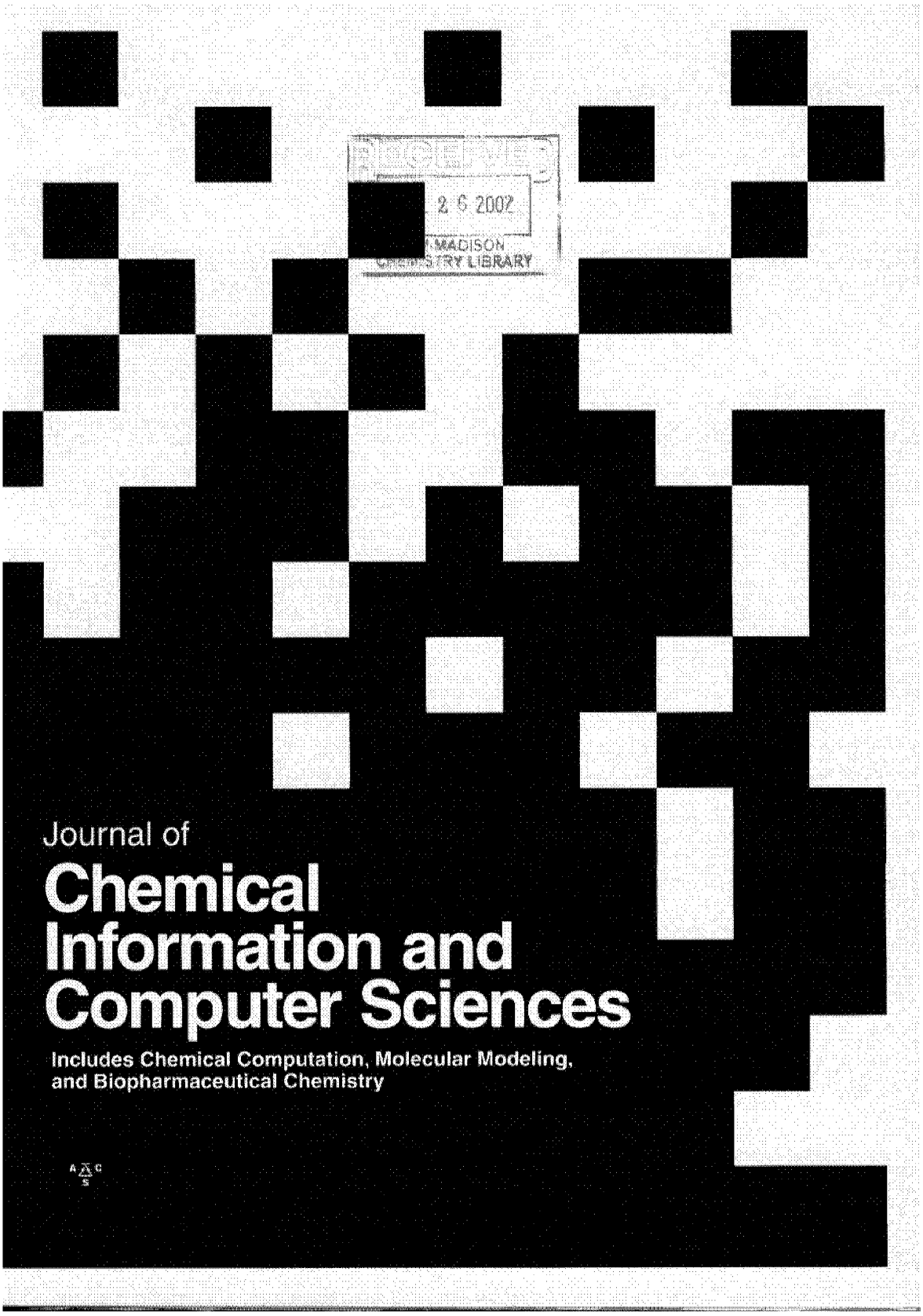

Journal of
Chemical
Information and
Computer Sciences

Journal of Chemical Information and Computer Sciences (ISSN 0095-2338) is published bimonthly by the American Chemical Society at 1155 16th St., N.W., Washington, DC 20036. Periodicals postage paid at Washington, DC, and additional mailing offices. POSTMASTER: Send address changes to *Journal of Chemical Information and Computer Sciences*, Member & Subscriber Services, P.O. Box 3337, Columbus, OH 43210.

American Chemical Society
1155 16th St., N.W.
Washington, DC 20036
(202) 872-4614
TDD (202) 872-6076
Fax (202) 776-8264

Journal Publications
American Chemical Society
2540 Olentangy River Road
P.O. Box 3330, Columbus, OH 43210
(614) 447-3665
Fax (614) 447-3745; Telex 6842086
E-mail acsproof@acs.org

Member & Subscriber Services
American Chemical Society
P.O. Box 3337, Columbus, OH 43210
(614) 447-3776; (800) 333-9511
Fax (614) 447-3671
E-mail service@acs.org

Advertising Office
Centcom, Ltd.
676 East Swedesford Road
Suite 202, Wayne, PA 19087-1612
(610) 964-8061

Publications Division
Robert D. Bovenschulte, Director

Publishing Operations
Mary E. Scanlan, Director

Journal Publications
Anne C. O'Melia, Manager; Diane E. Needham, Journals Editing Manager; Shana Sullivan, Associate Editor; Katherine S. Midnight, Production Associate

Composition Services
Teresa K. Lewandowski, Manager

Sales & Marketing
Justin R. Spence, Director
Dean J. Smith, Sales
Matthew J. Price, Marketing

Canadian GST Reg. No. 127571347
Printed in the USA

Editorial Information

Instructions for Authors and Copyright Status Form: See the first issue of each volume or visit the Publications Division Web site (http://pubs.acs.org) for instructions on preparing and submitting manuscripts and copyright forms to the Editor.

Manuscript Submission: Send manuscripts to the Editor or one of the Associate Editors or submit over the Web at http://pubs.acs.org/jcics. Include a signed ACS copyright status form, a copy of which appears in every issue.

Accepted Papers and Proofs: Direct correspondence to Journal Publications, Columbus, OH. Tel: (614) 447-3665. Fax: (614) 447-3745. E-mail: acsproof@acs.org.

Journal Policies: The American Chemical Society and its Editors assume no responsibility for the statements and opinions advanced by contributors. Registered names and trademarks, etc., used in this publication, even without specific indication thereof, are not to be considered unprotected by law.

Document Number: At the end of each document is a 9-character code that serves as a link between the printed and electronic products and facilitates retrieval of the document in electronic form.

Digital Object Identifier (DOI): The DOI identification system for digital media has been designed to provide persistent and reliable identification of digital objects. Information on the DOI and its governing body, the International DOI Foundation, can be found at http://www.doi.org. In the Web editions of ACS journals, the DOI appears at the top of the HTML version of an article and at the bottom of the first page in its PDF version; in the print editions, the DOI appears in the same location as in the PDF version.

2002 Subscription and Ordering Information

		North America	Outside North America[a]
Printed Edition	Members	$ 60	$ 95
	Student Members	$ 45	$ 80
	Nonmembers	$206	$241
	Institutional	$486	$521
Web Edition[b]	Members	$ 25	$ 25
Print/Web Combination[b]	Members	$ 85	$120

[a] Air service included. [b] For Institutional rates, call Member & Subscriber Services.

Web Edition: This journal is available to subscribers via the Internet. Contact Member & Subscriber Services [Tel: (614) 447-3776 or (800) 333-9511. Fax: (614) 447-3671. E-mail: service@acs.org] or visit the Publications Division Web site (http://pubs.acs.org) for additional details.

New and Renewal Subscriptions: Send with payment to American Chemical Society, P.O. Box 182426, Columbus, OH 43218-2426. Institutional subscribers in Japan must enter subscription orders with Maruzen Company Ltd., 3-10 Nihonbashi 2-chome, Chuo-ku, Tokyo 103, Japan. Tel: (03) 272-7211.

Subscription Donations: Members may donate/share their personal subscriptions to/with libraries but only after 5 years from publication.

Change of Address: Notify Member & Subscriber Services, ACS, Columbus, OH. Tel: (614) 447-3776 or (800) 333-9511. Fax: (614) 447-3671. E-mail: address@acs.org. Include both old and new addresses and a mailing label from a recent issue.

Microfilm, Microfiche, Back Issue, and Printed Edition Single Issue Orders: Send requests to Publications Support Services, ACS, Washington, DC. Tel: (202) 872-4376. Fax: (202) 872-6325. E-mail: pss@acs.org. Printed edition not available prior to 1998.

Bulk Reprint Orders: For quotes and information, contact Cadmus Reprints. Tel: (888) 257-2134 or (410) 819-3995. Fax: (410) 820-9765.

Claims for Issues Not Received: Claims will be honored only if submitted within 90 days of the issue date (subscribers in North America) or within 180 days of the issue date (all other subscribers). Send requests to Member & Subscriber Services, ACS, Columbus, OH. Tel: (614) 447-3776 or (800) 333-9511. Fax: (614) 447-3671. E-mail: service@acs.org.

Supporting Information (SI): SI from 1995 to present is available free of charge via the Internet (http://pubs.acs.org) from the journal's home page. For information on electronic access, send E-mail to si-help@acs.org. SI prior to 1995 is available, for a fee, from Publications Support Services. Tel: (202) 872-4376. Fax: (202) 872-6325. E-mail: pss@acs.org.

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,e,f,g,i,j]	71	butabarbital[f,i]
2	5-bromosalicylic acid acetate[g]	72	butacetin[f,g]
3	5-nitro-2propoxyacetanilide[c,h]	73	butaclamol[f,e,r]
4	acecarbromal[f,g]	74	butallylonal[a,d,g,h,i]
5	aceclofenac[e,f,h,i]	75	butanilicaine[b,c]
6	acefylline[c,i,d,e,g]	76	butibufen[f,g]
7	acetaminophen[b,i]	77	butidrine hydrochloride[d,e,i]
8	acetaminosalol	78	butoctamide[a,b,h,i,j]
9	acetanilide[a,e,f,i]	79	butofilolol[d,e,f,h]
10	acetazolamida[a,d,f,h,i]	80	caffeine[b,c,e,n,h,i]
11	acetobutolol[a,d,f,h,i]	81	capuride[f]
12	acetophenazine[b,e,f,n]	82	carazolol[a,h,i]
13	acetylsalicylic acid[c,d,e,g]	83	carbamazepine[a,b,h,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,b,i,j]	86	carbiphene[e,i]
17	alclofenac[e,g,j]	87	carbocloral[b,h,i]
18	alminoprofen[h,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,n,h,j]
22	aminopropylon[g]	92	carprofen[b,c,e,g,i]
23	aminopyrine[a,b,e,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	cellprolol[c,e,i]
27	antipyrine[f,h]	97	cetamolol[f,h,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[f,i]
31	arotinolol[b,c,e,h,i]	101	chloropyramine[e,r,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[n,r,f,i]
36	bamipine[e,i]	106	chlorthenoxacin[e,n]
37	beclofibrate[e,g,i]	107	chlorcyclizine[e,h]
38	befunolol[f,h]	108	cinchophen[b,c,e,g,n]
39	benfluorex[b,i]	109	cinmetacin[a,h,i,j]
40	benorylate[a,n,d,j]	110	cinnarizine[b,n,d,e,g,h]
41	benoxaprofen[f,h]	111	cinromida[a,e,j]
42	benserazide[d,e]	112	ciprofibrate
43	benzitramide[n,f,h]	113	ciramadol[e,h,i,j]
44	benzotropine mesylate[e]	114	clemastine[h,i]
45	benzpiperylon[f]	115	clenbuterol[c,d,e,j]
46	benzydamine[b,d,i]	116	clidanac[n,e]
47	bermoprofen[c,h]	117	clinofibrate[d,f,h,j]
48	betaxolol[a,e,j,i]	118	clocinizine[a,b,c,e,f,i,n]
49	bevantolol[b,g]	119	clofibrate[c,e,g,h,j]
50	bevonium methyl sulfate[a,b,d,f]	120	clofibric acid[f,i]
51	bezafibrate[e,g]	121	clometacin[b,h]
52	binifibrate[n,e]	122	clometiazol[d,f]
53	bisoprolol[a,e]	123	clonixin[b]
54	bitolterol[d,e,g,h,i]	124	clopirac[e,g]
55	bucloxic acid[a,b,c,g,h]	125	cloralsalicylamide[n,b,e,i,j]
56	bopindolol[c,f]	126	cloranolol[b,c,f,g,h,i,j]
57	bromfenac[a,h,j]	127	clordesmetildiazepam[f,n,e,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,i]	130	clozapine[f]
61	brotizolam[j]	131	codeine[b,e,f,i]
62	bucetin[b,h]	132	cropropamide[d,i]
63	bucolome[c,e]	133	crotethamide[n,c,d,e]
64	bucumolol[a,b,j]	134	deserpidine[f]
65	bufetolol[e,f,h,i,j]	135	diazepam[a,h,e,g]
66	bufexamac[f,g,i]	136	diclofenac[b,d,e]
67	bufuralol[e,g,h]	137	diethylbromoacetamide[b,c,g,h]
68	bumadizon[d,h]	138	difenamisole[e,g,i]
69	bunitrolol[h,i]	139	difenpiramide[b,i,j]
70	bupranolol[e,f,h]	140	diflunisal[b,d,g]

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.