Development of a Quantitative Structure−Activity Relationship Model for Inhibition of Gram-positive Bacterial Cell Growth by Biarylamides

David T. Stanton,*,† Prakash J. Madhav, Larry J. Wilson,‡ Timothy W. Morris,§ Paul M. Hershberger, and Christian N. Parker||

Health Care Research Center, Procter & Gamble Pharmaceuticals, 8700 Mason-Montgomery Road, Mason, Ohio 45040-9462

Received July 30, 2003

A set of compounds consisting of a new and diverse collection of biarylamides was examined using quantitative structure−activity relationship techniques for the purpose of developing a model to describe inhibition of gram-positive bacterial growth (minimum inhibition concentration). The model was sought in order to obtain insight for designing new molecules. A detailed analysis of the underlying structure−activity relationship helped provide insight concerning which structural features of the molecules modulated the activity of the compounds against gram-positive organisms.

INTRODUCTION

The rapid emergence of drug-resistant bacterial pathogens emphasizes the need for new classes of antibacterial agents.[1−4] The chemical diversity of compounds being generated using novel methods, such as combinatorial chemistry, and the use of high-throughput screening offers the opportunity to identify new classes of agents with antibacterial activity.[5] With this in mind, a number of different combinatorial libraries were screened to identify new lead compounds using a specific bacterial cell wall synthesis assay. The initial biological activity for the biarylamides was noted in an in vitro high-throughput screen measuring the inhibition of the synthesis of peptidoglycans in *Staphylococcus aureus* (*Staph. aureus*). This effort led to the identification of a series of biarylamide and related compounds that demonstrated antibacterial activity against a wide range of antibiotic resistant and sensitive bacterial pathogens in vitro.[6−9] The activity of these compounds appears to be related to their ability to inhibit peptidoglycan synthesis. Later, a second in vitro screen focused on measuring the minimum inhibition concentration (MIC) for a series of gram-positive organisms that included different strains of *S. aureus*, *Enterococcus faecium* (*E. faecium*), *E. faecali*, and *Streptococcus pneumoniae* (*S. pneumoniae*) and a number of gram-negative bacteria. This was done to help identify compounds that would have broad-spectrum activity against organisms of interest. While a number of the compounds did display activity against some of the gram-negative organisms, this class of compounds displayed only sporadic activity against this group of bacteria. The focus of subsequent work remained on the observed gram-positive activity.

Since combinatorial chemistry techniques were being employed in the hit follow-up process, methods for rapidly predicting the activity of large numbers of proposed synthetic targets were needed in order to facilitate the design of focused libraries. A further goal was the derivation of information from structure−activity relationships that could be used to design more potent compounds. Quantitative structure−activity relationship (QSAR) techniques were used to develop an equation that provided important SAR information.

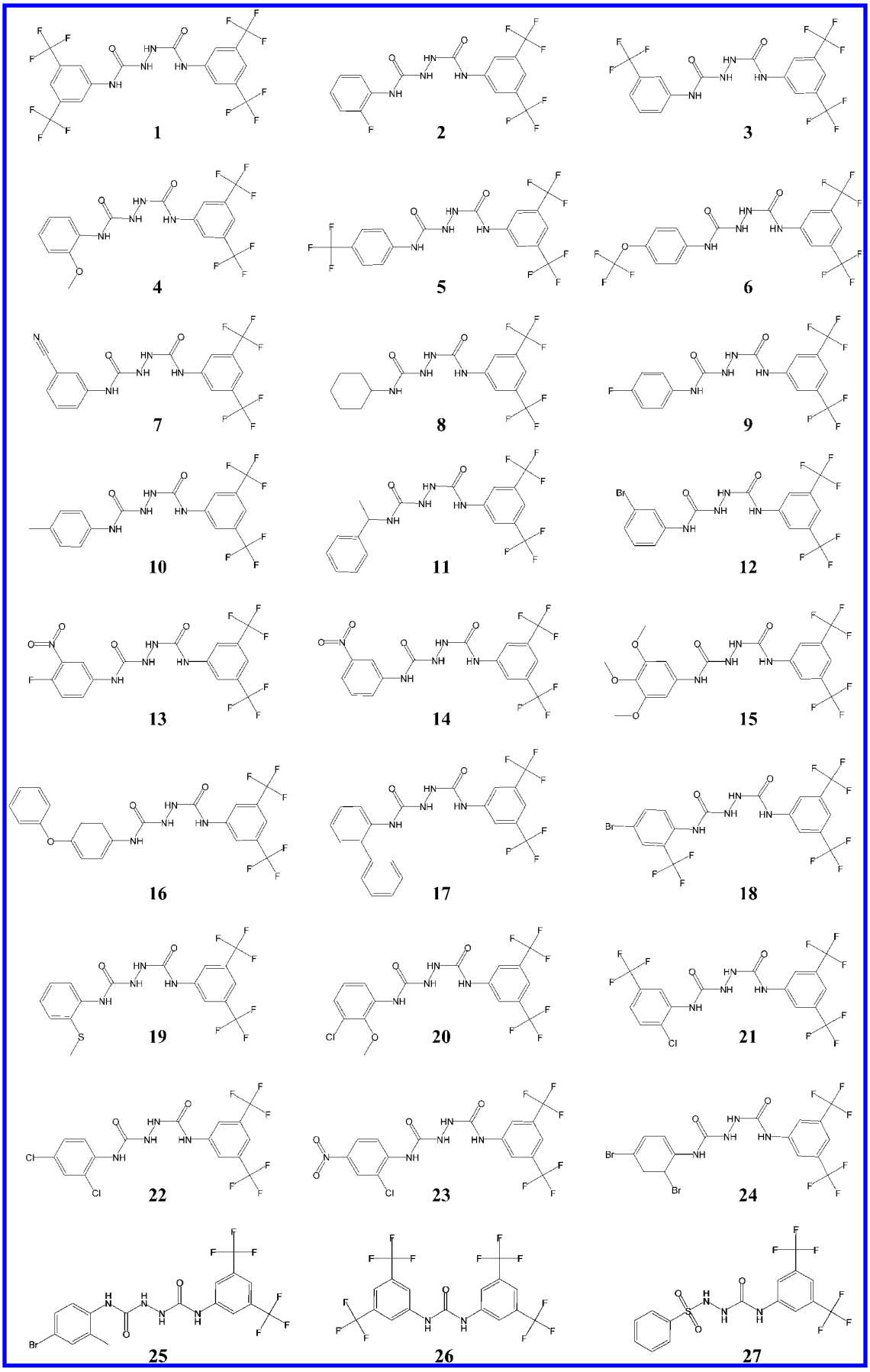
EXPERIMENTAL SECTION

Data Set. The structures for a set of 47 compounds were used as the training set for the MIC model. The structure diagrams for these compounds are shown in Figure 1. Table 1 lists the experimentally observed antibacterial activity for the training set, along with the fitted values obtained from the final model. Each structure was sketched into the computer and stored in a database using the Sybyl package.[10] An initial 3D conformation was generated for each structure using Concord.[11] The initial conformation was further refined using molecular mechanics energy minimization with use of the Tripos force field.[12] The minimization step included electrostatic terms, and the partial atomic charges required for this step were computed using the Gasteiger−Huckel method.[13] These final structures were then transferred to the ADAPT package[14,15] for the model development steps.

The biological data used in this study represent the geometric mean minimum growth inhibition concentration (MIC) for six gram-positive microorganisms: *Staph. aureus* strain MI273, *Staph. aureus* strain MI339, *E. faecium* strain EF12, *E. faecalis* strain STD44, *S. pneumoniae* strain STP51, and *S. pneumoniae* strain STP6301. This collection of microorganisms was selected because they represented a panel of clinically relevant gram-positive bacteria; including bacteria displaying resistance to examples of many important antibiotic classes, such as quinolones, vancomycin, and β-lactams. The inhibition data for the set were modeled as a mean value because it was of interest to evaluate the compounds in terms of a broad spectrum of activity, rather

* Corresponding author E-mail: stanton.dt@pg.com.
† Current address: Corporate Research Division, Chemical Technology Department, Miami Valley Laboratories, Procter & Gamble, 11810 East Miami River Rd., Cincinnati, OH 45252.
‡ Current address: Johnson & Johnson Pharmaceutical Research & Development, LLC, 920 Route 202, P.O. Box 300, Raritan, NJ 08869.
§ Current address: Cumbre, Inc., 1502 Viceroy Drive, Dallas, TX 75235-2304.
|| Current address: Novartis Institute of BioMedical Research Inc., 100 Technology Square, Room 2605, Cambridge, MA 02139.

10.1021/ci034158p CCC: $27.50
Published on Web 11/12/2003


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27

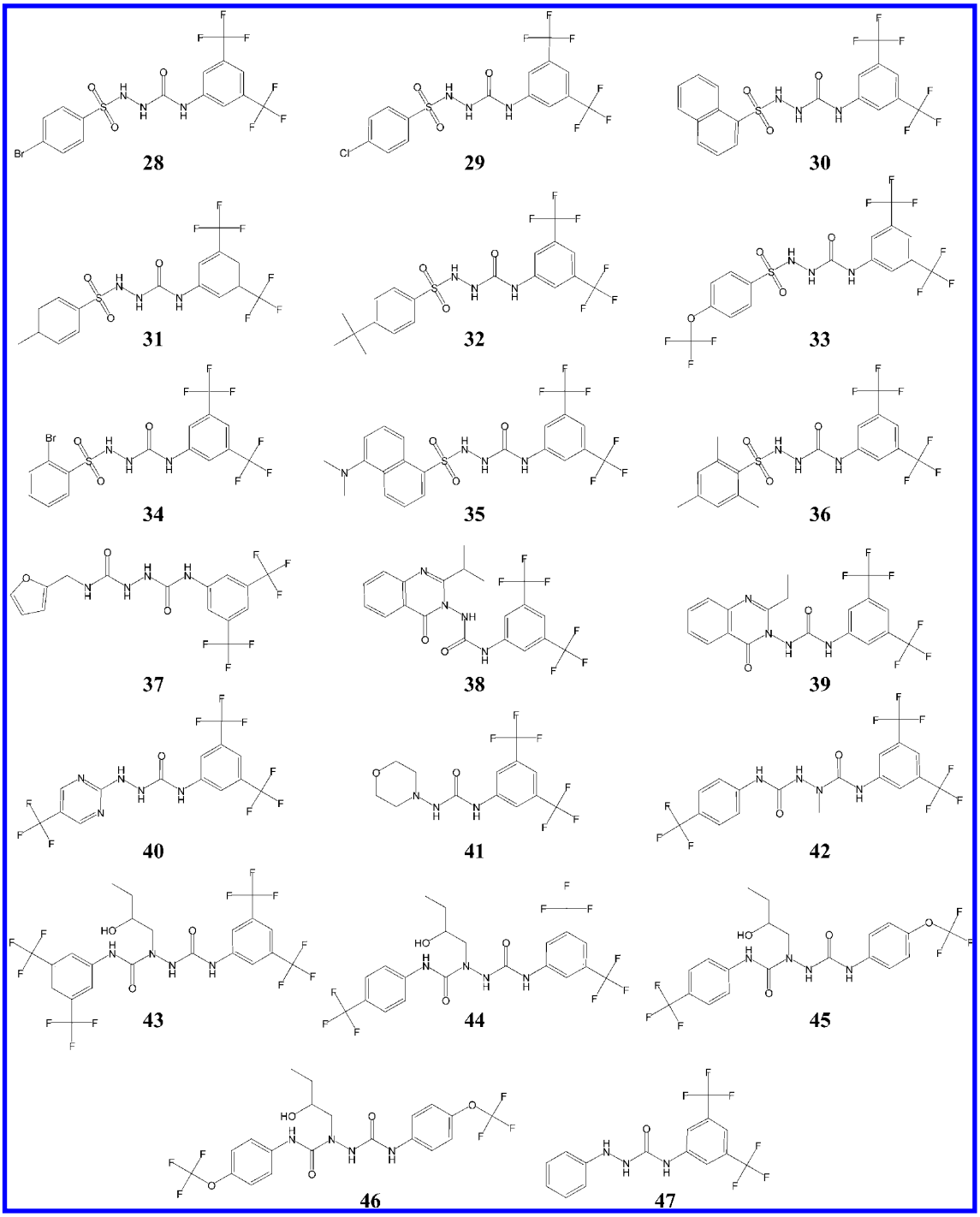


Figure 1. Structure diagrams for the 47 compounds that were combined to form the biarylamides MIC model training set.

than developing a separate model for each organism. The MIC was modeled as the log(base 10) of the reciprocal of the MIC (μM) (i.e., log(1/MIC)).

Descriptor Calculation. A set of 144 descriptors was calculated for each of the 47 structures in the data set using the ADAPT software. The descriptor set was chosen to capture important topological, geometric, and electronic structural features. The topological descriptors are derived using graph-theoretical approaches to defining chemical structures (chemical graph theory)[16] and are conformation-independent, whole-molecule descriptors. Additional conformation-independent information is captured as counts of specific structural fragments (i.e., counts of carbon and heteroatoms; counts of single, double, triple, and aromatic bonds; etc.). Geometric descriptors capture the conformation-dependent characteristics of structure, such as surface area and volume[17] and width, length, and thickness,[18] whereas electronic descriptors provide information concerning the distribution of charge in the molecule.[19] Additionally, some descriptors employ structural representations that capture two or more of these structural feature types (e.g., surface area and partial atomic charge). The CPSA descriptors[20,21] and the related hydrogen-bonding-specific descriptors[22] that have been shown to be useful in past studies represent this class of descriptor. The partial charges used in the calculation of the CPSA and related descriptors were those obtained using the Gasteiger−Huckel method computed during the strain-energy minimization step in SYBYL.

Table 1. Observed and Computed Activity Values for the 47-Observation MIC Model Training Set[a]

compd no.	obsd log (1/MIC, μM)	calcd log (1/MIC, μM)	residual log (1/MIC, μM)
1	-5.44×10^{-1}	-2.88×10^{-1}	-2.56×10^{-1}
2	-1.67	-1.30	-3.67×10^{-1}
3	-1.28	-9.32×10^{-1}	-3.47×10^{-1}
4	-1.75	-1.81	5.73×10^{-2}
5	-7.16×10^{-1}	-9.84×10^{-1}	2.68×10^{-1}
6	-8.13×10^{-1}	-7.95×10^{-1}	-1.76×10^{-2}
7	-1.40	-1.36	-4.03×10^{-2}
8	-2.00	-1.71	-2.90×10^{-1}
9	-1.64	-1.36	-2.74×10^{-1}
10	-1.25	-1.49	2.44×10^{-1}
11	-1.55	-1.38	-1.69×10^{-1}
12	-5.44×10^{-1}	-9.66×10^{-1}	4.22×10^{-1}
13	-1.05	-1.18	1.34×10^{-1}
14	-1.30	-1.36	6.35×10^{-2}
15	-1.90	-1.78	-1.18×10^{-1}
16	-5.44×10^{-1}	-6.69×10^{-1}	1.25×10^{-1}
17	-3.01×10^{-1}	-3.72×10^{-1}	7.10×10^{-2}
18	-6.02×10^{-1}	-5.26×10^{-1}	-7.63×10^{-2}
19	-1.55	-1.51	-4.10×10^{-2}
20	-9.44×10^{-1}	-1.24	2.97×10^{-1}
21	-7.16×10^{-1}	-7.06×10^{-1}	-9.85×10^{-3}
22	-6.43×10^{-1}	-6.35×10^{-1}	-7.56×10^{-3}
23	-9.19×10^{-1}	-9.08×10^{-1}	-1.14×10^{-2}
24	-5.44×10^{-1}	-4.59×10^{-1}	-8.49×10^{-2}
25	-8.92×10^{-1}	-9.16×10^{-1}	2.36×10^{-2}
26	-4.47×10^{-1}	-3.79×10^{-1}	-6.83×10^{-2}
27	-1.45	-1.82	3.71×10^{-1}
28	-1.45	-1.26	-1.91×10^{-1}
29	-1.45	-1.35	-9.50×10^{-2}
30	-8.45×10^{-1}	-8.63×10^{-1}	1.83×10^{-2}
31	-1.50	-1.61	1.09×10^{-1}
32	-1.25	-1.26	1.66×10^{-2}
33	-1.10	-1.19	9.34×10^{-2}
34	-1.94	-1.73	-2.12×10^{-1}
35	-1.50	-1.46	-3.46×10^{-2}
36	-1.52	-1.36	-1.61×10^{-1}
37	-1.85	-1.87	1.95×10^{-2}
38	-2.79×10^{-1}	-4.93×10^{-1}	2.14×10^{-1}
39	-8.06×10^{-1}	-6.41×10^{-1}	-1.65×10^{-1}
40	-6.13×10^{-1}	-8.48×10^{-1}	2.35×10^{-1}
41	-1.82	-1.72	-9.95×10^{-2}
42	-1.08	-1.37	2.97×10^{-1}
43	-4.10×10^{-2}	-3.12×10^{-2}	-9.84×10^{-3}
44	-5.19×10^{-1}	-5.93×10^{-1}	7.37×10^{-2}
45	-1.15	-1.12	-3.12×10^{-2}
46	-1.20	-1.13	-6.81×10^{-2}
47	-1.67	-1.76	9.44×10^{-2}

[a] The activity values represent the logarithm of the minimum inhibition concentrations (MICs) in micromolar units. The compound numbers are used to identify the structures throughout this paper.

Model Development and Validation. Model development was carried out in ADAPT using a generalized simulated annealing approach.[23] Additional statistical evaluation of the model was performed using linear regression methods in the Minitab package.[24] Prior to the acceptance of a final model, partial-least-squares (PLS) analysis[25] was performed using the Minitab program to ensure that the model was not overfitted. A model was considered to be overfitted if the PLS analysis showed the number of validated components to be less than the number of original descriptors in the model (e.g., a seven-variable model yielded six or fewer validated PLS components). A check for chance correlation was also performed that involved a 10-fold randomization of the dependent variable.[26] The set of the original descriptors from the model were then regressed against each of the 10 instances of the randomized dependent variable. If the average R^2 value for the 10-fold randomized regressions was substantially smaller than the original MIC, this was taken as evidence that the observed model was not due to random chance. Once a final model was obtained, PLS analysis was repeated to obtain the score plots and *X*-variable weights for the components that explained the majority of the variance in the observed property values (*Y*-variable). This information is used to develop the final physical interpretation of the model utilizing a procedure described previously.[27]

Table 2. Summary of the Biarylamide MIC Model[a]

descriptor	regression coeff	std dev of coeff	*t*-value	variance inflation factor
FPSA-1	-10.0	0.727	-13.8	3.7
WPSA-2	0.00173	1.88×10^{-4}	9.21	2.1
WTPT-2	39.6	4.15	9.54	2.4
WTPT-4	-0.0526	0.0145	-3.64	1.4
GEOH-5	0.00315	0.00101	3.11	1.3
L/B	-0.400	0.152	-2.64	1.3
1SP3	0.371	0.0519	7.15	1.7
Y-intercept	-73.8	7.83	-9.42	N/A

[a] $R^2 = 0.827$; $s = 0.197$; $N = 47$; F-value $= 38.0$.

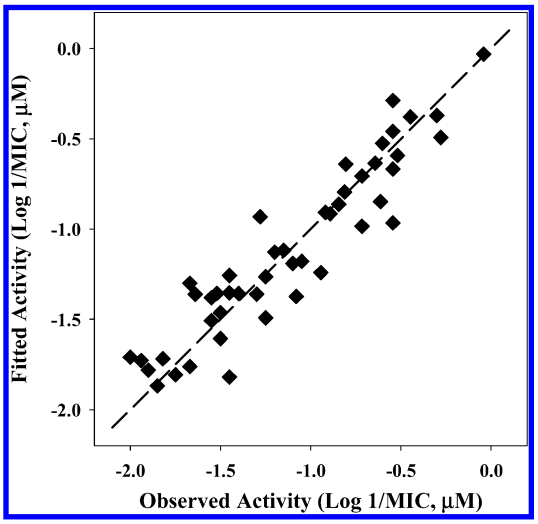


Figure 2. Scatter plot (fit plot) showing the correlation of the fitted (computed) and observed activity values for the 47-observation biarylamide training set.

RESULTS AND DISCUSSION

MIC Model. A model was developed that contained seven descriptors and yielded a good fit to the experimental data ($R^2 = 0.872$; $s = 0.197$). A summary of the model is provided in Table 2. The correlation of the fitted and observed antibacterial activity values is illustrated in the fit plot in Figure 2. The first step in model validation was to apply the overall F test.[28] This test showed the regression to be significant at the 95% level. The F-value (for analysis of variance) was 38.0, which compared well to the critical F-value of 2.255 (with 7 and 39 degrees of freedom, $\alpha = 0.05$). The individual descriptors were subjected to the partial F test.[29] The lowest partial F computed for any of the model descriptors was 6.970, which also compared favorably with the critical value of 4.091 (with 1 and 39 degrees of freedom, $\alpha = 0.05$). The model yielded a cross-validation R^2 (Q^2) of 0.828 using the leave-one-out method.[26] A check for chance correlation was performed that involved a 10-fold random-

Table 3

(a) Summary of the Results of the PLS Analysis of the Biarylamides MIC Model

component	residual sum of squared error	R^2 (cumulative)	PRESS	Q^2 (cumulative)
1	7.04	0.404	11.8	0.000
2	3.88	0.671	7.60	0.357
3	2.24	0.810	2.96	0.749
4	1.56	0.868	2.10	0.822
5	1.53	0.871	2.10	0.822
6	1.51	0.872	2.04	0.827
7	1.51	0.872	2.04	0.827

(b) *X*-Weights for Each Descriptor in the First Three Components of the PLS Analysis of the Biarylamides MIC Model

	X-weight		
descriptor	component 1	component 2	component 3
FPSA-1	−0.716	0.203	−0.586
WPSA-2	0.222	0.662	−0.147
WTPT-2	0.0539	0.623	0.339
WTPT-4	−0.394	0.125	0.111
GEOH-5	0.366	−0.0661	−0.350
L/B	−0.358	0.0170	0.609
1SP3	0.133	0.334	−0.125

ization of the dependent variable as described above. The average R^2 value for the 10-fold randomized regressions was 0.176, with a range from 0.037 to 0.288. The maximum correlation of any of the randomized MICs with the original MIC vector was 0.138. These results suggest that the good fit of the model is not a result of chance correlations. The subsequent PLS analysis of the original model showed that seven components are validated, suggesting the model is not overfitted. These results taken together provide evidence that the MIC model is statistically significant.

The next step involved the extraction of the underlying structure−activity relationship for the model. The model contains seven descriptors. The descriptors FPSA-1 and WPSA-2 are representatives from the class of CPSA descriptors.[20] The descriptor L/B is the length-to-breadth ratio and is considered a conformation-dependent shape descriptor.[30] The 1SP3 descriptor is a simple count of the number of sp^3-hybridized carbons that are attached (bonded) to only one other carbon. The GEOH-5 descriptor is a geometric descriptor that is computed as the ratio of the "length" and "thickness" of the molecule. In this case, the length is the longest axis through the structure, while the thickness is the third longest axis through the structure that is also perpendicular to the first two axes. The remaining two descriptors, WTPT-2, and WTPT-4, are both from the topological class of descriptors called the weighted-path descriptors as defined by Randić.[31] The contribution of each descriptor and its physical meaning in the model was examined next. This was done using PLS as has been described elsewhere.[27] Results obtained from the PLS analysis of the MIC model are provided in Table 3a,b. The first PLS component explains 40.4% of the variance in the original data. The FPSA-1 descriptor is by far the most important descriptor in this component, while WTPT-4, GEOH-5, and L/B also contribute, but to a much smaller extent. FPSA-1 captures information about the surface area of atoms with positive partial atomic charges, and it takes a negative weight in the first component. This means that for component 1, activity of the compounds decreases with increasing FPSA-1 values.

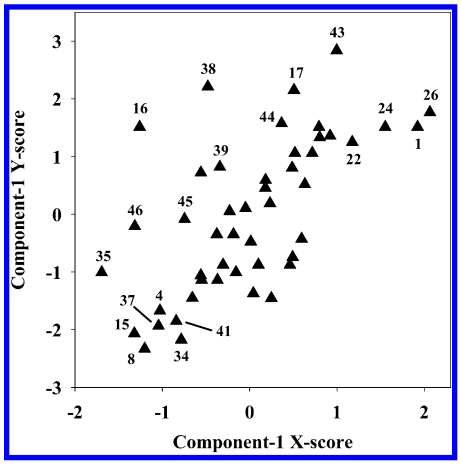


Figure 3. Score plot for component 1 obtained from the PLS analysis of the biarylamide MIC model. Points corresponding to compounds of interest are identified on the plot using the compound number.

The structural features that are the focus of this component can be identified by an examination of the component 1 score plot shown in Figure 3. Points representing structures of interest are identified in the score plot. This component is focusing on the number and type of substituents on the rings at either end of the molecule. One of the consistent differences between the active and inactive molecules is the presence or absence of trifluoromethyl groups on the rings at both ends of the molecule. Compounds **1** and **26** provide good examples of active molecules. The fluorine atoms of the trifluoromethyl groups possess a relatively large negative partial atomic charge, and they also possess a large amount of solvent-accessible surface area, resulting in smaller values for FPSA-1. This can be contrasted with the features of the less active molecules. In these cases, much more of the solvent-accessible surface area on the ends of the structures is associated with hydrogen atoms that take on a slightly positive partial atomic charge. This is particularly evident in compounds **8** and **15**, where most of the surface area of one end of the molecule is associated with hydrogen atoms. The WTPT-4 descriptor also brings attention to the same region of the molecule. The WTPT-4 takes on a negative weight in component 1 and captures information regarding the degree of complexity (branching) of the molecule in the vicinity of heteroatoms. Thus, increased presence of heteroatoms, and increased branching around them, is correlated with a decrease in activity of these molecules. Several of the structural modifications that yield a smaller FPSA-1 value do so as a result of the inclusion of an oxygen atom (see compounds **4**, **37**, and **41**) that possesses a negative partial atomic charge. It is clear that both FPSA-1 and, to a smaller extent, WTPT-4 are bringing attention to the nature of the ends of the molecules, with the characteristics of the trifluoromethyl groups being highly favored.

The second component from the PLS analysis provides further insight into the underlying structure−activity relationship for MIC. This component accounts for an additional 26.7% of the variance in the observed MIC (67.1% cumulative). The PLS score plot for component 2 is shown in Figure 4. Again, points representing structures of interest are identified on the plot. Two descriptors are highly weighted in this component, WPSA-2 and WTPT-2. Both descriptors

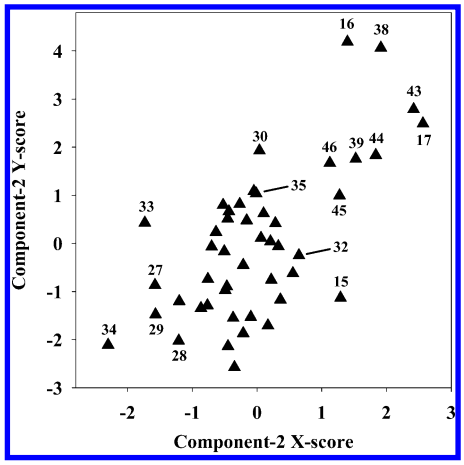


Figure 4. Score plot for component 2 obtained from the PLS analysis of the biarylamide MIC model.

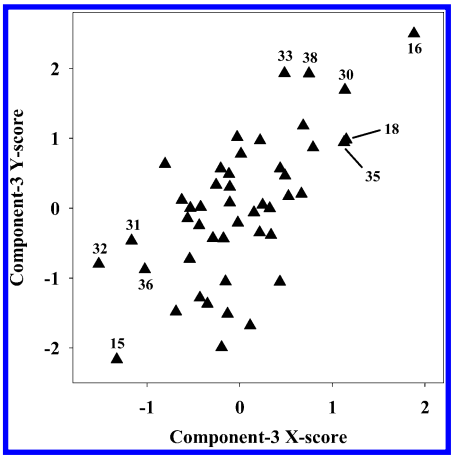


Figure 5. Score plot for component 3 obtained from the PLS analysis of the biarylamide MIC model.

take positive weights in this component, indicating that activity increases as the descriptor values increase. The WPSA-2 descriptor is another of the CPSA descriptors. It is related to FPSA-1 in that it captures information about the surface area of atoms with positive partial atomic charges. However, it is interesting to note that FPSA-1 and WPSA-2 yield a correlation coefficient of only 0.629, suggesting that they are capturing different structural information. The WTPT-2 descriptor is the molecular ID (sum of the path weights for the molecule) divided by the number of atoms in the molecule.[31] This time, both of these descriptors are focusing on changes in the middle of the molecules. An examination of the score plot shown in Figure 4 shows that the MICs for the compounds that are the focus of this component are those that were incorrectly estimated in component 1. In this case, the active molecules have a greater amount of positive surface area and increased overall branching. Both of these changes occur in the middle of the molecule. In component 1, increased positive surface area was associated with decreased activity. However, in component 2, the opposite is true. The primary difference is the part of the structure involved. Component 2 suggests that increased positive surface area and branching in the middle of the molecule has a positive effect on activity. The less active molecules also show the same type of correction. In the case of the more inactive molecules, there is a large proportion of the solvent-accessible surface area associated with atoms possessing negative partial-atomic charges due to the presence of the sulfone portion of the sulfonamide group in the linkage between the rings (see compounds **27**−**29** and **34**). In component 1, increases in these types of descriptors would have been correlated with increased activity. However, the additional negative surface area is not associated with the end of the molecules as was the case in component 1. These molecules are observed to be less active than estimated by component 1 and are corrected here in component 2. Thus, component 2 is bringing attention to the presence and characteristics of structural changes in the middle of the molecules.

Component 3 is the next to be considered and accounts for an additional 13.9% of the variance in the observed MIC values (81.0% cumulative). The score plot for component 3 is shown in Figure 5. The descriptors that are utilized in this component are L/B and FPSA-1. The L/B descriptor takes on a large and positive weight in this component, while the FPSA-1 descriptor takes a large and negative weight. These descriptors appear to be making adjustments to specific compounds that are incorrectly estimated in the previous two components. For example, the activity of **16** is underestimated in component 1 and slightly corrected in component 2. However, it takes the position of the most active molecule in the SAR trend captured in component 3. Similar corrections are made for **30** and **35**. On the other end of the scale, compounds **32** and **15** were overestimated in previous components. This is corrected in component 3, where they take the position as the two most inactive compounds.

Component 3 functions by emphasizing the combined measures of positive surface area and geometry changes. While some compounds containing sulfonamide groups would be considered less active by the SAR trend in component 2 (see **30** and **35**), they are actually more active than would be expected because they exhibit a larger L/B ratio compared to other compounds of that class and have less positive surface area. On the opposite end of the scale, compounds such as **31** and **32** have smaller L/B ratios and have more positive surface areas than similar molecules. Compound **15** is an interesting special case. While the L/B ratio for **15** is nearly as great as the value for compound **30**, 1.988 and 2.058, respectively, the FPSA-1 value for **15** is much larger. This is due to terminal methyl groups on each of the three methoxy substituents on the ring. This compound was carried along in the SAR trend in component 2, where large positive surface area was correlated to increased activity. Component 3 employs the combination of molecular shape and charged surface area to make a correction.

The remaining four components combine to account for the remaining 6.2% of the observed variance for which the whole model accounts. Since the bulk of the SAR information described by the model is accounted for by these first three components, it is possible at this point to consider the physical significance of the structural changes emphasized by the model. The most important feature in the largest component is a measure of positive surface area or, more accurately, the requirement to minimize the amount of positive surface area. The focus is on the ends of the molecules. Significantly, the compounds with the largest *X*-scores in component 1 have either trifluoromethyl groups or other halogens (Cl and Br) as substituents on the end rings.

Compounds at the other end of the scale have a variety of different groups, most of which involve more hydrogens. Since the CPSA descriptors were originally designed to capture information regarding structural features related to polar interactions between molecules, it was natural to look for a polar or electrostatic interpretation for the SAR in component 1. However, attempts to rationalize any type of polar or electrostatic interpretation for this trend were unproductive.

Attempts to visualize the properties of the solvent-accessible surface (SAS) led to an unexpected finding. The MOLCAD program[32] in the SYBYL package was employed to examine the nature of the SAS. MOLCAD provides the ability to color the displayed molecular surface based on several different properties. The first chosen was the electrostatic potential. However, as indicated above, this representation did not provide much insight. As an alternative, the lipophilic potential was also examined. This representation provided an important clue to understanding the SAR trends observed for the MIC model. The hydrophobicity of the ends of the molecule appears to be a primary factor determining the MIC activity of these molecules. Images of the MOLCAD surfaces of several example structures are provided in Figure 6. For compounds 1, **22**, **24**, and **26**, both of the end rings are substituted with very hydrophobic groups. Compounds with poor activity, for example **4**, **15**, and **41**, have substituents that are much less hydrophobic. This is reflected in the images of the SAS for these compounds in Figure 6.

To understand the images in Figure 6, it was instructive to determine the method used in MOLCAD to determine the lipophilic potential. The lipophilic potential in MOLCAD is computed using the atomic contribution values reported by Wildman and Crippen[33] that were designed to predict octanol−water partition coefficients (log(*P*)). These atomic contributions can be used to see the details of the SAR trend captured in component 1. For purposes of simplicity, and since many of the compounds share a common substructure on one end consisting of a phenyl ring and two trifluoromethyl groups, only changes on the other end of the molecule will be considered here. The contribution to log(*P*) for a single fluorine atom is 0.4202. Thus, the combined contribution of the fluorine atoms in two trifluoromethyl groups (see compounds **1** and **26** in Figure 4) is 2.5212. These two compounds have the two largest *X*-scores in component 1. The compound with the next largest *X*-score is **24**, which possesses two bromine atoms in place of the trifluoromethyl groups. The contribution to log(*P*) for a bromine atom is 0.8546 for a combined contribution of 1.6912. The compound with the next highest *X*-score is **22** in which chlorine atoms take the place of the bromine atoms. The atomic contribution for a single chlorine atom is 0.6895 for a combined contribution of 1.379. At the other end of the scale, the compounds possess more diverse groups on the end of the molecule, but one can still examine the atomic contributions to hydrophobicity. For example, the hydrogen atoms that dominate the SAS of the ring at the end of compound **8** each contribute only 0.1230 to log(*P*) (total = 1.353 for the whole ring). The contributions of the oxygen and nitrogen in the terminal ring on compound **41** are actually negative, with values of (−)0.0684 and (−)0.3187, respectively. Similar observations are made for compounds **4** and **37**. The picture that emerges from this analysis is one that strongly suggests that the hydrophobicity of the ends of the molecules is the primary determining factor for the activity of the biarylamides. It was surprising that the model captures this information using the FPSA-1, a CPSA-family descriptor, since they were originally designed to encode other structural characteristics.

A similar analysis can be applied to component 2. In this trend, the middle of the molecule is the focus. The features driving this SAR trend are the positive surface area and molecular complexity. Many of the active compounds have an additional 2-hydroxybutyl group added on one of the central nitrogen atoms (see compounds **43**−**46**). This changes the character of the whole molecule from being rigid and hydrophobic and provides a little more hydrophilicity at the center position due to the presence of the hydroxyl group. The less active structures are also modified toward the middle of the molecule with the addition of a sulfonamide group (e.g., **27**−**29** and **34**). This changes the geometry of the molecule, making it much less linear, and adds hydrogen bond acceptor groups. Component 3 provides corrections for compounds that are under- or overestimated by either of the first two trends, thus completing the SAR picture.

CONCLUSIONS

The biarylamides are novel structures exhibiting interesting antibacterial activity. The QSAR model described here provides details regarding how the changes in the structures of these molecules affect the observed antibacterial activity. The primary driving force appears to be the degree of hydrophobicity of the ends of the molecules. Yet the activity of the compounds cannot be due solely to some nonspecific hydrophobic interactions of the compounds with biological membranes as the ability of these compounds to lyse red blood cells was much less than the observed antibacterial activity, and these two properties showed little correlation (data not shown). Branching and polar moieties appear to be tolerated only at the center of the structure. While the trifluoromethyl groups appear to provide the largest observed effect on the ends of the molecule, the model suggests that other groups can be substituted while maintaining activity. To a lesser extent, the model also suggests that change is most tolerated in the center of the molecule, making this the potential focus for further modification to enhance activity, or perhaps improve other properties such as absorption or solubility.

There are some interesting general observations that can be drawn from the model. One of the most interesting features of the model is the use of the CPSA descriptor FPSA-1 to capture the information regarding the hydrophobicity of the ends of the molecule. Originally, the CPSAs were designed to capture information regarding polar intermolecular interactions. However, it is very clear in this case that an interpretation of the model on this basis would be misleading. This presents an excellent example of the advantage of allowing the modeling process to identify and highlight key structural features driving the SAR and then generating the physical interpretation at the end of the process. Often, preconceived notions concerning the physical meaning of a descriptor can result in its exclusion from consideration at the calculation step. By computing a diverse

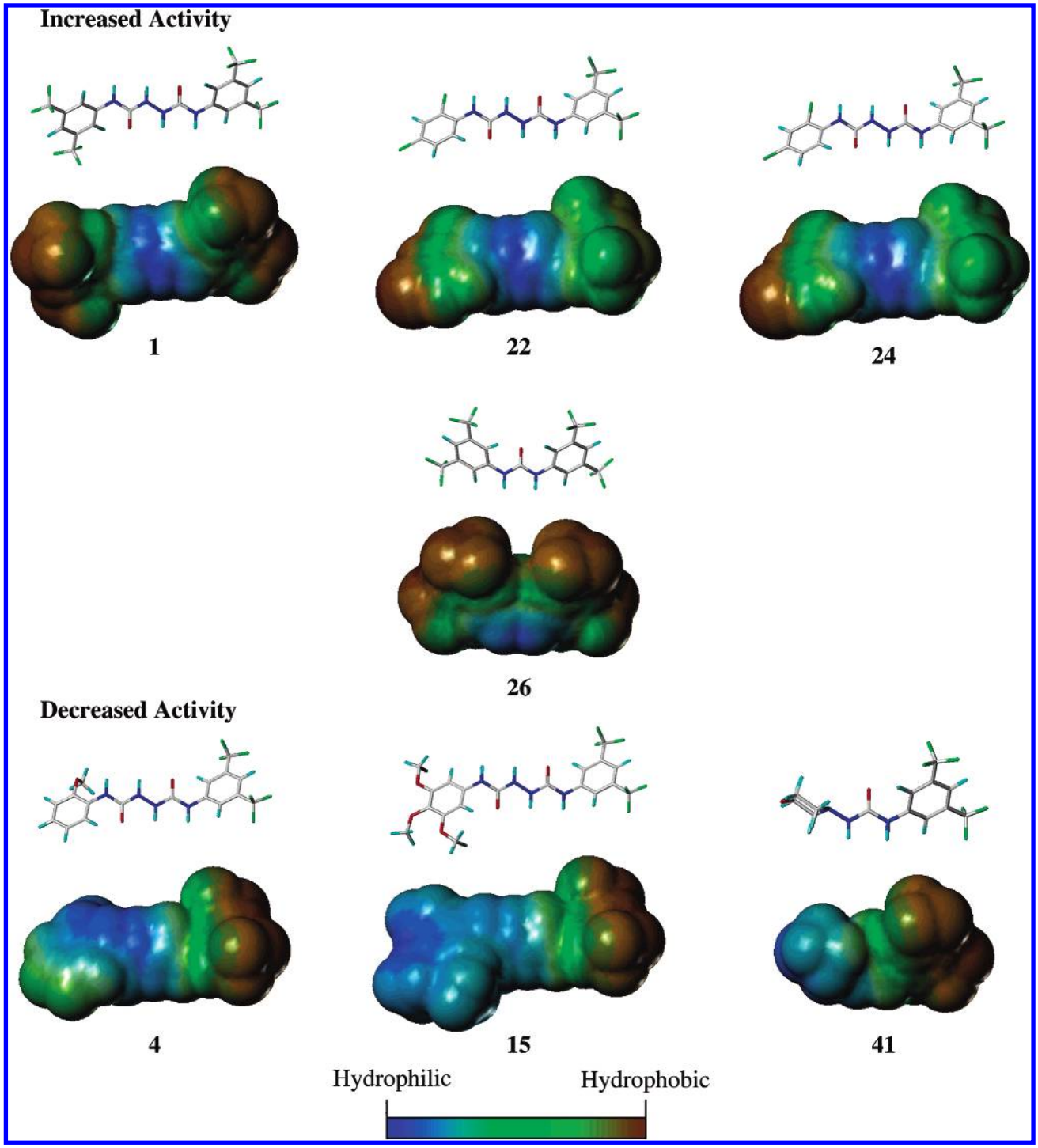


Figure 6. Illustration of the differences in the lipophilic potential surfaces for example compounds exhibiting increased and decreased activity. The conformation shown for the structures are those used in the model development process. The surfaces shown are the solvent accessible surfaces computed in MOLCAD using a probe radius of 1.4 Å.

set of descriptors and then allowing the modeling process to identify those that best explain the SAR, one stands the best chance of generating a superior model.

Last, this work suggests that structural features responsible for nonpolar or hydrophobic interactions may not be well enough characterized by any of the descriptor classes chosen. Although a diverse set of descriptors was selected, the best representative of the hydrophobicity of the structure came from a descriptor designed to account for polar interactions. At a minimum, this makes the physical interpretation more difficult. But it also suggests that better descriptors may be available or can be designed that better serve this purpose.

ACKNOWLEDGMENT

The authors gratefully acknowledge the contributions of G. P. Shrum, S. Sunder, A. G. Switzer, P. J. Renick, H. D. McKeever, and M. C. Davis, all of Procter & Gamble Pharmaceuticals, who were involved in the synthesis and biological evaluation of the compounds examined in this study.

REFERENCES AND NOTES

(1) Tomasz, A. Multiple-Antibiotic-Resistant Pathogenic Bacteria−A Report on the Rockefeller University Workshop. *N. Engl. J. Med.* **1994**, *330*, 1247−1251.

(2) Tenover, F. C.; Hughes, J. M. The Challenges of the Emerging Infectious Diseases−Development and Spread of Multiply Resistant Bacterial Pathogens. *JAMA, J. Am. Med. Assoc.* **1996**, *275*, 300−304.

(3) Novak, R.; Henriques, B.; Charpentier, E.; Normark, S.; Tuomanen, E. Emergence of Vancomycin Tolerance in *Streptococcus pneumoniae*. *Nature* **1999**, *399*, 590−593.

(4) Chen, D. K.; McGeer, A.; de Azavedo, J. C.; Low, D. E. Decreased Susceptibility of *Streptococcus pneumoniae* to Fluoroquinolones in Canada. Canadian Bacterial Surveillance Network. *N. Engl. J. Med.* **1999**, *341*, 233−239.

(5) Oldenburg, K. R.; Vo, K. T.; Ruhland, B.; Schatz, P. J.; Yuan, Z. A Dual Culture Assay for Detection of Antimicrobial Activity. *J. Biomol. Screening* **1996**, *1*, 123−130.

(6) Morris, T. W.; Renick, P. J.; Davis, M. G.; Switzer, A. G.; Sunder, S.; McKeever, H. D.; Wilson, L. J.; Hershberger, P. M.; Shrum, G.P.; Gavin, M.; Zoutendam, P. H.; Stanton, D. T.; Madhav, P. J.; Parker, C. N. Identification of a Biarylamides Class of Antibacterial Agents by Deconvolution of a Combinatorial Library. Book of Abstracts, 39th Interscience Conference on Antimicrobial Agents and Chemotherapy, San Francisco, CA, September 1999; poster F-1278.

(7) Madhav, P. J.; Shrum, G. P.; Sunder, S.; Switzer, A. G.; Wilson, L. J.; Renick, R. P.; McKeever, H. D.; Davis, M. C.; Parker, C. N.; Morris, T. W.; Hershberger, P. M.; Stella, M. E.; Stanton, D. T. Development of Quantitative Structure−Activity and Structure−Property Relationships of Gram-positive MIC, Cell Wall Synthesis Inhibition, and Aqueous Solubility for Biarylamides. Book of Abstracts, 39th Interscience Conference on Antimicrobial Agents and Chemotherapy, San Francisco, CA, September 1999; poster F-1279.

(8) Hershberger, P. M.; Shrum, G. P.; Wilson, L. J.; Switzer, A. G.; Sunder, S.; Seibel, W. L.; Parker, C. N.; Morris, T. W.; Renick, P. J.; Davis, M. C.; Koenigs, P. M.; McKeever, H. D.; Bierman, J. C.; Emig, J. E.; Stella, M. E.; Zoutendam, P. H.; Gavin, M.; Burt, T. M.; Canty, J. F.; Stanton, D. T.; Madhav, P. J. Synthesis and Biological Properties of Biarylamide Antibacterials. Book of Abstracts, 39th Interscience Conference on Antimicrobial Agents and Chemotherapy, San Francisco, CA, September 1999, poster F-1280.

(9) Wilson, L. J.; Morris, T. W.; Wu, Q.; Renick, P. J.; Parker, C. N.; Davis, M. C.; McKeever, H. D.; Hershberger, P. M.; Switzer, A. G.; Shrum, G.; Sunder, S.; Jones, D. R.; Soper, S. S.; Dobson, R. L. M.; Burt, T.; Morand, K. L.; Stella, M. The Identification and Characterization of Hydrazinyl Urea-Based Antibacterial Agents through Combinatorial Chemistry. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1149−1152.

(10) *SYBYL Molecular Modeling Software*, Version 6.8.1; Tripos, Inc.: St. Louis, MO, 2003.

(11) *Concord Users Manual*, Version 4.0; Tripos, Inc.: St. Louis, MO, 1998 (Feb.).

(12) *SYBYL, Version 6.8.1, Force Field Manual*; Tripos, Inc: St. Louis, MO, 2002 (May); p 203.

(13) Gasteiger−Huckel partial atomic charges are calculated using the Gasteiger−Marsili method to calculate the σ-electron contributions and the Huckel method for calculating the π-electron contributions. *Sybyl, Version 6.8.1, Force Field Manual*; Tripos: St. Louis, MO, 2002 (May); p 57.

(14) Stuper, A. J.; Jurs, P. C. ADAPT: A Computer System for Automating Data Analysis Using Pattern-Recognition Techniques. *J. Chem. Inf. Comput. Sci*. **1976**, *2*, 99−105.

(15) Jurs, P. C.; Chou, J. T.; Yuan, M. In *Computer-Assisted Drug Design*; Olson, R. C., Christoffersen, R. E., Eds.; American Chemical Society: Washington, DC, 1979; pp 103−129.

(16) Ivanciuc, O.; Balaban, A. T. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach: Amsterdam, 1999; pp 59−167.

(17) Pearlman, R. S. In *Physical Chemical Properties of Drugs*; Yalkowsky, S. H., Sinkula, A. A., Valvani, S. C., Eds.; Dekker: New York, 1980.

(18) Todeschini, R.; Consonni, V. In *Handbook of Molecular Descriptors*; Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Methods and Principles in Medicinal Chemistry, Vol. 11; Wiley-VCH: Weinheim, Germany, 2000; p 352.

(19) Dixon, S. L.; Jurs, P. C. Atomic Charge Calculations for Quantitative Structure−Property Relationships. *J. Comput. Chem*. **1992**, *13*, 492−504.

(20) Stanton, D. T.; Jurs, P. C. Development and Use of Charged Partial Surface Area Descriptors in Computer-Assisted Quantitative Structure−Property Relationship Studies. *Anal. Chem.* **1990**, *62*, 2323−2329.

(21) Stanton, D. T.; Dimitrov, S.; Grancharov, V.; Mekenyan, O. G. Charged Partial Surface Area (CPSA) Descriptors. QSAR Applications. *SAR QSAR Environ. Res.* **2002**, *13*, 341−351.

(22) Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G. Computer-Assisted Prediction of Normal Boiling Points of Pyrans and Pyrroles. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 306−316.

(23) Sutter, J. M.; Jurs, P. C. Selection of Molecular Descriptors for Quantitative Structure−Activity Relationships. *Data Handl. Sci. Technol.* **1995**, *15*, 111−132.

(24) *Minitab*, *Release 14*, Beta Version-3; Minitab, Inc.: State College, PA, 2003 (April).

(25) Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta* **1986**, *185*, 1−17.

(26) Wold, S.; Eriksson, L. Statistical Validation of QSAR Results. In *Chemometric Methods in Molecular Design*; van de Waterbeemd, H., Ed.; VCH: New York, 1995; pp 309−318.

(27) Stanton, D. T. On the Physical Interpretation of QSAR Models. *J. Chem. Inf. Comput. Sci*. **2003**, *43*, 1423−1433.

(28) Neter, J.; Wasserman, W.; Kutner, M. H. *Applied Linear Statistical Models*, 2nd ed.; Irwin: Homewood, IL, 1985; p 240.

(29) Neter, J.; Wasserman, W.; Kutner, M. H. *Applied Linear Statistical Models*, 2nd ed.; Irwin: Homewood, IL, 1985; p 281.

(30) Collantes, E. R.; Tong, W.; Welsh, W. J.; Zielinski, W. L. Use of Moment of Inertia in Comparative Field Analysis To Model Chromatographic Retention of Nonpolar Solutes. *Anal. Chem.* **1996**, *68*, 2038−2043.

(31) Randić, M. On Molecular Identification Numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164−175.

(32) *MOLCAD Graphics Manual, Version 6.8.1*; Tripos, Inc.: St. Louis, MO, 2002 (May).

(33) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868−873.

CI034158P