolConnZ descriptors in our previous study[12] were used for database mining. The training and test set sizes were 43 and 5, 39 and 9, and 38 and 10 compounds, respectively, and the optimal number of descriptors was in the range between 12 and 20 (see Table 6 in our previous paper;[12] this table is also included as Table 2 in the Supporting Information for reference).

**Database Mining.** As illustrated in Figure 2, validated QSAR models and descriptor pharmacophores derived from these models were used to mine chemical databases for novel lead anticonvulsant agents.

All compounds in the databases within the chosen similarity cutoff value (0.5 Euclidean distance units) of

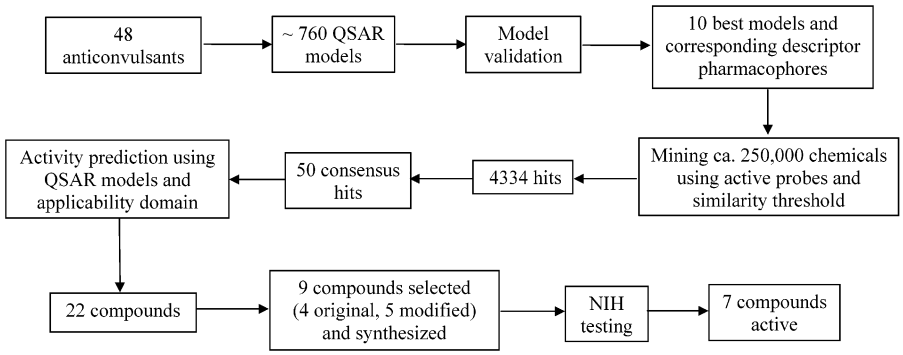


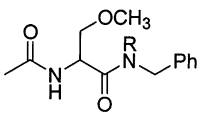

**Figure 2.** Consensus database-mining workflow based on QSAR modeling.

any of the probes were selected, resulting in 4334 initial mining hits total. Because of the differences in the descriptor pharmacophores, the resulting 10 hit lists were not identical. Thus, the initial list was additionally refined by selecting *consensus hits*, i.e., molecules found in all 10 individual hit lists, reducing the initial list to 50 compounds only.

The anticonvulsant activity of these 50 consensus hits was predicted using 10 QSAR models with applicability domains specific to each model. This procedure resulted in the final prediction of biological activity for only 22 compounds found within all 10 individual applicability domains. Table 1 shows the structures and averaged predicted activity of these 22 compounds. Four of these molecules were chosen for synthesis and experimental testing directly, and five compounds were designed as analogues of two of these hits (see below). Table 2 lists the chemical structures and predicted anticonvulsant activities for these nine compounds. Additional details of mining the individual databases and selecting hits for the experimental validation are described below.

**Hits from the NCI Database and Their Experimental Validation.** Application of the 10 best QSAR models to mining the NCI database resulted in 432, 175, 309, 766, 410, 202, 341, 443, 219, and 174 hits, respectively. Only 27 compounds were identified in all 10 searches, and their activities were predicted by all 10 QSAR models. Eleven compounds with consistently high predicted MES values ($ED_{50}$ less than 100 mg/kg) were selected as consensus hits.

The two compounds initially chosen for experimental evaluation featured a fully substituted terminal amide group, a functional unit not anticipated to provide an active compound based on prior experience (cf. compounds **3** and **5** in Table 1).[28] For example, we showed that the amide **30** displayed excellent activity in MES-




**30** R = H
**31** R = $CH_3$

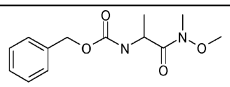
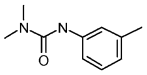
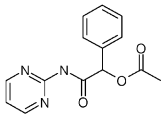
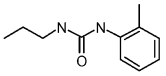
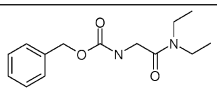
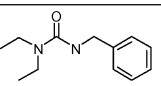
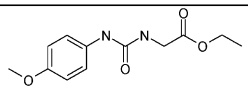
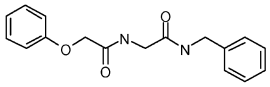
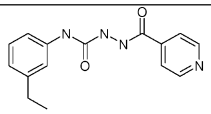
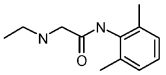
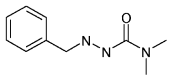
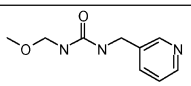
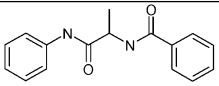
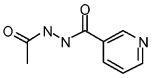
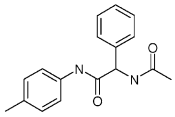
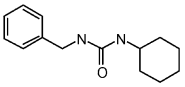
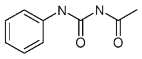
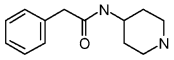
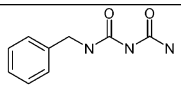
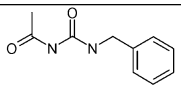
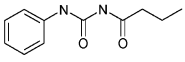
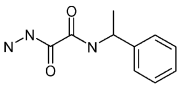
induced seizure test in mice (MES $ED_{50}$ = 8.3 mg/kg), while the corresponding fully *N*-substituted amide **31** exhibited little activity (MES $ED_{50}$ > 100, <300 mg/kg). Another interesting feature of these compounds was the presence of the carbobenzyloxy (Cbz) group, which was absent in any of the training set compounds.[12] The importance of this group was corroborated by independent observations by Geurts et al.,[29] who showed that compounds incorporating the Cbz moiety indeed displayed anticonvulsant activity. Finally, we purposively opted not to prepare compounds that contained the $PhCH_2N(H)$ moiety (compounds **8** and **12** in Table 1) since it has been already shown in FAAs (e.g., **2**) to be beneficial for anticonvulsant activity.[6−8] Several additional compounds (**25**−**29**), which were close analogues of computational hits **3** and **5**, were designed de novo. Where applicable, we prepared the racemic compound mixtures, since we did not know if either isomer would exhibit preferential activity. In total, seven compounds were synthesized and submitted to the NIH's ASP for evaluation using the MES test (a standard test for the anticonvulsant activity, which was used for the training set compounds as well). The results of testing available at this time are shown in Table 2.

The experimentally confirmed anticonvulsant activities for **3**, **5**, and **25**−**29** demonstrate the ability of our database-mining approach to identify molecules with chemical substructures that differed from those observed in the FAA training set. All seven compounds were tested in mice (ip) and rats (po). Each compound was tested at least at three doses, 300, 100, and 30 mg/kg in mice, and one dose, 30 mg/kg, in rats. The biological testing results (Table 2) indicate that upon biological screening in mice, five out of seven tested compounds showed anticonvulsant activity with MES $ED_{50}$ less than 100 mg/kg, which is considered promising by the NIH standard. Interestingly, all seven compounds, when evaluated in the MES test in rats, were active (MES $ED_{50}$ < 52 mg/kg) (experimental data on rats for the entire training set were not available, and therefore, no QSAR models were built).

**Hits from the Maybridge Database and Their Experimental Validation.** Mining of the Maybridge database using the same 10 best QSAR models that were employed in mining the NCI database yielded 156, 89, 127, 104, 93, 170, 138, 182, 113, and 141 hits, respectively. Only 23 compounds were identified in all 10 searches, and their activities were predicted by all 10 QSAR models using the respective applicability domain criteria. Eleven compounds with consistently high predicted MES values ($ED_{50}$ less than 100 mg/kg) were selected as consensus hits.

Two out of 11 consensus hits were synthesized and tested by the NIH, and both compounds were found to be active. Again, we elected not to prepare compounds that contained either $PhCH_2N(H)$ or a structurally related moiety (Table 1: **16**, **17**, **19**, **23**, **24**), because of

**Table 1.** Consensus Hits of Database Mining Found within All Individual Applicability Domains with High Predicted Average Activity Based on the 10 Best *k*NN-QSAR Models

| NCI Entry No. | Cpd. No. | Structure | Predicted Activity[a] ($ED_{50}$, mg/kg) | Maybridge Entry No. | Cpd. No. | Structure | Predicted Activity[a] ($ED_{50}$, mg/kg) |
|---|---|---|---|---|---|---|---|
| 655432[b] | **3** |  | 12.1 | PD00573[b] | **14** |  | 29.6 |
| 527069 | **4** |  | 16.5 | KM09906 | **15** |  | 32.6 |
| 125635[b] | **5** |  | 17.0 | JFD00126 | **16** |  | 35.0 |
| 216931 | **6** |  | 18.4 | JFD02071 | **17** |  | 35.0 |
| 241640 | **7** |  | 20.0 | SEW03864[b] | **18** |  | 37.5 |
| 137956 | **8** |  | 24.7 | CD00731 | **19** |  | 49.0 |
| 43159 | **9** |  | 28.3 | CD04013 | **20** |  | 51.7 |
| 43160 | **10** |  | 31.8 | JFD00125 | **21** |  | 52.0 |
| 60262 | **11** |  | 38.0 | RH00745 | **22** |  | 56.1 |
| 28653 | **12** |  | 45.4 | RJC03822 | **23** |  | 56.7 |
| 131980 | **13** |  | 55.8 | BTB12006 | **24** |  | 62.6 |

[a] Predicted mice MES (maximal electroshock seizure test) $ED_{50}$ value in mg/kg. [b] Database mining hits that were synthesized and sent to NIH for experimental test.

their apparent structural similarity to known FAA anticonvulsants such as **2**. Rather, we chose **14** and **18** as our test compounds, since the earlier SAR study on **2** showed that deletion of the methylene group within the $PhCH_2N(H)$ moiety resulting in a corresponding FAA arylamide led to a loss in anticonvulsant activity.[30] One of the compounds, **14**, exhibited significant anticonvulsant activity, providing MES $ED_{50}$ values between 30 and 100 mg/kg (in mice), and another, **18**, was found to be a highly potent anticonvulsant agent with a MES $ED_{50}$ of 18 mg/kg in mice (ip) that was comparable to the activity of phenobarbital (MES $ED_{50}$ = 22 mg/kg).[31] Both compounds were found to be very active in rats (po) as well, and the potency of **18** (MES $ED_{50}$ = 11 mg/kg) exceeded that of phenytoin (MES $ED_{50}$ = 30 mg/kg)[11] and was comparable to that of phenobarbital (MES $ED_{50}$ = 9.1 mg/kg)[31] (Table 2). An interesting finding is that **18** is a metabolite of lidocaine (lignocaine), which is widely used as a local anesthetic[32] and cardiac antiarrhythmic.[33] In vitro metabolism of lidocaine to **18** is mediated by CYP3A4.[34] Our results indicate that besides known anesthetic activity, **18** is likely to have an anticonvulsant activity as well.

## Conclusions

Our computational strategies afforded the identification of only 22 potentially active anticonvulsant agents that were selected rationally after computational screening of more than 250 000 compounds in two chemical databases. Nine compounds, including four computational hits and five molecules derived from two additional hits by minor chemical modifications, were selected for synthesis and evaluation. Most of these compounds (**3**, **5**, **14**, **18**, **27**-**29**) contained structural units (e.g., fully *N*-substituted amide, *N*-arylamide) that were previously shown to reduce activity in the MES-induced seizure test.[28,30] Seven of the nine compounds

**Table 2.** Results of Anticonvulsant Activity Testing from the Anticonvulsant Screening Project at the National Institutes of Health[g]

| | | | | Mice (ip)[a] | | | Rats (po)[b] | |
|---|---|---|---|---|---|---|---|---|
| | Structure | Cpd. No. | MP[c] | $MES_{exp}$,[d] $ED_{50}$ (mg/kg) | $MES_{calc}$, $ED_{50}$ (mg/kg) | Tox,[e] $TD_{50}$ (mg/kg) | MES,[d] $ED_{50}$ (mg/kg) | Tox,[e] $TD_{50}$ (mg/kg) |
| NCI – based compounds |  [f] | **3** | 63–64 | >100, <300 | 12.1 | >100, <300 | ~30 | >30 |
| |  [f] | **5** | 40–41 | **52 [0.25] (51–53)** | 17.0 | 91 [0.25] (76–114) | ~30 | >30 |
| |  | **25** | 105–106 | **>30, <100** | 64.3 | >100, <300 | <30 | >100 (ip) |
| |  | **26** | 99–100 | >100, <300 | 35.0 | >100, <300 | 52 [1.0] (28–78) | >500 |
| |  | **27** | 57–58 | **43 [0.25] (41–46)** | 41.3 | 130 [0.25] (120-140) | <30 | >30 |
| |  | **28** | 76–77 | **86 [0.25] (79–95)** | 22.4 | 140 [0.25] (130-160) | <30 | >30 |
| |  | **29** | oil | **74 [0.25] (69–79)** | 27.7 | 130 [0.25] (110-150) | <30 | >30 |
| Maybridge |  [f] | **14** | 128–129 | **>30, <100** | 29.6 | >100, <300 | ~30 | >30 |
| |  [f] | **18** | 48–49 | **18 [0.25] (13–24)** | 37.5 | 50 [0.25] (34–69) | 11 | >500 |

[a] The compounds were administered intraperitoneally. $ED_{50}$ and $TD_{50}$ values are in mg/kg. Numbers in parentheses are 95% confidence intervals. The dose effect data was obtained at the "time of peak effect" (indicated in hours in the brackets). Data for most promising compounds are shown in bold. [b] The compounds were administered orally. [c] Melting points (°C) are uncorrected. [d] MES = maximal electroshock seizure test. [e] Tox = neurologic toxicity determined from rotorod test. [f] Hit compounds selected from NCI and Maybridge databases.

were confirmed to be active by experimental studies, indicating an exceptionally high experimental hit rate. Two compounds, **3** and **26**, predicted to have significant MES-induced anticonvulsant activity in mice (ip), were found to be only moderately active ($ED_{50}$ > 100, <300 mg/kg). At this stage, we cannot provide an explanation for this finding but do note that both compounds when tested in rats (po) displayed appreciable anticonvulsant activity (Table 2).

The experimental findings reported in this paper are exciting in light of our long-term studies of FAA anticonvulsants. In 1985, the Kohn laboratory discovered the lead compound *N*-acetyl-D,L-alanine benzylamide (**2**, $R_2 = CH_3$)[35] with a MES $ED_{50}$ of 77 mg/kg in mice (ip), which is comparable with the activity of **28** (Table 2). The structure of this compound was later optimized in the Kohn laboratory using synthetic lead optimization strategies providing (*R*)-*N*-benzyl-2-acetamido-3-methoxypropionamide (**2**, $R_2 = CH_2OCH_3$).[8] This compound exhibited a MES $ED_{50}$ of 4.5 mg/kg in mice (ip) and is currently under phase II clinical studies for the treatment of epilepsy and neuropathic pain.[36] The work is now in progress to utilize comparable synthetic strategies in tandem with computational methods to optimize the structure of compound **28**. A similar approach is being employed for **18** as well.

The traditional approaches to database mining are based on chemical fragment or subfragment similarity searches. While this has been an efficient approach that has enjoyed some successes,[16−18] it limits the chemical diversity of selected compounds to those similar to existing ligands. One of the major goals of our database-mining experiments was to diversify the chemical repertoire of anticonvulsant agents. Our search methodologies were based on chemical similarity estimated by Euclidean distance in the *descriptor pharmacophore* subspace of the whole descriptor space as well as on quantitative predictions from existing QSAR models. Due to the nature of the descriptors (i.e., connectivity indices, which are derived from chemical structures but cannot be easily used in the inverse design experiments to design structures), such searches are more likely to result in the identification of novel compounds than traditional fragment-based search methodologies. The results reported in this paper demonstrate that our chemical similarity searches indeed afford the identification of potent anticonvulsants with chemical substructures that differ from those observed in the training set, such as Cbz, the fully *N*-substituted amide moiety, and the *N*-arylamide group.

The success of the computational and experimental studies reported in this paper lends firm support for all aspects of our computational drug discovery strategies. These strategies have been refined over several years of QSAR modeling research, including methods for model development,[26] training and test set selection,[27] validation,[19,20] and database mining.[16,17] The integration of these techniques and approaches into a comprehensive workflow reported for the first time in this paper (Figures 1 and 2) afforded a very high experimental hit rate of a very small collection of computational hits selected rationally from a very large compound collection. We believe that this enormous reduction of the experimental drug discovery search space is attributed to several important aspects of the workflow and its individual components. First, we place particular em-

phasis on the rigorous validation of QSAR models as well as conservative extrapolation limited to the applicability domain, which allows us to achieve the highest possible accuracy in predicted biological activity of compounds external to the training set. Second, the similarity metric used in database mining is defined by the Euclidean distances in the descriptor pharmacophore space as opposed to the full descriptor space. Third, we select only consensus hits obtained with multiple validated QSAR models as opposed to the predictions based on a single best model.

Future studies should provide additional enhancements of the workflow employed in this paper. Thus, we are currently enlarging the selection of data modeling algorithms and descriptor sets in the context of combinatorial QSAR modeling that we began to explore recently.[37] We suggest that the combined predictive QSAR modeling and database-mining approach developed in this paper can be applied to a variety of experimental datasets and exploited as a general tool for the design and discovery of novel, biologically active compounds.

## Experimental Methods

**General Procedure.** To a cold (−78 °C) THF solution (∼0.1−0.2 mmol *N*-(benzyloxycarbonyl) amino acid/mL of solvent) containing the *N*-(benzyloxycarbonyl)amino acid, the 4-methylmorpholine (1.1 equiv) was slowly added. After stirring (2 min), isobutyl chloroformate (1.1 equiv) was added. The reaction was stirred (2 min) and then the amine (1.1 equiv) was added dropwise. The reaction was stirred at −78 °C (15 min) and allowed to warm to room temperature (1 h). The reaction mixture was filtered and the filtrate evaporated in vacuo. The residue was purified by chromatography on $SiO_2$ gel to obtain the desired product. Using this general procedure, the following compounds were prepared.

**Synthesis of *N*-(Benzyloxycarbonyl)glycine-*N*-methylamide (25).**[29] *N*-(Benzyloxycarbonyl)glycine (5.00 g, 23.90 mmol), 4-methylmorpholine (2.9 mL, 26.29 mmol), isobutyl chloroformate (3.4 mL, 26.29 mmol), methylamine (2.0 M in THF, 13.2 mL, 26.40 mmol), and THF (150 mL) gave crude **25**. The product was purified by column chromatography ($SiO_2$; 1:19, MeOH:$CHCl_3$) to obtain 5.03 g (95%) of pure **25** as a white solid: mp 105−106 °C (lit.[29] mp 105−107 °C); $R_f$ 0.31 (1:19, MeOH:$CHCl_3$); $^1$H NMR ($CDCl_3$) $\delta$ 2.76 (d, $J$ = 4.5 Hz, NC$H_3$), 3.81 (d, $J$ = 5.7 Hz, C$H_2$C(O)), 5.10 (s, OC$H_2$), 5.86 (br t, $J$ = 5.7 Hz, N*H*), 6.48 (br s, N*H*), 7.27−7.33 (m, Ph*H*); $^{13}$C NMR ($CDCl_3$) 26.0 (N*C*H$_3$), 44.5 (*C*H$_2$C(O)), 67.0 (O*C*H$_2$), 127.9 (*C*$_{4'}$), 128.1 (2*C*$_{2'}$ or 2*C*$_{3'}$), 128.4 (2*C*$_{2'}$ or 2*C*$_{3'}$), 136.1 (*C*$_{1'}$), 156.6 (O*C*(O)NH), 169.7 (*C*(O)NH) ppm.

**Synthesis of *N*-(Benzyloxycarbonyl)glycine-*N*-ethylamide (26).** *N*-(Benzyloxycarbonyl)glycine (5.00 g, 23.90 mmol), 4-methylmorpholine (2.9 mL, 26.29 mmol), isobutyl chloroformate (3.4 mL, 26.29 mmol), ethylamine (2.0 M in THF, 13.2 mL, 26.40 mmol), and THF (150 mL) gave crude **26**. The product was purified by column chromatography ($SiO_2$; 1:49, MeOH:$CHCl_3$ to 1:19, MeOH:$CHCl_3$) to obtain 5.07 g (90%) of pure **26** as a white solid: mp 99−100 °C; $R_f$ 0.41 (1:19, MeOH:$CHCl_3$); IR (KBr) 3318, 3051, 2971, 1702, 1655, 1545, 1453, 1373, 1252, 1155, 732, 698 cm$^{-1}$; $^1$H NMR ($CDCl_3$) $\delta$ 1.01 (t, $J$ = 7.2 Hz, CH$_2$C$H_3$), 3.19−3.28 (m, C$H_2$CH$_3$), 3.81 (d, $J$ = 5.7 Hz, C$H_2$C(O)), 5.09 (s, OC$H_2$), 6.16 (t, $J$ = 5.7 Hz, N*H*), 6.76 (br s, N*H*), 7.28−7.32 (m, Ph*H*); $^{13}$C NMR ($CDCl_3$) 14.3 (CH$_2$*C*H$_3$), 34.1 (*C*H$_2$CH$_3$), 44.2 (*C*H$_2$C(O)), 66.7 (O*C*H$_2$), 127.7 (*C*$_{4'}$), 127.9 (2*C*$_{2'}$ or 2*C*$_{3'}$), 128.3 (2*C*$_{2'}$ or 2*C*$_{3'}$), 136.0 (*C*$_{1'}$), 156.6 (O*C*(O)NH), 168.9 (*C*(O)NH) ppm; MS (+CI) (rel intensity) 238 (12), 237 ($M^+$ + 1, 83), 194 (12), 193 (100); $M_r$ (+CI) 237.123 88 [$M^+$ + 1] (calcd for $C_{12}H_{17}N_2O_3$ 237.123 92). Anal. ($C_{12}H_{16}N_2O_3$) C, H, N.

**Synthesis of *N*-(Benzyloxycarbonyl)glycine-*N*,*N*-dimethylamide (27).** *N*-(Benzyloxycarbonyl)glycine (5.00 g, 23.90 mmol), 4-methylmorpholine (2.9 mL, 26.29 mmol), isobutyl chloroformate (3.4 mL, 26.29 mmol), dimethylamine (2.0 M in THF, 13.2 mL, 26.40 mmol), and THF (150 mL) gave crude **27**. The product was purified by column chromatography ($SiO_2$; 1:49, MeOH:$CHCl_3$) to obtain 4.24 g (75%) of pure **27** as a pale green solid: mp 57−58 °C; $R_f$ 0.36 (1:19, MeOH:$CHCl_3$); IR (KBr) 3261, 3052, 2929, 1725, 1643, 1546, 1407, 1248, 1170, 1045, 990, 741, 647 cm$^{-1}$; $^1$H NMR ($CDCl_3$) $\delta$ 2.96 (s, C$H_3$), 2.98 (s, C$H_3$), 4.01 (d, $J$ = 4.5 Hz, C$H_2$C(O)), 5.13 (s, OC$H_2$), 5.93 (br s, N*H*), 7.30−7.38 (m, Ph*H*); $^{13}$C NMR ($CDCl_3$) 35.4 (*C*H$_3$), 35.6 (*C*H$_3$), 42.5 (*C*H$_2$C(O)), 66.6 (O*C*H$_2$), 127.8 (*C*$_{4'}$), 127.9 (2*C*$_{2'}$ or 2*C*$_{3'}$), 128.3 (2*C*$_{2'}$ or 2*C*$_{3'}$), 136.4 (*C*$_{1'}$), 156.1 (O*C*(O)NH), 167.7 (*C*(O)NH) ppm; MS (+CI) (rel intensity) 238 (10), 237 ($M^+$ + 1, 75), 194 (14), 193 (100), 154 (31), 129 (67); $M_r$ (+CI) 237.123 88 [$M^+$ + 1] (calcd for $C_{12}H_{17}N_2O_3$ 237.123 92). Anal. ($C_{12}H_{16}N_2O_3$) C, H, N.

**Synthesis of *N*-(Benzyloxycarbonyl)glycine-*N*,*N*-diethylamide (5).** *N*-(Benzyloxycarbonyl)glycine (5.00 g, 23.90 mmol), 4-methylmorpholine (2.9 mL, 26.29 mmol), isobutyl chloroformate (3.4 mL, 26.29 mmol), diethylamine (3.0 mL, 26.29 mmol), and THF (150 mL) gave crude **5**. The product was purified by column chromatography ($SiO_2$; 3:2, EtOAc:hexanes) to obtain 5.68 g (90%) of pure **5** as a translucent solid: mp 40−41 °C; $R_f$ 0.31 (3:2, EtOAc:hexanes); IR (KBr) 3268, 3066, 2986, 2940, 1726, 1642, 1550, 1454, 1255, 1167, 1049, 1006, 743, 653 cm$^{-1}$; $^1$H NMR ($CDCl_3$) $\delta$ 1.13 (t, $J$ = 7.2 Hz, CH$_2$C$H_3$), 1.19 (t, $J$ = 7.2 Hz, CH$_2$C$H_3$), 3.26 (q, $J$ = 7.2 Hz, C$H_2$CH$_3$), 3.40 (q, $J$ = 7.2 Hz, C$H_2$CH$_3$), 4.02 (d, $J$ = 4.2 Hz, C$H_2$C(O)), 5.13 (s, OC$H_2$), 5.85 (br s, N*H*), 7.30−7.37 (m, Ph*H*); $^{13}$C NMR ($CDCl_3$) 12.7 (CH$_2$*C*H$_3$), 13.7 (CH$_2$*C*H$_3$), 40.2 (*C*H$_2$CH$_3$), 40.7 (*C*H$_2$CH$_3$), 42.3 (*C*H$_2$C(O)), 66.5 (O*C*H$_2$), 127.7 (*C*$_{4'}$), 127.8 (2*C*$_{2'}$ or 2*C*$_{3'}$), 128.2 (2*C*$_{2'}$ or 2*C*$_{3'}$), 136.4 (*C*$_{1'}$), 156.0 (O*C*(O)NH), 166.7 (*C*(O)NH) ppm; MS (+CI) (rel intensity) 266 (17), 265 ($M^+$ + 1, 100), 221 (32), 157 (16), 154 (11); $M_r$ (+CI) 265.154 61 [$M^+$ + 1] (calcd for $C_{14}H_{21}N_2O_3$ 265.155 22). Anal. ($C_{14}H_{20}N_2O_3$) C, H, N.

**Synthesis of *N*-(Benzyloxycarbonyl)glycine-*N*-methoxy-*N*-methylamide (28).**[38] *N*,*O*-Dimethylhydroxylamine was prepared by stirring *N*,*O*-dimethylhydroxylamine hydrochloride salt (1.57 g, 16.13 mmol), $K_2CO_3$ (4.46 g, 32.2 mmol), THF (30 mL), and $H_2O$ (3 mL) at room temperature (3 h). *N*-(Benzyloxycarbonyl)glycine (2.81 g, 13.44 mmol), 4-methylmorpholine (1.6 mL, 14.78 mmol), isobutyl chloroformate (1.9 mL, 14.78 mmol), the previously prepared *N*,*O*-dimethylhydroxylamine solution (33 mL), and THF (70 mL) gave crude **28**. The product was purified by column chromatography ($SiO_2$; 3:2, EtOAc:hexanes) and then crystallized from EtOAc (15 mL) to obtain 2.41 g (71%) of pure **28** as a translucent solid: mp 76−77 °C (lit.[38] mp 76−77 °C); $R_f$ 0.31 (3:2, EtOAc:hexanes); IR (KBr) 3291, 1725, 1661, 1545, 1469, 1252, 1169, 977, 741 cm$^{-1}$; $^1$H NMR ($CDCl_3$) $\delta$ 3.20 (s, NC$H_3$), 3.71 (s, OC$H_3$), 4.14 (d, $J$ = 4.8 Hz, C$H_2$C(O)), 5.12 (s, OC$H_2$), 5.61 (br s, N*H*), 7.30−7.38 (m, Ph*H*); $^{13}$C NMR ($CDCl_3$) 32.0 (N*C*H$_3$), 41.7 (*C*H$_2$C(O)), 61.0 (O*C*H$_3$), 66.4 (O*C*H$_2$), 127.6 (*C*$_{4'}$), 127.7 (2*C*$_{2'}$ or 2*C*$_{3'}$), 128.1 (2*C*$_{2'}$ or 2*C*$_{3'}$), 136.2 (*C*$_{1'}$), 156.1 (O*C*(O)NH), 169.5 (*C*(O)NH) ppm; MS (+CI) (rel intensity) 253 ($M^+$ + 1, 62), 210 (12), 209 (100), 192 (14), 145 (37), 119 (12); $M_r$ (+CI) 253.119 59 [$M^+$ + 1] (calcd for $C_{12}H_{17}N_2O_4$ 253.118 83). Anal. ($C_{12}H_{16}N_2O_4$) C, H, N.

**Synthesis of (*R*,*S*)-*N*-(Benzyloxycarbonyl)alanine-*N*-methoxy-*N*-methylamide ((*R*,*S*)-3).** *N*,*O*-Dimethylhydroxylamine was prepared by stirring *N*,*O*-dimethylhydroxylamine hydrochloride salt (1.57 g, 16.13 mmol), $K_2CO_3$ (4.46 g, 32.2 mmol), THF (30 mL), and $H_2O$ (3 mL) at room temperature (3 h). (*R*,*S*)-*N*-(Benzyloxycarbonyl)alanine (3.00 g, 13.44 mmol), 4-methylmorpholine (1.6 mL, 14.78 mmol), isobutyl chloroformate (1.9 mL, 14.78 mmol), the previously prepared *N*,*O*-dimethylhydroxylamine solution (33 mL), and THF (70 mL) gave crude **3**. The product was purified by column chromatography ($SiO_2$; 1:49, MeOH:$CHCl_3$) to obtain 3.16 g (88%) of pure **3** as an off-white solid: mp 63−64 °C; $R_f$ 0.42 (1:49, MeOH:$CHCl_3$); IR (KBr) 3267, 1709, 1650, 1532, 1254, 1071, 981, 740 cm$^{-1}$; $^1$H NMR ($CDCl_3$) $\delta$ 1.32 (d, $J$ = 6.9 Hz, C$H_3$CH), 3.18 (s, NC$H_3$), 3.74 (s, OC$H_3$), 4.73 (dq, $J$ = 8.1, 6.9 Hz,

$CH_3CH$), 5.03−5.13 (m, $OCH_2$), 5.84 (br d, $J$ = 8.1 Hz, N*H*), 7.28−7.34 (m, Ph*H*); $^{13}$C NMR ($CDCl_3$) 18.2 ($CH_3CH$), 31.9 ($NCH_3$), 46.9 ($CH_3CH$), 61.3 ($OCH_3$), 66.4 ($OCH_2$), 127.7 ($C_{4'}$), 127.8 ($2C_{2'}$ or $2C_{3'}$), 128.2 ($2C_{2'}$ or $2C_{3'}$), 136.3 ($C_{1'}$), 155.5 (O*C*(O)NH), 173.0 (*C*(O)NH) ppm; MS (+CI) (rel intensity) 268 (14), 267 ($M^+ + 1$, 100), 223 (42), 206 (38), 159 (26); $M_r$ (+CI) 267.135 15 [$M^+ + 1$] (calcd for $C_{13}H_{19}N_2O_4$ 267.134 48). Anal. ($C_{13}H_{18}N_2O_4$) C, H, N.

**Synthesis of (*R,S*)-*N*-(Benzyloxycarbonyl)alanine-*N,N*-diethylamide ((*R,S*)-29).** (*R,S*)-*N*-(Benzyloxycarbonyl)alanine (2.80 g, 12.54 mmol), 4-methylmorpholine (1.5 mL, 13.79 mmol), isobutyl chloroformate (1.8 mL, 13.79 mmol), diethylamine (1.4 mL, 13.79 mmol), and THF (100 mL) gave crude **29**. The product was purified by column chromatography ($SiO_2$; 1:49, MeOH:$CHCl_3$) to obtain 2.33 g (67%) of pure **29** as a clear oil: $R_f$ 0.26 (1:49, MeOH:$CHCl_3$); IR (KBr) 3410, 3282 (br), 2978, 1715, 1639, 1528, 1455, 1247, 1061, 745 cm$^{-1}$; $^1$H NMR ($CDCl_3$) $\delta$ 1.12 (t, $J$ = 7.2 Hz, $CH_2CH_3$), 1.24 (t, $J$ = 7.2 Hz, $CH_2CH_3$), 1.33 (d, $J$ = 6.9 Hz, $CH_3CH$), 3.20−3.56 (m, 2 $CH_2CH_3$), 4.62 (dq, $J$ = 7.8, 6.9 Hz, $CH_3CH$), 5.05−5.14 (m, $OCH_2$), 5.78 (br d, $J$ = 7.8 Hz, N*H*), 7.30−7.36 (m, Ph*H*); $^{13}$C NMR ($CDCl_3$) 12.7 ($CH_2CH_3$), 14.3 ($CH_2CH_3$), 19.4 ($CH_3CH$), 40.1 ($CH_2CH_3$), 41.5 ($CH_2CH_3$), 46.6 ($CH_3CH$), 66.4 ($OCH_2$), 127.7 ($C_{4'}$), 127.8 ($2C_{2'}$ or $2C_{3'}$), 128.3 ($2C_{2'}$ or $2C_{3'}$), 136.4 ($C_{1'}$), 155.4 (O*C*(O)NH), 171.5 (*C*(O)NH) ppm; MS (+CI) (rel intensity) 280 (18), 279 ($M^+ + 1$, 100), 117 (17); $M_r$ (+CI) 279.170 73 [$M^+ + 1$] (calcd for $C_{15}H_{23}N_2O_3$ 279.170 87). Anal. ($C_{15}H_{22}N_2O_3 \cdot 0.3H_2O$) C, H, N.

**Synthesis of *N*-(Ethyl)glycine-*N*-2,6-dimethylanilide (18).**[39] 2-Chloro-*N*-(2,6-dimethylphenyl)acetamide[40] (3.00 g, 15.19 mmol) was added to a THF solution containing $EtNH_2$ (2.0 M in THF, 20 mL, 40 mmol), and then the solution was stirred at 50 °C (18 h). The solvent was evaporated and the residue was purified by crystallization (hexanes) to obtain 2.20 g (70%) of pure **18** as a white solid: mp 48−49 °C (lit.[39] mp 49−50 °C); $^1$H NMR ($CDCl_3$) $\delta$ 1.41 (t, $J$ = 7.1 Hz, $CH_2CH_3$), 1.68 (br s, N*H*), 2.13 (s, $2CH_3$), 2.73 (q, $J$ = 7.1 Hz, $CH_2CH_3$), 3.40 (s, $CH_2$), 7.06 (s, 3 Ph*H*), 8.84 (br s, N*H*); $^{13}$C NMR ($CDCl_3$) 15.3 ($CH_2CH_3$), 18.3 ($2CH_3$), 44.6 ($CH_2CH_3$ or $CH_2$), 52.4 ($CH_2CH_3$ or $CH_3$), 126.8, 128.0, 133.8, 134.9 (Ph), 170.1 (*C*(O)NH) ppm; MS (+CI) (rel intensity) 208 (11), 207 ($M^+ + 1$, 100); $M_r$ (+CI) 207.149 83 [$M^+ + 1$] (calcd for $C_{12}H_{19}N_2O$ 207.149 74). Anal. ($C_{12}H_{18}N_2O$) C, H, N.

**Synthesis of *N,N*-Dimethyl-*N′*-(3-methylphenyl)urea (14).**[41] A THF (10 mL) solution of *m*-tosyl isocyanate (2.0 mL, 15.52 mmol) and dimethylamine (2.0 M in THF, 7.8 mL, 15.6 mmol) was stirred at room temperature (30 min). The solvent was evaporated and the solid residue was purified by crystallization (EtOH) to obtain 2.40 g (89%) of pure **14** as a white crystalline solid: mp 128−129 °C (lit.[41] mp 129−131 °C); $^1$H NMR ($CDCl_3$) $\delta$ 2.32 (s, Ph$CH_3$), 3.01 (s, $NCH_3(CH_3)$), 3.02 (s, $NCH_3(CH_3)$), 6.31 (br s, N*H*), 6.83−6.86 (m, 1 H, Ph*H*), 7.11−7.19 (m, 2 H, Ph*H*), 7.26−7.27 (m, 1 H, Ph*H*); $^{13}$C NMR ($CDCl_3$) 21.4 (Ph$CH_3$), 36.4 ($N(CH_3)_2$), 116.8, 120.5, 123.7, 128.6, 138.6, 139.1 (Ph), 155.7 (*C*(O)$NH_2$) ppm; MS (+CI) (rel intensity) 179 ($M^+ + 1$, 100); $M_r$ (+CI) 179.117 94 [$M^+ + 1$] (calcd for $C_{10}H_{15}N_2O$ 179.118 44). Anal. ($C_{10}H_{14}N_2O$) C, H, N.

**Pharmacology.** Procedures identical to those described in ref 12 were employed in this study.

**Acknowledgment.** We dedicate this paper to the memory of Dr. David W. Robertson, who contributed to the development of functionalized amino acid anticonvulsants while at the Lilly Research Laboratories (Indianapolis, IN). The authors thank the Anticonvulsant Screening Project at the National Institutes of Health, for kindly performing the pharmacological studies via the ASP's contract site at the University of Utah with Drs. H. Wolf, S. White, and K. Wilcox. This research was supported in part by the NIH research grant GM66940 awarded to A.T. The authors also acknowledge Tripos, Inc for the software grant.

**Supporting Information Available:** Tables listing the anticonvulsant activities of the 48 FAA compounds and the statistical characteristics of the 10 best QSAR models. This material is available free of charge via the Internet at http://pubs.acs.org.

## References

(1) Begley, C. E.; Annegers, J. F.; Lairson, D. R.; Reynolds, T. F.; Hauser, W. A. Cost of epilepsy in the United States: A model based on incidence and prognosis. *Epilepsia* **1994**, *35*, 1230−1243.
(2) Rogawski, M. A.; Porter, R. J. Antiepileptic drugs: Pharmacological mechanisms and clinical efficacy with consideration of promising development stage compounds. *Pharmacol. Rev.* **1997**, *42*, 223−286.
(3) Unverferth, K.; Rundfeldt, C. Antiepileptic drugs. *Pharmaceuticals* **2000**, *2*, 469−488.
(4) Mattson, R. H.; Cramer, J. A.; Collins, J. F.; Smith, D. B. Comparison of carbamazepine, phenobarbital, phenytoin, and primidone in partial and secondarily genealized tonic-clonic seizures. *New Eng. J. Med.* **1985**, *313*, 145−151.
(5) Pellock, J. M.; Willmore, L. J. A rational guide to monitoring in patients receiving anticonvulsants. *Neurology* **1991**, *41*, 961−964.
(6) Kohn, H.; Sawhney, K. N.; LeGall, P.; Conley, J. D.; Robertson, D. W.; Leander, J. D. Preparation and anticonvulsant activity of a series of functionalized α-aromatic and α-heteroaromatic amino acids. *J. Med. Chem.* **1990**, *33*, 919−926.
(7) Kohn, H.; Sawhney, K. H.; Bardel, P.; Robertson, D. W.; Leander, J. D. Synthesis and anticonvulsant activities of α-heteroaromatic-α-acetamido-*N*-benzylacetamide derivatives. *J. Med. Chem.* **1993**, *36*, 3350−3360.
(8) Choi, D.; Stables, J. P.; Kohn, H. Synthesis and anticonvulsant activities of *N*-benzyl-2-acetamidopropionamide derivatives. *J. Med. Chem.* **1996**, *39*, 1907−1916.
(9) Stables, J. P.; Kupferberg, H. J. In *Molecular and Cellular Targets for Antiepileptic Drugs*; Avanzini, G., Tanganelli, P., Avoli, M., Eds.; John Libbey: London, 1997; pp 191−198.
(10) Swinyard, E. A.; Woodhead, J. H.; White, H. S.; Franklin, M. R. In *Antiepileptic Drugs*, 3rd Edn.; Levy, R. H., Driefuss, F. E., Mattson, R. H., Meldrum, B. S., Perry, J. K., Eds.; Raven Press: New York, 1989; pp 85−102.
(11) Levy, R. H.; Mattson, R.; Meldrum, B. *Antiepileptic Drugs*, 4th Edn.; Raven Press: New York, 1995; Chapter 6.
(12) Shen, M.; LeTiran, A.; Xiao, Y.; Golbraikh, A.; Kohn, H.; Tropsha, A. Quantitative structure−activity relationship analysis of functionalized amino acid anticonvulsant agents using k nearest neighbor and simulated annealing PLS methods. *J. Med. Chem.* **2002**, *45*, 2811−2823.
(13) Güner, O. F. History and evolution of the pharmacophore concept in computer-aided drug design. *Curr. Top. Med. Chem.* **2002**, *2*, 1321−1332.
(14) Kurogi, Y.; Güner, O. F. Pharmacophore modeling and three-dimensional database searching for drug design using Catalyst. *Curr. Med. Chem.* **2001**, *8*, 1035−1055.
(15) Martin, Y. C. 3D database searching in drug design. *J. Med. Chem.* **1992**, *35*, 2145−2154.
(16) Tropsha, A.; Zheng, W. Identification of the descriptor pharmacophores using variable selection QSAR: Applications to database mining. *Curr. Pharm. Design* **2001**, *7*, 599−612.
(17) Tropsha, A.; Cho, S. J.; Zheng, W. "New Tricks For an Old Dog": Development and application of novel QSAR methods for rational design of combinatorial chemical libraries and database mining. In *Rational Drug Design: Novel Methodology and Practical Applications*; Parrill, A. L., Reddy, M. R., Eds.; ACS Symposium Series 719; American Chemical Society: Washington, DC, 1999; pp 198−211.
(18) Tropsha, A. Rational combinatorial library design and database mining using inverse QSAR approach. In *Combinatorial Library Design and Evaluation for Drug Discovery: Principles, Methods, Software Tools and Applications*. Ghose, A. K., Viswanadhan, V. N., Eds.; Marcel Dekker, Inc., New York, 2001; Chapter 12, pp 363−378.
(19) Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *Quant. Struct. Act. Relat. Comb. Sci.* **2003**, *22*, 69−77.
(20) Golbraikh, A.; Tropsha, A. Beware of $q^2$! *J. Mol. Graph. Model.* **2002**, *20*, 269−276.
(21) http://www.maybridge.com.
(22) http://dtp.nci.nih.gov/docs/3d_database/structural_information/smiles_strings.html.
(23) *MolConn-Z*, version 3.5; Hall Associates Consulting: Quincy, MA.

(24) Hoffman, B. T.; Cho, S. J.; Zheng, W.; Wyrick, S.; Nichols, D. E.; Mailman, R. B.; Tropsha, A. QSAR modeling of dopamine D1 agonists using comparative molecular field analysis, genetic algorithms-partial least squares, and *K* nearest neighbor methods. *J. Med. Chem.* **1999**, *42*, 3217−3226.
(25) Cho, S. J.; Zheng, W.; Tropsha, A. Rational combinatorial library design. 2. Rational design of targeted combinatorial peptide libraries using chemical similarity probe and the inverse QSAR approaches. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 259−268.
(26) Zheng, W.; Tropsha, A. A novel variable selection QSAR approach based on the K-nearest neighbor principle. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185−194.
(27) Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on rational division of experimental datasets into diverse training and test sets. *J. Comput. Aid. Mol. Des.* **2002**, *16*, 357−369.
(28) LeTiran, A.; Stables, J. P.; Kohn, H. Functionalized amino acid anticonvulsants: Synthesis and pharmacological evaluation of conformationally restricted analogues. *Bioorg. Med. Chem.* **2001**, *9*, 2693−2708.
(29) Geurts, M.; Poupaert, J. H.; Scriba, G. K. E.; Lambert, D. M. *N*-(Benzyloxycarbonyl)glycine Esters and Amides as New Anticonvulsants. *J. Med. Chem.* **1998**, *41*, 24−30.
(30) Robertson, D. W.; Leander, J. D. Private communication.
(31) Krall, R. L.; Perry, J. K.; White, B. G.; Kupferberg, H. J.; Swinyard, E. A. Antiepileptic drug development: II Anticonvulsant drug screening. *Epilepsia* **1978**, *19*, 409−428.
(32) Tucher, G. T. Pharmcokinetics of local anesthetics. *Brit. J. Anaesth.* **1986**, *58*, 717−731.
(33) Burney, R. G.; Difazio, C. A.; Peach, M. J.; Petrie, K. A.; Silvester, M. J. Antiarrhythmic effects of lidocaine metabolites. *Am. Heart J.* **1974**, *88*, 765−769.
(34) Sotaniemi, E. A.; Rautio, A.; Backstrom, M.; Arvela, P.; Pelkonen, O. CYP3A4 and CYP2A6 activities marked by the metabolism of lignocaine and coumarin in patients with liver and kidney diseases and epileptic patients. *Br. J. Clin. Pharmacol.* **1995**, *39*, 71−76.
(35) Cortes, S.; Liao, Z−K.; Watson, D.; Kohn, H. Effect of structural modification of the hydantoin ring on anticonvulsant activity. *J. Med. Chem.* **1985**, *28*, 601−606.
(36) Bialer, M.; Johannessen, S.; Kupferberg, H.; Levy, R.; Loiseau, P.; Perucca, E. Progress report on new antiepileptic drugs: A summary of the Fifth Eilat Conference (EILAT V). *Epilepsy Res.* **2001**, *43*, 11−58.
(37) Kovatcheva, A.; Golbraikh, A.; Oloff, S.; Xiao, Y.; Zheng, W.; Wolschann, P.; Buchbauer, G.; Tropsha, A. Combinatorial QSAR of ambergris fragrance compounds. *J. Chem. Inf. Comput. Sci.* **2004**. (jn press).
(38) Sawamura, M.; Nakayama, Y.; Kato, T.; Ito, Y. Gold(I)-catalyzed asymmetric aldol reaction of *N*-methoxy-*N*-methyl-α-isocyanoacetamide (α-isocyano weinreb amide). An efficient synthesis of optically active *β*-hydroxy α-amino aldehydes and ketones. *J. Org. Chem.* **1995**, *60*, 1727−1732.
(39) Nelson, S. D.; Breck, G. D.; Trager, W. F. In vivo metabolite condensations. Formation of *N*1-ethyl-2-methyl-*N*3-(2,6-dimethylphenyl)-4-imidazolidinone from the reaction of a metabolite of alcohol with a metabolite of lidocaine. *J. Med. Chem.* **1973**, *16*, 1106−1112.
(40) Lazar, J.; Bernath, G. Saturated heterocycles, Part 172. Synthesis of 2,6-disubstituted-5,6,7,8-tetrahydropyrido[4,3-*d*]pyrimidine Derivatives. *J. Heterocycl. Chem.* **1990**, *27*, 1885−1892.
(41) Martin, H.; Janiak, S.; Toepfl, W.; Duerr, D.; Rohr, O.; Pissiotas, G. Urea derivatives as selective herbicides. CIBA, DE 1905598, 1969.

JM030584Q