Structure-Based Classification of Antibacterial Activity

Mark T. D. Cronin,*,† Aynur O. Aptula,‡ John C. Dearden,† Judith C. Duffy,† Tatiana I. Netzeva,†,§ Hiren Patel,† Philip H. Rowe,† T. Wayne Schultz,|| Andrew P. Worth,†,# Konstantinos Voutzoulidis,† and Gerrit Schüürmann‡

School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, England, Department of Chemical Ecotoxicology, UFZ Centre for Environmental Research, Permoserstrasse 15, D-04318 Leipzig, Germany, Department of Chemistry, Faculty of Pharmacy, MU−Sofia, 2 Dunav Street, 1000 Sofia, Bulgaria, Department of Comparative Medicine, College of Veterinary Medicine, The University of Tennessee, 2407 River Drive, Knoxville, Tennessee 37996-4500, and ECVAM, Institute for Health & Consumer Protection, Joint Research Centre, European Commission, 21020 Ispra (VA), Italy

Received January 3, 2002

The aim of this study was to develop a simple quantitative structure−activity relationship (QSAR) for the classification and prediction of antibacterial activity, so as to enable *in silico* screening. To this end a database of 661 compounds, classified according to whether they had antibacterial activity, and for which a total of 167 physicochemical and structural descriptors were calculated, was analyzed. To identify descriptors that allowed separation of the two classes (i.e. those compounds with and without antibacterial activity), analysis of variance was utilized and models were developed using linear discriminant and binary logistic regression analyses. Model predictivity was assessed and validated by the random removal of 30% of the compounds to form a test set, for which predictions were made from the model. The results of the analyses indicated that six descriptors, accounting for hydrophobicity and inter- and intramolecular hydrogen bonding, provided excellent separation of the data. Logistic regression analysis was shown to model the data slightly more accurately than discriminant analysis.

INTRODUCTION

Combinatorial libraries are generally structurally diverse collections of chemicals and have become the starting place for many drug discovery programs. From such libraries, often of 100 000s of compounds, lead compounds may be identified.[1] High throughput and ultrahigh throughput screening (HTS and uHTS) are methods to identify such lead compounds.[2] However, while these methods are rapid, they are still relatively costly, and for many pharmacological activities, HTS endpoints may not be available. With that in mind, there has been considerable interest in the *in silico* identification of pharmacologically active compounds.[3−7]

In silico screening studies, particularly for use in conjunction with combinatorial chemistry, have taken a number of routes. A number of workers have attempted to assess how "drug-like" a molecule is. Perhaps best known among these studies is Lipinski's "rule of five" which has been applied by a large number of companies.[8] The "rule of five" is intended to identify qualitatively potential leads that are not capable of uptake following oral administration. The Lipinski rule has been used by many researchers to predict whether a lead drug will be bioavailable, and often novel combinatorial libraries are screened to eliminate compounds that would be considered not to be bioavailable. Several quantitative structure−activity relationships have also been developed to predict bioavailability per se. These attempts will always be limited by a number of factors including the limited availability of data (particularly for poorly bioavailable drugs), the considerable experimental error and interlaboratory variability present within the data, and the intrinsic problems associated with the modeling of such a complex phenomenon with relatively simplistic models. The area of ADME prediction has been well reviewed recently by Ekins et al.[9]

Other attempts to screen combinatorial libraries *in silico* have involved the identification of compounds with specific pharmacological activities. Over the past decade there have been many approaches including the use of pharmacophore searching of databases.[10] The search for compounds with a given activity is appropriate when the pharmacophore is known, and when pharmacological activity is related to interaction at a single receptor site. Many examples of the successful use of pharmacophore-based searching abound in the literature.[10−14] However, for some pharmacological activities, receptors have not been determined or activity may not result from binding at a particular receptor. In such instances, it is not generally possible to search for pharmacophores.

To circumvent the use of pharmacophores for database and combinatorial library screening, effort has also been

* Corresponding author phone: + 44 (0) 151 231 2066; fax: + 44 (0) 151 231 2170; e-mail: m.t.cronin@livjm.ac.uk.
† Liverpool John Moores University.
‡ UFZ Centre for Environmental Research.
§ MU−Sofia.
|| The University of Tennessee.
ECVAM, Institute for Health & Consumer Protection, Joint Research Centre, European Commission.

10.1021/ci025501d CCC: $22.00
Published on Web 06/07/2002

placed into the development of structure-based classification methods (or QSARs), utilizing pattern recognition techniques to predict biologically active molecules.[15] While these methods are often used to predict the potency of drugs (i.e. a continuous response), they may also be utilized to predict whether a compound is active (i.e. a categoric response).[16] Such approaches are commonplace in predictive toxicology for the estimation of endpoints such as carcinogenicity, eye irritation, etc.[17] Recent efforts and new trends in the use of QSARs to predict pharmacological activities have been reviewed recently.[18]

Compounds with antibacterial activity are a class of pharmaceuticals that clearly do not share a common mode of action. Furthermore, there is no "typical" antibacterial pharmaceutical compound, and they are structurally diverse.[19] Therefore development of a single pharmacophore and its use in screening are unlikely to be successful. Despite this, Tomás-Vert et al.[20] developed a classification model for the prediction of antibacterial compounds on the basis of a neural network trained upon 62 descriptors including numbers of particular atoms and topological information. There are a number of advantages and disadvantages to the use of neural networks in QSAR.[21] Neural networks are capable of modeling nonlinear relationships much better than are strictly linear techniques such as regression analysis or discriminant analysis. However, despite their utility, neural networks run a significant risk of over-training. Furthermore, they are not transparent; i.e., the developer is not able to view, comprehend, or interpret the relationship between activity and structural properties, as can be done when a more empirical technique such as regression or discriminant analysis is applied.[21]

The aim of this investigation, therefore, was to develop a classification model (or QSAR) for the prediction of antibacterial activity utilizing transparent and easily portable techniques. To accentuate further the simplicity of the approach, 2-D structural descriptors were utilized to develop models. Two complementary statistical techniques appropriate for the classification of categoric data, discriminant and binary logistic regression analysis, were applied to develop transparent and interpretable models.[22]

METHODS

Biological Activity. A total of 661 compounds were classified as either having or lacking antibacterial activity. These classifications were recorded originally by Tomás-Vert et al.[20] Biological data were taken as this fundamental categoric response; analyses with quantitative data were not attempted. The compounds are utilized, and their classifications are listed in Table 1. It should be noted that the data set is asymmetric with fewer active than inactive compounds.

Calculation of Physicochemical Descriptors. The names of the drugs were taken from Tomás-Vert et al.;[20] structures were converted into SMILES strings[23] for the calculation of 2-D physicochemical and structural descriptors. The logarithm of the octanol−water partition coefficient was obtained using the ClogP for Windows (version 1.0.0) software (Biobyte Corp., Claremont CA), measured values being used in preference to calculated values.

SMILES strings were then entered into the QSARis (ver1.1) software (SciVision − Academic Press, San Diego, CA) for the calculation of 230 2-D descriptors of the structural, steric, electronic, and hydrogen bonding features of the molecules. Examination of the descriptors calculated indicated that 64 had a value of 0 for more than 95% of the compounds (i.e. they contained little, if any, meaningful information). To ensure statistical validity, these descriptors were omitted from the analyses, so that 166 2-D descriptors in addition to log P were used. A summary of the descriptors calculated is given in Table 2.

Statistical Analyses. Preselection of Variables. Prior to statistical analysis for the classification of activity, descriptors were selected on the basis of their being able to discriminate between antibacterial and nonantibacterial activity. Selection may be performed by a number of approaches. In this study, one-way analysis of variance (ANOVA) was performed on each descriptor (prior to classification modeling) to determine those that were best able to separate the two classes of compounds. This technique was found to compare favorably to the stepwise determination of descriptors from, for instance, stepwise discriminant analysis. ANOVA was performed on each variable using SPSS for Windows statistical software (ver 10.0, SPSS Inc, Chicago, IL), and the F-ratios were recorded. The descriptors found to be the most significant from ANOVA were chosen for the development of the classification models described below.

Classification. Once significant descriptors for the classification of activity had been established, two statistical techniques were applied to develop functions to classify the compounds: linear discriminant analysis (LDA) and binary logistic regression (BLR). Both techniques are highly suited to the modeling of categoric data.[16,22] For the purposes of modeling, a value of 1 was assigned to compounds with antibacterial activity, and a value of 0 assigned to those lacking activity. Linear discriminant analysis was performed using the SPSS statistical software. Using this package, prior probabilities were computed from group size (0.623 and 0.377 for the nonantibacterials and antibacterials respectively) and within group covariance matrix was used. Results are reported as a percentage of correctly classified cases. Binary logistic regression was also performed, using the SPSS statistical software, on the variables deemed to be important for the classification of biological activity. Results are reported as a percentage of correctly classified cases at a probability cutoff value of 0.5.

Validation. The models were initially used to classify the compounds in the complete training set. While this provides some assessment of the goodness of fit of a model, it does not provide a thorough and independent assessment of how a model may predict new compounds. To assess such predictivity the use of a test set is essential.[24] With such a technique, a proportion of the data is removed before modeling (the test set), and the remaining data (the training set) are utilized to make predictions. To alleviate the problems of a priori identification of the test and training sets, in this study 30% of the compounds were randomly removed (to form the test set), and the remaining 70% of compounds were utilized to formulate a model. This process was repeated a total of 10 times to ensure that chance effects in the random selection of variables were eliminated.

Table 1. Names of Antibacterial and Nonantibacterial Compounds[a]

	nonantibacterials		nonantibacterials
1	2-amino-4-picoline[a,d,f,g,i,j]	71	butabarbital[b]
2	5-bromosalicylic acid acetate[g]	72	butacetin[f,g]
3	5-nitro-2propoxyacetanilide[c,h]	73	butaclamol[d,e,f]
4	acecarbromal[f,g]	74	butallylonal[a,d,g,h,j]
5	aceclofenac[a,f,h,i]	75	butanilicaine[b,d]
6	acefylline[c,d,e,g]	76	butibufen[d,g]
7	acetaminophen[b,i]	77	butidrine hydrochloride[d,e,i]
8	acetaminosalol	78	butoctamide[a,h,i,j]
9	acetanilide[a,e,f,i]	79	butofilolol[d,e,f,h]
10	acetazolamida[a,d,f,h,j]	80	caffeine[b,c,e,h,j]
11	acetobutolol[a,d,f,h,i]	81	capuride[f]
12	acetophenazine[c,f,h]	82	carazolol[a,h,j]
13	acetylsalicylic acid[c,d,e,g]	83	carbamazepine[a,b,d,e]
14	acrivastine[a,b,d,f,g,h]	84	carbidopa[b,f,g]
15	ahistan[e,g,h]	85	carbinoxamine[c,d,e,f]
16	albuterol[a,h,i,j]	86	carbiphene[c,i]
17	alclofenac[e,g,j]	87	carbocloral[b,h,i]
18	alminoprofen[b,f]	88	carbromal[b,f,j]
19	alphaprodine[e,g]	89	carbuterol[b,c]
20	alprenolol[f,h,i]	90	carfimate[c,f,j]
21	aminochlorthenoxazin[d,f,i]	91	carphenazine[a,c,e,h,j]
22	aminopropylon[j]	92	carprofen[b,c,e,g,i]
23	aminopyrine[a,b,c,f,i,j]	93	carsalam[a,b,e,f,h]
24	amosulalol[a,d,f]	94	carteolol[c,d,i]
25	amtolmetin guacil[b,d,h,i]	95	carvedilol[c,i]
26	anileridine[e,h]	96	celiprolol[c,e,i]
27	antipyrine[f,h]	97	cetamolol[b,e,f,i,j]
28	antrafenine[g,i]	98	cetirizine[b,c]
29	apazone[d,f,g]	99	chlorhexadol[a,d,e]
30	apronalide[e]	100	chlorobutanol[h]
31	arotinolol[b,c,e,h,f]	101	chloropyramine[c,e,f]
32	atenolol[f,h]	102	chlorothen[d,f]
33	atropine[b,c,f,j]	103	chlorpheniramine[d,e,f,i,j]
34	bambuterol[a,b,f]	104	chlorpromazine[c,g,h]
35	bamifylline[h,i,j]	105	chlorprothixene[a,c,f,i]
36	bamipine[e,i]	106	chlorthenoxacin[e,h]
37	beclofibrate[c,g,j]	107	chlorcyclizine[e,h]
38	befunolol[h]	108	cinchophen[b,c,e,g,i]
39	benfluorex[b,i]	109	cinmetacin[a,b,i,j]
40	benorylate[a,b,d,j]	110	cinnarizine[b,c,d,e,g,h]
41	benoxaprofen[f,h]	111	cinromida[a,e,j]
42	benserazide[d,e]	112	ciprofibrate
43	benzitramide[a,f,h]	113	ciramadol[e,h,i,j]
44	benzotropine mesylate[e]	114	clemastine[a,j]
45	benzpiperylon[i]	115	clenbuterol[c,d,e,j]
46	benzydamine[b,d,i]	116	clidanac[a,c]
47	bermoprofen[c,h]	117	clinofibrate[d,f,h,j]
48	betaxolol[a,c,j,i]	118	clocinizine[a,b,c,e,f,h]
49	bevantolol[b,g]	119	clofibrate[c,e,g,h,j]
50	bevonium methyl sulfate[a,b,d,f]	120	clofibric acid[i,j]
51	bezafibrate[c,e]	121	clometacin[b,h]
52	binifibrate[a,e]	122	clometiazol[d,f]
53	bisoprolol[a,e]	123	clonixin[b]
54	bitolterol[d,e,g,h,i]	124	clopirac[c,g]
55	bucloxic acid[a,b,c,g,h]	125	cloralsalicylamide[a,b,e,h,j]
56	bopindolol[c,f]	126	cloranolol[b,c,f,g,h,i,j]
57	bromfenac[a,h,j]	127	clordesmetildiazepam[b,c,d,i]
58	bromisovalum[b]	128	clorprenaline
59	bromodiphenhydramine[b,e,h]	129	clothiapine[b,c,d,h,i]
60	brompheniramine[a,b,f,g,h,j]	130	clozapine[j]
61	brotizolam[j]	131	codeine[b,c,f,i]
62	bucetin[b,h]	132	cropropamide[d,h]
63	bucolome[c,e]	133	crotethamide[a,c,d,e]
64	bucumolol[a,b,j]	134	deserpidine[j]
65	bufetolol[c,f,h,i,j]	135	diazepam[a,b,c,d]
66	bufexamac[f,g,i]	136	diclofenac[b,d,e]
67	bufuralol[c,g,h]	137	diethylbromoacetamide[b,d,g,h]
68	bumadizon[d,h]	138	difenamisole[c,g,j]
69	bunitrolol[h,j]	139	difenpiramide[b,i,j]
70	bupranolol[e,f,h]	140	diflunisal[b,d,g]

Table 1 (Continued)

	nonantibacterials		nonantibacterials
141	dilevalol[e]	211	haloperidida[h]
142	dioxadrol[c,h]	212	haloperidol[e,g,i]
143	diphenhydramine[c,d,g]	213	hexapropymate[c,e,g,i,j]
144	diphenylpyraline[b,d,h,i,j]	214	hexobarbital[a,d]
145	dipyrocetyl[a,b,f,g]	215	hexoprenaline[a,c,d,e,i]
146	dipyrone[a,d,f,h]	216	histapyrrodine[c]
147	disulfiram[e,i,j]	217	hydroxyethylpromethazine[f,g]
148	doxefacepam[a,g,h,j]	218	hydroxyzine[f,g,i]
149	doxofylline[e,g,j]	219	ibufenac[b,f]
150	doxylamine[b,f,g,i]	220	ibuprofen[c,g,h,i]
151	droperidol[c,d,h,j]	221	ibuproxam[b,c,d,e,h,i,j]
152	droxicam[e,g,h]	222	indenolol[a,b,g,i]
153	dyphylline[a,h]	223	indomethacin[d,f]
154	ectylurea[d,e]	224	indoprofen[b,g,i]
155	embramine[e,f,j]	225	ipratropium bromide[d,g,i]
156	emorfazone[b,f,h]	226	isoetharine[a,j]
157	enfenamic acid[b,d,j,i]	227	isofezolac[b,i]
158	enprofylline[a,e,g]	228	isoladol[b,d,g,h]
159	epanolol[b,c,e,f,i]	229	isonixin[e,i]
160	ephedrine[a,b,c,g]	230	isopromethazine[f,i]
161	epirizole[f,g]	231	isoproterenol[b,f]
162	eprozinol[d,j]	232	isoxicam[a,c,e,g,h]
163	esmolol[b,d,g]	233	ketoprofen[a,b,g,h,i]
164	estazolam[a,c]	234	ketorolac[b,i,j]
165	etafedrine[a,e,i,j]	235	labetalol[c]
166	etamiphyllin[b,d,j]	236	lefetamine[g,h,j]
167	etaqualone[c,d,j]	237	levobunolol[a,i]
168	eterobarb[i]	238	lorazepam[c,h,i]
169	etersalate[i]	239	lornoxicam[a,b,d,j,h,j]
170	ethclorvinol[b,j]	240	lovastatin[c,e]
171	ethenzamide[c,f,g,j]	241	loxapina[h,j]
172	ethinamate[d,g,h]	242	loxoprofen[a,g]
173	ethoheptazine[b,c,f,h]	243	mabuterol[b,c,e,f]
174	ethoxazene[a,d,f,g]	244	mazindol[i]
175	etodolac[e,f,j]	245	meclofenamic acid[f]
176	etofibrate[h,j]	246	mecloqualone[c,f]
177	etofylline[a,c,d]	247	mecloxamine[a,c,d,e,h,j]
178	etomidate([a,c,d,f,j])	248	medibazine[a,i]
179	etymemazine[a,d,f,g,h,i]	249	medrylamine[b,i]
180	felbinac[a,d,e,f,g,h,j]	250	mefenamic acid[a,h]
181	fenadiazole[a,b,d,f,g,j]	251	mefexamide[f]
182	fenbufen[f,h]	252	meparfynol[a,e,j]
183	fenclofenac([b,f,h])	253	meperidine[a,e,f,i]
184	fenethazine[a,d,e,i]	254	mephobarbital[d,e,f,i]
185	fenofibrate[d,e]	255	mepindolol[b,e,g]
186	fenoprofen	256	meprobamate[a,f,h]
187	fenoterol[a,b,c,d]	257	mequitazine[e]
188	fentanyl[e,f]	258	methafurylene[e,g]
189	fentiazac[f,g,i,j]	259	methaphenilene[a,b,g,i]
190	feprazone[c,j]	260	methapyrilene[c,f,j]
191	floctafenine[h]	261	methotrimeprazine[c,f]
192	flufenamic acid[a,b]	262	methoxyphenamine[b,c,e]
193	flunitrazepam[b,d,j]	263	methyldopa[c,g,h]
194	fluoresone[e,g,j]	264	methyltyrosine[c,e,f,g]
195	fluphenazine[b,d]	265	methyprylon[a,b,e,g]
196	flupirtine[b,c,d,f,i]	266	metiapine[b,c,d,e,g]
197	fluproquazone[a,b,j]	267	metipranolol[a,d,e,g,j]
198	flurazepam[c,e,f,i]	268	metofoline[d]
199	flurbiprofen[a,b,f]	269	metoprolol[b,c,h]
200	fluspirilene[b,f]	270	metron[a]
201	flutropium bromide[g,j]	271	mexiletine[c,g,h,j]
202	formoterol[f,h,j]	272	mofezolac[a,c,i]
203	fosazepam[c,e,f,g,i,j]	273	molindone[a]
204	fosfosal[d,e,h]	274	moperone[b,d,h]
205	fusaric acid[c,g]	275	moprolol[a,d,e,f]
206	gemfibrozil[c,g]	276	morazone[g,j]
207	gentisic acid[f,i]	277	morphine[a,b,c,h]
208	glafenine[a,d,g]	278	moxastine[d,j]
209	glucametacin[a,c,e,f,i]	279	nadolol[g,i]
210	glutethimide[j]	280	nadoxolol[c,e]

Table 1 (Continued)

	nonantibacterials		nonantibacterials
281	naproxen[e,f]	347	propyphenazone[a,c,e,g]
282	narcobarbital	348	protokylol[c,d,g]
283	nefopam[b,d]	349	proxibarbital[b,e,f]
284	niceritrol[b,j]	350	proxyphylline[d,i,j]
285	nicoclonate[c,h]	351	pyrilamine[a,d,f,g,j]
286	nicofibrate[a,b,c,e,f,j]	352	pyrrobutamine[d,e,i]
287	nifenalol[b,e,f,g,i,j]	353	quazepam[b,i]
288	nifenazone[a,b,e,i]	354	ramifenazone[a,b]
289	niflumic acid[a,b,e,g,i]	355	reproterol[c,g,i]
290	nimetazepam	356	rimiterol[d]
291	nipradilol[a,e,h,i]	357	ronifibrate[a,f,h,i]
292	nitrazepam[a,b,g,j]	358	salacetamide[b,c,d,g,h]
293	nordiazepam[c,h,j]	359	salicylamide[b,d,e,h,i]
294	novonal[a,c]	360	salicylamide O-acetic acid[d,g,h,i]
295	octopamine[a,b,c]	361	salsalate[b,d,e,g,h,j]
296	orphenadrine[f]	362	salverine[c,j]
297	oxaceprol[b,c,e,h,i]	363	scopolamine[a,d,j]
298	oxametacine[d,h,i]	364	secobarbital[a,c,d]
299	oxanamide[a,e,i,j]	365	setastine[g]
300	oxaprozin[e]	366	simetride[b,e,f,g,h]
301	oxipertine[a,i]	367	simfibrate[b,d]
302	oxitropium bromide[g,j]	368	simvastatin
303	oxprenolol[b,f,g,i,j]	369	sotalol[b,c,e,h]
304	paraldehyde	370	soterenol[b,d,e,f,i]
305	paramethadion[d,e]	371	sulfinalol[h,j]
306	parsalmide[g]	372	sulindac[e]
307	*p*-bromoacetanilide[a,b,e]	373	sulpiride[g,i]
308	pemoline[c,e,f,g,j]	374	suprofen[g,j]
309	penbutolo[b,d,e,g]	375	talastine[a]
310	penfluridol[a,f,g]	376	talbutal[d,g,i,j]
311	perisoxal[f,i,j]	377	talinolol[a,c,d,g]
312	perphenazine[b,c,e,i,g]	378	talniflumate[c,i,j]
313	phenacemide[d,e,i,j]	379	temazepam[c,d,j]
314	phenacetin[b,c,j]	380	tenoxicam[f,j]
315	pheniramine[a,h,j]	381	terbutaline[c,d,g]
316	phenocoll[c,d,e,f,g]	382	tertatolol[c,d,g,h,i,j]
317	phenoperidine[f,g,j]	383	tetrabarbital[b,c,d,e,g,i]
318	phenopyrazone[b,d,f]	384	thenaldine[h,i,j]
319	phenylbutazone[b,f,i]	385	thenyldiamine[c]
320	phenyltoloxamine[a,c,g]	386	theobromine[c,e,h]
321	phenyramidol[d,i]	387	theofibrate[b,f,i]
322	pimozide[a,c,e,h,i]	388	theophylline[f,h,i,j]
323	pindolol[a,c,e]	389	thioridazine[b,i]
324	pipebuzone[a,h,j]	390	thiothixene[a,d,e,h,j]
325	piperacetazine[h]	391	thonzylamine[e,f,g,i]
326	piperidione[a,b,c,d,h]	392	tiaprofenic acid[a,b,j]
327	piperylone[a,d,e]	393	timolol[g,j]
328	pirbuterol[j]	394	toliprolol[a,c,f,h]
329	pirifibrate[g,h]	395	tolmetin[a,d,i]
330	piroxicam[a,d]	396	tolpropamine[d,i]
331	pirprofen[a,b,d,f]	397	tretoquinol[f,h]
332	p-lactophenetidine[a,g]	398	triazolam[c,e,i,j]
333	*p*-methyldiphenhydramine[f,j]	399	triclofos[d,f,g,h]
334	practolol[a,g,h,i]	400	trifluoperazine[b,c,g,i]
335	pravastatin[f,g]	401	trifluperidol[c,d]
336	prazepam[f,j]	402	trimethadione[a,d,f]
337	primidone[d,e,f,i,j]	403	triparanol[a,c,j]
338	probucol[a,d,i]	404	tripelennamine[c,d,i]
339	procaterol[a,b,h]	405	triprolidine[c,e,h]
340	proglumetacin[d,j]	406	tulobuterol[a,d]
341	proheptazine[a,b,c,e,f,j]	407	viminol[h]
342	prolintane[a,j]	408	vinylbital[a,b,e,f]
343	promazine[g,h]	409	xenbucin[c,i,j]
344	pronethalol[b,g]	410	xibenolol[d]
345	propacetamol[h,i,j]	411	zolamine[d,e,h,i]
346	propanolol[a,h,j]	412	zomepirac[a,d,i]

	antibacterials		antibacterials
1	4″-(methylsulfamoyl)sulfanilanilide[e,f,j]	4	acediasulfone[b,d,f,j]
2	4′-formylsuccinanilic acid thiosemicarbazone[b,c,d,f]	5	acetyl sulfamethoxypyrazine[a,d]
3	4-sulfanilamidosalicylic acid[b,e]	6	acetyl sulfisoxazole[b,c]

Table 1 (Continued)

	antibacterials		antibacterials
7	amidinocillin[d,i,j]	77	cephaloridine[b,c,g,i]
8	amidinocillin pivoxil[c,e,g,h,i]	78	cephalosporin c[b]
9	amifloxacin[c]	79	cephalothin[b,e]
10	amikacin[b,e]	80	cephapirin sodium[f]
11	amoxicillin[a,b,c,h,j]	81	cephradine[c,h,j]
12	ampicillin[d,g]	82	chloramphenicol[i]
13	apalcillin[b,d,f,g]	83	chloramphenicol palmitate[b,d,i,j]
14	apicycline[a,d,e,h,j]	84	chloramphenicol pantothenate[f]
15	apramycin[a,c,g,h]	85	chlortetracycline[d,e,j]
16	arbekacin[a,c,f,g,h]	86	cinoxacin[c,g]
17	aspoxicillin[c]	87	ciprofloxacin[e]
18	azidamfenicol[c]	88	clinafloxacin[a,c,d,h,f]
19	azidocillin[d,f,g,i]	89	clindamycin[a,b,d,f]
20	azlocillin[f,g]	90	clometocillin[b]
21	azosulfamide[c,i]	91	clomocycline[a,b,c,d,j]
22	aztreonam[c,d,e,f,j]	92	cloxacillin[b,c,d,g,h]
23	bacampicillin[c,h]	93	cyclacillin[h,i]
24	benzylpenicillinic acid[g,i,j]	94	demeclocycline[b,e,g,h]
25	benzylsulfamide[g]	95	dibekacin[a,b,c,d,e]
26	biapenem[h]	96	dichloramine[e,h]
27	brodimoprim[c,e,i]	97	dicloxacillin[b,c,f,j]
28	butirosin[d,e,i,j]	98	difloxacin[a,e,h]
29	carbenicillin[a,e]	99	diphenicillin sodium[f,i]
30	carfecillin sodium[a,b,d]	100	doxycycline[b,d,e,g,h,i]
31	carindacillin[a,e,f,i]	101	enoxacin[a,b,d,h,j]
32	carumonam[a,b,c,d,j]	102	enrofloxacin[b]
33	cefaclor[d,e,g,i,j]	103	epicillin[a,e,h]
34	cefadroxil[b,d,g,i,j]	104	fenbenicillin[c,e,j]
35	cefamandole[b,i]	105	fleroxacin[a,e,h]
36	cefatrizine[a,g,j]	106	flomoxef[b,e]
37	cefazedone[d,e,f,h,j]	107	florfenicol[b,e,j]
38	cefazolin[a,c,i,j]	108	floxacillin[b,f,i,j]
39	cefbuperazone[g,h,i,j]	109	flumequine[b,c]
40	cefcapene pivoxil[b,j]	110	fortimicin a[b,c]
41	cefclidin[a,i,j]	111	fortimicin b [a]
42	cefdinir[e,i]	112	furaltadone[f,j]
43	cefditoren[a,f,g,i]	113	gentamicin c1[b,g]
44	cefepime[b,c,d,h]	114	gentamicin c2[f]
45	cefetamet[a,c,g]	115	gentamicin c3[a]
46	cefixime[b,e,h,i]	116	grepafloxacin[d,e,f]
47	cefmenoxime[d,e,f,i]	117	guamecycline[g,h]
48	cefmetazole[b,e,g,h]	118	hetacillin[a,d,i,j]
49	cefminox[g,i]	119	imipenem[d,e,f,i]
50	cefodizime[c,f]	120	isepamicin[b,d,h,i]
51	cefonicid[g,h]	121	kanamycin a[a,e,j]
52	cefoperazone[b,c,e,f]	122	kanamycin b[c,e,i]
53	ceforanide[a,c]	123	kanamycin c[b,d,e,h]
54	cefotaxime[b,d,e,h,i]	124	lenampicillin[b,f,g]
55	cefotetan[g,i]	125	lincomycin[d,i]
56	cefotiam[d,e,g]	126	lomefloxacin[a]
57	cefoxitin[a,e]	127	loracarbef[d,h,j]
58	cefozopran[b,c,d,f,h,i]	128	lymecycline[d,j]
59	cefpimizole[c]	129	mafenide[g]
60	cefpiramide[a,d,f,j]	130	meclocycline[a,b,c,d,j]
61	cefpirome[g,h]	131	meropenem[d,h,i,j]
62	cefpodoxime proxetil[d,h,j]	132	metampicillin[b,e,g,i]
63	cefprozil[c,d,e,f,g,j]	133	methacycline[b,f,j]
64	cefroxadine[e]	134	methicillin sodium[e,i]
65	cefsulodin[e,i,j]	135	mezlocillin[e,f,g]
66	ceftazidime[g,h]	136	micronomicin[a,j]
67	cefteram[c,e,f]	137	miloxacin[f,g]
68	ceftezole[d,e,g,h,i,j]	138	minocycline[a,d,g,h,i]
69	ceftibuten[a,c,e,g,i,j]	139	moxalactam[a,g,i]
70	ceftizoxime[e,f,i]	140	n2-formylsulfisomidine[a,b,c,g]
71	ceftriaxone[e,g]	141	n4-sulfanilylsulfanilamide[a,b,c,d,f,j]
72	cefuroxime[d,j]	142	nadifloxacin[f]
73	cefuzonam[a,c,f,i]	143	nafcillin sodium[c,e]
74	cephacetrile sodium[f]	144	nalidixic acid[c,j]
75	cephalexin[b,c,j]	145	neomycin a[c,i,j]
76	cephaloglycin[j]	146	neomycin b[a,d,h,i]

Table 1 (Continued)

	antibacterials		antibacterials
147	netilmicin[a,b,g,h,j]	199	sulbenicillin[c,g,j]
148	nifuradene[a,b]	200	sulfabenzamide[a,e,h,i,j]
149	nifuratel[d]	201	sulfacetamide[b,c,f,h]
150	nifurfoline[f]	202	sulfachlorpyridazine[a,c,g]
151	nifurpirinol[h,j]	203	sulfachrysoidine[a,b]
152	nifurprazine[a,c,e,i]	204	sulfacytine[a,i,j]
153	nifurtoinol[c,h,j,f]	205	sulfadiazine[a,f,g,h,i,j]
154	nitrofurantoin[e,i]	206	sulfadicramide[a,b,d,e]
155	noprylsulfamide[a,d,g,i]	207	sulfadimethoxine[a,b,c,d,j]
156	norfloxacin[a,f,j]	208	sulfadoxine [a,f]
157	N-sulfanilyl-3,4-xylamide[d]	209	sulfaethidole[b]
158	ofloxacin[b,f,i,j]	210	sulfaguanidine[d]
159	oxacillin[b,f,g,h]	211	sulfaguanol[g]
160	oxolinic acid[c,d,f]	212	sulfalene[d,e,i,j]
161	oxytetracycline[f]	213	sulfaloxic acid[c,d,e,f,j]
162	panipenem[b,e,j]	214	sulfamerazine[h]
163	paromomycin[a,c,h,j]	215	sulfamethazine[f,j]
164	parsiniazide[f,g,i]	216	sulfamethizole[a,c,d,g,i,j]
165	pazufloxacin[d,g,h]	217	sulfamethomidine[d,f,h,i]
166	pefloxacin[a]	218	sulfamethoxazole[g,i]
167	penamecillin[a,c,h]	219	sulfamethoxypyridazine[f,h]
168	penethamate hydriodide[j]	220	sulfametrole[e,h]
169	penicillin G potassium[a,e,h]	221	sulfamidochrysoidine[f,g,h,i]
170	penicillin N[f,h]	222	sulfamoxole[d,e,f,h,i]
171	penicillin O[c,d,f,j]	223	sulfanilamide[a,b]
172	penicillin V[e,f]	224	sulfanilic acid[f,h,i]
173	phenethicillin potassium[g,j]	225	sulfanilylurea[d,g,h]
174	phthalylsulfathiazole[g,i]	226	sulfanitran[h,i]
175	pipacycline[c,d,e,f]	227	sulfaperine[c,e,h,i]
176	pipemidic acid[b,d,e,j]	228	sulfaphenazole[b]
177	piperacillin[a,h]	229	sulfaproxyline[g]
178	piromidic acid[b,c]	230	sulfapyrazine[d,f]
179	pivampicillin[f]	231	sulfathiazole[b,g,i]
180	pivcefalexin[a,c]	232	sulfathiourea[h]
181	p-nitrosulfathiazole[a,e]	233	sulfisomidine[d,g,j]
182	propicillin[a,d,e,f,h]	234	sulfisoxazole[g,i,j]
183	p-sulfanylbenzylamine[c]	235	sultamicillin[b,c,d,g,h,j]
184	quinacillin[a,c,d,e,i]	236	talampicillin[d,e,g,h]
185	ribostamycin[d,f,h,i,j]	237	temafloxacin[e,g,h]
186	rifamide[a,b,c,g]	238	temocillin[a,g,j]
187	rifamycin sv[b,h,j]	239	tetracycline[b,f,g,h,i]
188	rifaximin[b,f,g]	240	tetroxoprim[a,e]
189	ritipenem[a,d,i]	241	thiamphenicol[a,c,f,i]
190	rolitetracycline[c,g,i,j]	242	thiazolsulfone[b,c,f]
191	rosoxacin[a,b,c]	243	ticarcillin[a,f,h]
192	rufloxacin[f,g]	244	tigemonam[g,h]
193	salazosulfadimidine[a,c,g]	245	tobramycin[b]
194	sancycline[a,e,f]	246	tosufloxacin[b,c,h,j]
195	sisomicin[a,i,j]	247	trimethoprim[b,c,e,g,h,i]
196	sparfloxacin[d,e,h,i]	248	trospectomycin[a,f]
197	spectinomycin[c,d,f,g,h]	249	trovafloxacin[b]
198	succinylsulfathiazole[b,c,e,f]		

[a] The letters in parentheses following the names indicate which, if any, of the random test sets the drugs were assigned to as described in Table 4.

RESULTS

The antibacterial activity of a total of 661 compounds was obtained from the study of Tomás-Vert et al.[20] Three compounds were noted as having duplicate entries (see Table 1), and the duplicate entry was discarded. The drugs in the data set were an extremely large and structurally heterogeneous group of compounds, representing a great number of types of pharmaceuticals.

Preselection of Variables. One-way analysis of variance identified certain descriptors as being important for classification of activity. Input of the descriptors into both LDA and BLR indicated that six descriptors were required to model activity optimally. Fewer than six descriptors resulted in decreased statistical fit, while the inclusion of more than six descriptors failed to improve statistical fit. The six descriptors found to be most efficient were as follows: octanol/water partition coefficient (ClogP); intramolecular hydrogen bond descriptors for the product of E-state indices for hydrogen bond donor and hydrogen bond acceptors, separated by two bond lengths (i.e., E-state of hydrogen bond donor * hydrogen bond acceptor) (SHBint2); the count of the product of E-state indices of hydrogen bond donor and hydrogen bond acceptor separated by three bond lengths (i.e., integer count of how many of these descriptors are present in the molecule) (SHBint3_acnt); the count of the product of E-state indices of hydrogen bond donor and hydrogen

Table 2. Summary of Calculated Physicochemical and Structural Descriptors

descriptors calculated
logarithm of the octanol/water partition coefficient (ClogP) molecular connectivities simple ($^0\chi$, $^1\chi$, $^2\chi$); path ($^3\chi_p$,$^4\chi_p$,$^5\chi_p$,$^6\chi_p$,$^7\chi_p$,$^8\chi_p$,$^9\chi_p$,$^{10}\chi_p$) cluster ($^3\chi_c$,$^4\chi_c$) path-cluster ($^4\chi_{pc}$) chain ($^3\chi_{ch}$,$^4\chi_{ch}$,$^5\chi_{ch}$,$^6\chi_{ch}$,$^7\chi_{ch}$,$^8\chi_{ch}$,$^9\chi_{ch}$,$^{10}\chi_{ch}$); valence ($^0\chi^v$,$^1\chi^v$,$^2\chi^v$,); valence-path ($^3\chi^v_p$,$^4\chi^v_p$,$^5\chi^v_p$,$^6\chi^v_p$,$^7\chi^v_p$,$^8\chi^v_p$,$^9\chi^v_p$,$^{10}\chi^v_p$); valence-cluster ($^3\chi^v_c$,$^4\chi^v_c$); valence-path-cluster ($^4\chi^v_{pc}$); valence-chain ($^3\chi^v_{ch}$,$^4\chi^v_{ch}$,$^5\chi^v_{ch}$,$^6\chi^v_{ch}$,$^7\chi^v_{ch}$,$^8\chi^v_{ch}$,$^9\chi^v_{ch}$,$^{10}\chi^v_{ch}$); electrotopological-state indices (for SsCH3, SdCH2, SssCH2, SdsCH, SaaCH, SddC, SdssC, SaasC, SaaaC, SssssC, SsNH2, SssNH, SdsN, SaaN, SsssN, SdaaN, SsOH, SdO, SssO, SaaO, SsF, SssS, SaaS, SdddsS, SsCl, SsBr); HE-state categories (HsOH, HsNH2, HssNH, Hother, Hmax, Gmax, Hmin, Gmin, Hmaxpos); counts for number of elements (SsCH3_acnt, SssCH2_acnt, SdsCH_acnt, SaaCH_acnt, SsssCH_acnt, SdssC_acnt, SaasC_acnt, SaaaC_acnt, SssssC_acnt, SsNH2_acnt, SssNH_acnt, SdsN_acnt, SaaN_acnt, SsssN_acnt, SdaaN_acnt, SsOH_acnt, SdO_acnt, SssO_acnt, SaaO_acnt, SsF_acnt, SssS_acnt, SaaS_acnt, SdddsS_acnt, SsCl_acnt, SsBr_acnt); dipolar descriptors (Qs, Qsv, Qv), dipole moment, the sum of absolute values of the charges on each atom of a molecule, in electrons (ABSQ), sum of absolute charges on nitrogen and oxygen atoms in a molecule (ABSQon), the largest, positive charge on a hydrogen atom (MaxHp); the largest negative charge over the atoms in a molecule (MaxNeg); the largest positive charge over the atoms in a molecule (MaxQp); ovality of a molecule, molecular surface, molecular volume, molecular polarizability, specific polarizability (polarizability/volume), components of dipole moments along the inertial *X*, *Y*, and *Z*-axis (Px, Py, Pz) magnitude of dipole moment (P), magnitude of principal quadrupole moment (Q); the principal moment of inertia along the *X*, *Y*, and *Z*-axis (Ix, Iy, Iz); the component of the displacement between the center-of-mass and center-of-dipole along the inertial *X*, *Y*, and *Z*-axis (Dx, Dy, Dz); the xx (yy) component of the second rank tensor in the inertial coordinate frame of reference translated so that its origin (Qxx, Qyy); kappa shape indices (κ0, κ1, κ2, κ3, $\kappa\alpha$1, $\kappa\alpha$2, $\kappa\alpha$3); internal H-bond E-state indices (SHBint2, SHBint3, SHBint4, SHBint5, SHBint6, SHBint7, SHBint2_Acnt, SHBint3_Acnt, SHBint4_Acnt, SHBint5_Acnt, SHBint6_Acnt, SHBint7_Acnt, SHBint8_Acnt, SHBint9_Acnt, SHBint10_Acnt); molecular properties (fw, nvx, nelem, nrings, ncirc, phia, knotp, knotpv); number of H-bond acceptors (numHBa); number of H-bond donors (numHBd); descriptors SHBa, SHHBd, SHwBd

Table 3. Correlation Coefficients between the Six Most Significant Variables

descriptor	ClogP	SHBint2	SHBint3_acnt	SHBint6_acnt	NumHBa	SHHBd
ClogP	1.000					
SHBint2	0.048	1.000				
SHBint3_acnt	−0.302	0.112	1.000			
SHBint6_acnt	−0.278	−0.004	0.441	1.000		
NumHBa	−0.308	0.086	0.553	0.352	1.000	
SHHBd	−0.592	−0.019	0.452	0.494	0.551	1.000

bond acceptor separated by six bond lengths (SHBint6_acnt); the number of H-bond acceptors (NumHBa); and the sum of E-state indices of hydrogen bond donors (SHHBd). The six descriptors were also chosen to ensure that the correlations between descriptors were as low as possible (see Table 3).

Linear Discriminant Analysis. Following preselection of variables, the following significant canonical linear discriminant function was obtained

$$p = -0.193\ \text{ClogP} + 0.033\ \text{SHBint2} - 0.101\ \text{SHBint3\_acnt} + 0.335\ \text{SHBint6\_acnt} + 0.235\ \text{NumHBa} - 0.104\ \text{SHHBd} - 1.856$$

where p is the classification score or probability of group membership.

The ability of the model to classify the compounds, and its predictivity, are recorded in Table 4.

Binary Logistic Regression. Following preselection of variables, the following logistic function was obtained

$$\ln(p_1/p_2) = -0.645\ \text{ClogP} + 0.104\ \text{SHBint2} - 0.325\ \text{SHBint3\_acnt} + 0.696\ \text{SHBint6\_acnt} + 0.894\ \text{NumHBa} - 0.255\ \text{SHHBd} - 7.428$$

where (p_1/p_2) is the odds ratio calculated from the probability of an object belonging to group 1 (p_1) and group 2 (p_2) where group 1 represents antibacterial activity and group 2 represents lack of antibacterial activity.

Table 4. Classification from Original Models and Predictivity Based on the Six Selected Parameters Following the Use of Ten Randomly Selected Subsets of Linear Discriminant Analysis (LDA) and Binary Logistic Regression (BLR) Models

	correctly predicted nonantibacterials (%)		correctly predicted antibacterials (%)		total correctly predicted (%)	
	LDA	BLR	LDA	BLR	LDA	BLR
classification	97.1	96.4	85.1	92.0	92.6	94.7
random sample (a)	96.8	94.3	85.1	94.6	92.4	94.4
random sample (b)	96.8	96.8	85.1	93.2	92.4	95.4
random sample (c)	98.4	98.4	81.1	82.4	91.9	92.4
random sample (d)	96.0	91.9	90.5	95.9	93.9	93.4
random sample (e)	97.6	96.0	83.8	91.9	92.4	94.4
random sample (f)	96.8	95.2	87.8	93.2	93.4	94.4
random sample (g)	95.2	95.2	89.2	90.5	92.9	93.4
random sample (h)	98.4	97.6	89.2	93.2	94.9	95.6
random sample (i)	98.4	97.6	87.8	90.5	94.4	94.9
random sample (j)	98.4	94.3	94.6	94.6	97.0	94.4
mean	97.3	95.7	87.4	92.0	93.6	94.3
standard deviation	1.15	1.94	3.83	3.79	1.54	0.97

The ability of the model to classify the compounds, and its predictivity, is recorded in Table 4.

DISCUSSION

A highly diverse data set of drug compounds has been assembled for this study. These compounds have been classified according to whether they have antibacterial activity. The study describes the development of models, based on 2-D physicochemical and structural properties, for

the classification of antibacterial activity. The purpose of these models is to allow screening and assessment of larger libraries of compounds.

A large number of 2-D physicochemical and structural descriptors were calculated for each drug substance in this study. Selection of significant variables for classification of antibacterial activity was required. This was performed in this study by one-way analysis of variance. This is a relatively novel method of selecting variables but allows for a more rational selection of descriptors than is possible with methods such as stepwise selection (i.e. stepwise discriminant analysis). Further, selection of variables in this manner resulted in the selection of relatively uncorrelated variables for modeling purposes (see Table 3). A process of trial and error, comparing prediction rates for both statistical techniques, indicated that six was the optimal number of descriptors required to model antibacterial activity.

A prerequisite of a good predictive model for any biological activity is that it should be transparent and mechanistically interpretable.[25] To achieve transparency and mechanistic interpretability, the physicochemical meaning of parameters utilized for the modeling needs to be elucidated. Mechanistic interpretation of QSAR descriptors is seldom easy, especially for heterogeneous data sets. In this analysis six descriptors were found to be important. The most significant parameter selected was log P, a fundamental descriptor of molecular hydrophobicity.[26] Both LDA and BLR models suggest that lower hydrophobicity is required for antibacterial activity. While there is no immediate mechanistic reason for this finding, it suggests that antibacterial compounds have, in general, a lower log P than other drugs. Log P has been demonstrated to be important in classifying drug substances, with compounds with drug activity having a lower mean log *P* value than nondrugs.[4,6] The other parameters in the models all account for the effects of inter- and intramolecular hydrogen bonding.[27] Hydrogen bonding is found to be important on two levels. First, hydrogen bonding may be associated with the mechanism of action of particular groups of antibiotics. Second, certain classes of antibiotics (e.g. penicillins) have structures associated with large numbers of functional groups associated with hydrogen bonding. Generally, however, the hydrogen bonding ability of large numbers of functional groups is not associated with drug-like molecules.[8] The most common reason for this is perceived to be problems with uptake following oral administration.

Both the LDA and BLR models were highly successful at predicting antibacterial activity (the results are summarized in Table 4). As expected, the statistics for classification are generally better than for the prediction of the test sets. However, the very small decrease in predictivity from classification (i.e. predicting within the model) to the use of test sets (i.e. predicting outside of the model) indicates that stable statistical models were formed by both statistical techniques. The overall average predictions from the 10 randomly chosen tests suggest that BLR provides a slightly more predictive model than does LDA. Further examination of the data reveals that LDA predicts compounds to have no antibacterial activity very well (approximately 97% correctly predicted) and significantly better than BLR. This point indicates that few, if any, compounds with no antibacterial activity are predicted as being active (i.e. few false positives). BLR, however, is slightly better at predicting compounds with antibacterial activity (approximately 92% correctly predicted as compared to 87% from LDA). This is in agreement with the findings of Worth and Cronin[22] using LDA and BLR to model eye irritation. Thus, both statistical techniques, and BLR in particular, provide good methods to screen databases, and there is a high degree of confidence in the predicted activity for compounds predicted to be active. It should be noted, however, that the asymmetric nature of the data set (i.e. more inactive than active compounds) may also influence the results and the relative predictivity of the classes.

The models reported in this paper compare favorably with the neural network approach described by Tomás-Vert et al.,[20] using the same data set, and previous models based on fewer compounds.[28,29] Tomás-Vert et al.[20] achieved a similar overall classification rate for activity of 94.8% compounds correctly classified from a testing set comprising 30% of the original data set. The neural network proposed by Tomás-Vert et al.[20] was slightly less successful at predicting compounds with no activity (95.9%), though rather better at predicting compounds with activity (93.6%), than are the models reported in this paper. The former point indicates that there will be an increased probability of false positives if the method is used for screening. There are a number of clear advantages in the use of the LDA and BLR models reported in this study, as opposed to a neural network approach. The neural network is based upon a total of 62 descriptors, as opposed to only six in this study. Clearly there must be redundancy in the original data set used in the neural network. Further, due to the high number of variables in the neural network it is difficult to ascribe any mechanistic basis to the descriptors, and thus interpretation of the model is virtually impossible. Also, for predictions to be made from a neural network, it must be retrained, and as such it is not easily portable or transferable.[21] The models reported in this paper are simple statistical techniques from which predictions can easily be made.

CONCLUSION

This study has examined a large, diverse group of drug substances that have been classified according to their antibacterial activity. Highly significant models for the prediction of antibacterial activity have been developed from these data using six preselected 2-D physicochemical and structural variables. The models confirm the role of molecular hydrophobicity and hydrogen bonding in controlling antibacterial activity. There are a number of novel aspects to this study. It provides a means of screening databases to identify compounds with antibacterial activity on the basis of a small number of easily calculated properties. Further, it provides a direct comparison of two statistical techniques, and use of a neural network, for the classification of biological activity.

ACKNOWLEDGMENT

This work was supported in part by the European Union IMAGETOX Research Training Network (HPRN-CT- 1999-00015).

REFERENCES AND NOTES

(1) Warr, W. A. Combinatorial chemistry and molecular diversity. An overview. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 134−140.

(2) Wolcke, J.; Ullmann, D. Miniaturized HTS technologies − uHTS. *Drug Discuss. Today* **2001**, *6*, 637−646.
(3) Ajay; Bemis, G. W.; Murcko, M. A. Designing libraries with CNS activity. *J. Med. Chem.* **1999**, *42*, 4942−4951
(4) Ajay; Walters, W. P.; Murcko, M. A. Can we learn to distinguish between "drug-like" and "nondrug-like" molecules? *J. Med. Chem.* **1998**, *41*, 3314−3324.
(5) Estrada, E.; Uriarte, E. Recent advances in the role of topological indices in drug discovery research. *Curr. Med. Chem.* **2001**, *8*, 1573−1588.
(6) Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 251−264.
(7) Sadowski, J.; Kubinyi, H. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* **1998**, *41*, 3325−3329.
(8) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **1997**, *23*, 3−25.
(9) Ekins, S.; Waller, C. L.; Swaan, P. W.; Cruciani, G.; Wrighton, S. A.; Wikel, J. H. Progress in predicting human ADME parameters *in silico*. *J. Pharmac. Toxicol. Methods* **2000**, *44*, 251−272.
(10) Mason, J. S.; Good, A. C.; Martin, E. J. 3-D pharmacophores in drug discovery. *Curr. Pharmaceut. Des.* **2001**, *7*, 567−597.
(11) Beno, B. R.; Mason, J. S. The design of combinatorial libraries using properties and 3-D pharmacophore fingerprints. *Drug Discuss. Today* **2001**, *6*, 251−258.
(12) Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. A novel approach for the virtual screening and rational design of anticancer compounds. *J. Med. Chem.* **2000**, *43*, 1975−1985.
(13) Kurogi, Y.; Guner, O. F. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr. Med. Chem.* **2001**, *8*, 1035−1055.
(14) Tropsha, A.; Zhang, W. F. Identification of the descriptor pharmacophores using variable selection QSAR: applications to database mining. *Curr. Pharmaceut. Des.* **2001**, *7*, 599−612.
(15) Langer, T.; Hoffmann, R. D. Virtual screening: An effective tool for lead structure discovery? *Curr. Pharmaceut. Des.* **2001**, *7*, 509−527.
(16) Livingstone, D. J. *Data analysis for chemists. Applications to QSAR and chemical product design*; Oxford University Press: Oxford, 1995; p 239.
(17) Cronin, M. T. D.; Dearden, J. C. QSAR in toxicology. 3. Prediction of chronic toxicities. *Quant. Struct.-Act. Relat.* **1995**, *14*, 329−334.
(18) de Julian-Ortiz, J. V. Virtual Darwinian Drug Design: QSAR inverse problem, virtual combinatorial chemistry, and computational screening. *Comb. Chem. High T. Scr.* **2001**, *4*, 295−310.
(19) McDonnell, G.; Russell, A. D. Antiseptics and disinfectants: activity, action, and resistance. *Clin. Microbiol. Rev.* **1999**, *12*, 147−179.
(20) Tomás-Vert, F.; Pérez-Giménez, F.; Salabert-Salvador, Ma. T.; Garcia-March, F. J.; Jaén-Oltra, J. Artificial neural network applied to the discrimination of antibacterial activity by topological methods. *J. Mol. Struct. − Theochem* **2000**, *504*, 249−259.
(21) Cronin, M. T. D.; Schultz, T. W. Development of quantitative structure−activity relationships for the toxicity of aromatic compounds to *Tetrahymena pyriformis*: comparative assessment of the methodologies. *Chem. Res. Toxicol.* **2001**, *14*, 1284−1295.
(22) Worth, A. P.; Cronin, M. T. D. The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects. *J. Mol. Struct. − Theochem* **2002**, in press.
(23) Weininger, D. SMILES, a chemical language and information-system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31−36.
(24) Golbraikh, A.; Tropsha, A. Beware of q2! *J. Mol. Graphics Mod.* **2002**, *20*, 269−276.
(25) Cronin, M. T. D.; Schultz T. W. Pitfalls in QSAR. *J. Mol. Struct. − Theochem* **2002**, in press.
(26) Dearden, J. C. Partitioning and lipophilicity in quantitative structure−activity relationships. *Environ. Health Persp.* **1985**, *61*, 203−228.
(27) Kier, L. B.; Hall, L. H. *Molecular Structure Description. The Electrotopological State*; Academic Press: San Diego, CA, 1999; p 288.
(28) Garcia-Domenech, R.; de Julián-Ortiz, J. V. Antimicrobial activity characterization in a heterogeneous group of compounds. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 445−449.
(29) Mishra, R. K.; Garcia-Domenech, R.; Galvez, J. Getting discriminant functions of antibacterial activity from physicochemical and topological parameters. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 387−393.

CI025501D