Identification and Prediction of Promiscuous Aggregating Inhibitors among Known Drugs

James Seidler,[†,⊥] Susan L. McGovern,[†,⊥] Thompson N. Doman,[‡] and Brian K. Shoichet*[,§]

Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, 303 East Chicago Avenue, Chicago, Illinois 60611, and Pfizer Incorporated, 4901 Searle Parkway, Skokie, Illinois 60062, and Department of Pharmaceutical Chemistry, University of California, San Francisco, Genentech Hall, 600 16th Street, San Francisco, California 94143

Received April 21, 2003

Some small molecules, often hits from screening, form aggregates in solution that inhibit many enzymes. In contrast, drugs are thought to act specifically. To investigate this assumption, 50 unrelated drugs were tested for promiscuous inhibition via aggregation. Each drug was tested against three unrelated model enzymes: β-lactamase, chymotrypsin, and malate dehydrogenase, none of which are considered targets of these drugs. To be judged promiscuous, the drugs had to inhibit all three enzymes, do so in a time-dependent manner, be sensitive to detergent and to enzyme concentration, and form particles detectable by light scattering. Of the 50 drugs tested, 43 were nonpromiscuous by these criteria. Surprisingly, four of the drugs showed promiscuous, aggregation-based inhibition at concentrations below 100 μM: clotrimazole, benzyl benzoate, nicardipine, and delavirdine. Three other drugs also behaved as aggregation-based inhibitors, but only at high concentrations (about 400 μM). To investigate possible structure−activity relationships among promiscuous drugs, five analogues of the antifungal clotrimazole were studied. Three of these, miconazole, econazole, and sulconazole, were promiscuous but the other two, fluconazole and ketoconazole, were not. Using recursive partitioning, these experimental results were used to develop a model for predicting aggregate-based promiscuity. This model correctly classified 94% of 111 compounds—47 aggregators and 64 nonaggregators—that have been studied for this effect. To evaluate the model, it was used to predict the behavior of 75 drugs not previously investigated for aggregation. Several preliminary points emerge. Most drugs are not promiscuous, even at high concentrations. Nevertheless, at high enough concentrations (20−400 μM), some drugs can aggregate and act promiscuously, suggesting that aggregation may be common among small molecules at micromolar concentrations, at least in biochemical buffers.

Introduction

The efficiency of high throughput and virtual screening is compromised by hits that later prove to be false positives. Some of these false positives are promiscuous compounds that act noncompetitively and show little relationship between structure and function. Recently, we have found that one mechanism of inhibitor promiscuity is aggregate formation: individual molecules group together to form particles 30−1000 nm in diameter, and these aggregates are the active inhibitory species.[1]

Although aggregation as a mechanism for inhibition is a recent proposal, the problem of false positives from screening is well-known. There is an extensive literature on distinguishing compounds that may act as false positives from those that are "lead-like" and "drug-like."[2−9] Disconcertingly, we recently found that eight widely studied lead inhibitors of kinases, such as quercetin and rottlerin, also acted as promiscuous, aggregation-based inhibitors.[10] If these highly studied molecules can act promiscuously, we wondered whether drugs might do so as well.

To investigate this question, we tested 50 diverse drugs for inhibition of three unrelated model enzymes: β-lactamase, chymotrypsin, and malate dehydrogenase. These enzymes recognize dissimilar ligands, and none of them are considered targets for the drugs tested here. To be judged as a promiscuous, aggregation-based inhibitor, a drug had to meet the following criteria. First, it had to inhibit all three enzymes. Second, inhibition had to have properties characteristic of aggregation-based inhibition, including time-dependence, sensitivity to detergent, and sensitivity to enzyme concentration. Finally, the drug had to form particles detectable by light scattering. To compensate for these fairly strict criteria of promiscuity, we initially tested the drugs at concentrations between 10 and 400 μM.

As expected, most of the drugs were inactive against all three enzymes, consistent with the "drug-like" behavior of these compounds. Surprisingly, several of the drugs acted as promiscuous, aggregation-based inhibitors of the model enzymes. To investigate this behavior further, we explored structure−activity (SAR) and structure−property relationships (SPR) among a larger set of over 100 known aggregators and nonaggregators. We also studied the SAR of one family of

* Corresponding author: shoichet@cgl.ucsf.edu; phone: 415-514-4126; fax: 415-502-1411.
⊥ These authors contributed equally to this work.
† Northwestern University.
‡ Pfizer Incorporated.
§ University of California.

10.1021/jm030191r CCC: $25.00
Published on Web 09/13/2003

Table 1. Inhibition by Nonpromiscuous Drugs of β-Lactamase, Chymotrypsin, and MDH[a]

compound	% inhibition		
	β-lactamase	chymotrypsin	MDH
4-aminophenyl sulfone	<5	<5	11
5-aminosalicylic acid	<5	<5	18[c]
amrinone	<5	<5	<5[e]
azelaic acid	<5	<5	8
carbamazepine	<5	<5	8
carisoprodol	<5	<5	<5
chlorambucil	<5	23	44
chlorthalidone	<5	<5	<5[c]
cinoxacin	<5	<5	<5[e]
deferoxamine mesylate	<5	<5	<5
diflunisal	<5	13	<5
diphenhydramine	<5	<5	<5
etodolac	<5	<5	9
flutamide	<5	<5	<5[d]
furosemide	<5	<5	21
gemfibrozil	<5	<5	14[c]
guaifenesin	<5	<5	<5
guanabenz	83	<5	20
indomethacin	<5	<5	7
lamotrigine	<5	<5	11
lansoprazole	<5	<5	<5
leflunomide	<5	<5	12
lidocaine	<5	<5	<5
mebendazole	<5	<5	15
nitrofurantoin	<5	91	97
omeprazole	<5	<5	<5
ondansetron	<5	<5	<5
phenazopyridine	<5	96	34
primidone	<5	<5	13
prazocin	13	<5	14[d]
propantheline bromide	<5	96[b]	<5
proptriptyline	<5	<5	<5
riboflavin	<5	<5	<5[d]
riluzole	<5	<5	<5[c]
sulfadiazine	<5	<5	<5[c]
sulfinpyrazone	<5	27	83
tacrine	<5	<5	83
tetraethylene pentamine	<5	<5	<5
thalidomide	<5	<5	<5
torasemide	<5	<5	14
triamterene	<5	<5	21[e]
trimethoprim	<5	<5	<5
tripelennamine	<5	<5	<5

[a] Drugs were assayed for β-lactamase inhibition at 100 μM, chymotrypsin inhibition at 250 μM, and MDH inhibition at 400 μM, unless otherwise noted. [b] Inhibition is time dependent. [c] 200 μM. [d] 100 μM. [e] 50 μM.

promiscuous drugs in more detail and considered how promiscuity related to structural changes among related compounds.

Results

Fifty diverse drugs were tested for inhibition of three unrelated model enzymes: β-lactamase, chymotrypsin, and malate dehydrogenase (MDH). The drugs were initially tested at 100 μM against β-lactamase, 250 μM against chymotrypsin, and 400 μM against MDH. As expected, most drugs, 35 of those tested, did not significantly inhibit any of the enzymes (Table 1). These 35 were not considered further as candidates for promiscuity.

Fifteen drugs inhibited at least one of the model enzymes at the initial concentrations and were tested against the other enzymes at concentrations up to 400 μM. Two of these, guanabenz and propantheline bromide, did not inhibit either of the other two enzymes significantly and were thus classified as nonpromiscuous (Table 1). Another, nitrofurantoin, though relatively potent against chymotrypsin and MDH, did not inhibit β-lactamase measurably even at 400 μM. Moreover, nitrofurantoin showed few of the enzymological features characteristic of aggregate-forming promiscuous compounds and did not form particles detectable by dynamic light scattering (DLS) (data not shown); it was therefore considered nonpromiscuous (Table 1). Five drugs—chlorambucil, sulfinpyrazone, indomethacin, tacrine, and phenazopyridine—showed significant inhibition of all three model enzymes at a concentration of 400 μM. However, even at 400 μM, inhibition never reached 50% for all three enzymes, and for none of these five were particles measurable by DLS. Thus, although these five drugs showed activity that suggested promiscuous behavior, they failed to meet our criteria for aggregation-based inhibition, and were considered nonpromiscuous, nonaggregate formers (Table 1). Of the 50 drugs initially tested, 43 (86%) were classified as nonpromiscuous, nonaggregating molecules (Figure 1).

Three of the drugs tested—glyburide, mefenamic acid, and oxaprozin—were classified as weakly promiscuous, aggregate forming molecules (Figure 1). Although none of these drugs were active against any of the model enzymes at 100 μM concentrations, all three showed enzymological features characteristic of aggregation-based promiscuity[1,10] at 400 μM. This is clearest for glyburide and mefanamic acid, which inhibited all three enzymes significantly at this concentration (Table 2). Inhibition for both of these drugs increased significantly on preincubation with the enzyme, and diminished dramatically when enzyme concentration was increased 10-fold (Table 3). Both compounds showed significant loss of inhibition in the presence of the mild, nonionic detergent saponin (the mechanism of this effect will be discussed elsewhere) (Table 4). Both drugs also formed clearly measurable particles by DLS (Table 5). For oxaprozin the classification is less certain, in that even at 400 μM this drug did not inhibit β-lactamase significantly (Table 2), nor was inhibition of chymotrypsin sensitive to preincubation (Table 3). However, inhibition did diminish significantly when the concentration of

Table 2. Inhibition by Promiscuous Drugs of β-Lactamase, Chymotrypsin, and MDH

compound	original target	therapeutic concentration (μM)	IC_{50} (μM)		
			β-lactamase	chymotrypsin	MDH
clotrimazole	sterol 14α-demethylase[35]	30[a 36]	20	85	35
benzyl benzoate	scabies[35]	1.2 M[b 37]	90	250	125
nicardipine	Ca^{2+} channel[35]	0.184[c 38]	20	175	50
delavirdine	reverse transcriptase[39]	0.038[d 40]	90	225	85
glyburide	K^+ channel[35]	0.215[c 41]	300	360	400
mefanamic acid	cyclooxygenase[35]	83[c 40]	350	>400[e]	225
oxaprozin	cyclooxygenase[35]	372[c 42]	>400	200	175

[a] Fungicidal concentration. [b] Topical dosage. [c] Mean maximum plasma concentration. [d] IC_{50}. [e] 41% inhibition at 400 μM.

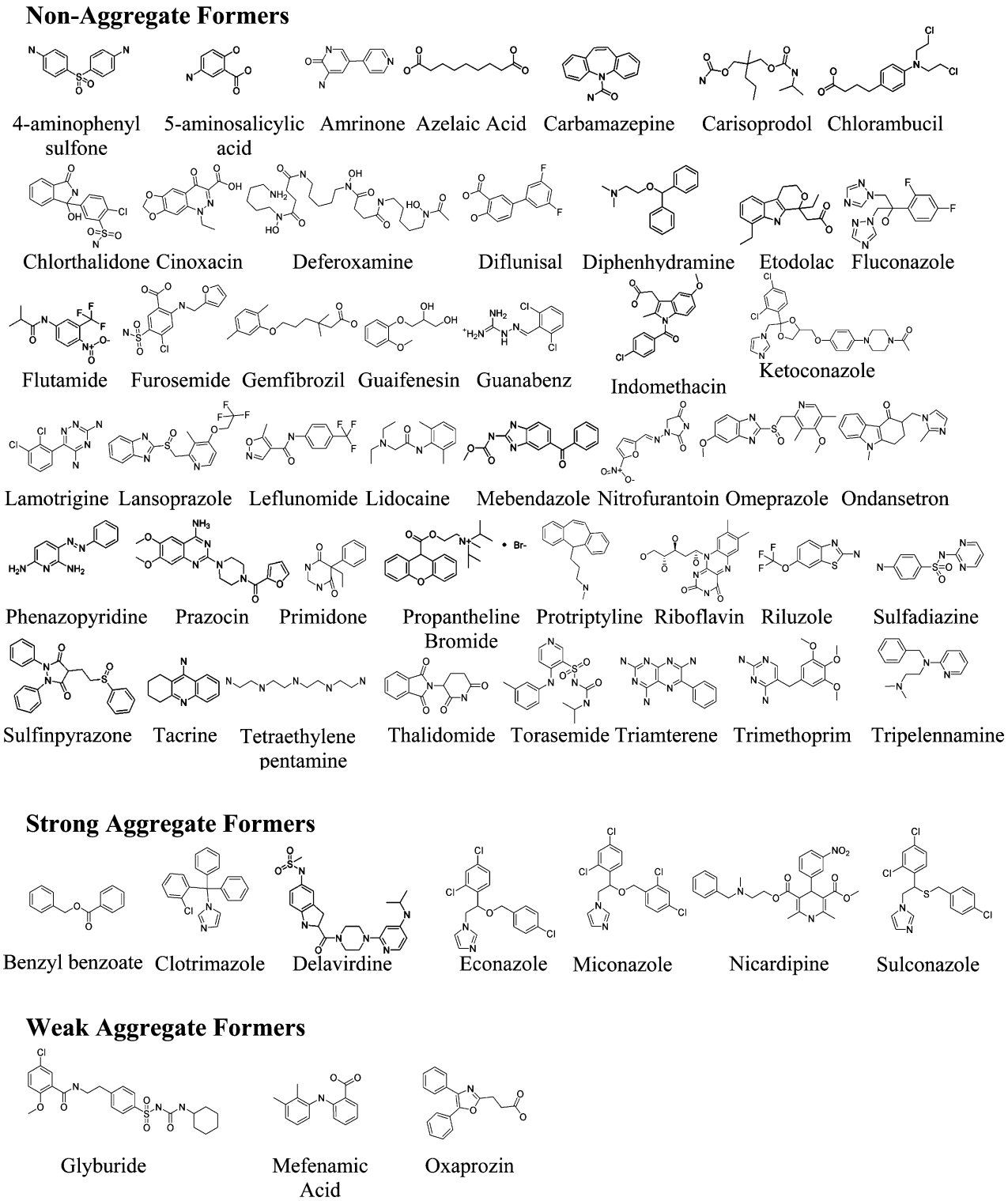

Figure 1. Drugs assayed for promiscuous inhibition.

chymotrypsin was raised 10-fold (Table 3) or when saponin was used in the assay (Table 4), and oxaprozin did form particles observable by DLS (Table 5).

Four of the 50 drugs initially tested—clotrimazole, benzyl benzoate, nicardipine, and delavirdine—were unambiguously promiscuous, aggregate-forming inhibitors at concentrations in the 100 μM range (Figure 2). All four compounds inhibited the three model enzymes (Table 2). All four had a pronounced incubation effect and were highly sensitive to a 10-fold increase in enzyme concentration (Table 3). For all four drugs, saponin greatly reduced their potency against the three model enzymes (Table 4). Finally, DLS experiments on each of these drugs yielded high intensity, well-defined autocorrelation functions, consistent with the presence of particles 200−900 nm in diameter (Table 5, Figure 2). These features contrast markedly with negative control compounds, such as benzo[*b*]thiophene-2-boronic acid (BZBTH2B), a well-characterized, specific inhibitor of β-lactamase that is unaffected by incubation, the presence of saponin, or increases in enzyme concentration (Table 5). BZBTH2B, as well as two other control

Table 3. Effect of Incubation or Enzyme Concentration on Inhibition

compound	↓ IC_{50} with incubation			↑ IC_{50} vs 10×
	β-lactamase	chymotrypsin	MDH	β-lactamase
BZBTH2B[a]	none	N.A.	N.A.	none
clotrimazole	6-fold	increase	2-fold	24-fold
benzyl benzoate	>50-fold	3-fold	24-fold	>50-fold
nicardipine	>50-fold	3-fold	no change	>50-fold
delavirdine	19-fold	3-fold	>50-fold	3-fold
glyburide	5-fold	6-fold	3-fold	5-fold
mefanamic acid	3-fold	2-fold	3-fold	>50-fold
oxaprozin	N. A.	increase	>50-fold	>50-fold[b]

[a] A specific, competitive, and reversible inhibitor of AmpC β-lactamase.[32] [b] Assay performed with chymotrypsin.

Table 4. Effect of Saponin on Inhibition of β-Lactamase, Chymotrypsin, and MDH

compound	↑ IC_{50} with saponin		
	β-lactamase	chymotrypsin	MDH
BZBTH2B[a]	none	N.A.	N.A.
clotrimazole	>50-fold[b]	>50-fold[b]	>50-fold[c]
benzyl benzoate	>50-fold[c]	19-fold[d]	7-fold[c]
nicardipine	>50-fold[b]	2-fold[d]	49-fold[c]
delavirdine	24-fold[c]	3-fold[c]	>50-fold[c]
glyburide	49-fold[c]	8-fold[c]	24-fold[c]
mefanamic acid	3-fold[c]	2-fold[b]	increase
oxaprozin	N.A.	2-fold[b]	>50-fold[c]

[a] A specific, competitive, and reversible inhibitor of AmpC β-lactamase.[32] [b] 0.1 mg/mL saponin. [c] 1.0 mg/mL saponin. [d] 0.2 mg/mL saponin.

Table 5. Dynamic Light Scattering of Several Promiscuous Drugs[a]

compound	IC_{50} vs β-lactamase (μM)	DLS conc (μM)	intensity (kcps)	diameter (nm)
50 mM KPi			0.1 ± 0.01	no particles
BZBTH2B[b]	0.1	100	0.9 ± 0.2	no particles
ANS[c]	>1600	1000	0.5 ± 0.1	no particles
clotrimazole	20	50	35.9 ± 12.6	323.2 ± 31.8
benzyl benzoate	90	250	30.5 ± 2.3	893.1 ± 56.4
nicardipine	20	60	23.8 ± 3.3	514.7 ± 67.8
delavirdine	90	100[d]	43.3 ± 2.3	207.2 ± 15.2
glyburide	300	400	24.4 ± 4.2	302.1 ± 9.9
mefanamic acid	350	200[d]	29.9 ± 7.2	>1000
oxaprozin	>400	1000[d]	6.7 ± 1.6	>1000

[a] DLS performed in 50 mM KPi at the concentration given under "DLS conc"; kcps, kilocounts per second. [b] A specific, competitive, and reversible inhibitor of AmpC β-lactamase.[32] [c] ANS is a dye that is known not to aggregate.[43] [d] Assay in 5 mM KPi.

compounds, the dye 8-anilino-1-naphthyl-sulfonic acid (ANS) and the drug fluconazole (see below) did not scatter light even at millimolar concentrations (Table 6, Figure 2).

To investigate possible structure−activity relationships (SAR) within a family of promiscuous drugs, we considered the behavior of five azole antifungal drugs related to clotrimazole: miconazole, sulconazole, econazole, ketoconazole, and fluconazole (Figure 3). Like clotrimazole, three of these—miconazole, sulconazole, and econazole—were strongly promiscuous, aggregate-forming molecules (Table 6), with IC_{50} values against β-lactamase, chymotrypsin, and MDH in the 10 to 150 μM range. All three showed a strong incubation effect and were highly sensitive to enzyme concentration and the presence of saponin (Table 6), and all three strongly scattered light (Figure 2, Table 6). Conversely, the related antifungals fluconazole and ketoconazole (Figure 3) were not promiscuous (Table 6) and did not form detectable aggregates in solution, even at millimolar concentrations (Table 6, Figure 2). Whereas the number of compounds considered here is small by SAR standards, making conclusions tentative, the principle distinguishing feature among these drugs appears to be their hydrophobicity, with the aggregating promiscuous azoles having clogP values above 5 and the nonpromiscuous azoles having clogP values well below 5 (below).

To investigate possible patterns among aggregating, promiscuous inhibitors more broadly, we looked for patterns among 47 molecules that have been shown explicitly to form promiscuous aggregates in solution (this work and also refs 1 and 10). We were especially interested in features that would distinguish them from 64 molecules that have been shown explicitly not to aggregate at micromolar concentrations (this work and also refs 1, 10, and 11). First, we performed clustering by structural similarity.[12] Though there were a few small clusters, we did not find this particularly informative or predictive.

Because aggregation may be considered a solubility phenomenon, we asked if simple solubility parameters would be sufficient to differentiate aggregate-formers from nonaggregate-formers. In general, aggregators typically have higher clogP values and lower solubility than nonaggregating compounds. With a Gao[13] solubility cutoff of 22.065 μM, 97 of the 111 compounds (87.4%) are correctly classified. Correspondingly, a clogP value of 3.633 successfully classifies 90 of 111 compounds (81.1%). These results suggest that consideration of either solubility or clogP alone may provide useful differentiation between these classes of compounds, but neither of these criteria, by themselves, is sufficient to robustly distinguish aggregators from nonaggregators.

To develop a more precise model, we explored more complex patterns with recursive partitioning (RP) analysis of molecular descriptors.[14] As used in chemical informatics, RP explores the relationship between a dependent variable, such as a biological activity or a physical property (here aggregation) with various physicochemical descriptors. There are many variants of RP analysis, but the one we used here required a categorical (e.g., "aggregator," "nonaggregator") dependent variable. Descriptors and thresholds are sought to optimally partition the data into two sets (nodes), at least one of which is enriched in a particular category. Once a suitable split is found, the two resulting nodes are then themselves split, and this continues recursively until the nodes are very small. This decision tree is ultimately pruned back so that the final, unsplit nodes (terminal nodes) are of some reasonable size as controlled by the user. Simple measures such as "percent of categories predicted correctly" are used to assess the validity of the tree. The goal of this procedure was to select a small number of properties to account for aggregation behavior. For the 111 compounds, we considered a total of 260 physicochemical properties (see Methods), covering a wide range of descriptor types. We included a descriptor of our own design to reflect conjugation within a molecule because we had noticed that many of the

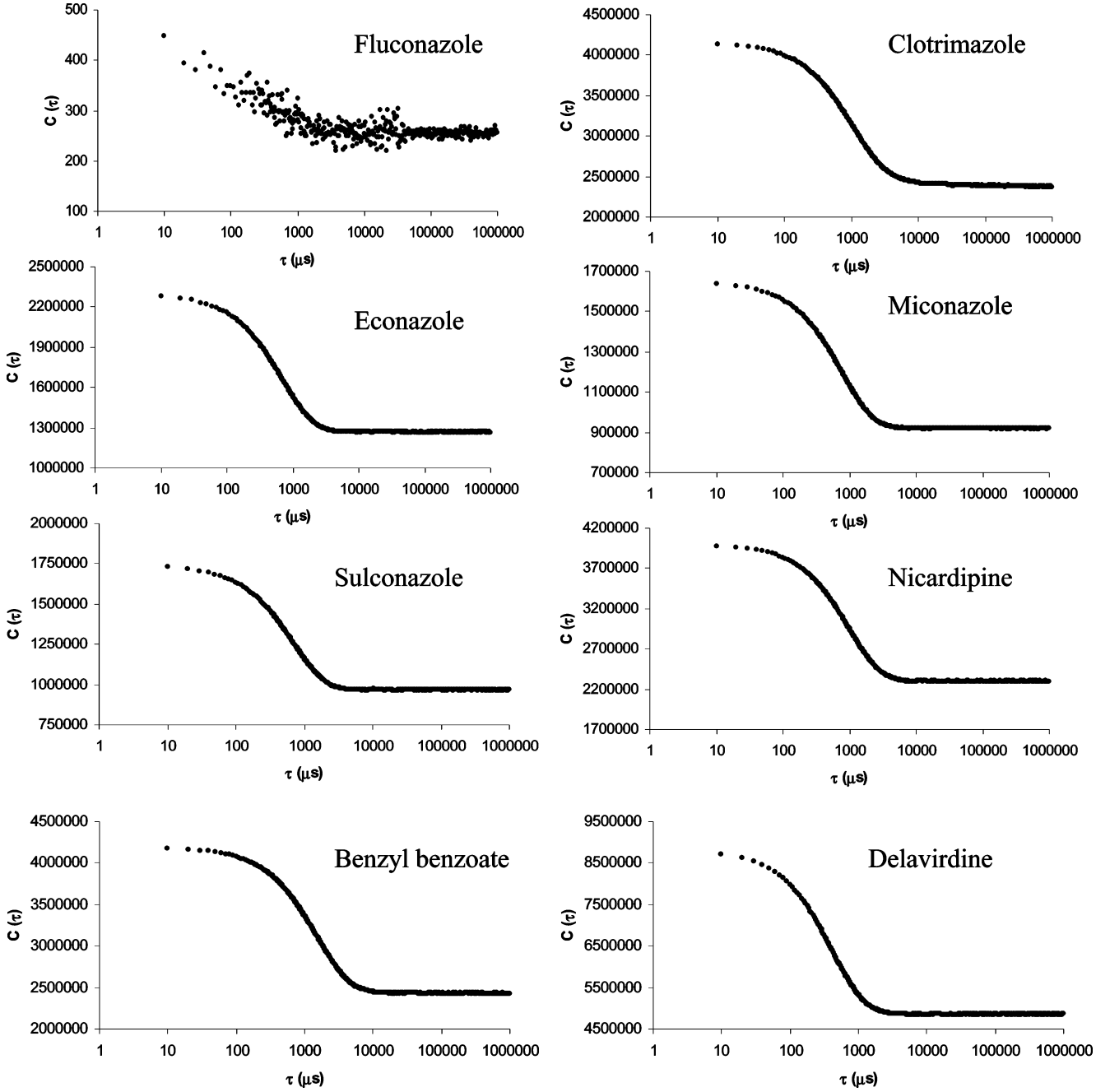

Figure 2. Autocorrelation functions from DLS of strong aggregate formers and fluconazole, a negative control.

Table 6. Inhibition and Aggregate Formation by Azole Antifungals

	clotrimazole	miconazole	sulconazole	econazole	fluconazole	ketoconazole
therapeutic concentration (μM)	30[a 20]	40[a 28]	<80[a 28]	26[a 29]	22[b]	6[b]
IC_{50} (μM) vs β-lactamase	20	40	14	25	>400	>150[c]
IC_{50} (μM) vs chymotrypsin	85	125	110	150	>400	>150[c]
IC_{50} (μM) vs MDH	35	25	20	25	>400	>150[c]
↓ IC_{50} with incubation[d]	6-fold	2-fold	13-fold	19-fold	none	none
↑ IC_{50} vs 10× β-lactamase	24-fold	5-fold	> 50-fold	24-fold	none	none
↑ IC_{50} with saponin[d]	>50-fold[e]	>50-fold[e]	16-fold[f]	>50-fold[f]	none	none
dynamic light scattering data						
DLS concentration (μM)	50	20	20	35	1000	150
intensity (kcps)	35.9 ± 12.6	18.7 ± 2.6	20.0 ± 2.4	21.9 ± 1.9	0.3 ± 0.0	0.3 ± 0.1
diameter (nm)	323.2 ± 31.8	390.4 ± 30.4	331.6 ± 15.2	284.4 ± 16.1	no particles	no particles

[a] Fungicidal concentration. [b] Mean maximum plasma concentration. [c] Ketoconazole is insoluble at higher concentrations. [d] Assay conducted with β-lactamase. [e] 0.1 mg/mL saponin. [f] 1.0 mg/mL saponin.

aggregate-formers were extensively conjugated (max_conj_path, see Methods).

We then used RP analysis, as implemented in the Accelrys' Cerius-2 modeling package[15] to identify a few

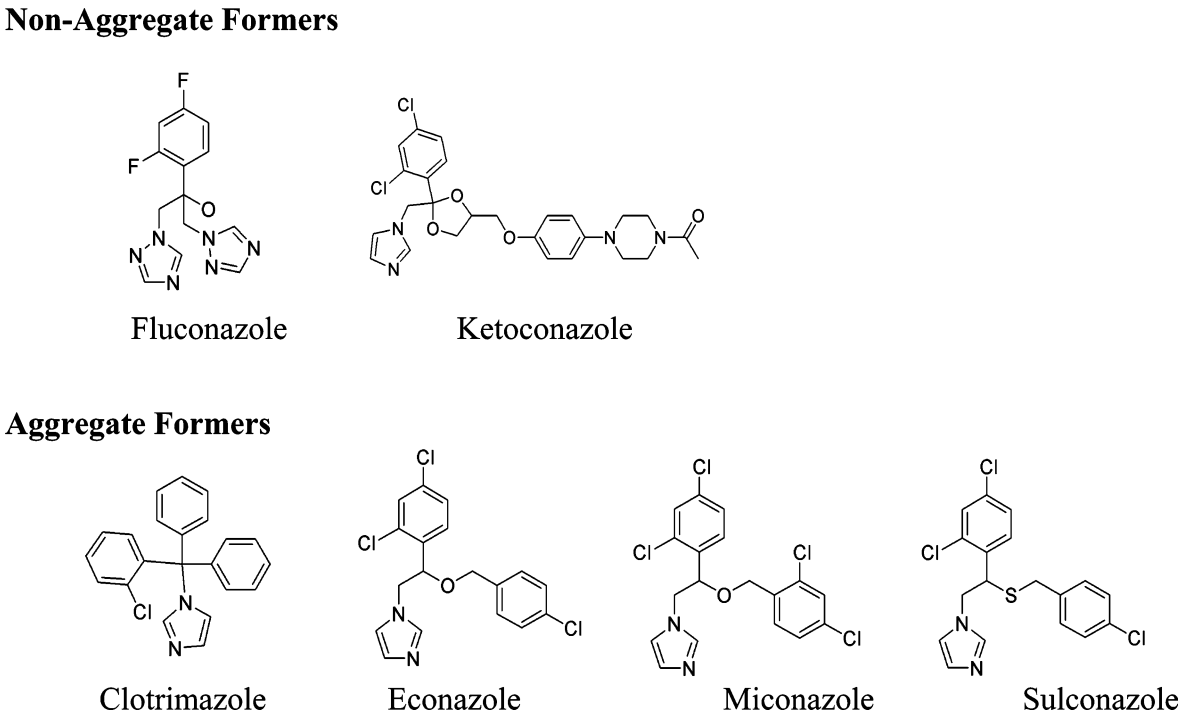

Figure 3. Comparison of the structures of fluconazole and ketoconazole, nonaggregate formers, with the aggregate formers clotrimazole, econazole, miconazole, and sulconazole.

descriptors that correlate with aggregation and to build a predictive model. Over the course of several dozen RP experiments, we systematically adjusted the RP parameters and assessed which descriptors were used most often; we routinely generated RP models that predicted 90−97 of the 111 compounds correctly. Regardless of our changes, we found that the first split was always based on the Gao aqueous solubility.[13] When we removed the Gao solubility as an independent variable, we found that either another calculated solubility measure, the "Lee solubility," or the Daylight clogP[16] were used for the first split. Interestingly, the ACD/Labs[17] "sol. in pure water (mol/L)" or QikProp[18] "aq solubility in micrograms/mL" were never used for any split in any RP experiment we performed. Since both solubility models with which we built reasonable RP models are not commercially available, and the two commercial methods we tried were not useful in our studies, we excluded our two solubility models to build a model that could be more easily implemented by others. In doing so, we found that excluding solubility actually allowed us to build RP models with a higher prediction success rate by using clogP as the first split, even though a model with just a single split on clogP performs less well than the Gao solubility alone.

The model that emerged from the RP analysis correctly classifies the aggregation behavior of 104 of 111 compounds (93.7%) (Figure 4). The descriptors used were the Daylight clogP; an electropological state index[19] denoted as "S_sssN," which corresponds to nitrogen single bonded to three heavy atoms; our max_conj_path descriptor; and the presence or absence of carboxylic acid. From our model, we identify six RP nodes, three of which consist predominantly or completely of aggregators (nodes 2, 4, and 5), and three of which are primarily or completely nonaggregators (nodes 1, 3, and 6). Of the three nodes in which aggregators predominate, node 2 consists of higher clogP compounds

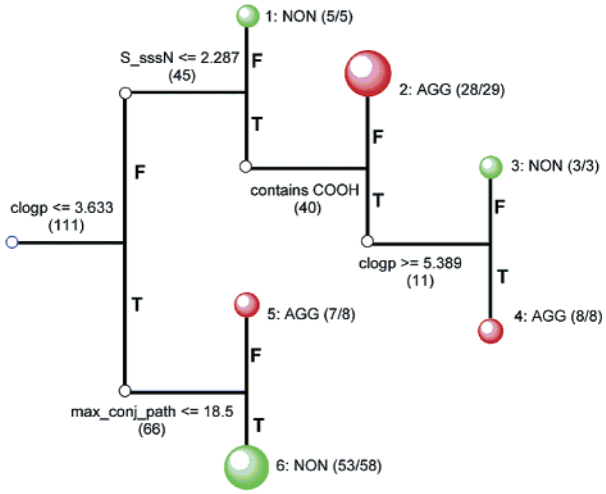

Figure 4. Recursive partitioning (RP) analysis of 111 aggregators and nonaggregators. Each branch contains the physicochemical criterion used to split a group of compounds; T indicates compounds that satisfy this criterion, and F indicates compounds that do not. Terminal nodes are green and coded as "NON" if they consist predominantly or completely of nonaggregators, red and coded as "AGG" if they are predominantly or completely aggregators. Nodes with more compounds are identified by larger circles.

without ionizable tertiary nitrogens or carboxylic acids, node 4 consists of carboxylic acids with very high clogP values, and node 5 consists of lower clogP compounds with extensive conjugation.

Seven of 111 compounds are classified incorrectly by this model including (Figure 5): compound **1**[1] (17-(2-amino-ethylamino)-3-methoxy-10,13-dimethyl-hexadecahydro-cyclopenta(a)henanthren-11-ol), compound **2**[1] (3-(4-(dimethylamino)-benzylidene)-1,3-dihydro-indol-2-one), delavirdine, quercetin, and U0126, which have been experimentally shown to form aggregates (this work and ref 10), and protriptyline and SB203580, both

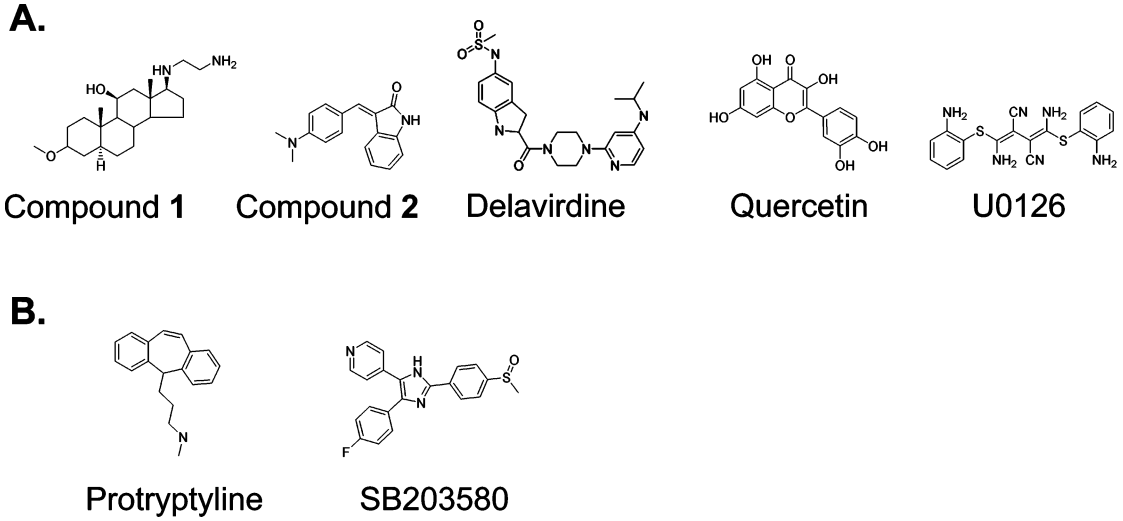

Figure 5. Structures of compounds that were incorrectly classified by the RP model. (A) Compounds that were incorrectly predicted as nonaggregate formers. (B) Compounds that were incorrectly predicted as aggregate formers.

Table 7. 75 Drugs in Negative Control Set with RP Predictions

drug	predicted node	predicted behavior
hydroxyzine, meclizine, nefazodone, trazodone	1	non-aggregator
celecoxib, desogestrel, estradiol, ethinylestradiol, felodipine, fenofibrate, glimepiride, glyburide, irbesartan, mestranol, nortriptyline, sertraline, simvastatin	2	aggregator
ibuprofen, valsartan	3	non-aggregator
montelukast	4	aggregator
alprazolam, clonazepam	5	aggregator
acetaminophen, acyclovir, allopurinol, amoxicillin, atenolol, benzonatate, bupropion, captopril, cefprozil, cephalexin, cetirizine, cimetidine, ciprofloxacin, clindamycin, clonidine, diazepam, digoxin, divalproexsodium, doxycycline, folic acid, gabapentin, gatifloxacin, glipizide, hydrochlorothiazide, isosorbidemononitrate, levofloxacin, levothyroxine, lisinopril, lorazepam, metaxalone, methylprednisolone, metoclopramide, metoprololsuccinate, metoprololtartrate, metronidazole, mirtazapine, nabumetone, naproxen, nifedipine, olanzapine, penicillinvk, phenytoin, prednisone, rabeprazole, ramipril, risperidone, rofecoxib, spironolactone, sulfamethoxazole, temazepam, tetracycline, triamcinolone, warfarin	6	non-aggregator

of which do not aggregate (this work and ref 10). Our RP model predicts SB203580 to be a node 5 aggregator. However, we would not expect a conformation with the coplanar ortho phenyl rings to be of sufficiently low energy to actually allow them to be in conjugation, so that the max_conj_path of 23 we assign to it is excessive; this compound may actually belong in node 6. Conversely, compound **2** and quercetin we predict to be node 6 nonaggregators. We note that the max_conj_path for each compound is 17, which is very near the decision point of 18.5, above which compounds are assigned as node 5 aggregators. We can currently offer no rationalization for the other incorrectly classified compounds—compound **1** (node 6), delavirdine (node 6), U0126 (node 6), and protriptyline (node 2).

Since we did not have a readily available set of untested compounds on which to gauge the predictiveness of our RP model, we decided to construct a "negative control" set of known drugs. We term this a negative control set because we assume that the number of aggregators in such a set will be relatively small (Figure 1). We chose 75 frequently prescribed, orally available drugs[20,21] that were not included among the 111 used in constructing our RP model. The RP model predicts that 16 of these 75 drugs (21%) are aggregate-formers (Table 7). Coincidentally, we included the drug Glyburide in the negative control set and found that our model predicts it to be a node 2 aggregator. Though it was not included in the set of 111 compounds used to build the model, we have found that it behaves as a weak promiscuous aggregator (Tables 2−5). This encouraged us to use our RP to predict the behavior of two other drugs that also fall in this borderline region, mefenamic acid and oxaprozin. Our model predicts both oxaprozin and mefenamic acid to be node 3 nonaggregators, though the latter is very close to the split between nodes 3 and 4.

Discussion

Of the 50 diverse drugs initially tested, 43 (86%) were nonpromiscuous, nonaggregate forming compounds, consistent with the expectation that drugs have been optimized to be well-behaved, highly specific molecules. Surprisingly, four drugs (8%) showed strongly promiscuous, aggregation-based inhibition of the three model enzymes. How can this non-"drug-like" behavior be reconciled with the fact that these four molecules are, in fact, drugs? If drugs can behave promiscuously, what conclusions can we draw about promiscuous screening hits? What, if anything, can be concluded about the sort of molecules that are likely to behave in this manner?

In reconciling the promiscuous activity of these molecules with their status as drugs, two aspects of their behavior bear consideration. First, the drugs are

being tested in a simple buffer that is very different from the in vivo environment. The protein-rich environment of the latter may well obviate the aggregation of these compounds. For topical drugs such as clotrimazole and benzyl benzoate, their formulation as creams may serve the same role. Second, the drugs were tested at concentrations that, certainly for systemic drugs such as delavirdine, are orders of magnitude higher than their IC_{50} values against their canonical targets. Although a role for aggregation in the physiological actions and side reactions of these drugs cannot be completely discounted at this time, the easier interpretation of these results is that, under screening conditions, even some drugs can act promiscuously, through aggregation.

Aggregation-based inhibition may be a common feature among false positive screening hits, but if drugs can also behave this way, how useful is this property for separating true from false positives in screening? We contend that a molecule that is active in a screen because it is an aggregator is unlikely to be an interesting hit. The observation that some drugs can aggregate and act nonspecifically does not imply there is reason to believe that a new hit, active through aggregation, may be a specific ligand for the target of that screen. Conversely, aggregate-mediated inhibition at micromolar concentrations does not preclude specific, monomeric inhibition at nanomolar concentrations (e.g., nicardipine and delavirdine).

The promiscuous behavior of these drugs, though admittedly a small percentage of the total number of drugs tested, does raise the question of whether "drug-like" has any strong physicochemical or chemical informatics meaning. To explore this question, we analyzed the structures of all molecules that we have investigated for aggregation-based promiscuity. At least 47 compounds have been explicitly shown to be aggregation-based, promiscuous inhibitors (this paper and refs 1 and 10) and another 64 compounds have been explicitly shown to be nonaggregators at micromolar concentrations (mostly this paper but also refs 1, 10, and 11). Using chemical similarity, we were unable to reliably separate compounds that inhibit promiscuously through aggregation from those that do not (data not shown). Physical chemical criteria were more dependable; recursive partitioning analysis based on these criteria produced a model that classified compounds according to descriptors including clogP, the presence or absence of ionizable groups, and the extent of conjugation (Figure 4). This model correctly classifies 104 of the 111 total compounds, and can distinguish the four promiscuous azole antifungals, miconazole, sulconazole, econazole, and clotrimazole (node 2), from the nonpromiscuous azoles fluconazole (node 6) and ketoconazole (node 1).

There is always the danger in classification or QSAR of overfitting. The RP model has the advantage of simplicity, only five splits are made on four descriptors (Figure 4). Certain aspects of the model are of concern. Some of the nodes are small; node 1 contains five compounds, while node 3 has only three. Node 2 contains 29 compounds, almost all aggregators, yet the physicochemical characteristics of the node (high clogP without tertiary nitrogens or carboxylic acids) could be characteristic of many drug-like compounds. For this reason, we decided to choose another set of known drugs as a "negative control", a set of compounds within which we expected few aggregate-formers (Table 7).

The model predicts that 16 of the 75 negative control drugs would aggregate (Table 7). One of these, glyburide, is a weakly promiscuous, aggregating compound that we tested earlier (Table 2) but did not include among the 111 compounds used for building the RP model. The untested 15 drugs in the negative control set that the model classifies as aggregators constitute testable predictions. Of these, 13 are strong predictions; alprazolam and clonazepam, predicted to fall in node 5, are benzodiazepines, a structure type in which the three-ring fused system has been shown to be significantly noncoplanar.[22] This suggests that max_conj_path values for these compounds (19 and 20, respectively), which are already barely above the split value, are excessive. Among the more interesting predictions are the four steroids predicted to aggregate and the five steroids predicted not to aggregate; our training set of 111 includes only three steroids, all aggregators.[1] We have provided the full list of 111 molecules, and their categorization as aggregators or nonaggregators, in SMILE string format downloadable from the ACS website (Supporting Information, table S1).

In beginning this investigation, it was our hope that drugs would not show aggregation-based promiscuity, allowing us to draw a clear, physical distinction between "drug-like" molecules, which do not aggregate at micromolar concentrations, and non-"drug-like" molecules, which do. Indeed, this was the case for most of the drugs tested here, but not all. Aggregate formation has now been shown to occur among screening hits,[1] among true lead compounds,[10] and now drugs. The fact that aggregate-based inhibition occurs even in drugs, which have presumably been optimized for specificity, suggests that aggregate formation may be common at micromolar levels. It may be interesting to consider larger studies to investigate just how common such molecules are, especially among pharmaceutically and biologically relevant compound libraries.

Experimental Section

Materials. AmpC β-lactamase was purified as described.[23] Loracarbef was a gift from Larry Blaszczak at Eli Lilly (Indianapolis, IN). Lamotrigine was purchased from Kemprotec (Middlebrough, UK), and tacrine from Butt Park (Bath, UK). Ondansetron HCl and torasemide were purchased from Sequoia (Oxford, UK), oxalacetic acid from Fluka (St. Louis, MO), saponin from Calbiochem (La Jolla, CA), nitrocefin from Oxoid, Ltd (Basingstoke, UK), riluzole from Matrix (Columbia, SC), oxaprozin from Maybridge (Cornwall, UK), mefenamic acid from Lancaster (Boston, MA), indomethacin from Cayman (Ann Arbor, MI), gemfibrozil from Spectrum (Gardena, CA), and benzyl benzoate from Chemserve-AS (West Chester, PA). Delavirdine, ketconazole, and glyburide were purchased from Biomol (Plymouth Meeting, PA), and flutamide from LKT-Labs (St. Paul, MN). Azelaic acid and triamterene were purchased from Alfa (Ward Hill, MA). Tetraethylene pentamine, 4-aminophenyl sulfone, and 5-aminosalicylic acid were purchased from Acros (Pittsburgh, PA). Carbamazepine, carisoprodol, chlorambucil, clotrimazole, diflunisal, diphenhydramine, fluconazole, lidocaine, miconazole, nicardipine, primidone, propantheline bromide, riboflavin, sulfinpyrazone, and trimethoprim were purchased from IGN (Pittsburgh, PA). α-Chymotrypsin, *N*-succinyl-Ala-Ala-Pro-Phe *p*-nitroanilide, malate dehydrogenase, β-nicotinamide adenine dinucleotide, chlorthalidone, cinoxacin, deferoxamine mesylate, econazole, etodolac,

furosemide, guaifenesin, guanabenz, amrinone, lansoprazole, leflunomide, mebendazole, nitrofurantoin, omeprazole, phenazopyridine, prazocin, protriptyline, sulfadiazine, sulconazole, thalidomide, and tripelennamine were purchased from Sigma-Aldrich (St. Louis, MO). All materials were used as supplied by the manufacturer, without further purification.

Enzyme Assays. Compounds were tested for inhibition of β-lactamase, chymotrypsin, and malate dehydrogenase. Unless otherwise stated, assays were performed in 50 mM potassium phosphate (KPi) buffer, pH 7.0, at room temperature. Stocks of substrates and inhibitors were typically prepared at 10 mM in dimethyl sulfoxide (DMSO), with the following exceptions: chlorambucil, glyburide, indomethacin, mefenamic acid, oxaprozin, and sulfinpyrazone were prepared at 50 mM in DMSO, whereas 5-aminosalicylic acid, deferoxamine, diphenhydramine, flutamide, guaifenesin, guanabenz, lidocaine, propantheline bromide, protriptyline, tacrine, and tripelennamine were prepared at 10 mM in 50 mM KPi buffer. No more than 5% DMSO was present in any assay, and results were controlled for the presence of DMSO. All reactions were monitored on a HP8453 spectrophotometer.

For most β-lactamase assays, inhibitor and 1 nM enzyme were incubated for 5 min, and the reaction was initiated with 200 μM nitrocefin. For β-lactamase assays without incubation, inhibitor and 200 μM nitrocefin were mixed, and the reaction was initiated with 1 nM enzyme. For all assays with a 10-fold increase in β-lactamase, inhibitor and 10 nM enzyme were incubated for 5 min, and the reaction was initiated with 100 μM loracarbef. Loracarbef was used because it was a slower substrate for the enzyme and allowed for the measurement of reaction rate over a 5-min interval, even with a 10-fold increase in enzyme concentration. Hydrolysis was monitored at 260 nm for loracarbef and at 482 nm for nitrocefin.

For chymotrypsin assays, inhibitor and 28 nM enzyme were incubated for 5 min and the reaction was initiated with 200 μM *N*-succinyl-Ala-Ala-Pro-Phe *p*-nitroanilide. For chymotrypsin assays without incubation, inhibitor and 200 μM *N*-succinyl-Ala-Ala-Pro-Phe *p*-nitroanilide were mixed, and the reaction was initiated with 28 nM enzyme. Reaction progress was monitored at 410 nm. For malate dehydrogenase assays, inhibitor and 2 nM enzyme were incubated for 5 min, and the reaction was initiated with 200 μM oxalacetic acid and 200 μM β-NADH. For malate dehydrogenase assays without incubation, inhibitor, 200 μM oxalacetic acid, and 200 μM β-nicotinamide adenine dinucleotide were mixed, and the reaction was initiated with 2 nM enzyme. Reaction progress was monitored at 340 nm. Several drugs at 400 μM absorbed light at 340 nm, and it was necessary to test them at lower concentrations.

When used, saponin was present at 0.1, 0.2, or 1.0 mg/mL, and was added before the 5 min enzyme−inhibitor incubation period. Results were controlled for the presence of saponin.

Dynamic Light Scattering (DLS). Drugs were dissolved to 10 or 50 mM in DMSO or 50 mM KPi buffer, and diluted with filtered 5 or 50 mM KPi. All compounds were analyzed with a 3 W argon-ion laser at 514.4 nm with a BI-9000 and BI-200 optical systems from Brookhaven Instrument Corporation. The laser power and integration times were comparable for all experiments on this instrument. Calculation of mean particle diameter was performed by the cumulant analysis tool of a 400-channel BI9000AT digital autocorrelator, with the last eight channels used for baseline calculation. The detector angle was 90°. Each diameter and intensity value represents four or more independent measurements at room temperature. For several drugs, light scattering data was also measured on a DynaPro MS/X instrument, which gave qualitatively similar results, although, as expected, the intensities of the scattered light differed between the two instruments.

Recursive Partitioning (RP). From the list of compounds studied explicity for aggregation behavior, we excluded one aggregator, tris(dicarboxymethylene)fullerene-C3[1] and three nonaggregators, deferoxamine, BZBTH2B,[1] and suramin.[10] These four compounds were excluded from our analysis because of difficulties in generating three-dimensional structures or physicochemical descriptors. We were left with a set of 111 compounds for our study.

A total of 260 descriptors were used in the statistical analyses. These are listed here by category, with the total number of descriptors for each category following in parentheses. Accelrys' Cerius-2 descriptors:[15] Jurs surface area descriptors (29), electrotopological state indices (34), Ghose-Crippen AlogP atom types (78), shadow indices (10), subgraph counts (5), Kier & Hall chi connectivity indices (10), InfoContent (5), kappa shape indices (6), HOMO/LUMO energies (4), principal moments of inertia (mag, X, Y, Z) (4), miscellaneous descriptors (25): charge, apol, dipole-mag, V-ADJ-mag, V-DIST-mag, E-ADJ-mag, E-DIST-mag, RadOfGyration, area, density, Hbond acceptor, Hbond donor, JX, Fh2o, foct, MolRef, Hf, MW, AlogP98, PHI, Wiener, Zagreb, Sr, Vm, HF_MOPAC. ACD/Labs descriptors[17] (7):

pK_a, logP, sol. in pure water (mol/L), sum of N+O, sum of NH+OH, total score, rotatable bonds. Daylight clogP;[16] QikProp descriptors[18] (37); Miscellaneous descriptors (4): Gao aqueous solubility,[13] Lee solubility (Lee, P., Pharmacia Corp., unpublished), Barone − Chanon complexity ("bcComplexity"),[24] bcComplexity/number of heavy atoms ("bcAComplexity"), topological polar surface area,[25] max_conj_path.

The max_conj_path variable was calculated as follows: By inspecting the graph of each molecule, we counted the maximum number of heavy atoms that can be in conjugation with each other, and we refer to this as "max_conj_path." Supporting Information Figure S1 shows examples of simple compounds (not part of our study) for illustration of our procedure. Compound "a" has a max_conj_path value of 6 because the geometry of the sulfonamide group prevents conjugation beyond that of the isolated phenyl rings. By contrast, compounds b−f all have rings or functional groups which can assume planar conformations and thus conjugate more extensively. We also consider compounds such as benzophenone (compound "d") to conjugate extensively, even though it has contiguous single bonds. Note, however, that we do not do any conformational analysis when assigning max_conj_path, and there may be examples (such as with benzodiazepines, discussed above) where conjugation apparently allowed from the chemical graph is probably not feasible.

RP analysis was performed with the Accelrys' Cerius-2 modeling package.[15]

Acknowledgment. This work was supported by GM59759 from the NIH and by Pfizer (to B.K.S.). S.L.M. is partially supported by T32-GM08061 (K. Mayo, PI), the PhRMA Foundation, and is a Northwestern University Presidential Scholar. We thank the Keck Biophysics Facility at Northwestern University for use of their facilities and DynaPro for a partial funding of a MS/X DLS instrument. We thank B. Feng, A. Graves, and J. Horn for reading this manuscript.

Supporting Information Available: Calculated max_conj_path values for six representative compounds (Figure S1) and table of aggregator and nonaggregator compounds (Table S1) are available free of charge via the Internet at http://pubs.acs.org.

References

(1) McGovern, S. L.; Caselli, E.; Grigorieff, N.; Shoichet, B. K. A common mechanism underlying promiscuous inhibitors from virtual and high-throughput screening. *J. Med. Chem.* **2002**, *45*, 1712−1722.

(2) Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E. M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjogren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von Der Saal, W.; Zimmermann, G.; Schneider, G. Development of a Virtual Screening Method for Identification of "Frequent Hitters" in Compound Libraries. *J. Med. Chem.* **2002**, *45*, 137−142.

(3) 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. Deliv. Rev.* **1997**, *23*, 3−25.

(4) Muegge, I.; Heald, S. L.; Brittelli, D. Simple selection criteria for drug-like chemical matter. *J. Med. Chem.* **2001**, *44*, 1841−1846.
(5) Frimurer, T. M.; Bywater, R.; Naerum, L.; Lauritsen, L. N.; Brunak, S. Improving the odds in discriminating "drug-like" from "non drug-like" compounds. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1315−1324.
(6) Xu, J.; Stevenson, J. Drug-like index: a new approach to measure drug-like compounds and their diversity. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1177−1187.
(7) Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 251−264.
(8) Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discovery* **2002**, *1*, 882−894.
(9) Rishton, G. M. Reactive compounds and in vitro false positives in HTS. *Drug Discovery Today* **1997**, *2*, 382−384.
(10) McGovern, S. L.; Shoichet, B. K. Kinase Inhibitors: Not Just for Kinases Anymore. *J. Med. Chem.* **2003**, *46*, 1478−1483.
(11) Powers, R. A.; Morandi, F.; Shoichet, B. K. Structure-based discovery of a novel, noncovalent inhibitor of AmpC beta-lactamase. *Structure (Camb)* **2002**, *10*, 1013−1023.
(12) Doman, T. N.; Cibulskis, J. M.; Cibulskis, M. J.; McCray, P. D.; Spangler, D. P. Algorithm5: A Technique for Fuzzy Similarity Clustering of Chemical Inventories. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1195−1204.
(13) Gao, H.; Shanmugasundaram, V.; Lee, P. Estimation of Aqueous Solubility of Organic Compounds with QSPR Approach. *Pharm. Res.* **2002**, *19*, 497−503.
(14) Hawkins, D. M.; Young, S. S.; Rusinko, A. I. Analysis of Large Structure−Activity Data Set Using Recursive Partitioning. *Quant. Struct.-Act. Relat.* **1997**, *16*, 296−302.
(15) *Cerius-2*; 4.2 ed.; Accelrys, Inc., San Diego.
(16) *Daylight Toolkit*; Daylight Chemical Information Systems, Mission Viejo.
(17) *ACD/Labs*; Advanced Chemistry Development Inc., Toronto.
(18) *QikProp*; 2.0 ed.; Schrödinger, Inc., New York City.
(19) Hall, L. H.; Kier, L. B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1039−1045.
(20) http://mosbysdrugconsult.com/DrugConsult/Top_200/; Mosby's Drug Consult.
(21) http://www.fda.gov/cder/orange/obreadme.htm; FDA Orange Book.
(22) Camerman, A.; Camerman, N. Stereochemical Basis of Anticonvulsant Drug Action. II. Molecular Structure of Diazepam. *J. Am. Chem. Soc.* **1972**, *94*, 268−272.
(23) Weston, G. S.; Blazquez, J.; Baquero, F.; Shoichet, B. K. Structure-based enhancement of boronic acid-based inhibitors of AmpC beta-lactamase. *J. Med. Chem.* **1998**, *41*, 4577−4586.
(24) Barone, R.; Chanon, M. A New and Simple Approach to Chemical Complexity. Application to the Synthesis of Natural Products. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 269−272.
(25) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43*, 3714−3717.
(26) Molinoff, P.; Ruddon, R. *Goodman's and Gilman's The Pharmacological Basis of Therapeutics*, 9th ed.; Hardman, J., Limbird, L., Eds.; McGraw-Hill: 1996.
(27) Sawyer, P.; Brogden, R.; Pinder, R.; Speight, T.; Avery, G. Clotrimazole: A Review of its Antifungal Activity and Therapeutic Efficacy. *Drugs* **1975**, *9*, 424−447.
(28) Walton, S.; Myerscough, M.; Currie, B. Studies in vitro on the relative efficacy of current acaricides for *Sarcoptes scabiei* var. *hominis*. *Trans. R. Soc. Trop. Med. Hyg.* **2000**, *94*, 92−96.
(29) Roche Laboratories Cardene (nicardipine hydrochloride) capsules: Complete Product Information; Roche, 1999.
(30) Barry, M.; Mulcahy, F.; Merry, C.; Gibbons, S.; Back, D. Pharmacokinetics and potential interactions amongst antiretroviral agents used to treat patients with HIV infection. *Clin. Pharm.* **1999**, *36*, 289−304.
(31) PDR Electronic Library; Thomson MICROMEDEX, 2002.
(32) Pharmacia Glynase PresTab: Micronized Glyburide Tablets; Pharmacia, 2002.
(33) Chiang, S.; Knowles, J.; Hubsher, J.; Ruelius, H.; Walker, B. Effects of Food on Oxaprozin Bioavailability. *J. Clin. Pharm.* **1984**, *24*, 381−385.
(34) Stopa, B.; Gorny, M.; Konieczny, L.; Piekarska, B.; Rybarska, J.; Skowronek, M.; Roterman, I. Supramolecular ligands: monomer structure and protein ligation capability. *Biochimie* **1998**, *80*, 963−968.

JM030191R