Quantitative Structure−Activity Relationship Analysis of Functionalized Amino Acid Anticonvulsant Agents Using *k* Nearest Neighbor and Simulated Annealing PLS Methods

Min Shen,[†] Arnaud LeTiran,[‡] Yunde Xiao,[†] Alexander Golbraikh,[†] Harold Kohn,*,[†] and Alexander Tropsha*,[†]

Division of Medicinal Chemistry and Natural Products, School of Pharmacy, CB# 7360, University of North Carolina, Chapel Hill, North Carolina 27599-7360, and Department of Chemistry, University of Houston, Houston, Texas 77204-5641

We report the development of rigorously validated quantitative structure−activity relationship (QSAR) models for 48 chemically diverse functionalized amino acids with anticonvulsant activity. Two variable selection approaches, simulated annealing partial least squares (SA-PLS) and *k* nearest neighbor (kNN), were employed. Both methods utilize multiple descriptors such as molecular connectivity indices or atom pair descriptors, which are derived from two-dimensional molecular topology. QSAR models with high internal accuracy were generated, with leave-one-out cross-validated R^2 (q^2) values ranging between 0.6 and 0.8. The q^2 values for the actual dataset were significantly higher than those obtained for the same dataset with randomly shuffled activity values, indicating that models were statistically significant. The original dataset was further divided into several training and test sets, with highly predictive models providing q^2 values greater than 0.5 for the training sets and R^2 values greater than 0.6 for the test sets. These models were capable of predicting with reasonable accuracy the activity of 13 novel compounds not included in the original dataset. The successful development of highly predictive QSAR models affords further design and discovery of novel anticonvulsant agents.

Introduction

Epilepsy is a cluster of symptoms arising from dysfunctional events in the brain that are characterized by recurrent seizures produced by paroxysmal excessive neuronal discharges.[1,2] A diverse set of antiepileptic agents exist that exert seizure control by targeting different neurological pathways.[3−5] Unfortunately, these therapies are ineffective for more than a third of patients with epilepsy,[6] documenting the need for new agents with different mechanisms of action.

In recent years, we[7] and others[8,9] have advanced a novel series of anticonvulsant agents that are termed functionalized amino acids (FAA). The target(s) of FAA function has not been identified. Nearly 250 FAA have been prepared and evaluated in animal seizure models. Of these, 12 provided protection against maximal electroshock (MES) induced seizures[10,11] at doses comparable to or better than those for phenytoin.[12] The MES test is a proven method for the identification of new drug candidates for partial and generalized seizures. While our studies have provided structural patterns beneficial for FAA activity, it has become increasingly difficult to formulate a useful SAR because the diversity of molecular structures has increased. Accordingly, a need exists for reliable molecular models that 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 the design of new FAA.

Several QSAR studies correlating structure with anticonvulsant activity have been reported.[13] Most have focused on specific drug classes including analogues of valproic acid,[13] propofol,[14] 2-benzothiazolamines,[15] barbiturates,[16] hydantoins,[17] carbamates,[18] and arylpiperazines.[19] In 2000, Estrada and Peña reported a study based on a topological substructural molecular design (TOSS-MODE) approach for 235 compounds of diverse structure that included 87 clinically tested anticonvulsants.[20] The method discriminated between active and inactive anticonvulsant compounds with 88% accuracy and permitted identification of structural fragments important for bioactivity. However, this method would not allow ranking of active anticonvulsant agents with respect to their relative potency. We sought a computational approach that would not only provide a higher degree of predictive power but would also afford a quantitative prediction of the anticonvulsant activity. Thus, we chose to apply a QSAR approach to a limited set of FAA compounds that were likely to function by similar biochemical pathways.

Many QSAR approaches have been developed over the years. The rapid increase in structural three-dimensional (3D) information of bioorganic molecules,[21,22] coupled with the development of fast methods for 3D structure generation (e.g., CONCORD[23,24]) and alignment (e.g., active analogue approach[25,26]), has led to the development of 3D structural descriptors and associated

* To whom correspondence should be addressed. Phone: 919-966-2955 (A.T.); 919-966-2680 (H.K.). Fax: 919-966-0204. E-mail for H.K.: harold_kohn@unc.edu. E-mail for A.T.: alex_tropsha@unc.edu.
† University of North Carolina.
‡ University of Houston.

3D QSAR methods (several recent reviews on 3D QSAR can be found in ref 27).

Most 3D QSAR methods, such as comparative molecular field analysis (CoMFA),[28] implement the pharmacophore concept, which is central to the rational drug design and discovery process. Traditionally, a pharmacophore is defined by the specific and characteristic 3D arrangement of chemical functional groups found in active molecules. However, the detection of a unique 3D pharmacophore and the unambiguous alignment of active molecules necessary to conduct a CoMFA study are difficult, if not impossible, for structurally diverse compounds. Recently, by analogy with 3D pharmacophores, we have introduced a more general concept of the descriptor pharmacophore (summarized in ref 29). The descriptor pharmacophores are defined by the means of variable selection QSAR as a subset of molecular descriptors that afford the most statistically significant structure–activity correlation. The q^2-GRS[30] method developed earlier in our laboratory is an example of variable selection applied to 3D descriptors (i.e., molecular fields) of molecular structure. We have also demonstrated that QSAR analysis methods with 2D descriptors (termed 2D QSAR) can provide a powerful alternative to 3D QSAR.[31–33]

A major benefit of 2D compared to 3D QSAR methods is that the former neither requires conformational search nor structural alignment. Accordingly, 2D QSAR methods are easily automated and adapted to the task of database searching, or virtual screening.[34] These considerations led us to develop two variable selection QSAR algorithms: genetic algorithm or simulated annealing partial least squares (GA/SA-PLS)[31,34,35] and k-nearest-neighbor (kNN)[32,34] analyses. Typically, these methods 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 achieve a robust QSAR model that is characterized by the highest value of cross-validated R^2 (q^2). By default, the descriptor pharmacophore represents an optimal selection of descriptors types (cf., invariant selection of traditional pharmacophoric elements); however, for structurally dissimilar molecules, descriptor values are generally different. These methods are computationally efficient and automated and have been used to produce predictive models[32,35] that are comparable to, or better than, those obtained with CoMFA.

In this paper, we have applied kNN and SA-PLS QSAR approaches to a dataset of 48 FAA anticonvulsants that were synthesized previously (cf., Table 1). Our objective was to develop robust, validated QSAR models suitable for the discovery and design of new anticonvulsant agents. High q^2 value for the training set has been often considered sufficient criteria of QSAR model accuracy. However, we have demonstrated recently[36] that a high value of q^2 alone does not guarantee the acceptable predictive ability of a QSAR model. All QSAR models developed herein have been extensively validated using several criteria of robustness and accuracy.[36] The models were successfully tested by accurate prediction of anticonvulsant activity for a new

Scheme 1. Preparation of **49** and **50**
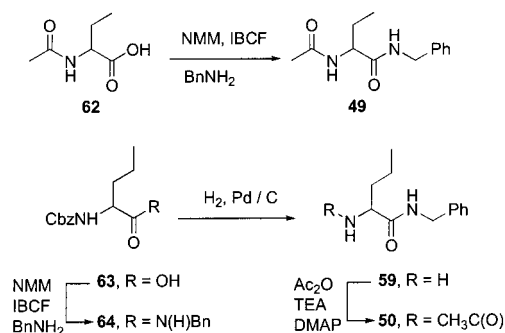
set of 13 FAA that formed an external validation set. The QSAR models developed and validated in this study can be used to screen large databases or virtual libraries for new anticonvulsant agents.

Methods

Chemistry and Anticonvulsant Activities. The structures and in vivo activities of 48 racemic FAA anticonvulsant agents used in this study are listed in Table 1. We used the experimentally determined ED_{50} values (mg/kg) obtained in the MES-seizure-induced assay in mice (ip). No allowances were made for potential differences in the transport or metabolic properties of FAA, although these features may influence the measured ED_{50} values. Table 2 shows the structures and activities for compounds in the external test set that were not known prior to this QSAR study. For QSAR calculations, the compound weight per animal weight (mg/kg) values were converted to μmol/animal weight (μmol/kg) units (Tables 1 and 2).

The synthetic route for FAA **49** and **50** is outlined in Scheme 1. N-Acetyl-DL-α-aminobutyric acid[37] (**62**) was coupled to benzylamine using the mixed-anhydride coupling (MAC) procedure[7j,38] to afford **49** in 90% yield. The preparation of **50** and its amine precursor **59** started with commercially available Cbz-DL-norvaline (**63**) (Cbz-DL-Nva). Treatment of Cbz-Nva (**63**) with benzylamine using MAC[7j,38] methodology provided **64**. Removal of the Cbz group (H$_2$, Pd/C) followed by N-acetylation of **59** with Ac$_2$O, TEA, and DMAP (cat.) afforded **50** in 79% overall yield (three steps).

Scheme 2 details the preparative route for FAA **56–58**. Treatment of **65**[39] with propionyl chloride afforded **57** in 71% yield, while addition of acryloyl chloride to **65** gave **58** in 83% yield. Formylation of **65** with N-(diethylcarbamoyl)-N-methoxyformamide[40] provided **56** in 93% yield.

Computational Details

Generation of Molecular Descriptors. All chemical structures were generated using SYBYL software.[41] Multiple descriptors derived from 2D molecular topology were used. Molecular topological indices[42,43] were generated with the MolConnZ program (MZ descriptors).[44] Overall, MolConnZ produces over 400 different descriptors. Most of these descriptors characterize chemical structure, but several depend on the arbitrary numbering of atoms in a molecule and are introduced solely for bookkeeping purposes. In our study, only 312 chemically

Table 1. Structures and in Vivo Anticonvulsant Activities of 48 FAA Compounds Used in This Study

compd	ref	R$_1$	R$_2$	R$_3$	MES, mice ip ED$_{50}$ (mg/kg)	log ED$_{50}$ (μmol/kg)
1	7b,c	Ac	CH$_3$	CH$_2$Ph	76.5 (66.5−89.0)[a]	2.54
2	b	Ac	CH$_3$	(CH$_2$)$_3$Ph	87.6[c]	2.55
3	7c	Ac	CH$_3$	CH$_2$−3-F−C$_6$H$_4$	77.3 (62.5−91.0)[a]	2.51
4	7e	Ac	2-furanyl	CH$_2$Ph	10.3 (9.1−11.6)[c]	1.58
5	7e	Ac	2-furanyl	CH$_2$−2-F−C$_6$H$_4$	40.0[c]	2.14
6	7e	Ac	2-furanyl	CH$_2$−3-F−C$_6$H$_4$	13.3 (11.5−15.3)[c]	1.66
7	7e	Ac	2-furanyl	CH$_2$−4-F−C$_6$H$_4$	12.7 (10.4−15.1)[c]	1.64
8	7e	Ac	2-furanyl	CH$_2$−2,5-di-F−C$_6$H$_3$	23.8 (20.2−28.4)[c]	1.89
9	7c	Ac	Ph	CH$_2$Ph	20.3 (16.8−24.4)[a]	2.06
10	d	Ac	2-allyl	CH$_2$Ph	33.6 (28.4−45.1)[c]	2.16
11	7g	Ac	2-tetrahydrofuranyl	CH$_2$Ph	51.7 (44.4−59.9)[c]	2.27
12	7j	Ac	CH$_2$OCH$_3$	CH$_2$Ph	8.3 (7.9−9.8)[a]	1.52
13	7j	Ac	CH$_2$OC$_2$H$_5$	CH$_2$Ph	17 (15−19)[a]	1.81
14	7e	Ac	2-furanyl-5-CH$_3$	CH$_2$Ph	19.2 (16.4−23.8)[c]	1.83
15	7e	Ac	2-pyrrolyl	CH$_2$Ph	16.1 (13.2−19.9)[c]	1.77
16	7e	Ac	2-pyrrolyl-5-CH$_3$	CH$_2$Ph	36.5 (30.6−57.1)[c]	2.12
17	7e	Ac	2-thienyl	CH$_2$Ph	44.8 (38.9−51.4)[c]	2.19
18	7e	Ac	3-thienyl	CH$_2$Ph	87.8 (69.9−150)[c]	2.48
19	7g	Ac	1-pyrrole	CH$_2$Ph	80.2 (66.6−101)[c]	2.47
20	7g	Ac	1-pyrazole	CH$_2$Ph	16.5 (14.1−22.5)[c]	1.78
21	7i	Ac	2-pyridyl	CH$_2$Ph	10.8 (9.1−12.1)[a]	1.58
22	7i	Ac	2-pyrazinyl	CH$_2$Ph	14.8 (12.5−17.2)[a]	1.72
23	7i	Ac	2-pyrimidyl	CH$_2$Ph	8.1 (5.5−11.5)[a]	1.46
24	7g	Ac	2-oxazole	CH$_2$Ph	10.4 (9.2−11.6)[c]	1.58
25	7g	Ac	2-thiazole	CH$_2$Ph	12.1 (9.5−14.5)[c]	1.62
26	7g	Ac	C(S)NH$_2$	CH$_2$Ph	86.2 (75.4−101)[c]	2.51
27	e	Ac	C(NH)NHOAc	CH$_2$Ph	140 (45.7−715)[c]	2.66
28	e	Ac	C(NH)NHOH	CH$_2$Ph	112 (85.7−151)[c]	2.63
29	7f	Ac	NH$_2$	CH$_2$Ph	65.1 (56.2−75.3)[c]	2.47
30	7f	Ac	NHCH$_3$	CH$_2$Ph	44.5 (37.0−52.4)[c]	2.28
31	7f	Ac	NHC$_2$H$_5$	CH$_2$Ph	42.4 (37.2−47.8)[c]	2.23
32	7f	Ac	N(CH$_3$)$_2$	CH$_2$Ph	45.3[c]	2.26
33	7f	Ac	NH(OCH$_3$)	CH$_2$Ph	6.2 (5.4−7.2)[c]	1.39
34	7f	Ac	N(CH$_3$)OCH$_3$	CH$_2$Ph	6.7 (5.7−7.7)[c]	1.40
35	7f	Ac	*N*-isoxazoline	CH$_2$Ph	31.4 (26.7−37.8)[c]	2.05
36	f	Ac	NPhNH$_2$	CH$_2$Ph	42.8 (34.9−54.3)[c]	2.14
37	7f	Ac	NHNHCbz	CH$_2$Ph	55.6 (49.3−63.9)[c]	2.18
38	7f	Ac	OH	CH$_2$Ph	80.1 (70.6−91.0)[c]	2.56
39	7f	Ac	OCH$_3$	CH$_2$Ph	98.3 (84.4−114)[c]	2.62
40	7f	Ac	OC$_2$H$_5$	CH$_2$Ph	62.0 (51.1−78.4)[c]	2.39
41	g	Ac	NH−3-NH$_2$−C$_6$H$_4$	CH$_2$Ph	98.6 (76.8−120)[c]	2.50
42	g	Ac	NH−4-NH$_2$−C$_6$H$_4$	CH$_2$Ph	∼176[c]	2.75
43	7g	CH$_3$C(S)	2-furanyl	CH$_2$Ph	18.4 (15.9−22.0)[c]	1.80
44	7c				270[c]	3.06
45	f				60.1(51.9−68.0)[c]	2.50
46	8a,b	CH$_3$	CH$_3$	CH$_2$Ph	31.2 (21.4−40.8)[a]	2.21
47	9				53 (46−61)[a]	2.38
48	8c				29.1 (14.6−41.0)[a]	2.09

[a] This compound was evaluated at the NIH Epilepsy Branch (Bethesda, MD) under the direction of Dr. Harvey Kupferberg and Mr. James P. Stables. [b] D. W. Robertson, J. D. Leander. Unpublished results. [c] This compound was evaluated at the Eli Lilly Laboratories (Indianapolis, IN) under the direction of Drs. J. David Leander and David W. Robertson. [d] K. N. Sawhney, H. Kohn. Unpublished results. [e] P. Bardel, H. Kohn. Unpublished results. [f] H. Kohn. Unpublished results. [g] Choi, D. Ph.D. Thesis, University of Houston, Houston, TX, 1995.

relevant descriptors were initially calculated and 189 descriptors were eventually used (after deleting descriptors with zero value or zero variance).

Atom pair descriptors (AP)[45] were obtained with the GenAP[46] program developed in this laboratory using an approach proposed by Carhart et al.[45] AP descriptors are defined by their atom types and topological distance bins. An AP is a substructure defined by two atom types and the shortest path separation (or graph distance) between the atoms. The graph distance is defined as the smallest number of atoms along the path connecting two atoms in a molecular structure. The general form of an atom pair descriptor is as follows:

$$\text{atom type } i \cdots (\text{distance}) \cdots \text{atom type } j$$

In our study, the following 15 atom types were used, selected from SYBYL atom types (mol2 format):[41] (1) negative charge center, NCC; (2) positive charge center, PCC; (3) hydrogen bond acceptor, HA; (4) hydrogen bond donor, HD; (5) aromatic ring center, ARC; (6) nitrogen atoms, N; (7) oxygen atoms, O; (8) sulfur atoms, S; (9) phosphorus atoms, P; (10) fluorine atoms, FL; (11) chlorine, bromine, iodine atoms, HAL; (12) carbon

Table 2. Structures and in Vivo Anticonvulsant Activities for Compounds in the External Set

compd	ref	R$_1$	R$_2$	R$_3$	R$_4$	MES, mice ip ED$_{50}$ (mg/kg)	log ED$_{50}$ (μmol/kg)
49		Ac	H	C$_2$H$_5$	H	>100, <300[a]	2.63–3.12
50		Ac	H	nC$_3$H$_7$	H	38.4 (34.9–45.0)[a]	2.19
51	39	Ac	CH$_3$	CH$_3$	H	>30, <100[a]	2.10–2.63
52	39	Ac	CH$_3$	C$_2$H$_5$	H	>30, <100[a]	2.08–2.60
53	39	Ac	C$_2$H$_5$	CH$_3$	H	85.0 (74.3–99.1)[a]	2.53
54	39	Ac	H	CH$_3$	CH$_3$	>100, <300[a]	2.63–3.11
55	39	Ac	H	CH$_2$OCH$_3$	CH$_3$	>100, <300[a]	2.58–3.06
56		HC(O)	H	CH$_2$OCH$_3$	H	>30, <100[a]	2.10–2.63
57		EtC(O)	H	CH$_2$OCH$_3$	H	>100, <300[a]	2.58–3.06
58		CH$_2$=CHC(O)	H	CH$_2$OCH$_3$	H	>100, <300[a]	2.58–3.06
59		H	H	nC$_3$H$_7$	H	>30, <100[a]	2.16–2.69
60	39	H	CH$_3$	C$_2$H$_5$	H	>30, <100[a]	2.16–2.69
61	7b	H	H	CH$_3$	H	>100, <300[a]	2.75–3.23

[a] This compound was evaluated at the NIH Epilepsy Branch (Bethesda, MD) under the direction of Dr. Harvey Kupferberg and Mr. James P. Stables.

Scheme 2. Preparation of **56–58**
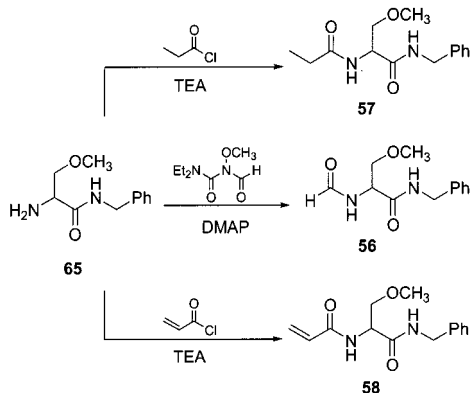
atoms, C; (13) all other elements, OE; (14) triple bond center, TBC; (15) double bond center, DBC. The total number of pairwise combinations of all 15 atom types is 120. Furthermore, 15 distance bins were defined in the interval from graph distance zero (i.e., zero atoms separating an atom pair) to 14 and greater. Thus, a total of 1800 (120 × 15) AP descriptors were generated for each molecular structure. Many of the AP descriptors frequently have zero value (when certain atom types or atom pairs are absent in molecular structures). Thus, for our data set of 48 anticonvulsant FAA, only 273 descriptors with nonzero value and nonzero variance were generated with the GenAP program.

MZ descriptors were range-scaled prior to distance calculations, since the absolute scales for MZ descriptors can differ by orders of magnitude. Accordingly, our use of range scaling avoided giving descriptors with significantly higher ranges a disproportional weight upon distance calculations in multidimensional MZ descriptor space. No scaling was needed for AP descriptors, since they are integers ranging from zero to no more than a couple of dozens of AP counts. All calculations were performed on an SGI Octane at University of North Carolina's Laboratory for Molecular Modeling.

kNN QSAR Method. The kNN QSAR method[32] employs the kNN classification principle[47] and the variable selection procedure. Briefly, a subset of *nvar* (number of selected variables) descriptors is selected randomly as a hypothetical descriptor pharmacophore

(HDP). The *nvar* is set to different values to obtain the best q^2 possible. The HDP is validated by leave-one-out cross-validation, where each compound is eliminated from the training set and its biological activity is predicted as the average activity of k most similar molecules ($k = 1$–5). The similarity is characterized by the Euclidean distance between compounds in multidimensional descriptor space. A method of simulated annealing with the Metropolis-like acceptance criteria is used to optimize the variable selection. Further details of the kNN method implementation, including the description of the simulated annealing procedure used for stochastic sampling of the descriptor space, are given elsewhere.[32]

The original kNN method[32] was enhanced in this study by using weighted molecular similarity. In the original method, the activity of each compound was predicted as the algebraic average activity of its k-nearest-neighbor compounds in the training set. However, in general, the Euclidean distances in the descriptor space between a compound and each of its k nearest neighbors are not the same. Thus, the neighbor with the smaller distance from a compound was given a higher weight in calculating the predicted activity as follows.

$$w_i = \frac{\exp(-d_i)}{\sum_{k \text{ nearest neighbors}} \exp(-d_i)} \quad (1)$$

$$\hat{y} = \sum w_i y_i \quad (2)$$

where d_i is the Euclidean distance between the compound and its k nearest neighbors, w_i is the weight for every individual nearest neighbor, y_i is the actual activity value for nearest neighbor i, and $\hat{y}$ is the predicted activity value.

In summary, the kNN QSAR algorithm generates both an optimum k value and an optimal *nvar* subset of descriptors, which afford a QSAR model with the highest value of q^2. Figure 1 shows the overall flowchart of the current implementation of the kNN method.

SA-PLS Method. The SA-PLS method is based on the GA-PLS method.[34,35] While we employ the same
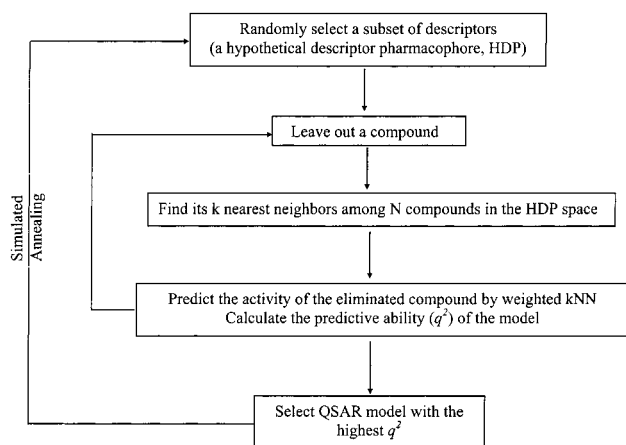
Figure 1. Flowchart of the kNN method.

Table 3. Frequently Used α Values and the Corresponding Critical Values of Z_c for One-Tail Test

$$\alpha = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-z^2/2} \quad \text{for} \quad z \geq 4$$

α	Z_c
0.10	1.28
0.05	1.64
0.01	2.33
0.001	3.10

stochastic descriptor sampling used in the kNN QSAR method, the QSAR model for HDP is developed using PLS[28,48] analysis. The $[1 - (n - 1)(1 - q^2)/(n - c)]$ expression (where q^2 is the cross-validated R^2, n is the number of compounds, and c is the optimal number of components) is used as a fitting function to guide the SA optimization.

Robustness of QSAR Models. The q^2 values for the models for experimental training sets were compared to those derived for so-called random datasets, which are generated by random shuffling of compound activities. The statistical significance of QSAR models for training sets was evaluated with the standard hypothesis testing method.[49] In this approach, two alternative hypotheses are formulated: (1) for H_0, $h = \mu$; (2) for H_1, $h > \mu$, where μ is the average value of q^2 for random datasets and h is the q^2 value for the actual dataset. Thus, the null hypothesis, H_0, states that the QSAR model for the actual dataset is not significantly better than random models whereas the alternative hypothesis, H_1, assumes the opposite (i.e., that the actual model is significantly better than the random models). The decision-making is based on a standard one-tail test, which involves the following procedure.

(1) Determine the average value of q^2 (μ) and its standard deviation (σ) for random datasets.

(2) Calculate the Z score that corresponds to the q^2 value for the actual dataset:

$$Z = (h - \mu)/\sigma \qquad (3)$$

(3) Compare this Z score with the tabular critical values of Z_c at different levels of significance (α)[49] to determine the level at which H_0 should be rejected. If the Z score is higher than tabular values of Z_c (cf., Table 3), one concludes that at the level of significance that corresponds to that Z_c, H_0 should be rejected and,

therefore, H_1 should be accepted. In this case, it is concluded that the result obtained for the actual dataset is statistically better than those obtained for random datasets at the given level of significance.

Model Validation: Training and Test Set Compound Selection. A q^2 value greater than 0.5 has been considered to be standard proof of the high predictive ability of the model. This axiom is not always true. We have demonstrated[36] that while a high q^2 is a necessary condition for a model to have a high predictive power, it is *not* a sufficient condition. To obtain a truly predictive and validated model, it is necessary to split the whole dataset into training and test sets, to develop a training set model, and to validate it by accurately predicting the activity of the test set compounds. To obtain training and test sets, we have used a diversity of sampling algorithms described in detail elsewhere.[36] This algorithm allows construction of training and test sets of various sizes that cover the entire descriptor space of all compounds.

Conservative Activity Prediction for the Test Set Compounds. In the kNN method, kNN QSAR models are developed from chemical similarity calculations by interpolating the activity of the test set compounds. Potentially, certain molecules from external test sets could be too dissimilar from the test set compounds to afford accurate prediction of their activity. Thus, we have introduced a distance cutoff value D_c that defines a similarity threshold for external compounds:

$$D_c = \bar{y} + Z\sigma \qquad (4)$$

where $\bar{y}$ is the average Euclidean distance of k nearest neighbors of each compound in the training set, σ is the standard deviation of Euclidean distances of k nearest neighbors for each compound in the training set, and Z is an empirical parameter to control the significance level, with a default value of 0.5.

If the distance from an external compound to any of its nearest neighbors in the training set is above the distance cutoff D_c, we consider it impossible to evaluate its activity accurately and exclude this compound from consideration. Otherwise, the activity of an external compound is predicted from eqs 1 and 2. In our study, all compounds in the external data set were found to be within the default distance cutoff (corresponding to the Z value of 0.5) of their nearest neighbors in the training set.

Results and Discussion

kNN and SA-PLS QSAR analyses of 48 anticonvulsant agents were performed independently using MZ and AP descriptors. Four different types of models were built using various combinations of descriptors and optimization algorithms: AP_kNN, MZ_kNN, AP_SA-PLS, and MZ_SA-PLS. The results obtained with these methods are discussed in terms of the optimized q^2 values, variable selection, actual vs predicted activities, and statistical significance of the resulting QSAR models.

QSAR Models and Their Robustness. In the kNN QSAR method, *nvar* can be set to any value that is less than the total number of descriptors. Since the optimal number for *nvar* is not known a priori, multiple models have to be generated to examine the relationship
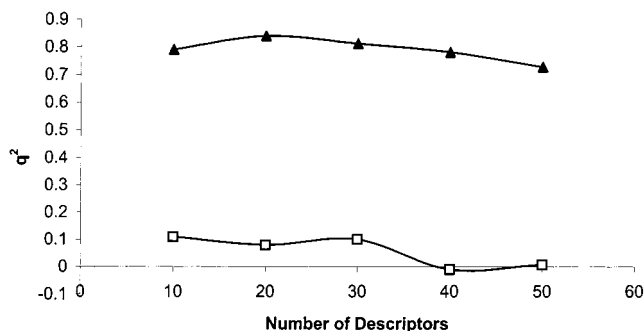
Figure 2. Plots of q^2 vs the number of descriptors selected for the best AP_kNN QSAR models for 48 anticonvulsant α-amino acid derivatives. The results for both actual and random (with shuffled activity values) data sets are shown. Every q^2 value is the average of 10 independent calculations. Triangles represent the actual data set, and squares represent the random data set.

Table 4. Standard One-Tail Hypothesis Testing for a 20-Descriptor QSAR Model for 48 FAA Compounds Using AP_kNN, MZ_kNN, AP_SA-PLS, and MZ_SA-PLS Methods

| QSAR model	h (q^2 of actual model)	μ (av q^2 of 10 random models)	σ	Z score
AP_kNN	0.84	0.08	0.16	4.73
MZ_kNN	0.71	0.01	0.15	4.57
AP_SA-PLS	0.66	0.08	0.10	5.92
MZ_SA-PLS	0.71	0.06	0.16	4.21

between q^2 and *nvar*. As previously discussed, the robustness of a QSAR model should be established by comparing results for the actual data set with those for data sets with randomized activity values. Thus, 10-, 20-, 30-, 40-, and 50-descriptor models were generated. Figure 2 shows a plot of q^2 vs *nvar* for the actual and random data sets obtained with AP_kNN calculations (similar plots, which are not shown, were also generated as a result of MZ_kNN, AP_SA-PLS, and MZ_SA-PLS QSAR analyses as well). Every q^2 value is the average of 10 independent computations. Overall, we obtained consistently higher q^2 values for the actual data set compared to those for randomized activity data sets. The q^2 values for the real data set were in the range 0.60–0.85 compared to −0.10 to 0.10 for the random data sets.

The statistical examination of the results was performed with one-tail hypothesis testing as described in Computational Details. The q^2 values for 20-descriptor AP_kNN QSAR models obtained for 10 different randomized data sets are shown in Table 4, which also lists the average q^2 value, the standard deviation of the q^2 values, and the Z score for the most significant 20-descriptor AP_kNN QSAR model for the actual data set. A Z score of 4.73 indicates that the probability that the AP_kNN QSAR model constructed for the real FAA data set is random is approximately 10^{-5}. The similar results for this statistical significance test were obtained for our MZ_kNN, AP_SA-PLS, and MZ_SA-PLS QSAR models; they are provided in Table 4 as well. These findings document that all QSAR models for the actual data set were nonspurious.

kNN QSAR Model Validation. To obtain reliable and truly predictive QSAR models, it is necessary to demonstrate that the training set models can accurately predict activities of compounds in external test sets. Generally, we accept models with q^2 values for the training set greater than 0.5 and R^2 values for predicted
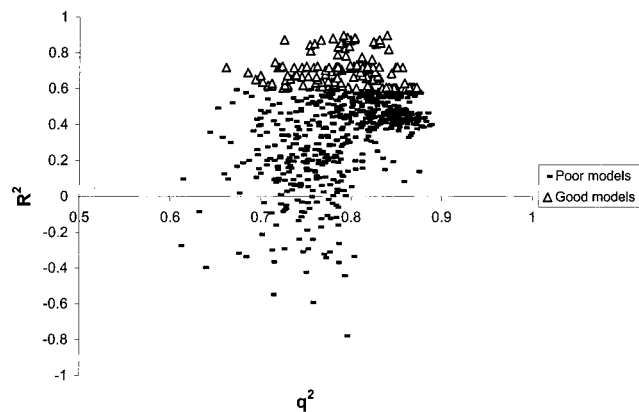

Figure 3. Predictive R^2 vs q^2 for AP_kNN models with $q^2 > 0.6$. The models with $R^2 > 0.6$ are indicated by triangles.
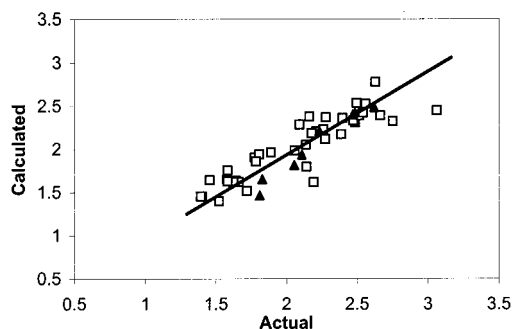
Table 5. Ten Best AP_kNN QSAR Models

model no.	size of the training set (no. of cpds)	size of the test set (no. of compds)	no. of descriptors	q^2 (training set)	R^2 (test set)
1	41	7	16	0.79	0.90
2	41	7	26	0.84	0.90
3	41	7	16	0.80	0.88
4	40	8	12	0.82	0.76
5	40	8	14	0.81	0.73
6	40	8	10	0.81	0.69
7	39	9	18	0.81	0.67
8	37	11	10	0.78	0.64
9	35	13	22	0.73	0.65
10	34	14	26	0.77	0.64

vs actual activities of the test set compounds greater than 0.6.[36] Certainly, models of the greatest utility are those obtained with the smallest possible training set, which are still capable of accurately predicting the activities of large test sets. By use of the diversity selection algorithm referenced in Methods, the entire 48-compound dataset characterized by the AP descriptors was divided into different training and test sets, with the training and test set sizes being 41 and 7, 40 and 8, 39 and 9, 37 and 11, 35 and 13, and 34 and 14. Multiple variable selection models with the high q^2 values (greater than 0.5) were generated. However, as in earlier studies,[36] no correlation was found between q^2 and R^2 (Figure 3). On the basis of our criteria, acceptable models with both high statistical significance ($q^2 > 0.5$) and predictive power ($R^2 > 0.6$) represented only a small fraction of all models with $q^2 > 0.5$ (Figure 3). These results provide further evidence that the use of the term "predictive" to describe QSAR models with $q^2 > 0.5$ should be generally avoided.

Table 5 presents the 10 best models obtained from multiple AP_kNN analyses. The trajectory of the SA-driven optimization of q^2 values in developing the best AP_kNN QSAR model 1 (Table 5) is shown in Figure 4. Figure 5 shows actual vs calculated activity values for the training and test sets based on this model. The number of descriptors for the 10 best models varied between 12 and 26. As expected, we found that as the size of the training set decreased and the size of the test set correspondingly increased, the predictive power of the models (R^2) decreased as well (cf., Table 5). Nonetheless, an acceptable model was obtained with the test set as large as 14 compounds (nearly one-third of the entire data set) with $q^2 = 0.77$ and $R^2 = 0.64$.
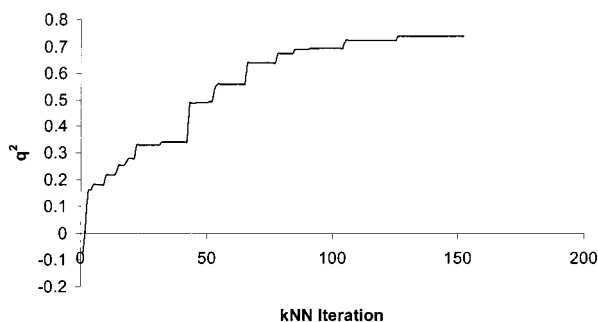
**Figure 4.** Trajectory of the SA-driven optimization of q^2 values in developing the AP_kNN model 1.

Figure 5. Actual vs calculated activity values for the training (squares) and test (triangles) set using AP_kNN model 1 ($q^2 = 0.79$, $R^2 = 0.90$).

Table 6. Ten Best MZ_kNN QSAR Models

model no.	size of the training set (no. of compds)	size of the test set (no. of compds)	no. of descriptors	q^2 (training set)	R^2 (test set)
1	43	5	12	0.64	0.81
2	43	5	22	0.71	0.77
3	43	5	22	0.77	0.71
4	43	5	20	0.78	0.71
5	43	5	16	0.76	0.71
6	43	5	12	0.75	0.71
7	39	9	20	0.65	0.72
8	39	9	14	0.65	0.72
9	39	9	12	0.69	0.71
10	38	10	18	0.72	0.67

Table 6 presents the 10 best models obtained from multiple MZ_kNN analyses. Since the training/test set selection algorithm is sensitive to the type of descriptors used, the original 48-compound dataset was divided into different training and test sets from those employed when AP descriptors were used. Here, the training and test set sizes were 43 and 5, 39 and 9, and 38 and 10 compounds. Multiple QSAR models were generated independently for all training sets and validated using the test sets. The best models with the highest predictive power were obtained for the test sets with five and nine compounds, with the optimal number of descriptors ranging between 12 and 20 (Table 6). Figure 6 shows actual vs calculated activity values for the training and test sets using this model.

We have attempted to increase the size of the test sets to include 11 (37 compounds in the training set), 12 (36), 13 (35), and 14 (34) compounds. However, as the number of compounds in the test set increased above 10, the predictive ability of the models (R^2) decreased dramatically. Indeed, the best R^2 values for the test sets with 11, 12, 13, and 14 compounds were only 0.36, 0.13, 0.34, and 0.22, respectively, whereas q^2 values for the respec-
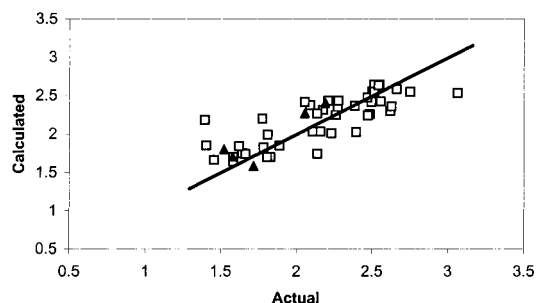
Figure 6. Actual vs calculated activity values for the training (squares) and test (triangles) set using MZ_kNN model 1 ($q^2 = 0.64$, $R^2 = 0.81$).

tive training sets were still in the range 0.70−0.80. These results demonstrate once more that q^2 alone does not serve as an estimate of the predictive power of kNN models.

SA-PLS Model Validation. The results obtained with the AP_SA-PLS analysis are summarized in Table 7. The test and training sets were the same as reported above for the AP_kNN analysis. Multiple QSAR models with $q^2 > 0.5$ were obtained independently for 10, 20, 30, 40, and 50 descriptors, and the 10 best models were selected using the R^2 value as the criterion (Table 7). We found that the most predictive models were obtained with only 7 or 8 compounds in the test set. When the test sets contained 9, 11, 13, and 14 compounds, high q^2 values were still observed for the respective training sets; however, the R^2 values obtained for these test sets did not exceed 0.5. These findings again illustrate that high q^2 values for the training sets do not imply an acceptable QSAR model. The fitted (non-cross-validated) R^2 values are also provided in Table 7. We obtained no models that gave an acceptable R^2 value using MZ descriptors.

In general, similar to our kNN analysis, the AP_SA-PLS method appeared to provide models with higher predictive power than models resulting from MZ_SA-PLS calculations. The AP_SA-PLS models with the highest predictive power were produced using training sets corresponding to the test sets with 7 and 8 compounds. The range of the optimum number of descriptors varied from 20 to 50 for AP_SA-PLS. As the number of compounds in the test set further increased, the predictive ability of the models decreased dramatically. The trajectory of the SA-driven optimization of q^2 values in developing the best model 1 (Table 7) is shown in Figure 7. Figure 8 shows the actual vs calculated activity values for the training and test sets using this AP_SA-PLS model.

The detailed results are listed in the table in Supporting Information, which compares predicted vs actual biological activity values for each compound in the training and test sets, as determined by three types of QSAR models. Models obtained with AP descriptors appeared to reproduce experimental data better than those generated with MZ descriptors.

Prediction of Anticonvulsant Activity for an External Data Set. QSAR models validated with the test sets were used to predict the anticonvulsant activity of 13 new compounds (Table 2), which were not available prior to our QSAR studies of 48 FAA. For the majority of these compounds (Table 2), the experimental ED_{50}

Table 7. Ten Best AP_SA-PLS QSAR Models

model no.	size of the training set (no. of compds)	size of the test set (no. of compds)	no. of descriptors	q^2 (training set)	fitted R^2 (training set)	R^2 (test set)
1	41	7	20	0.66	0.89	0.77
2	41	7	40	0.71	0.91	0.64
3	41	7	50	0.68	0.94	0.73
4	41	7	50	0.55	0.91	0.73
5	40	8	20	0.64	0.86	0.64
6	40	8	30	0.68	0.89	0.67
7	40	8	40	0.62	0.92	0.65
8	40	8	40	0.69	0.91	0.65
9	40	8	50	0.64	0.92	0.63
10	40	8	50	0.55	0.91	0.63
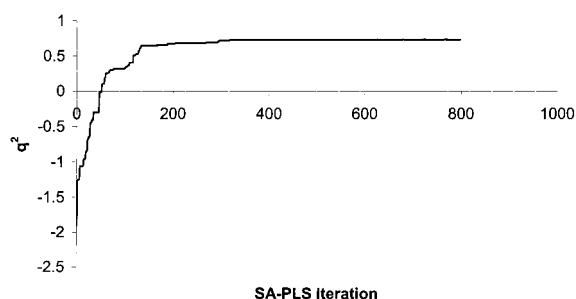
Figure 7. Trajectory of the SA-driven optimization of q^2 values in developing the AP_SA-PLS model 1.
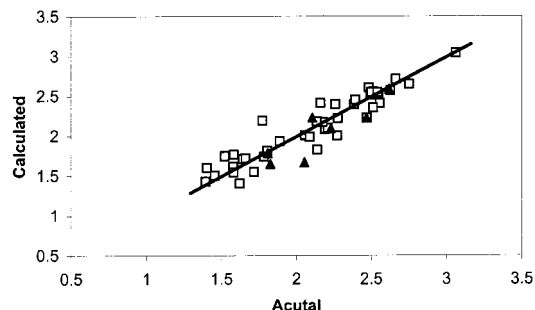
Figure 8. Actual vs calculated activity values for the training (squares) and test (triangles) sets using AP_SA-PLS model 1 ($q^2 = 0.66$, $R^2 = 0.77$).

Table 8. Comparison of Actual and Predicted Activities (log ED_{50})[a] for the External Test Set Compounds

compd	actual activity	AP_kNN predicted	MZ_kNN predicted	AP_SA-PLS predicted
49	2.63−3.12	2.33	2.37	2.41
50	2.19	2.33	2.37	2.41
51	2.10−2.63	2.33	2.06	2.32
52	2.08−2.60	2.48	2.02	2.16
53	2.53	2.38	2.03	2.16
54	2.63−3.11	2.84	2.60	2.67
55	2.58−3.06	2.83	2.76	2.68
56	2.10−2.63	2.40	2.42	2.68
57	2.58−3.06	2.52	2.81	2.57
58	2.58−3.06	2.52	2.19	2.36
59	2.16−2.69	2.32	2.42	2.75
60	2.16−2.69	2.32	2.33	2.08
61	2.75−3.23	2.24	2.39	3.16

[a] ED_{50} values are given in μmol/kg units.

values were determined as a range of values (e.g., 100−300 mg/kg). All of the external compounds displayed moderate activity that ranged between 2.08 and 3.22 log ED_{50} units (Table 8), while the original training set compounds (table in Supporting Information) featured log ED_{50} activity values that ranged between 1.39 and

3.06 log ED_{50} units (smaller log ED_{50} value means more active compound).

Previous studies with the external validation of QSAR models showed that the most reliable results are usually obtained by averaging predictions from multiple QSAR models. Table 8 lists the averaged predicted activity values for the external data set obtained from the 10 best AP_kNN, MZ_kNN, and AP_SA-PLS models. Since the experimental ED_{50} values were determined as a range of values, we discuss the accuracy of prediction using our models in qualitative terms.

Analysis of the data reported in Table 8 suggests that all three models afforded reasonable results. The activities predicted with the AP_kNN method fell within the observed range of activities for compounds **51, 52, 54−57, 59**, and **60** and close to the lowest boundary of activity predicted for compound **58**. Furthermore, AP_kNN models predicted the activity for compounds **50** and **53**, with an absolute value of residuals less than 0.20. Both MZ_kNN and AP_SA-PLS models performed slightly worse with respect to the same compounds, a result that mirrored the findings for the original 48 FAA. The activities for the least active compounds **49** and **61** were somewhat better predicted with the SA-PLS method compared to the kNN method. This result may be due to the difference in the two QSAR approaches in terms of their predictive ability. The kNN method determines the activity of an external compound by interpolating the activity of selected compounds from the training set. Accordingly, the predicted activity always lies within the range of activities for the training set. Alternatively, SA-PLS is a linear optimization approach that permits extrapolation of the predicted activity outside the training set values. These considerations may explain why the SA-PLS method was able to predict the activity of compounds **49** and **61** (the log ED_{50} value for the latter compound is actually higher than for any compound in the original dataset). In summary, results for the external datasets paralleled those obtained in model building and validation using internal training and test sets, with the AP_kNN method being the most successful. This observation emphasizes that QSAR models should be validated internally to choose the most successful method before embarking on the prediction of an external data set.

It was interesting to analyze the performance of QSAR models with respect to distinctive chemical modifications implied in the design of the external set compounds. The FAA compounds can be divided into three substructures: the N-terminus, the central amino acid with the C(2) substituent, and the C-terminus.

Inspection of the external data set showed that the QSAR models successfully predicted trends in the activity change resulting from modifications within two structural FAA substructures. In compounds **56**−**58**, the acetyl moiety (N-terminus) in **12** was either decreased in size (**56**) or increased (**57**, **58**). We observed that either change led to a loss in biological activity, a result in agreement with the QSAR models (Table 8). Similarly, we tested whether nitrogen alkylation (N- and C-termini) affected anticonvulsant activity. Our QSAR models (Table 8) predicted that N(1) alkylation (N-terminus) (R_2 = alkyl) would not appreciably alter drug activity for those compounds with or without an N(1) acetyl unit (**51**−**53**, **60**) while N(2) alkylation (C-terminus) (R_4 = alkyl) would adversely affect anticonvulsant activity (**54**, **55**). These results mirrored those obtained in the MES-seizure assay (Table 2). Less satisfactory results were obtained for compounds containing C(2)-alkyl substitutions (R_3). The QSAR models predicted moderate activities for **49** and **50** (Table 8). We observed moderate activity for **49** in the MES seizure test but excellent activity for **50** (ED_{50} = 38.4 mg/kg), showing that a substituted hereroatom one atom removed from the chiral FAA carbon was not essential for anticonvulsant activity.

Visualization of the QSAR Results. At the outset, we discussed advantages (such as speed and automation) of 2D QSAR methods compared with 3D QSAR approaches. However, QSAR models developed with 2D descriptors generally lack the ability of 3D QSAR based models (such as CoMFA) to visualize the results of modeling with respect to underlying chemical structures (cf., the graphical 3D contours typically obtained with CoMFA). This insufficiency is largely due to the different nature of descriptors used in 2D QSAR studies. Molecular field descriptors used in 3D QSAR modeling are derived directly from 3D representation of chemical structures and consequently are easy to visualize. Uncovering chemical structural features that give rise to particular values of molecular topological descriptors used in our studies (i.e., descriptors calculated for the entire molecular structure as opposed to atom-based descriptors such as atom E-state indices[50]) is practically impossible. However, we can indeed uncover substructures, which are essential with respect to underlying biological activity, using AP descriptors. Unlike molecular topological indices, AP descriptors represent counts of specific molecular features, or chemical subgraphs, found in different molecules. Therefore, it should be feasible to map specific AP descriptors selected by a successful variable selection AP-kNN model onto underlying chemical structures and then to analyze these AP features in terms of their relative contribution to a compound's activity.

An example of such an analysis is shown in Figure 9 for descriptors implicated in the best AP_kNN model 1 (Table 5). This model is based on 16 selected AP descriptors, which afforded both high internal (q^2 = 0.79) and external (R^2 = 0.90) accuracy. These descriptors represent absolute counts of 16 specific types of atom pairs that occur in molecules with different anticonvulsant activities. The descriptors can be analyzed in terms of the number of times they are present in the more active vs the less active molecules, and those
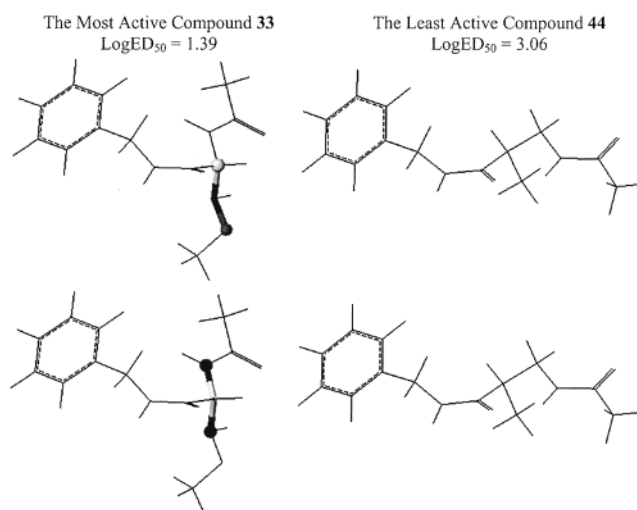
Figure 9. Visualization of important AP descriptors. The most active compound **33** features two atom pairs, O−C_03 (top panel, oxygen atom separated from carbon atom by a graph path with the length of 3) and N−N_03 (bottom panel, two nitrogen atoms separated by a path of 3), which are absent in the least active compound **44**.

descriptors that occur with the highest frequency in the most active molecules can then be visualized. Several of these AP descriptors are shown in Figure 9 for the most active compound **33** compared to the least active compound **44.** For instance, the O−C_03 atom pair (oxygen and sp^3 carbon atoms separated by the chemical graph path of length 3) is found in the most active compound **33** but not in **44**. Similarly, the N−N_03 atom pair occurs in compound **33** but not in **44**. These observations suggest that increasing the number of these molecular features may potentially improve a given compound's activity. The example provides an illustration of how 2D QSAR results can be visualized and even used to design new potentially active compounds.

Conclusions and Future Studies

In this study, we have developed and thoroughly validated QSAR models for a series of FAA anticonvulsants. In agreement with our earlier observations, we have demonstrated that the high value of leave-one-out cross-validated R^2 (q^2) obtained for the training set does not ensure the predictive power of the QSAR model. We showed that the AP_kNN approach was particularly successful in generating models with high internal and external accuracy. These models can be further exploited for the design and discovery of new, potent anticonvulsant agents.

Database mining is an obvious future application of our QSAR models. We plan to search available chemical databases (such as the National Cancer Institute database NCISMA99[51]) for compounds with high predicted anticonvulsant activity. In addition, we will extend our analysis of AP features selected by the kNN method to aid in the design of new FAA anticonvulsants. The approaches employed in this paper can be adopted for other anticonvulsants as well as for QSAR studies of other datasets.

Experimental Section

General Methods. Melting points were determined in open capillary tubes using a Thomas-Hoover melting point apparatus and are uncorrected. Infrared spectra (IR) were run on an ATI Mattson Genesis FTIR spectrometer. Absorption values are expressed in wavenumbers (cm^{-1}). Nuclear magnetic resonance spectra were measured at 300 MHz for ^{1}H NMR and at 75 MHz for ^{13}C NMR on either General Electric QE-300 NMR or Varian Gemini 2000 spectrometers. Chemical shifts (δ) are reported in parts per million (ppm) downfield from tetramethylsilane. Low-resolution mass spectra (CI+) were obtained with a Varian MAT CH-5 spectrometer by Dr. M. Moini at the University of Texas—Austin. The high-resolution chemical ionization mass spectra were obtained from a Finnigan MAT TSQ-70 by Dr. M. Moini at the University of Texas—Austin. Microanalyses were provided by Atlantic Microlab, Inc. (Norcross, GA). Analytical thin-layer chromatography (TLC) was performed on precoated silica gel GHLF microscope slides (2.5 cm × 10 cm; Analtech no. 21521). All column chromatography separations were performed on Merck silica gel (SiO$_2$) (grade 9385, 230–400 mesh, 60 Å). Yields reported are for purified products and were not optimized.

Chemical Synthesis. General Procedure for the Preparation *N*-Benzylamide Amino Acid Derivatives Using the Mixed Anhydride Coupling (MAC) Method (Method A).[7j,38] A dry THF solution of carboxylic acid (0.5–2.0 M) was cooled to −78 °C under Ar, and 4-methylmorpholine (1.1–1.25 equiv) was added. After the mixture was stirred (2 min), isobutyl chloroformate (1.1–1.25 equiv) was added, leading to the precipitation of a white solid. The reaction was allowed to proceed for an additional 2 min, and then benzylamine (1.1–1.25 equiv) was added at −78 °C. The reaction mixture was allowed to stir at room temperature (30 min to 3 h), and then the insoluble salts were filtered. The organic layer was concentrated in vacuo, and the product was purified by column chromatography on SiO$_2$ gel.

(*R,S*)-*N*-Benzyl-2-acetamidobutanamide (49). Utilizing method A and using **62** (3.50 g, 24.1 mmol), THF (50 mL), 4-methylmorpholine (3.2 mL, 29.0 mmol), isobutyl chloroformate (3.8 mL, 29.0 mmol), and benzylamine (3.8 mL, 29.0 mmol) gave crude **49**. The product was purified by column chromatography (SiO$_2$; 3:1 hexanes/EtOAc) to obtain 4.12 g (90%) of pure **49** as a white solid: mp 133–134 °C; R_f 0.57 (1:9 MeOH/CHCl$_3$); IR (KBr) 3288, 3062, 2962, 2931, 1633, 1541, 1297 cm^{-1}; ^{1}H NMR (CDCl$_3$) δ 0.89 (t, J = 7.5 Hz, CH$_2$**CH$_3$**), 1.58–1.67 (m, C**H**H'CH$_3$), 1.73–1.87 (m, CH**H'**CH$_3$), 1.88 (s, C**H$_3$**C(O)), 4.28 (dd, J = 5.4, 14.9 Hz, C**H**H'Ph), 4.40 (dd, J = 5.9, 14.9 Hz, CH**H'**Ph), 4.54 (br app q, J = 7.5 Hz, C**H**), 7.20–7.33 (m, 5 Ph**H** and N**H**), 7.84–7.88 (m, N**H**CH$_2$); ^{13}C NMR (CDCl$_3$) δ 10.2 (CH$_2$**CH$_3$**), 22.9 (**C**H$_3$C(O)), 26.1 (**C**H$_2$-CH$_3$), 43.4 (**C**H$_2$Ph), 54.5 (**C**H), 127.3 (**C$_{4'}$**), 127.7 (2**C$_{2'}$** or 2**C$_{3'}$**), 128.6 (2**C$_{2'}$** or 2**C$_{3'}$**), 138.2 (**C$_{1'}$**), 170.6 (CH$_3$**C**(O) or CH**C**(O)), 172.3 (CH$_3$**C**(O) or CH**C**(O)) ppm; MS (+CI) (rel intensity) 236 (18), 235 (M$^+$ + 1, 100); M_r (+CI) 235.143 77 [M$^+$ + 1] (calcd for C$_{13}$H$_{19}$N$_2$O$_2$ 235.144 65). Anal. (C$_{13}$H$_{18}$N$_2$O$_2$) C, H, N.

(*R,S*)-*N*-Benzyl-2-*N*-(benzyloxycarbonyl)-2-aminopentanamide (64). Utilizing method A and using a 1:1 mixture of D-**63** (1.76 g, 7.02 mmol (Novabiochem)) and L-**63** (1.76 g, 7.02 mmol (Novabiochem)), 4-methylmorpholine (1.9 mL, 17.6 mmol), isobutyl chloroformate (2.3 mL, 17.6 mmol), and benzylamine (1.9 mL, 17.6 mmol) gave crude **64**. The product was purified by column chromatography (SiO$_2$; 1:1 EtOAc/hexanes) to obtain 4.37 g (91%) of pure **64** as a white solid: mp 138–139 °C; R_f 0.49 (1:1 EtOAc/hexanes); IR (KBr) 3296, 2958, 2930, 1688, 1641, 1537, 1464, 1256, 1229, 1056 cm^{-1}; ^{1}H NMR (CDCl$_3$) δ 0.93 (t, J = 7.8 Hz, CH$_2$**CH$_3$**), 1.38 (app sext, J = 7.8 Hz, C**H$_2$**CH$_3$), 1.58–1.66 (m, CHC**H**H'), 1.82–1.87 (m, CHCH**H'**), 4.14 (app q, J = 7.4 Hz, C**H**), 4.44 (d, J = 5.5 Hz, C**H$_2$**Ph), 5.10 (s, OC**H$_2$**Ph), 5.19–5.22 (m, N**H**), 6.20–6.23 (m, N**H**), 7.23–7.34 (m, 10 Ph**H**); ^{13}C NMR (CDCl$_3$) δ 13.7 (**C**H$_3$), 18.8 (**C**H$_2$CH$_3$), 34.8 (CH**C**H$_2$), 43.6 (**C**H$_2$Ph), 55.1

(**C**H), 67.1 (O**C**H$_2$Ph), 127.6, 127.7, 128.1, 128.3, 128.6, 128.8, 136.2, 137.9 (2 **C$_6$**H$_5$), 156.2 (N**C**(O)O), 171.7 (CH**C**(O)) ppm; MS (+CI) (rel intensity) 342 (21), 341 (M$^+$ + 1, 100), 233 (29); M_r (+CI) 341.186 51 [M$^+$ + 1] (calcd for C$_{20}$H$_{25}$N$_2$O$_3$ 341.186 52). Anal. (C$_{20}$H$_{24}$N$_2$O$_3$) C, H, N.

(*R,S*)-*N*-Benzyl-2-aminopentanamide (59). A methanolic solution (75 mL) of **64** (4.30 g, 12.6 mmol) was hydrogenated (1 atm) in the presence of 5% Pd–C (∼300 mg) at room temperature (4 h). The mixture was filtered over a bed of Celite (521), and the clear filtrate was evaporated in vacuo to yield **59** (2.68 g, 99%) as a clear oil: R_f 0.39 (1:19 MeOH/CHCl$_3$); IR (neat) 3299 (br), 3066, 2954, 2872, 1654, 1531, 1455, 1252 cm^{-1}; ^{1}H NMR (CDCl$_3$) δ 0.93 (t, J = 7.2 Hz, CH$_2$**CH$_3$**), 1.32–1.55 (m, C**H$_2$**CH$_3$, N**H$_2$**, and CHC**H**H'), 1.77–1.88 (m, CHCH**H'**), 3.37 (dd, J = 4.4, 8.0 Hz, C**H**), 4.41 (d, J = 6.3 Hz, C**H$_2$**Ph), 7.23–7.34 (m, 5 Ph**H**), 7.62–7.75 (br s, N**H**); ^{13}C NMR (CDCl$_3$) δ 13.9 (**C**H$_3$), 19.1 (**C**H$_2$CH$_3$), 37.3 (CH**C**H$_2$), 43.1 (**C**H$_2$Ph), 55.1 (**C**H), 127.3 (**C$_{4'}$**), 127.7 (2**C$_{2'}$** or 2**C$_{3'}$**), 128.6 (2**C$_{2'}$** or 2**C$_{3'}$**), 138.7 (**C$_{1'}$**), 175.3 (**C**(O)NH) ppm, the ^{13}C NMR assignments were in agreement with the DEPT experiment; MS (+CI) (rel intensity) 208 (16), 207 (M$^+$ + 1, 100); M_r (+CI) 207.149 60 [M$^+$ + 1] (calcd for C$_{12}$H$_{19}$N$_2$O 207.149 74). Anal. (C$_{12}$H$_{18}$N$_2$O·0.25H$_2$O) C, H, N.

General Procedure for the Preparation of *N*-Acyl-Substituted Amino Acid Derivatives (Method B). To a dry THF solution of amine (∼0.1–1 M) were successively added TEA (1–1.2 equiv) and the acyl chloride or the acid anhydride (1–2 equiv). The reaction mixture was stirred at room temperature (5–60 min for acyl chlorides, 1 h to 1 day for acid anhydrides). The organic layer was concentrated in vacuo, and the product was purified by column chromatography on SiO$_2$ gel.

(*R,S*)-*N*-Benzyl-2-acetamidopentanamide (50). Utilizing method B and using **59** (1.42 g, 6.89 mmol), THF (70 mL), TEA (2.0 mL, 1.4 mmol), Ac$_2$O (2.0 mL, 21.4 mmol), and a catalytic amount of DMAP (∼100 mg) gave crude **50** (room temperature, 14 h). The product was purified by column chromatography (SiO$_2$; 1:19 MeOH/CHCl$_3$) to obtain 1.50 g (88%) of pure **50** as a white solid: mp 138–139 °C; R_f 0.62 (1:19 MeOH/CHCl$_3$); IR (KBr) 3284, 3072, 2957, 1637, 1546, 1438, 1383, 1275 cm^{-1}; ^{1}H NMR (CDCl$_3$) δ 0.88 (t, J = 7.2 Hz, CH$_2$**CH$_3$**), 1.26–1.38 (m, C**H$_2$**CH$_3$), 1.52–1.64 (m, CHC**H**H'), 1.69–1.83 (m, CHCH**H'**), 1.91 (s, C**H$_3$**C(O)), 4.31 (dd, J = 5.5, 14.9 Hz, C**H**H'Ph), 4.40 (dd, J = 5.9, 14.9 Hz, CH**H'**Ph), 4.48–4.56 (m, C**H**), 6.79 (d, J = 8.4 Hz, N**H**CH), 7.21–7.34 (m, 5 Ph**H** and N**H**CH$_2$); ^{13}C NMR (CDCl$_3$) δ 14.0 (**C**H$_3$), 19.0 (**C**H$_2$CH$_3$), 23.1 (**C**H$_3$C(O)), 35.0 (CH**C**H$_2$), 43.6 (**C**H$_2$Ph), 53.2 (**C**H), 127.6 (**C$_{4'}$**), 127.8 (2**C$_{2'}$** or 2**C$_{3'}$**), 128.8 (2**C$_{2'}$** or 2**C$_{3'}$**), 138.2 (**C$_{1'}$**), 170.5 (CH$_3$**C**(O) or CH**C**(O)), 172.3 (CH$_3$**C**(O) or CH**C**(O)) ppm, the ^{13}C NMR assignments were in agreement with the HETCOR experiment; MS (+CI) (rel intensity) 250 (14), 249 (M$^+$ + 1, 100), 108 (45); M_r (+CI) 249.159 32 [M$^+$ + 1] (calcd for C$_{14}$H$_{21}$N$_2$O$_2$ 249.160 30). Anal. (C$_{14}$H$_{20}$N$_2$O$_2$) C, H, N.

(*R,S*)-*N*-Benzyl-2-propionamido-3-methoxypropionamide (57). Compound **57** was prepared utilizing method B and using **65**[39] (2.01 g, 9.6 mmol), THF (100 mL), TEA (0.92 mL, 10.6 mmol), and propionyl chloride (1.50 mL, 10.6 mmol). The reaction mixture was stirred at room temperature (10 min), and then the precipitated salts were filtered and the organic layer was concentrated in vacuo. The residue was purified by column chromatography (SiO$_2$; 1:33 MeOH/CHCl$_3$) to give 1.80 g (71%) of **57** as a white solid: mp 119–120 °C; R_f 0.63 (1:9 MeOH/CHCl$_3$); IR (KBr) 3289, 3067, 2927, 2883, 1637, 1547, 1455, 1226, 1129 cm^{-1}; ^{1}H NMR (CDCl$_3$) δ 1.12 (t, J = 7.6 Hz, CH$_2$**CH$_3$**), 2.24 (q, J = 7.6 Hz, C**H$_2$**CH$_3$), 3.35 (s, OC**H$_3$**), 3.45 (dd, J = 7.1, 9.2 Hz, C**H**H'OCH$_3$), 3.76 (dd, J = 4.4, 9.2 Hz, CH**H'**OCH$_3$), 4.41 (dd, J = 5.7, 15.0 Hz, C**H**H'NH), 4.48 (dd, J = 5.7, 15.0 Hz, CH**H'**NH), 4.58–4.64 (m, C**H**), 6.58 (d, J = 6.9 Hz, N**H**CH), 7.02–7.06 (br s, N**H**CH$_2$), 7.23–7.34 (m, 5 Ph**H**); ^{1}H NMR (DMSO-d_6) δ 0.98 (t, J = 7.5 Hz, CH$_2$**CH$_3$**), 2.17 (q, J = 7.5 Hz, C**H$_2$**CH$_3$), 3.25 (s, OC**H$_3$**), 3.45–3.55 (m, C**H$_2$**OCH$_3$), 4.29 (d, J = 6.0 Hz, C**H$_2$**NH), 4.47–4.53 (m, C**H**), 7.20–7.33 (m, 5 Ph**H**), 7.94 (d, J = 8.1 Hz, N**H**CH), 8.42 (t, J = 6.0 Hz, N**H**CH$_2$), the assignments were in agreement with

the ^{1}H$-^1$H COSY experiment; ^{13}C NMR (CDCl$_3$) δ 9.8 (CH$_2$**C**H$_3$), 29.6 (**C**H$_2$CH$_3$), 43.6 (**C**H$_2$Ph), 52.4 (**C**H), 59.2 (O**C**H$_3$), 72.0 (**C**H$_2$OCH$_3$), 127.6 (**C**$_{4'}$ and 2**C**$_{2'}$ or 2**C**$_{3'}$), 128.8 (2**C**$_{2'}$ or 2**C**$_{3'}$), 138.1 (**C**$_{1'}$), 170.3 (CH$_2$**C**(O) or CH**C**(O)), 174.2 (CH$_2$**C**(O) or CH**C**(O)) ppm; MS (+CI) (rel intensity) 266 (14), 265 (M$^+$ + 1, 100); M_r (+CI) 265.155 45 [M$^+$ + 1] (calcd for C$_{14}$H$_{21}$N$_2$O$_3$ 265.155 22). Anal. (C$_{14}$H$_{20}$N$_2$O$_3$) C, H, N.

(R,S)-N-Benzyl-2-formylamino-3-methoxypropiona-mide (56). N-(Diethylcarbamoyl)-N-methoxyformamide (TCI America) (1.27 g, 7.3 mmol) and a catalytic amount of DMAP (~100 mg) were added to a THF (65 mL) solution of **65**[39] (1.38 g, 6.6 mmol) and TEA (925 μL, 6.6 mmol), and the reaction solution was stirred at 40 °C (6 h). The solvent was evaporated in vacuo, and the residue was purified by column chromatography (SiO$_2$; 1:33 MeOH/CHCl$_3$). The isolated pale-yellow solid was recrystallized from EtOH to give 1.45 g (93%) of **56** as a white solid: mp 119−120 °C; R_f 0.45 (1:9, MeOH/CHCl$_3$); IR (KBr) 3286, 3101, 3039, 2885, 1635 (br), 1558, 1450, 1389, 1219, 1126 cm^{-1}; ^{1}H NMR (CDCl$_3$) δ 3.39 (s, OC**H**$_3$), 3.46 (dd, J = 8.1, 8.8 Hz, C**H**H′OCH$_3$), 3.85 (dd, J = 4.1, 8.8 Hz, CH**H**′OCH$_3$), 4.45 (dd, J = 5.4, 14.9 Hz, C**H**H′NH), 4.51 (dd, J = 6.2, 14.9 Hz, CH**H**′NH), 4.59−4.65 (m, C**H**), 6.58−6.61 (m, N**H**CH), 6.71−6.75 (br s, N**H**CH$_2$), 7.24−7.36 (m, 5 Ph**H**), 8.24 (s, **H**C(O)); ^{13}C NMR (CDCl$_3$) δ 43.9 (**C**H$_2$Ph), 51.4 (**C**H), 59.4 (O**C**H$_3$), 71.7 (**C**H$_2$OCH$_3$), 127.7 (**C**$_{4'}$), 127.9 (2**C**$_{2'}$ or 2**C**$_{3'}$), 129.0 (2**C**$_{2'}$ or 2**C**$_{3'}$), 138.0 (**C**$_{1'}$), 161.2 (H**C**(O)), 169.6 (**C**(O)-NH) ppm; MS (+CI) (rel intensity) 238 (12), 237 (M$^+$ + 1, 100), 209 (14); M_r (+CI) 237.124 19 [M$^+$ + 1] (calcd for C$_{12}$H$_{17}$N$_2$O$_3$ 237.123 92). Anal. (C$_{12}$H$_{16}$N$_2$O$_3$) C, H, N.

(R,S)-N-Benzyl-2-acryloylamino-3-methoxypropiona-mide (58). Utilizing method B and using a THF solution (130 mL) of **65**[39] (1.80 g, 8.65 mmol), TEA (1.45 mL, 10.4 mmol), and acryloyl chloride (844 μL, 10.4 mmol) gave a reaction mixture that was stirred at 0 °C (1 h). The precipitated salts were filtered, and the organic layer was concentrated in vacuo to provide a pale-yellow residue. The product was purified by column chromatography (SiO$_2$; 1:19 MeOH/CHCl$_3$) and then triturated with hot Et$_2$O to give 1.88 g (83%) of pure **58** as a white solid: mp 138−139 °C; R_f 0.63 (1:19 MeOH/CHCl$_3$); IR (KBr) 3294, 3032, 2931, 1736, 1651, 1543, 1365, 1219, 1126 cm^{-1}; ^{1}H NMR (CDCl$_3$) δ 3.36 (s, OC**H**$_3$), 3.50 (dd, J = 7.1, 9.2 Hz, C**H**H′OCH$_3$), 3.81 (dd, J = 4.1, 9.2 Hz, CH**H**′OCH$_3$), 4.41 (dd, J = 5.8, 15.1 Hz, C**H**H′Ph), 4.49 (dd, J = 6.0, 15.1 Hz, CH**H**′Ph), 4.67−4.73 (m, C**H**), 5.65 (dd, J = 1.8, 9.9 Hz, C**H**H′=CH), 6.16 (dd, J = 9.9, 16.8 Hz, CH**H**′=CH), 6.28 (dd, J = 1.8, 16.8 Hz, CH**H**′=CH), 6.85 (d, J = 7.2 Hz, N**H**CH), 7.05−7.10 (m, N**H**CH$_2$), 7.26−7.33 (m, 5 Ph**H**); ^{13}C NMR (CDCl$_3$) δ 43.7 (**C**H$_2$NH), 52.7 (**C**H), 59.2 (O**C**H$_3$), 72.0 (**C**H$_2$OCH$_3$), 127.5 (**C**H=CH$_2$), 127.6, 128.8 (**C**$_{4'}$, 2**C**$_{2'}$, and 2**C**$_{3'}$), 131.4 (CH=**C**H$_2$), 138.0 (**C**$_{1'}$), 165.7 (**C**(O)CH=CH$_2$), 170.1 (**C**(O)CH) ppm, the ^{13}C NMR assignments were in agreement with the DEPT experiment; MS (+CI) (rel intensity) 264 (17), 263 (M$^+$ + 1, 100); M_r (+CI) 263.139 88 [M$^+$ + 1] (calcd for C$_{14}$H$_{19}$N$_2$O$_3$ 263.139 57). Anal. (C$_{14}$H$_{18}$N$_2$O$_3$·0.2H$_2$O) C, H, N.

Pharmacology. New compounds were screened under the auspices of the National Institutes of Health's Anticonvulsant Screening Project. Experiments were performed with male albino Carworth Farms no. 1 mice (intraperitoneal route, ip). The mice weighed between 18 and 25 g. All animals had free access to feed and water except during the actual testing period. Housing, handling, and feeding were all in accordance with recommendations contained in the "Guide for the Care and Use of Laboratory Animals". All of the test compounds were administrated in suspensions of 0.5% (w/v) methyl-cellulose in water. The volumes administered were 0.01 mL/g of body weight. Anticonvulsant activity was established using the maximal electroshock (MES) test.[52] For the MES test, a drop of electrolyte solution with an anesthetic (0.5% butacaine hemisulfate in 0.9% sodium chloride) was placed in the eyes of the animals prior to positioning the corneal electrodes and delivery of a nonlethal current. A 60-cycle alternating current was administered for 0.2 s, utilizing 50 mA. Protection end points were defined as the abolition of the hind limb tonic extensor component of the induced seizure.[53] In the mouse

identification screens, all compounds were administered at three dose levels (30, 100, 300 mg/kg) and two time periods (0.5 and 4 h). Typically, in the MES seizure test, one animal was used at 30 and 300 mg/kg and three animals were used at 100 mg/kg. The quantitative determination of the median effective (ED$_{50}$) dose was conducted at a previously calculated time of peak effect using the ip route in mice. Groups of at least eight animals were tested using different doses of test compound until at least two points were determined between 100% and 0% protection and minimal motor impairment. The dose of the candidate substance required to produce the desired end point (abolition of hindlimb tonic extensor component) in 50% of the animals in each test and the 95% confidence interval were calculated by a computer program based on methods described by Finney.[54]

Supporting Information Available: Table showing the predicted vs actual biological activity values for each compound in the training and test sets, as determined by all three types of QSAR models. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Evans, J. H. Post-traumatic epilepsy. *Neurology* **1962**, *12*, 665−674.

(2) Lindsay, J. M. Genetics and epilepsy. *Epilepsia* **1971**, *12*, 47−54.

(3) Rogawski, M. A.; Porter, R. J. Antiepileptic drugs: pharmacological mechanisms and clinical efficacy with consideration of promising developmental stage compounds. *Pharmacol. Rev.* **1997**, *42*, 223−286.

(4) Brodie, M. J.; Dichter, M. A. Antiepileptic drugs. *N. Engl. J. Med.* **1996**, *334*, 168−175.

(5) Dichter, M. A.; Brodie, M. J. New Antiepileptic drugs. *N. Engl. J. Med.* **1996**, *34*, 1583−1590.

(6) Mattson, R. H.; Cramer, J. A.; Collins, J. F.; Smith, D. B. Comparison of carbamazepine, phenobarbital, phenytoin, and primidone in partial and secondary generalized tonic-clonic seizures. *N. Engl. J. Med.* **1985**, *313*, 145−151.

(7) (a) Kohn, H.; Conley, J. D. New antiepileptic agents. *Chem. Br.* **1988**, *24*, 231−233. (b) 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. (c) Conley, J. D.; Kohn, H. Functionalized DL-amino acid derivatives. Potent new agents for the treatment of epilepsy. *J. Med. Chem.* **1987**, *30*, 567−574. (d) Kohn, H.; Conley, J. D.; Leander, J. D. Marked stereospecificity in a new class of anticonvulsants. *Brain Res.* **1988**, *457*, 371−375. (e) 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. (f) Kohn, H.; Sawhney, K. N.; LeGall, P.; Robertson, D. W.; Leander, J. D. Preparation and anticonvulsant activity of a series of functionalized α-heteroaromatic-substituted amino acids. *J. Med. Chem.* **1991**, *34*, 2444−2452. (g) 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. (h) Kohn, H.; Sawhney, K. N.; Robertson, D. W.; Leander, J. D. Anticonvulsant properties of N-substituted α,α-diamino acid derivatives. *J. Pharm. Soc.* **1994**, *83*, 689−691. (i) Bardel, P.; Bolanos, A.; Kohn, H. Synthesis and anticonvulsant activities of α-acetamido-N-benzylacetamide derivatives containing an electron-deficient α-heteroaromatic substituent. *J. Med. Chem.* **1994**, *37*, 4567−4571. (j) Choi, D.; Stables, J. P.; Kohn, H. Synthesis and anticonvulsant activities of N-benzyl-2-acetamidopropionamide derivatives. *J. Med. Chem.* **1996**, *39*, 1907−1916.

(8) (a) Paruszewski, R.; Rostafinska-Suchar, G.; Strupinska, M.; Jaworski, P.; Stables, J. P. Synthesis and anticonvulsant activity of some amino acid derivatives. *Pharmazie* **1996**, *3*, 145–148. (b) Paruszewski, R.; Rostafinska-Suchar, G.; Strupinska, M.; Jaworski, P.; Winiecka, I.; Stables, J. P. Synthesis and anticonvulsant activity of some amino acid derivatives. *Pharmazie* **1996**, *51*, 212–215. (c) Paruszewski, R.; Rostafinska-Suchar, G.; Strupinska, M.; Winiecka, I.; Stables, J. P. Synthesis and anticonvulsant activity of some amino acid derivatives. *Pharmazie* **2000**, *55*, 27–30.

(9) Ho, B.; Venkatarangan, P. M.; Cruse, S. F.; Hinko, C. N.; Andersen, P. H.; Crider, A. M.; Adloo, A. A.; Roane, D. S.; Stables, J. P. Synthesis of 2-piperidinecarboxylic acid derivatives as potential anticonvulsants. *Eur. J. Med. Chem.* **1998**, *33*, 23–31.

(10) 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.

(11) Swinyard, E. A.; Woodhead, J. H.; White, H. S.; Franklin, M. R. In *Antiepileptic Drugs*, 3rd ed.; Levy, R. H., Driefuss, F. E., Mattson, R. H., Meldrum, B. S., Perry, J. K., Eds.; Raven Press: New York, 1989; pp 85–102.

(12) Levy, R. H.; Mattson, R.; Meldrum, B. *Antiepileptic Drugs*, 4th ed.; Raven Press: New York, 1995; Chapter 6.

(13) Netzeva, T.; Doytchinova, I.; Natcheva, R. 2D and 3D QSAR analysis of some valproic acid metabolites and analogues as anticonvulsant agents. *Pharm. Res.* **2000**, *17*, 727–732.

(14) Trapani, G.; Latrofa, A.; Franco, M.; Altomare, C.; Sanna, E.; Usala, M.; Biggio, G.; Liso, G. Propofol analogues. Synthesis, relationships between structure and affinity at $GABA_A$ receptor in rat brain, and differential electrophysiological profile at recombinant human $GABA_A$ receptors. *J. Med. Chem.* **1998**, *41*, 1846–1854.

(15) Hays, S. J.; Rice, M. J.; Ortwine, D. F.; Johnson, G.; Schwarz, R. D.; Boyd, D. K.; Copeland, L. F.; Vartanian, M. G.; Boxer, P. A. Substituted 2-benzothiazolamines as sodium flux inhibitors: Quantitative structure–activity relationships and anticonvulsant activity. *J. Pharm. Sci.* **1994**, *83*, 1425–1432.

(16) Bikker, J. A.; Kubanek, J.; Weaver, D. F. Quantum pharmacologic studies applicable to the design of anticonvulsants: Theoretical conformational analysis and structure–activity studies of barbiturates. *Epilepsia* **1994**, *35*, 411–425.

(17) (a) Lopez-Rodriguez, M. L.; Rosado, M. L.; Benhamu, B.; Morcillo, M. J.; Fernandez, E.; Schaper, K. J. Synthesis and structure–activity relationships of a new model of arylpiperazines. 2. Three-dimensional quantitative structure–activity relationships of hydantoin–phenylpiperazine derivatives with affinity for $5-HT_{1A}$ and α_1 receptors. A comparison of CoMFA models. *J. Med. Chem.* **1997**, *40*, 1648–1656. (b) Brown, M. L.; Zha, C. C.; Van Dyke, C. C.; Brown, G. B.; Brouillette, W. J. Comparative molecular field analysis of hydantoin binding to the neuronal voltage-dependent sodium channel. *J. Med. Chem.* **1999**, *42*, 1537–1545.

(18) Knight, J. L.; Weaver, D. F. A computational quantitative structure–activity relationship study of carbamate anticonvulsants using quantum pharmacological methods. *Seizure* **1998**, *7*, 347–354.

(19) Lopez-Rodriguez, M. L.; Morcillo, M. J.; Fernandez, E.; Rosado, M. L.; Pardo, L.; Schaper, K. J. Synthesis and structure–activity relationships of a new model of arylpiperazines. 6. Study of the $5-HT_{1A}/\alpha_1$-adrenergic receptor affinity by classical Hansch analysis, artificial neural networks, and computational simulation of ligand recognition. *J. Med. Chem.* **2001**, *44*, 198–207.

(20) Estrada, E.; Peña, A. In silico studies for the rational discovery of anticonvulsant compounds. *Bioorg. Med. Chem.* **2000**, *8*, 2755–2770.

(21) Allen, F. H.; Davies, J. E.; Galloy, J. J.; Johnson, O.; Kennard, O.; Macrae, C. F.; Mitchell, E. M.; Mitchell, G. F.; Smith, J. M.; Watson, D. G. The Development of Versions 3 and 4 of the Cambridge Structural Database System. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 187–204.

(22) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. The Cambridge Crystal Data Centre: Computer-Based Search, Retrieval, Analysis, and Display of Information. *Acta Crystallogr., Sect B: Struct., Crystallogr. Cryst. Chem.* **1979**, *B35*, 2331–2339.

(23) Rusinko, A., III; Skell, J. M.; Balducci, R.; McGarity, C. M.; Pearlman, R. S. *Concord, a Program for the Rapid Generation of High Quality Approximate 3-Dimensional Molecular Structures*; The University of Texas at Austin and Tripos Associates: St. Louis, MO, 1988.

(24) Pearlman, R. S. Rapid generation of high quality approximate 3D molecular structures. *Chem. Des. Autom. News* **1987**, *2*, 1–6.

(25) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. In *Computer-Assisted Drug Design*; Olson, E. C. Christoffersen, R. E.. Eds.; American Chemical Society, Washington, DC, 1979; Vol. 112, pp 205–226.

(26) Marshall, G. R.; Cramer, R. D., III Three-dimensional structure–activity relationships. *TIPS Rev.* **1988**, *9*, 285–289.

(27) Kubinyi, H., Folkers, G., Martin, Y. C., Eds. 3D QSAR in drug design, Kluwer/ESCOM: Dordrecht, The Netherlands, 1998; Vols. 2, 3.

(28) Cramer, R. D., III; Patterson, D. E.; Bunce, J. D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

(29) Tropsha, A.; Zheng, W. Identification of the descriptor pharmacophores using variable selection QSAR: Applications to database mining. *Curr. Pharm. Des.* **2001**, *7*, 599–612.

(30) Cho, S. J; Tropsha, A. Cross-validated R^2-guided region selection for comparative molecular field analysis: A simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060–1066.

(31) Hoffman, B. T.; Cho, S. J.; Zheng, W.; Wyrick, S.; Nichols, D. E.; Mailman, R. B.; Tropsha, A. QSAR modeling of dopamine D_1 agonists using comparative molecular field analysis, genetic algorithms–partial least squares, and *K* nearest neighbor methods. *J. Med. Chem.* **1999**, *42*, 3217–3226.

(32) 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.

(33) Golbraikh, A.; Bonchev, D.; Tropsha, A. Novel chirality descriptors derived from molecular topology. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 147–158.

(34) 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.

(35) (a) 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. (b) The method is described in detail and can be executed on the QSAR Web server at http://mmlin1.pha.unc.edu/~jin/QSAR.

(36) Golbraikh, A.; Tropsha, A. Beware of q²! *J. Mol. Graphics Modell.* **2002**, *20*, 269–276.

(37) Synge, R. L. M. Experiments on amino-acids. I. The partition of acetamino-acids between immiscible solvents. *Biochem. J.* **1939**, *33*, 1913–1917.

(38) Anderson, G. W.; Zimmerman, J. E.; Callahan, F. M. A reinvestigation of the mixed carbonic anhydride method of peptide synthesis. *J. Am. Chem. Soc.* **1967**, *87*, 5012–5017.

(39) 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.

(40) Akikusa, N.; Mitsui, K.; Sakamoto, T.; Kikugawa, Y. A new formylating reagent: *N*-(diethylcarbamoyl)-*N*-methoxyformamide. *Synthesis* **1992**, 1058–1060.

(41) The program Sybyl is available from Tripos Associates, St. Louis, MO.

(42) Randic, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.

(43) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.

(44) *Molconn-Z*, version 3.5; Hall Associates Consulting: Quincy, MA.

(45) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. Atom pairs as molecular features in structure–activity studies: definition and applications. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64–73.

(46) *GenAP Program Manual, Laboratory for Molecular Modeling*; University of North Carolina: Chapel Hill, NC.

(47) Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics*; John Wiley & Sons: New York, 1986.

(48) Dunn, W. J., III; Wold, S.; Edlund, U.; Hellberg, S.; Gasteiger, J. Multivariate structure–activity relationships between data from a battery of biological tests and an ensemble of structure descriptors: the PLS method. *Quant. Struct.–Act. Relat.* **1984**, *3*, 131–137.

(49) Gilbert, N. *Statistics*; W. B. Saunders, Co.: Philadelphia, PA, 1976.

(50) Hall, L. H.; Mohney, B. K.; Kier, L. B. The electrotopological state: an atom index for QSAR. *Quant. Struct.–Act. Relat,* **1991**, *10*, 43–51.

(51) http://search.nci.nih.gov/search97cgi/s97_cgi

(52) (a) 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. (b) Stables, J. P.; Kupferberg, H. J. The NIH anticonvulsant drug development (ADD) program: preclinical anticonvulsant screening project. In *Molecular and Cellular Targets for Antiepileptic Drugs*; Avanzini, G., Tanganelli, P., Avoli, M., Eds.; John Libbey: London, 1997; pp 191–198.

(53) White, H. S.; Woodhead, J. H.; Franklin, M. R. In *Antiepileptic Drugs*; Levy, R. H., Meldrum, B. S., Eds.; Raven: New York, 1998; pp 99−100.

(54) Finney, D. J. *Probit Analysis*, 3rd ed.; Cambridge University Press: London, 1971.