Assessment of in Vitro and in Vivo Activities in the National Cancer Institute's Anticancer Screen with Respect to Chemical Structure, Target Specificity, and Mechanism of Action

Ruili Huang,[†] Anders Wallqvist,[‡] and David G. Covell*[,†]

Developmental Therapeutics Program, Screening Technologies Branch, Laboratory of Computational Technologies, National Cancer Institute−Frederick, Frederick, Maryland 21702, and Science Applications International Corporation, National Cancer Institute−Frederick, Frederick, Maryland 21702

This paper examines two biological models of anticancer activity, cytotoxicity and hollow fiber (HF) activity, for chemotherapeutic agents evaluated as part of the National Cancer Institute's (NCI's) drug screening effort. Our analysis proposes strategies to globally assess compounds tested in the NCI's 60-cell (NCI_{60}) in vitro anticancer screen in terms of structural features, biological activity, target specificity, and mechanism of action by data integration via our self-organizing maps of structural and biological response patterns. We have built statistical models to predict compound potency and HF activity based on physicochemical properties. Our results find that it is the combination of different structural properties that determines a compound's biological activity. A direct correlation is also found between compound potency and specificity, indicating that specific targeting, rather than promiscuous poisoning, gives rise to potency. Finally, we offer a strategy to exploit this relationship for future mining of novel anticancer candidates.

Introduction

The foundations for our present understanding of cancer began over five decades ago with the observation that tumor-derived cell lines proliferate indefinitely,[1] and since that time this result has served as the basis for establishing numerous in vitro anticancer drug discovery and testing initiatives.[2] Notable among these endeavors is the massive anticancer screening effort of the National Cancer Institute (NCI[a]), launched in 1990, that profiles small molecule compounds for their cytotoxicity and proliferation inhibition on more than 60 cultured tumor cell lines (NCI_{60}), representative of the major histologic types of cancer prevalent in the United States.[3,4] Initially the NCI's tumor screening panel was widely viewed as a black box providing phenotypic readouts limited to growth and viability profiles. Over time, however, this as well as other cell-based screens have established themselves as multifaceted tools offering the potential to assess relevant details about intracellular target specificity, target-related cytotoxicity, metabolic stability, and bioavailability. Building on this history, latest generation screening cell lines have been developed by numerous labs that are transgenic and engineered to study specific targets and/or inhibitors.[5,6]

While scientific and technological advancements have enhanced our understanding of the molecular physiology of cancer, most cancer chemotherapy protocols have been established empirically, without the aid of a modern perspective.[7] A compelling feature of cell-based screening is that tumor cell lines exhibit diverse sensitivities to chemotherapeutic agents, and it is this feature that has been proposed as a basis for rational drug selection and therapy design.[2,8] An equally exciting facet of this observation is that the capacity to associate molecular structures with sensitivity as a means to predict tumor-cell responsiveness is more possible today than ever before.

This paper proposes a novel analysis of chemotherapeutic agents based on an evaluation of cytotoxic screening data and hollow fiber screening data obtained within the NCI's anticancer discovery effort. Our analysis presents a strategy for relating structural features within chemical classes as a basis for cellular responsiveness in the screening and hollow fiber data. Our perspective will be derived by integrating our recent global assessments of the NCI's screening data with evaluations about activity in the NCI's hollow fiber assay. Our primary goal is to utilize the information resources within the NCI, derived over the past 15 years, combined with more recent functional genomic technologies, to facilitate the understanding of chemistries associated with effective agents and use this information to propose the development of investigative strategies to discover new potential oncologic agents.

Since its inception in 1990, ~85 000 compounds have been tested against the NCI_{60} for in vitro anticancer activity.[9,10] These compounds are part of the NCI chemical database, which contains samples of compounds from both organic synthesis and natural product extracts, provided by academic, government, and other nonprofit and industrial laboratories, collected by the NCI since 1955 for testing in anticancer, and more recently anti-AIDS, assays of various types. The NCI database is one of the largest and structurally most diverse databases, containing ~250 000 compounds, >90% of which are unique, having minimal overlap with most commercial databases.[11,12] The NCI_{60} is composed of immortalized cancer cell lines reflecting diverse cell lineages, as they are derived from lung, renal, colorectal, ovary, breast, prostate, central nervous system, melanoma, and hematological malignancies. The 50% cancer cell growth inhibition concentration (GI_{50}) for any particular cell line is an index of cytotoxicity or cytostasis. Within the complete set of

* Corresponding author. Phone: 301-846-5785. Fax: 301-846-6978. E-mail: covell@ncifcrf.gov.
† National Cancer Institute−Frederick.
‡ Science Applications International Corp.

[a] Abbreviations: NCI, National Cancer Institute; NCI_{60}, NCI's 60-cancer cell screen; HF, hollow fiber; MOA, mechanism of action; SOM, self-organizing map; GI_{50}, 50% cancer cell growth inhibition concentration; MW, molecular weight; HBA, hydrogen bond acceptors; HBD, hydrogen bond donors; PAC, parent atom count; RB, rotatable bonds; PSA, polar surface area; PMW, parent molecular weight; LSUFC, Leadscope unique feature count; LSFC, Leadscope feature count; SD, standard deviation; ROC, receiver operating characteristic; DYFP, Daylight fingerprint; CSA, consistent structure−activity; PLS, partial least squares; NIPALS, nonlinear iterative partial least squares; SVD, singular value decomposition; TP, true positive; FP, false positive; TN, true negative; FN, false negative.

screening compounds, GI_{50} measures for over 40 000 of them are publicly available. GI_{50} growth patterns across the 60 tumor cells have been found to be an information-rich resource for establishing a compound's mechanism of action (MOA).[13−19] Moreover, this in vitro cell line model has been shown to be predictive of the phase II clinical trial performance of cancer drugs for certain cancer types.[20]

The NCI also developed the hollow fiber (HF) model in 1995 to be employed as a routine preliminary screening assay for identification of in vivo activity of potential anticancer compounds.[21] The HF assay assesses the pharmacologic capacity of compounds to reach two physiologic compartments, intraperitoneal (ip) and subcutaneous (sc), within the nude mouse and is a practical means of quantifying viable tumor cell mass in an animal model. The HF assay has been shown to have good correlation with in vivo xenograft activity.[22−24] A recent evaluation of the prognostic value of this assay indicates that good HF activity is predictive of good activity in NCI's xenograft models.[22] This connection is particularly relevant for anticancer drug discovery, since the human xenograft model has been found useful in postpredicting the phase II clinical trial performance of certain cancer drugs in non-small-cell lung and ovarian cancers.[20] Here we focus our attention on the HF data, under the assumption that the HF model can serve as a valuable pipeline for selecting compound entry into more costly in vivo models.[22] To date, about 3000 compounds have been screened in the HF assay. A strong correlation between potency in the 60-cell line screen and activity in the HF assay has also been reported.[24] The lines of evidence connecting the initial cell-based screening data to phase II outcome through HF and xenograft follow-up assays serve as the primary motivation for joint reassessment of the tumor screen and HF data sets.

We have previously organized the GI_{50} data into self-organizing maps (SOMs).[17,25,26] SOM clustering of the GI_{50} data segregates compounds into nine major response categories: mitosis (M), membrane function (N), nucleic acid metabolism (S), metabolic stress and cell survival (Q), kinases/phosphatases and oxidative stress (P), and four uncharacterized regions R, F, J, and V.[17−19] In addition, we have clustered the Daylight fingerprints (bit vectors of length 2048) derived from the structures of these same screened compounds using the SOM method to group structurally similar compounds. Here we will develop strategies to combine structural features and physicochemical parameters with GI_{50} and HF activity data of all screened compounds, with the aid of our GI_{50} and structural SOMs, to achieve a global perspective on connections between compound structure, biological activity, target specificity, and MOA. Various quantitative structure−activity relationship (QSAR) studies have been previously carried out on small subsets of potential anticancer agents screened in the NCI_{60}.[27−29] Those efforts were based primarily on the pioneering work by Hansch and Leo as a model for correlating different physicochemical parameters with biological activities.[30−32] Following this earlier work, we propose to extend the analysis to include HF data and GI_{50} data for all screened compounds. In this study, we will evaluate the feasibility of deriving valid statistical models from structural/physicochemical parameters to predict the potency of compounds in the 60-cell screen as well as their in vivo HF activity. Each model will be validated through rigorous statistical procedures. As we are aware of the limitations brought by the large data size and large structural variations therein,[33] we do not expect the models to appear as accurate as what one can obtain from a small, focused set of compounds with only minor structural variations;[34] however, as a proof of

concept, we will show that models derived from large data sets are statistically valid and have excellent predictive power for proposing novel compounds with potential anticancer activity. Furthermore, since these models are built from large data sets, the trends captured are statistically more robust and can be applied to general populations of compounds.

Results

Structural-Activity Assessment of All Screened Compounds. Several common molecular descriptors including ALogP, Lipinski score, molecular weight (MW), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), parent atom count (PAC), rotatable bonds (RB), polar surface area (PSA), and parent molecular weight (PMW) for ∼43 000 publicly accessible compounds that have been screened in the NCI_{60} for GI_{50} were calculated and exported from Leadscope. Each compound is also encoded in a vector (fingerprint) of length 27 000 in Leadscope, where each element represents a structural feature of the compound, which can be a molecular fragment, a functional group, or a certain bond or atom type. If a feature is present in a compound, then the element is a nonzero number, which is the number count of that feature in the compound; the element is zero if the feature is absent in the compound. The number of unique features (multiple occurrences of the same feature are counted as one) present in a compound can be used as a measure of structural complexity, and this number (Leadscope unique feature count or LSUFC) as well as the total number of features (Leadscope feature count or LSFC) are also calculated for each of the 43 000 compounds and included as additional molecular descriptors.

The activity measures of the screened compounds are based on the arithmetic mean of the 60 GI_{50} measurements (negative log of 50% growth inhibition concentration) for all tumor cells. (Throughout the paper, we will use "GI_{50}" in place of $-\log(GI_{50})$ for convenience.) Since approximately 2.2% of the compounds have a mean GI_{50} of ≤ 4 and 2.6% have a mean GI_{50} of ≥ 7, we arbitrarily selected these values as cutoffs for active and inactive compounds, respectively. The distributions of each molecular descriptor within the active and inactive compounds, as well as for all 43 000 compounds, are shown in Figure 1, and the mean values for each compound group are listed in Table 1. All average descriptor values are higher in the active compound set than the inactive compound set, and this difference is statistically significant (*t*-test, $p < 0.05$). LSUFC shows the greatest difference between the active and inactive compound sets, followed by PAC and MW, and the number of RB shows the least difference, albeit significant. The average descriptor values for the inactive compounds are generally close to the overall compound averages.

Structural Characterization of the GI_{50} SOM. Our GI_{50} SOM segregates compounds into clusters and clusters into regions. These divisions define compounds that share similar GI_{50} response patterns and, putatively, mechanisms of action (MOA). Taking our prior analysis one step further, by introducing a means of relating compound molecular features to MOA, we have chosen to examine the structure−activity relationship within every SOM region, each representing different sets of MOAs and possibly unique sets of molecular descriptors. For compounds within each node (SOM cluster), the average descriptor values and mean GI_{50} are calculated. These values are then used to generate the region averages for SOM regions M, N, P, Q, R, S, F, J, and V. The distributions of each molecular descriptor value and mean GI_{50} across all nine SOM regions are displayed in Figure 2. Clearly, these values vary
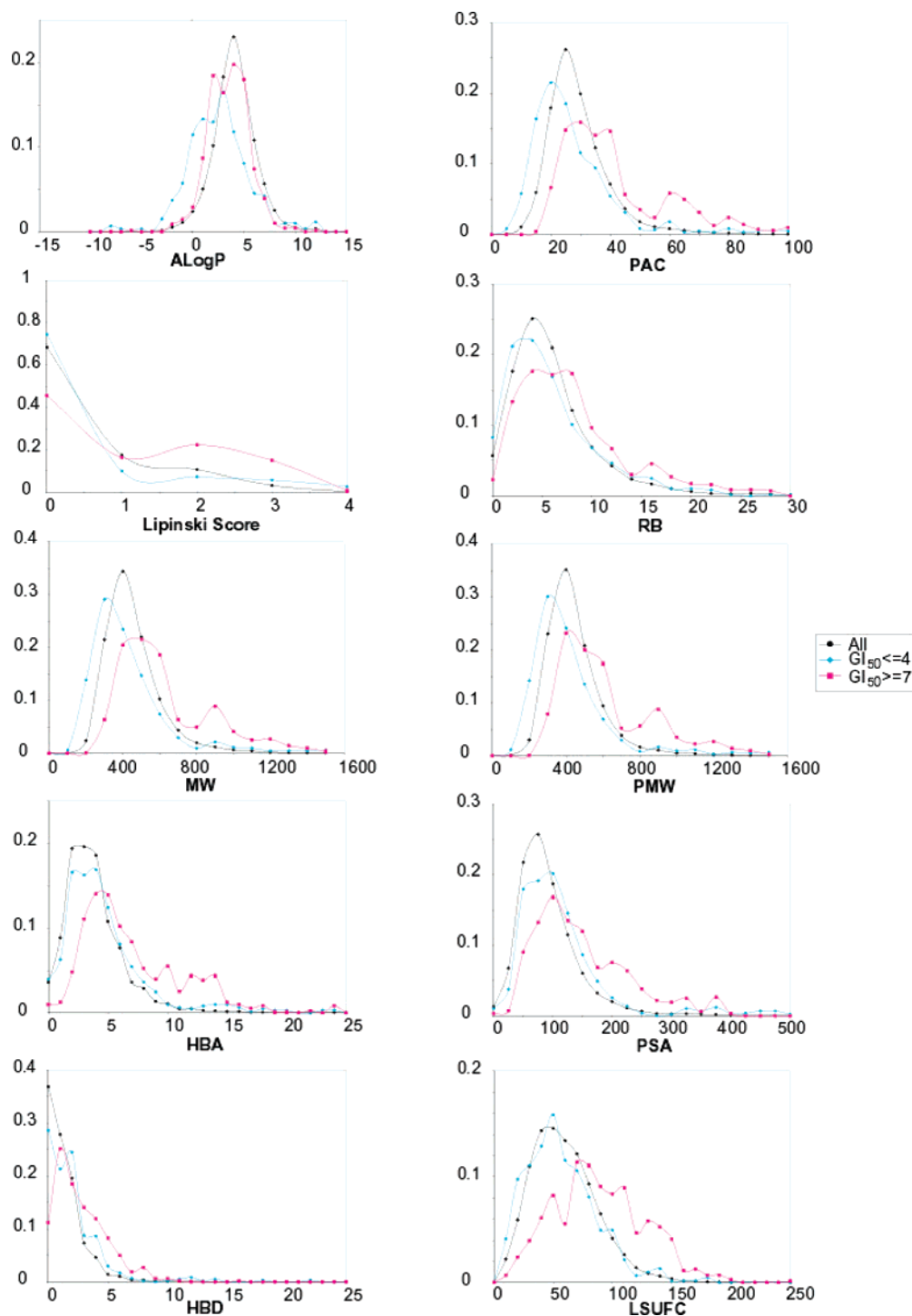
Figure 1. Distributions of common molecular descriptors ALogP, Lipinski score, molecular weight (MW), hydrogen bond acceptors (HBA), hydrogen bond donors (HBD), parent atom count (PAC), rotatable bonds (RB), polar surface area (PSA), parent molecular weight (PMW), and Leadscope unique feature count (LSUFC) for compounds that are active (mean $GI_{50} \geq 7$) and inactive (mean $GI_{50} \leq 4$) and for all 43 000 compounds screened in the NCI_{60}. The y-axis shows the compound fractions. All average descriptor values are significantly higher (t-test, $p < 0.05$) in the active compound set than the inactive compound set, and the average descriptor values for the inactive compounds are close to the overall compound averages.

for each descriptor, as well as for compound activity (mean GI_{50}), for compounds clustered in different SOM regions, showing that distinctions exist, not only in terms of putative MOA, but also in molecular features. These nine regions are then ranked according to their descriptor and mean GI_{50} values, such that the region having the smallest value is assigned a rank of one and the largest value is assigned the highest rank of nine. These rankings are listed in Table 2. A general correlation is

evident between region descriptor values and compound potency for most regions.

The M-region, which contains compounds that act by interfering with mitosis, such as tubulin active agents, has the highest ranking in every molecular descriptor (except for ALogP) and has the most potent compounds overall. This is consistent with our earlier finding that large and more complex molecules tend to be more potent. The M-region contains many

Table 1. Average Molecular Descriptor Values for Compounds that Are Active (mean $GI_{50} \geq 7$) and Inactive (mean $GI_{50} \leq 4$) and for All 43 000 Compounds Screened in the NCI_{60}[a]

descriptor	all compds with measured GI_{50}	$GI_{50} \leq 4$	$GI_{50} \geq 7$	$[(GI_{50} \geq 7) - (GI_{50} \leq 4)]/all$
LSUFC	55.18	52.56	84.84	0.59
parent atom count	27.53	25.78	40.18	0.52
molecular weight	408.6	383.6	581.2	0.48
parent molecular weight	397.9	373.4	567.1	0.49
hydrogen bond acceptors	3.91	4.823	7.12	0.59
Lipinski score	0.50	0.52	1.09	1.15
ALogP	3.54	2.01	3.00	0.28
polar surface area	81.01	105.6	136.9	0.39
hydrogen bond donors	1.36	2.01	2.72	0.52
rotatable bonds	5.76	6.05	7.65	0.28

[a] The table also lists the difference between the active and inactive compounds in each descriptor relative to the global averages. Differences in descriptor values between the active and inactive compound sets are all statistically significant (*t*-test, $p < 0.05$), with LSUFC being the most significant ($p = 9.92 \times 10^{-57}$), followed by PAC and MW, and the RB showing the least significant difference ($p = 2.30 \times 10^{-6}$).

compounds that are of natural origins, such as taxanes and vinca alkaloids, which are generally large and contain many functional groups. Furthermore, the observation that these molecules are clustered in the M-region may imply that large and functional molecules are more likely to be active against the mitotic cell cycle or, alternatively, that compounds active against mitosis are more likely to be potent. Compounds in regions S and P are second and third, respectively, in terms of their molecular descriptor rankings, as they share many descriptor values except for ALogP, which has its lowest ranking in S and its third highest ranking in P. In terms of potency as measured by mean GI_{50}, however, regions S and P show a large difference, S being the second most potent region and P only showing medium

potency when compared to other regions. The S-region contains compounds that exert their activity mainly through interfering with DNA synthesis and metabolism, such as alkylators, intercalators, and DNA/RNA antimetabolites. The low lipophilicity and high functionality of the S-region compounds may have contributed to their ability to interact with DNA, which would account for their high potency. The P-region, on the other hand, contains compounds postulated to interfere with kinase/phosphatase activity, such as CDK inhibitors, compounds that can cause oxidative stress, and other compounds with unknown MOAs. A reasonable amount of lipophilicity may be required for these activities, which are mostly not as lethal as DNA targeting. Even though the overall potency of P-region compounds is only average, evidence exists that highly potent compounds can still be found in the P-region, since this region contains the largest number of compounds that exhibit the most diverse MOAs. By contrast, compounds that cause oxidative stress are generally not potent, whereas kinase-targeting agents, such as staurosporine, can be very potent.

Compounds in regions J, N, and V also have mediocre descriptor rankings and medium potencies. The N-region contains putative membrane disruptive agents and the MOAs of compounds in the other two regions, J and V, are mostly unknown. The R-region appears to be special, having the least number of hydrogen bond donors and acceptors and the smallest polar surface area, but the highest lipophilicity (ALogP), and it is the second least potent region. Region R has agents that can effectively disrupt the mitochondrial respiration chain and thus oxidative phosphorylation. The lipophilic aspect of these agents may contribute to their activity in mitochondrial membrane and toward proteins involved in the respiration chain. However, the MOA for most of the R-region compounds is largely unknown. Region F, the MOA of which is still unknown, is the least potent
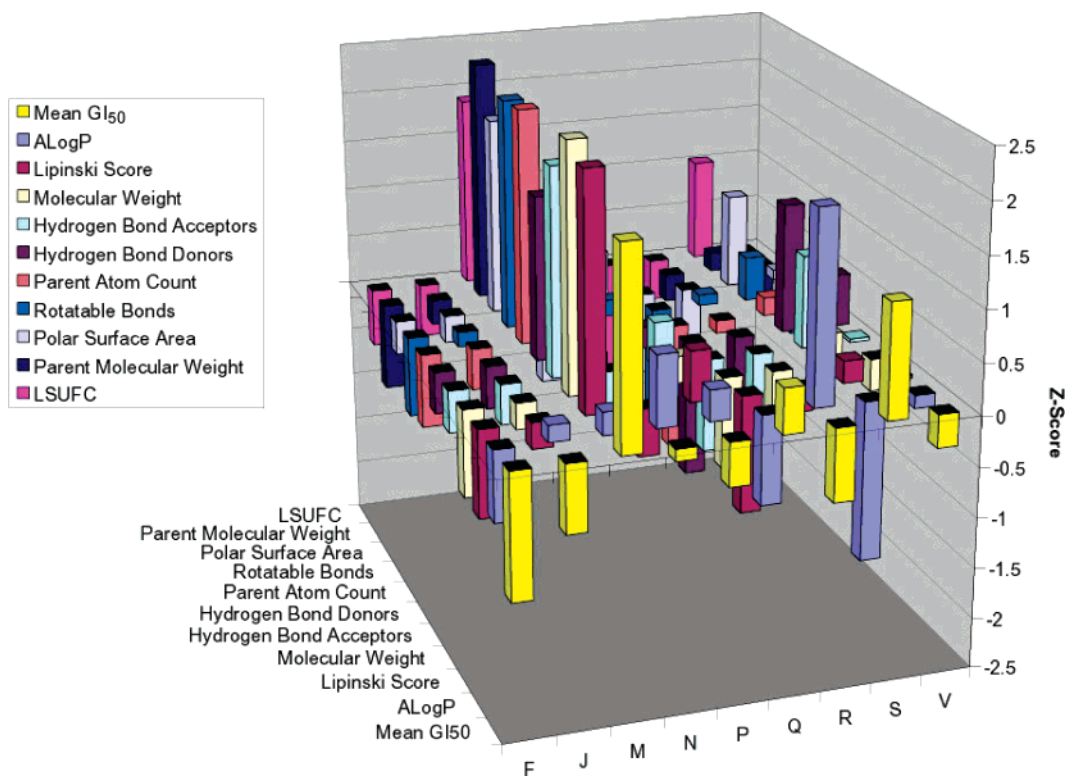

Figure 2. Distributions of molecular descriptor values and mean GI_{50} across all nine GI_{50} SOM regions. For compounds within each SOM cluster, the average descriptor values and mean GI_{50} are calculated. These values are then used to generate the region averages for SOM regions M, N, P, Q, R, S, F, J, and V. The figure shows the Z-score normalized values for each descriptor, by subtracting the mean and then dividing by the standard deviation of values across the nine regions. Clearly, these values vary for each descriptor as well as for compound activity (mean GI_{50}), showing that distinctions exist between compounds clustered in different SOM regions, not only in terms of putative MOA but also in molecular features.

Table 2. Ranks of Molecular Descriptor and Mean GI$_{50}$ Values for the Nine SOM Regions[a]

SOM region	ALogP	Lipinski	MW	HBA	HBD	PAC	RB	PSA	PMW	LSUFC	sum of ranks	mean GI$_{50}$
M	6	9	9	9	9	9	9	9	9	9	87	9
S	1	7	8	8	8	7	8	8	7	8	70	8
Q	2	1	1	2	1	1	1	2	1	1	13	7
N	8	3	6	3	5	6	4	3	6	6	50	6
V	4	6	4	6	7	4	3	6	3	5	48	5
P	7	8	7	7	6	8	7	7	8	7	72	4
J	5	4	5	5	3	3	5	5	5	3	43	3
R	9	5	3	1	2	5	6	1	4	4	40	2
F	3	2	2	4	4	2	2	4	2	2	27	1

[a] The region with the smallest parameter value is assigned a rank of one and the largest value is assigned the highest rank of nine. A positive correlation is evident between region descriptor ranks and compound potency for most regions except for the Q-region.

Table 3. Correlations (r) and Significance Levels (p-values) Calculated between Each Molecular Descriptor and Mean GI$_{50}$ Using SOM Node Average Values[a]

descriptor	SOM		Q-region	
	r	p	r	p
LSUFC	0.38	8.09×10^{-47}	−0.27	9.15×10^{-4}
molecular weight	0.35	5.26×10^{-39}	0.36	9.16×10^{-6}
LSFC	0.34	3.65×10^{-38}	−0.21	1.02×10^{-2}
parent molecular weight	0.32	5.34×10^{-34}	0.32	1.02×10^{-4}
parent atom count	0.32	1.19×10^{-32}	0.07	4.11×10^{-1}
hydrogen bond acceptors	0.31	1.42×10^{-30}	−0.21	1.07×10^{-2}
polar surface area	0.28	7.21×10^{-25}	−0.18	2.77×10^{-2}
Lipinski score	0.25	6.76×10^{-21}	0.16	6.00×10^{-2}
hydrogen bond donors	0.24	8.43×10^{-20}	−0.23	6.11×10^{-3}
rotatable bonds	0.13	2.82×10^{-6}	0.14	9.83×10^{-2}
ALogP	−0.11	1.03×10^{-4}	0.15	7.32×10^{-2}

[a] The first two columns list the correlations calculated using all SOM nodes. All but one parameter correlates positively with mean GI$_{50}$, with feature counts and MW having the strongest and ALogP having the weakest and negative correlation. The last two columns list the correlations calculated using only nodes within the Q-region. Notably, the parameters for structural feature counts, hydrogen bond donors and acceptors, and PSA are negatively correlated with mean GI$_{50}$ in the Q-region, in contrast to the results obtained with the entire SOM.

region with generally lower than average parameter rankings. This is consistent with our earlier observation that small and simple molecules are less likely to be potent. The Q-region, however, appears to be the most peculiar, because it has the lowest rankings in almost all descriptors but is also the third most potent region following regions M and S. This may be explained by the fact that the Q-region has the largest content of metal-containing compounds, especially heavy metal compounds, and other electrophiles that can readily form covalent bonds with protein targets.[18,19] These compounds are unique in the sense that, even though they do not have many hydrogen bond donors or acceptors or other structural features (Q ranks the lowest in the number of features), they can still be very potent because they interact with protein targets differently from other organic molecules by covalently modifying their targets instead of forming hydrogen bonds. Another possible reason is that many molecular descriptors for metal-containing compounds cannot be accurately calculated. Taken together, we have shown that compounds segregated by their MOA also exhibit distinct physiochemical properties, indicating a differential impact of these properties on cancer cell growth.

Table 3 lists the correlations and significance levels (p-values) calculated between each molecular descriptor and mean GI$_{50}$ using SOM node average values. It is not surprising that all but one parameter correlates positively, feature counts and molecular weight being the strongest, with mean GI$_{50}$, whereas ALogP has the weakest and negative correlation. These parameters are not all independent, because a large molecule is more likely to have a large value in all size-related molecular descriptors (e.g., MW, PAC, HBD, HBA, RB, PSA) and a large Lipinski score as well, because the chance of a large molecule violating the Lipinski rules will be increased. In fact, ALogP is the only descriptor that is not largely determined by the size of the molecule and is, therefore, relatively independent of the other descriptors. Correlations between descriptors calculated using SOM region average values show that all the other descriptors are highly correlated with each other and with mean GI$_{50}$, having an average correlation coefficient (r) of 0.86 and a minimum r of 0.66. The descriptor that has the best correlation with ALogP is polar surface area, which correlates negatively with ALogP, as expected, but only has a correlation of $r = -0.36$. The correlations between ALogP and the other descriptors are even weaker, with both positive and negative values. Therefore, these descriptors appear to contain only two major, independent components, molecule size and lipophilicity (ALogP).

As mentioned earlier, the structure−activity relationship in the Q-region appears to be distinctly different from the other regions. Excluding this region, the sum of parameter ranks appears to be a very good predictor of region potency, having a correlation with mean GI$_{50}$ rank of $r = 0.85$, which means that >70% of the variation in the GI$_{50}$ ranks can be explained by the variation in the molecular descriptor ranks. With the inclusion of the Q-region, however, the correlation with GI$_{50}$ is reduced to 0.49, which means that only 24% of the variation in GI$_{50}$ ranks can be explained. Correlations of each molecular descriptor with mean GI$_{50}$ are calculated for each SOM region and are shown in Figure 3 together with the correlations calculated using all SOM nodes (shown as black histograms). Most regions display similar trends in their structure−activity relations, which are also similar to the global trend exhibited when all compounds in the SOM are taken into consideration; that is, most parameters are positively correlated with mean GI$_{50}$, with ALogP and the number of rotatable bonds showing the largest variations across different regions. The Q-region (cyan histograms in Figure 3), however, is a clear exception. The molecular descriptor−GI$_{50}$ correlation results within the Q-region are listed in the last two columns of Table 3. The most notable difference, when compared to correlations using all SOM nodes, is that the descriptors for structural feature counts, hydrogen bond donors and acceptors, and PSA are negatively correlated with mean GI$_{50}$ in the Q-region. As discussed earlier, the Q-region contains compounds that are distinct from the other regions, being characterized by unique features (heavy metals, reactive electrophillic groups, etc.) that make them exhibit different behavior in their structure−activity relationships.[18] A detailed feature assessment of the compounds in the Q-region is not within the scope of this study. Negative correlations with mean GI$_{50}$ are also observed for some descriptors in the

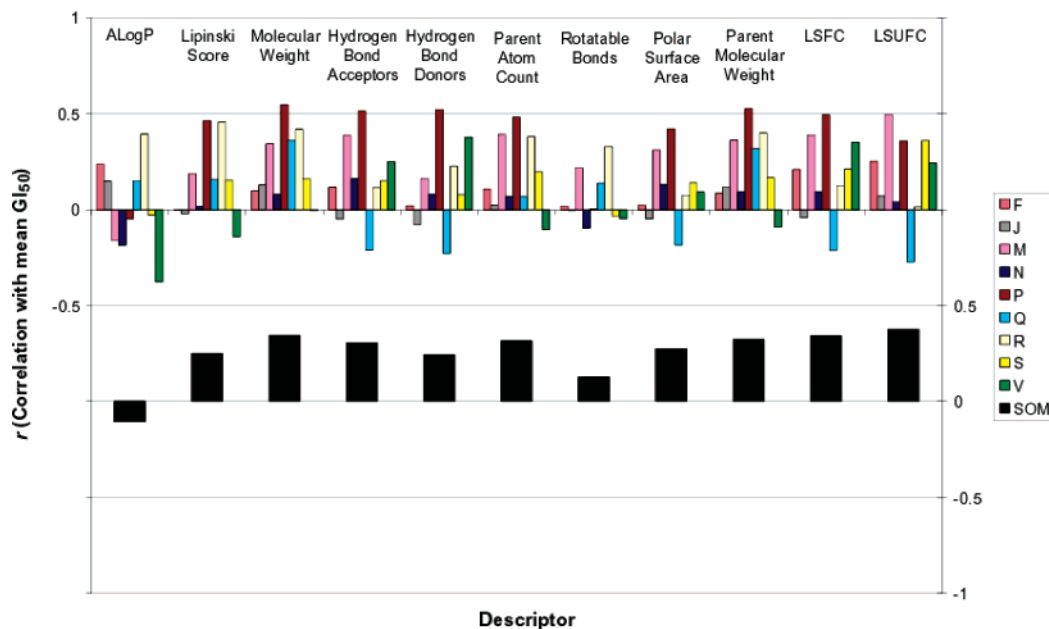**Figure 3.** Structure−activity relationships for compounds clustered in each SOM region: correlations of each molecular descriptor with mean GI_{50}. Top histograms: Correlations calculated for each SOM region using only nodes within that region. Bottom histograms: Correlations calculated using all SOM nodes. For most regions, as well as the entire SOM, the majority of descriptors are positively correlated with mean GI_{50}, with the exception of the Q-region (cyan histograms), where the descriptors for structural feature counts, hydrogen bond donors and acceptors, and PSA are negatively correlated with mean GI_{50}.

V-region; most of them, however, are not statistically significant. No significant correlation between molecular descriptors and mean GI_{50} is observed for compounds in the J-region. Taken as a whole, the results obtained here indicate that the relationship between molecular descriptors and compound activity is, at least in part, dependent on compound type and MOA.

Activity Prediction: GI_{50} Modeling. Our analysis finds that many molecular descriptors show good correlations with compound activity, specifically, mean GI_{50}, raising the possibility for predicting compound activity with a suitable combination of only a few molecular descriptors. Our goal is to build a global statistical model that will enable us to predict the activity of any compound with a known 2D structure, from which our set of molecular descriptors can be calculated. Such a model can be used as a prescreening tool to filter compounds in large databases and select a smaller set of compounds with good predicted activity for further testing and validation. Limited by the data size and quality as well as large structural variations between compounds, the model is not expected to be highly accurate; however, a valid model with good predictive power that allows quick activity prediction for large sets of compounds would suffice for our purpose.

The final GI_{50} prediction model included 108 descriptors that are statistically significant, the combination of which can be used to estimate compound potency (see Methods section for details). Descriptors such as MW, HBD, certain structural features containing heteroatoms (N, P, O, S, halides, etc.), and ADME properties PSA, BBB (blood−brain barrier), solubility, and hepatoxicity are among the descriptors that have the most significant influence on GI_{50}. (All model parameters can be found in the Supporting Information.) The GI_{50} values predicted by the model are highly correlated with the measured GI_{50} values, as shown in Figure 4a (see Methods section for details). This correlation is only slightly weaker in the testing (data not used to build the model) than the training data set (data used to build the model), indicating that this model can be applied to a new data set without significant loss of accuracy. Figure 4b shows the receiver operating characteristic (ROC) curve of the

model, which exhibits the characteristics of a valid and predictive model (see Methods section for details). The sensitivity and specificity of the model in selecting truly potent compounds are maximized when the mean GI_{50} value is around 6. The probability of finding active compounds (measured GI_{50} ≥ 6) within the compound set predicted as active (predicted GI_{50} ≥ 6) is 64%, which is significantly improved when compared to the 6% probability of finding active compounds by random chance. Moreover, the average predicted GI_{50} value for a set of known anticancer drugs (collected by Leadscope) is shown higher than that of all screened compounds (4.9 vs 4.7; *t*-test, $p = 1.26 \times 10^{-5}$). These results indicate that the biological activity, such as GI_{50}, of a compound can be related to a combination of its physiochemical properties.

Activity Prediction: Hollow Fiber Activity Modeling. About 3000 compounds have been tested in the in vivo HF model developed by the NCI for antitumor effects.[22] In the standard assay, 12 human tumor cell lines growing in polyvinylidene fluoride "hollow fibers", representing six different histologies, selected owing to the expected behavior of their corresponding xenografts, are placed in both ip and sc compartments of nude mice. Compounds can then be dosed at intervals over a 4−5 day period, and then the fibers are removed from the mice, and the effect of compound action on the proliferation of tumor cells in the HF is assessed by colorimetric assays. Agents are considered to have a HF effect if there is a 50% or greater reduction in net cell growth compared with the controls. The hollow fiber assay is scored by assigning 2 points to a compound for each fiber in which the compound has shown such an effect. The score is typically recorded as the score in the ip fibers plus the score in the sc fibers. The greater the number of fibers demonstrating evidence of an antiproliferative effect, the greater the likelihood that a compound will display activity against xenografts. The criteria for activity in the standard HF assay were based, and statistically validated, on the scores achieved by clinically used anticancer agents. Historically, the criterion for compounds to be considered for
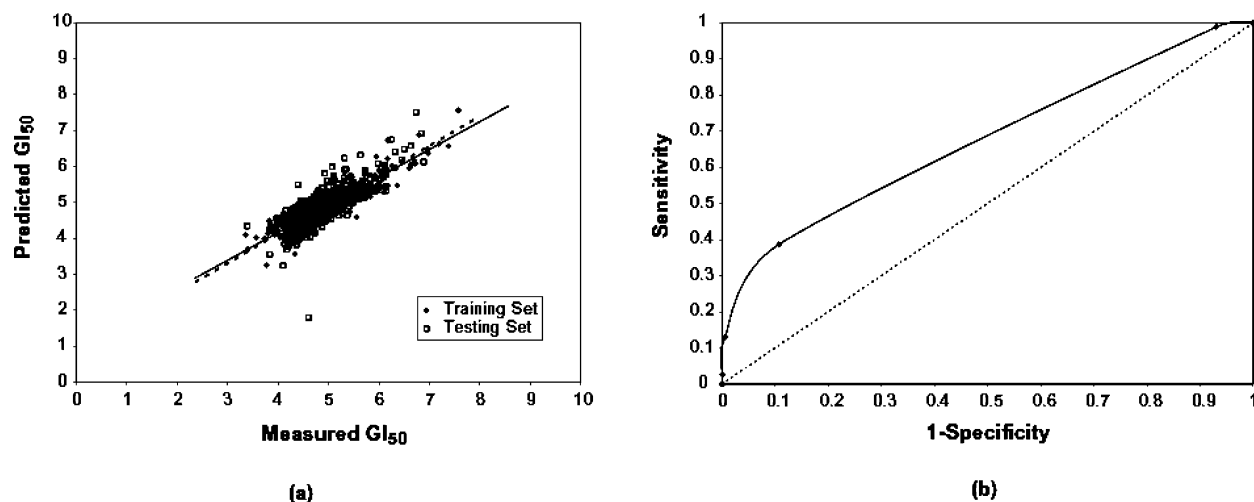
(a)

(b)

Figure 4. Validation of the GI_{50} prediction model built with Cerius[2] fast descriptors. (a) Plots of predicted GI_{50} values against measured GI_{50} values for the training (data used to build the model) and testing (data not used to build the model) sets, yielding $r^2 = 0.77$ for the training set and cross-validation $r^2 = 0.67$ for the testing set. (b) Receiver operating characteristic (ROC) curve of the model (see Methods section for details). The ROC curve exhibits the characteristics of a valid and predictive model. The area under the ROC curve is maximized when the mean GI_{50} value is around 6, where maximum sensitivity and specificity are achieved. The probability of finding active compounds (measured $GI_{50} \geq 6$) within the compound set predicted as active (predicted $GI_{50} \geq 6$) is 64%, which is significantly improved when compared to the 6% probability of finding active compounds by random chance.

follow-up xenograft testing was a total HF (ip + sc) score of 20 or greater.

Using the total HF score for each compound (in cases where a compound has been tested repeatedly in the hollow fiber assay and thus has multiple scores, we pick the maximum score as its HF score), the 3119 compounds tested have a mean optimal HF score of 11.7 with a standard deviation (SD) of 9.8. Approximately 17% of the compounds have a HF score of 20 (mean + SD) or greater, which we will consider as HF active, and 14% have a HF score of 2 (mean − SD) or less, which we will consider as HF inactive. A compound's HF score has been found to be significantly correlated with its mean GI_{50} ($r = 0.31$, $p = 4.60 \times 10^{-61}$), indicating that a compound showing potency in the 60-cell screen is more likely to be active in the HF assay. Analogous to our earlier effort to build a model with molecular descriptors to predict a compound's mean GI_{50}, here we propose a similar model using the same descriptors to predict a compound's HF activity. Mean GI_{50} is included as a parameter as well.

The final HF activity prediction model (see Table 4) kept 20 structural descriptors that significantly contribute to a compound's HF activity. Mean GI_{50}, not surprisingly, is the most significant parameter. The structural descriptor HBD; information-content descriptors, such as the entropy of vertex adjacency matrix (V_ADJ_mag) and the entropy of edge distance matrix (E_DIST_mag); ADME properties, such as absorption, protein plasma binding (PPB), and AlogP; and certain structural features and atom types, as well as the number of unique structural features (LSUFC), are also among the most significant molecular descriptors influencing compound activity in the HF assay. Dependent variable random permutation tests show that the HF model is predictive and statistically significant ($p < 0.01$; see Methods section for details). The predicted HF scores show strong correlation with the measured HF scores in both the training and testing data sets, as shown in Figure 5a, indicating that the model has good predictive power and is applicable to other data sets. Figure 5b shows the ROC curve of the model, which exhibits the characteristics of a valid and predictive model. The area under the ROC curve is maximized when the HF score is approximately between 20 and 25, where maximum

Table 4. HF Activity Prediction Model Built with Cerius[2] Fast Descriptors, LSUFC, and Mean GI_{50} as Independent Predictors (see Methods section for details)[a]

descriptor	parameter estimate	standard error	p
mean GI_{50}	2.61	0.24	$< 1 \times 10^{-4}$
Hbond donor	1.21	0.20	$< 1 \times 10^{-4}$
V_ADJ_mag	−0.03	4.86×10^{-3}	$< 1 \times 10^{-4}$
E_DIST_mag	4.01×10^{-4}	7.54×10^{-5}	$< 1 \times 10^{-4}$
ADME_Absorption_T2_2 D	−0.15	0.03	$< 1 \times 10^{-4}$
S_dssS	−13.58	3.02	$< 1 \times 10^{-4}$
Atype_S_109	−21.73	5.42	$< 1 \times 10^{-4}$
ADME_Absorption_Leve l_2D	1.33	0.33	$< 1 \times 10^{-4}$
CHI_V_3_CH	21.26	5.74	2.00×10^{-4}
LSUFC	0.04	0.01	4.00×10^{-4}
Atype_Cl_90	−2.82	0.81	5.00×10^{-4}
ADMET_PPB	−1.01	0.30	7.00×10^{-4}
ADME_Unknown_AlogP 98	−1.89	0.57	1.00×10^{-3}
Atype_O_57	−1.14	0.39	3.70×10^{-3}
Atype_F_81	9.32	3.60	9.80×10^{-3}
Atype_C_17	0.80	0.33	1.41×10^{-2}
Atype_C_28	0.54	0.23	1.89×10^{-2}
Atype_N_68	1.06	0.46	2.26×10^{-2}
Atype_H_49	0.85	0.38	2.37×10^{-2}
Atype_C_43	2.41	1.08	2.57×10^{-2}
Atype_C_34	0.73	0.38	5.45×10^{-2}
S_ddC	8.21	4.40	6.22×10^{-2}
S_sNH2	−0.20	0.11	6.86×10^{-2}
S_sCH3	0.11	0.07	8.72×10^{-2}
model intercept	−0.63	1.55	6.86×10^{-1}

[a] The table lists the parameter estimates, standard errors, and the significance levels of for all parameters significant at $p < 0.15$ and the intercept of this linear regression model. Mean GI_{50}; the structural descriptor HBD; information-content descriptors V_ADJ_mag and E_DIST_mag; ADME properties, such as absorption, protein plasma binding (PPB), and AlogP; certain structural features and atom types; and the number of unique structural features (LSUFC) are among the most significant predictors of HF activity.

sensitivity and specificity are achieved. The 60% probability of finding HF active compounds (measured HF score ≥ 20) in the compound set predicted as active (predicted HF score ≥ 20) is significantly improved when compared to the 13% probability of finding active compounds by random chance. Using this model, we have predicted the HF activity for more than 27 000 compounds with available GI_{50} data, and over 600 of them are
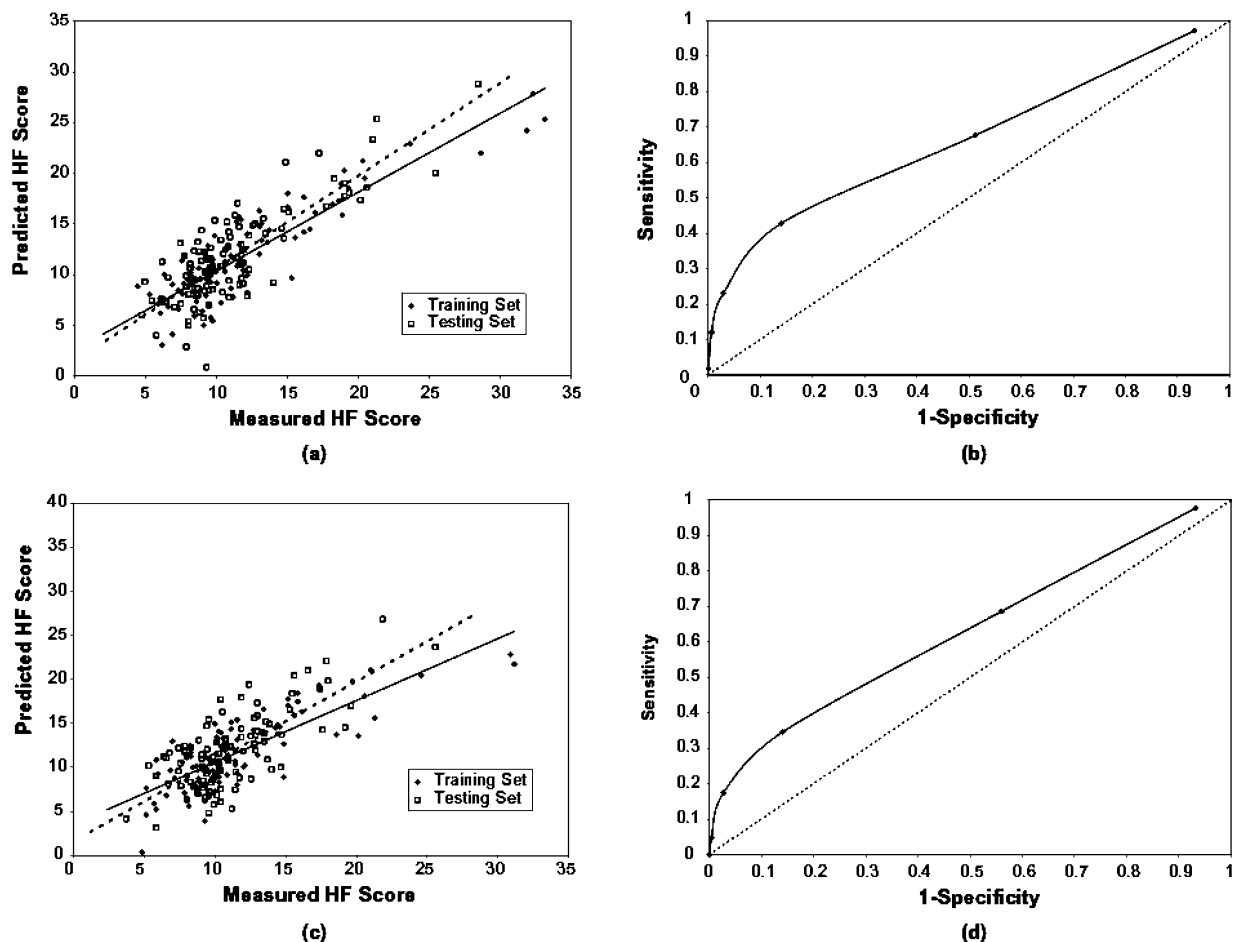
Figure 5. Validation of the hollow fiber (HF) activity prediction models built with Cerius[2] fast descriptors. (a) Model built with GI_{50} as an additional predictor. Plots of predicted HF scores against measured HF scores for the training (data used to build the model) and testing (data not used to build the model) sets, yielding $r^2 = 0.78$ for the training set and cross-validation $r^2 = 0.61$ for the testing set. (b) Receiver operating characteristic (ROC) curve of the model built with GI_{50} as an additional predictor (see Methods section for details). The ROC curve exhibits the characteristics of a valid and predictive model. The area under the ROC curve is maximized when the HF score is approximately between 20 and 25, where maximum sensitivity and specificity are achieved. The 60% probability of finding HF active compounds (measured HF score ≥ 20) in the compound set predicted as active is significantly improved when compared to the 13% probability of finding active compounds by random chance. (c) Model built without using GI_{50} as an additional predictor. Plots of predicted HF scores values against measured HF scores for the training (data used to build the model) and testing (data not used to build the model) sets, yielding $r^2 = 0.68$ for the training set and cross-validation $r^2 = 0.49$ for the testing set. (d) ROC curve of the model built without using GI_{50} as an additional predictor (see Methods section for details). The ROC curve exhibits the characteristics of a valid and predictive model. The area under the ROC curve is maximized when the HF score is approximately 20, where maximum sensitivity and specificity are achieved. HF active compounds (measured HF score ≥20) are significantly enriched in the compound set predicted as active when compared to the rest of the compounds (55% vs 14%). Both models are statistically valid and predictive. The model built without using GI_{50} as an additional predictor is only slightly inferior.

predicted to be active (HF score ≥20) in the HF assay. It is noteworthy that the average predicted HF score for a set of known anticancer drugs (collected by Leadscope) is significantly higher than that of all screened compounds (14.2 vs 8.9; *t*-test, $p = 2.96 \times 10^{-23}$).

Our model can be used to virtually screen compounds with GI_{50} data available and test those predicted to be active in the HF assay. For compounds that have not been tested in the NCI_{60}, our original model, built using GI_{50} as a parameter for prediction, is then not applicable. Even though GI_{50} is found to be the best-correlated parameter for HF activity, other descriptors are also shown to contribute significantly to a compound's HF activity; therefore, building a separate model for HF activity without using GI_{50} as a predictor can be achieved. The HF activity prediction model built using the same set of data, excluding GI_{50}, and similar procedures, is slightly inferior but still statistically significant ($p < 0.01$) based on randomization tests (see Methods section for details). In this model 24 descriptors are shown to significantly contribute to HF activity (All model

parameters can be found in the Supporting Information). The few extra descriptors included by this GI_{50} absent model are presumably needed to compensate for its loss as a model parameter. Nevertheless, the results suggest that GI_{50} measurements can be supplanted by other molecular descriptors without causing a significant loss in the reliability and predictive power of the model. The most significant parameters influencing compound HF activity again include the number of unique structural features (LSUFC) and ADME properties such as solubility, absorption, and protein plasma binding (PPB). The training and cross-validation results for the model are shown in Figure 5c. The correlation between the predicted and experimentally measured HF scores is still strong but considerably lower than what obtained for the HF model using GI_{50} as a parameter. Figure 5d shows the ROC curve of the GI_{50}-free model. Despite the area under the ROC curve being less than that for the original model (Figure 5b), the curve still exhibits the characteristics of a valid predictive model, indicating a HF score of around 20 as the optimal activity cut off. HF active

Table 5. Correlations of Molecular Descriptors and Mean GI$_{50}$ with Target Specificity[a]

descriptor	r	p
mean GI$_{50}$	0.32	4.36×10^{-33}
molecular weight	0.15	3.63×10^{-8}
parent molecular weight	0.13	3.07×10^{-6}
LSFC	0.09	9.12×10^{-4}
parent atom count	0.09	1.52×10^{-3}
LSUFC	0.08	2.64×10^{-3}
rotatable bonds	0.08	3.47×10^{-3}
hydrogen bond acceptors	0.08	4.89×10^{-3}
polar surface area	0.06	2.06×10^{-2}
Lipinski score	0.05	8.04×10^{-2}
ALogP	-0.03	2.26×10^{-1}
hydrogen bond donors	0.01	8.38×10^{-1}

[a] Within-node (DYFP SOM) GI$_{50}$ response pattern correlation strength (intra-node r) is used as a measure of target specificity for structurally similar compounds. Correlations are calculated using node average values of common molecular descriptors and mean GI$_{50}$ within each DYFP SOM cluster. Potency (intranode mean GI$_{50}$) is found to be the parameter that is the most significantly correlated ($r = 0.32$, $p = 4.36 \times 10^{-33}$) with target specificity, followed by MW ($r = 0.15$, $p = 3.63 \times 10^{-8}$).

compounds (measured HF score ≥ 20) are significantly enriched in the compound set predicted as active when compared to the rest of the compounds (55% vs 14%). Our model predicts that more than 1200 of the 43 000 public compounds screened in the NCI$_{60}$, most of which have not yet advanced to HF testing, will be active in the HF assay (predicted HF score ≥ 20). Taken together, our results show that biological activity in different cancer models, either in vitro or in vivo, can be modeled on the basis of compound physiochemical properties.

Compound Structural Similarity, Target Specificity, and HF Activity Relationships. Our results demonstrate that compound activity or potency is closely related to and predictable by, even though not perfectly, a set of structure-based molecular descriptors. We have also observed that compounds having different MOAs or targets may exhibit different structure–activity dependencies. As others have observed, structurally similar compounds may or may not share the same MOA. We would consider structurally similar compounds that also exhibit similar activity (not in terms of potency but in terms of targets) as being specific and otherwise as being promiscuous. Building from this premise, we will examine the relationship between structural features, activity or potency, and specificity, in a global sense. The virtual selection of potent compounds that are also specific would be of high interest in choosing compounds for further testing.

SOM clustering of the 43 000 screened compounds using Daylight fingerprints (DYFP) segregated the compounds into 1334 clusters (nodes). Compounds clustered in the same node on the DYFP SOM are considered structurally similar. The degree of similarity in GI$_{50}$ activity for structurally similar compounds can be defined as the average correlation coefficient (intranode r) between the GI$_{50}$ data vectors of these compounds. Within-node r-values can then be used as a measure of similarity in compound MOA, i.e., target specificity. Average values of some common molecular descriptors and compound mean GI$_{50}$'s are calculated for each DYFP SOM cluster as well. Correlations between these parameters and intranode r are listed in Table 5. Interestingly, the descriptor found to be the most significantly correlated ($r = 0.32$, $p = 4.36 \times 10^{-33}$) with target specificity (intranode r) is compound potency (intranode mean GI$_{50}$), the correlation of which is much stronger than that of the second most significantly correlated descriptor, MW ($r = 0.15$, $p = 3.63 \times 10^{-8}$). Other structural descriptors such as LSFC, PAC, LSUFC, RB, HBA, and PSA also show some degree of correlation with target specificity, though much weaker. None-

theless, the underlying implication that target specificity might be one of the important factors contributing to compound potency is very intriguing. Since mean GI$_{50}$ has been found to be one of the most important determinants of HF activity (Table 4), it was reasonable to determine the correlation strength between target specificity and HF activity. Using the average intranode HF scores for DYFP SOM nodes that contain compounds with measured HF activity finds a weak correlation between target specificity and HF activity (significant at $p < 0.10$).

Compounds that are both potent and target specific are generally of great interest. Our observation that target-specific compounds are more likely to be potent, or vice versa, can be exploited for mining novel anticancer candidates. Here we propose a strategy to find compounds that have a consistent structure–activity (CSA) relationship, that is, structurally similar compounds that also show similar GI$_{50}$ patterns (i.e., target specificity), through the combined use of the structural (DYFP) and activity (GI$_{50}$) SOMs. A selection criterion is used such that, for each node in the DYFP SOM that contains at least five compounds that are also clustered in the same node on the GI$_{50}$ SOM, these compounds are included in the CSA set. This procedure identifies 1541 compounds occupying 167 GI$_{50}$ SOM nodes and 153 DYFP SOM nodes (i.e., structural classes). The average potency for the CSA compounds (mean GI$_{50}$ = 5.5) is significantly larger (t-test, $p < 10^{-10}$) than an average screened compound (mean GI$_{50}$ = 4.7). The average HF score of the 176 CSA compounds with measured HF activity is 13.1, slightly higher than the average of all compounds tested in the HF assay (mean HF = 11.7). This difference, however, is only significant at the 90% confidence level (t-test, $p = 0.06$).

HF activities for all screened compounds including the 1541 CSA compounds are predicted using the prediction model built from molecular descriptors and mean GI$_{50}$ (Table 4). Of the CSA compounds, 138 are predicted to be active (HF ≥ 20) in the HF assay. The average predicted HF activity for the CSA compound set (mean predicted HF = 11.7) is significantly larger than that of an average screened compound (mean predicted HF = 8.8) (t-test, $p = 3.29 \times 10^{-51}$). The number of predicted active compounds (9% with HF ≥ 20) is also significantly enriched in the CSA set compared to a random set of screened compounds (2% with HF ≥ 20) (Fisher's exact, $p = 1.12 \times 10^{-21}$). The predicted HF activity for the 176 CSA compounds that have been tested in the HF assay correlates well with their measured activity ($r = 0.39$, $p = 7.08 \times 10^{-8}$). These results show that compounds selected on the basis of target specificity, e.g., the CSA compounds, are more likely to be potent and active in the HF assay.

To compare the CSA compounds, in terms of structural descriptors, with compounds showing different levels of HF activity, several common descriptor values as well as mean GI$_{50}$ are averaged for HF active (HF ≥ 20), inactive (HF ≤ 2), and the CSA compounds. Each parameter is then Z-score normalized, and the Z_x scores, where x = HF active, HF inactive, and CSA, for ALogP, Lipinski score, MW, HBA, HBD, PAC, RB, PSA, PMW, LSUFC, and mean GI$_{50}$, are depicted in Figure 6a (see Figure 6 caption for details). A compound set with positive Z_x scores has above average parameter values and a compound set with negative Z_x scores has below average parameter values. Consistent with earlier findings, the HF active compounds have above average values for all structural descriptors (except for ALogP), and the HF inactive compounds have below average values for all structural descriptors. Moreover, all three compound sets show higher potency than an average screened
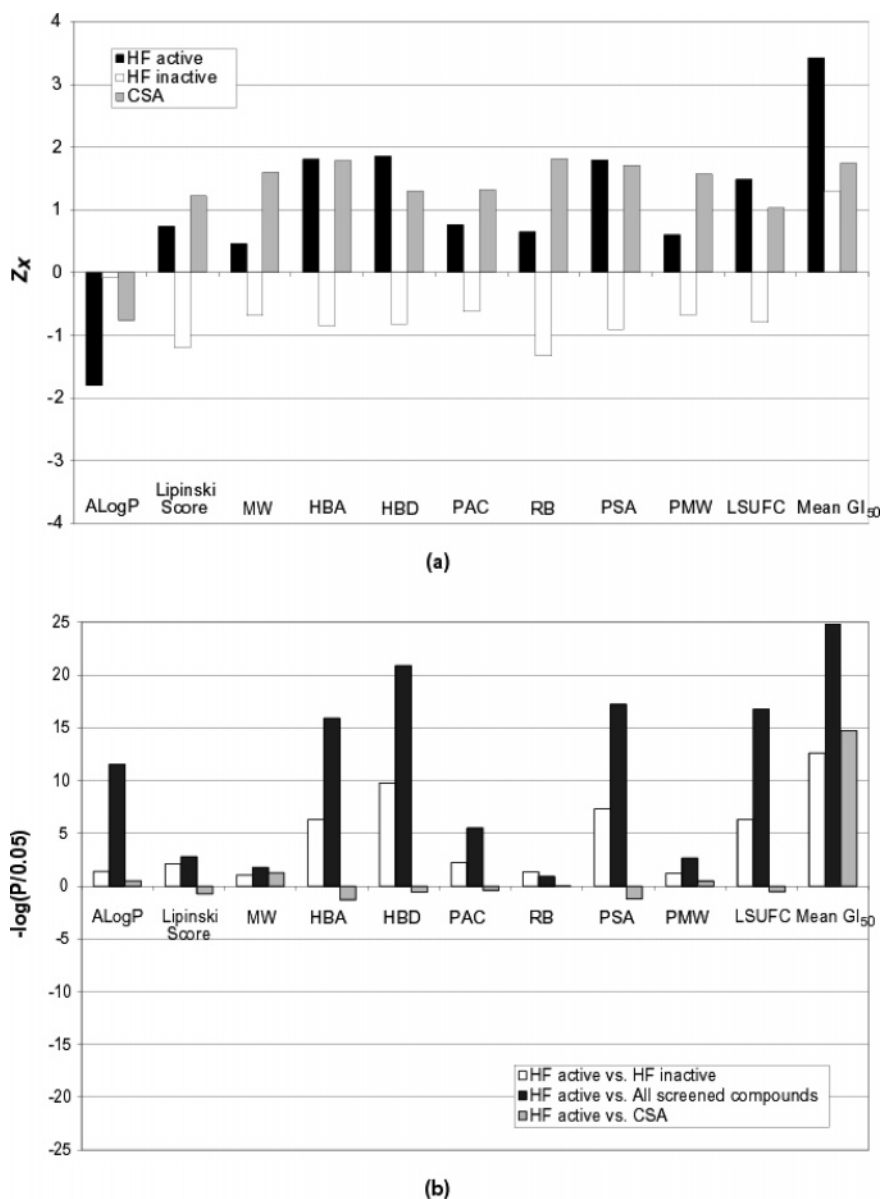
Figure 6. Comparison of the CSA, HF active (HF score ≥ 20), and inactive (HF score ≤ 2) compound sets in terms of molecular descriptors and mean GI_{50}. For each compound set, the average descriptor values are calculated; each descriptor is then normalized using the descriptor values for all screened compounds as reference points ($Z_x = (\text{mean}_x - \text{mean}_{ref})/SD_{ref}$, where Z_x is the normalized value of a descriptor for compound set x, x is the compound set of interest, and ref is the set of all screened compounds). Z_x then indicates the degree of deviation for compound set x from an average screened compound. A compound set with positive Z_x scores has above average descriptor values and a compound set with negative Z_x scores has below average descriptor values. (a) The Z_x scores, where $x =$ HF active, HF inactive, and CSA, for ALogP, Lipinski score, MW, HBA, HBD, PAC, RB, PSA, PMW, LSUFC, and mean GI_{50}. The HF active compounds have above average values for all structural descriptors (except for ALogP), and the HF inactive compounds show below average values for all structural descriptors. All three compound sets show higher potency than an average screened compound. The Z_x scores for the CSA compounds most closely resemble those of the HF active compounds. (b) The significance levels of differences in descriptor values between compound sets. t-Tests are performed to compare the descriptor values between HF active and inactive, HF active and an average screened compound, and HF active and the CSA compound sets. The histograms are $-\log(P/0.05)$ values calculated for each descriptor, such that a positive value indicates a statistical significance at $p < 0.05$. All descriptors are significantly different between the HF active and inactive compound sets and an average screened compound. No significant difference is found between the descriptor values of the CSA compounds and the HF active compounds, except that the latter has lower MW and ALogP and shows slightly higher potency.

compound, as their Z_x scores for mean GI_{50} are all large positive values. The Z_x scores for the CSA compounds most closely resemble those of the HF active compounds. The significance levels of these differences are shown in Figure 6b (see Figure 6 caption for details). All parameters are significantly different between the HF active and inactive compound sets and an average screened compound. There is no significant difference, however, between the descriptor values of the CSA compounds and the HF active compounds, except that the latter has lower MW and ALogP and shows slightly higher potency.

Figure 7a depicts the Z_x scores for compounds in the nine GI_{50} SOM regions together with the HF active and inactive and the CSA compound sets. Close resemblance in descriptor patterns can be seen between the HF active compounds, the CSA compounds, and the compounds in SOM region M, as well as between the HF inactive compounds and compounds in region Q. This again shows that even though potency or GI_{50} is a good predictor of HF activity, it is certainly not the sole determinant, as the compounds in Q are clearly not the least potent, yet they most resemble HF inactive compounds.
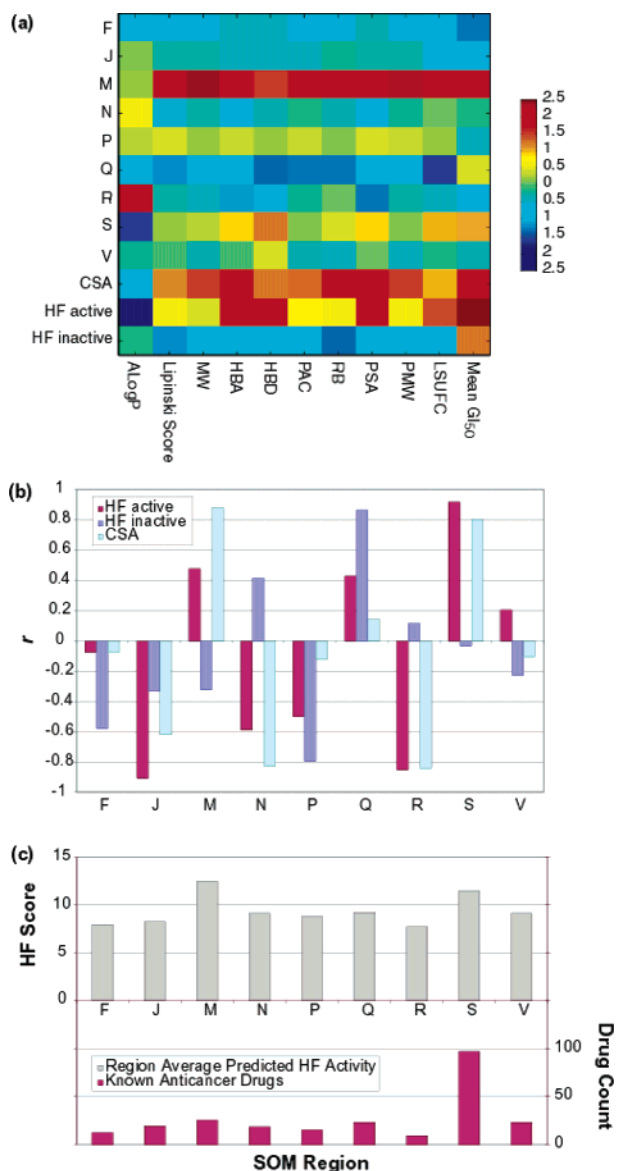
Figure 7. Comparison of the CSA, HF active (HF score ≥ 20) and inactive (HF score ≤ 2) compound sets and compounds clustered in each SOM region in terms of molecular descriptors, mean GI_{50}, and HF activity. (a) Z_x scores (see Figure 6 caption for details) for compounds in the nine GI_{50} SOM regions together with the HF active and inactive and the CSA compound sets. Close resemblance in descriptor patterns can be seen between the HF active compounds, the CSA compounds, and the compounds in SOM region M and between the HF inactive compounds and compounds in region Q. (b) Descriptor Z_x score (see Figure 6 caption for details) correlations between each compound set and the nine SOM regions. A large positive r with a SOM region indicates a close similarity between a compound set and the compounds in that SOM region in terms of chemical properties and potency. A large negative r indicates incompatibility. These relationships can be used as a way to roughly estimate the MOA of a compound set. HF active compounds are most likely to reside in the SOM regions that have higher "affinity" for the HF active compounds than the HF inactive compounds. (c) HF activity and known drug content in the nine SOM regions. Top histograms: Average predicted HF activity for each SOM region. Regions M and S have the highest predicted HF activity, and the rest of the regions show comparable activities. Bottom histograms: Distribution of a set of known anticancer drugs collected by Leadscope on the SOM. The S-region has the most known drugs, and the other drugs are distributed fairly evenly across the other regions. The distribution of the known drugs shows a good correlation with their predicted HF activity ($r = 0.61$).
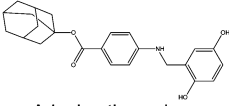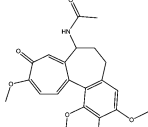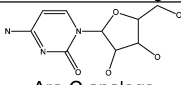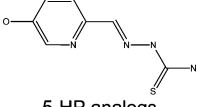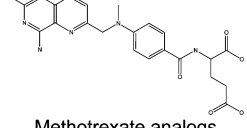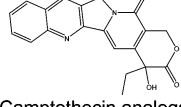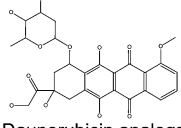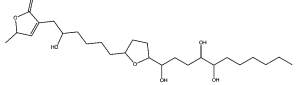
Moreover, even the HF inactive compounds have significantly above average mean GI_{50} values. The relationship between the

compound groups and each SOM region is better illustrated in terms of correlations in Figure 7b. Correlations are calculated for the parameter Z_x scores of each compound group with those of each SOM region. A large positive r with a SOM region indicates a close similarity of the compounds in x and the compounds in that SOM region in terms of chemical properties and potency. A large negative r, on the other hand, indicates incompatibility. Since SOM regions can be used as surrogates for compound MOAs, the relationships illustrated in Figure 7b can be viewed as a way to roughly estimate the MOA of a compound set. The idea of predicting MOA based solely on compound structural descriptors (GI_{50} can be used as an additional descriptor but is not necessary) via the SOM is intriguing. The difference in the correlation patterns of the different compound sets with the SOM reveals differences in their MOAs. The compounds in the SOM region that most strongly correlate with a compound set in terms of structural descriptors are also most likely to share their primary MOA with the compound set. This method can be expanded to assess individual compounds.

Figure 7b can be used alternatively to find new HF active compounds, which are most likely to reside in the SOM regions that have higher "affinity" for the HF active compounds than the HF inactive compounds. Therefore, even though regions S, M, Q, and V all have positive correlations with HF active compounds, the Q-region has an even stronger positive correlation with HF inactive compounds, which leaves regions S, M, and V with selective affinity for HF active compounds. Conversely, the regions to avoid would be N, Q, and possibly R, where HF inactive compounds are likely to be prevalent. The average predicted HF activities for compounds in each SOM region, using the prediction model built from molecular descriptors and mean GI_{50} (Table 4), are shown in Figure 7c. Regions M and S have the highest predicted HF activity, and the rest of the regions show comparable activities, with the R-region compounds predicted to have the lowest activities. The results are consistent with those indicated by Figure 7b. The CSA compounds show clear promise, since they are selected as target specific compounds and are also predicted to be potent and active in the HF assay. Their structural descriptors have the strongest positive correlations with regions M and S, which are the two regions predicted to have the best HF activity.

Table 6 lists a few examples of the CSA compound classes derived from our integrated structure−activity analysis, their predicted and measured mean GI_{50} values and HF scores, and the SOM regions where these compounds are located. The activity values are averaged over each compound class. The predicted activity values agree well with the measured values for most of these compound classes. The compound classes clustered in the antimitotic (M) and nucleic acid synthesis and metabolism (S) regions show clearly above average GI_{50} and HF scores. Noteworthy is that these compound classes have included several standard anticancer agents, such as the antimitotic agents colchicines; the DNA antimetabolites Ara-C, 5-HP, and methotrexate; and the topoisomerase inhibitors camptothecin and daunorubicin. Furthermore, our procedure also finds adaphostin and its analogues as a CSA compound class, which has been actively pursued as p210bcr-abl protein tyrosine kinase inhibitors that induce apoptosis in human leukemia cells.[35−37] Identified as another CSA class are the acetogenins, which are known inhibitors of mitochondrial complex I and oxidative phosphorylation.[38,39] These results indicate that other novel CSA compound classes discovered by our procedure may also be interesting agents that warrant further investigation.

Table 6. Examples of Several CSA Compound Classes Derived from Our Integrated Structure−Activity Analysis[a]

CSA Compound Class	Predicted GI_{50}	Measured GI_{50}	Predicted HF Score	Measured HF Score	SOM Region
Adaphostin analogs	4.65	5.44	10.87	10.83	J
Colchicine analogs	5.82	5.80	18.35	19.60	M
Ara-C analogs	4.74	5.00	13.10	11.00	S
5-HP analogs	4.78	5.92	15.21	15.00	S
Methotrexate analogs	5.01	5.55	14.48	16.80	S
Camptothecin analogs	5.12	5.99	16.55	13.67	S
Daunorubicin analogs	5.87	6.53	19.59	20.67	S
Acetogenins	5.19	5.53	10.22	N/A	R

[a] The table lists the representative structural motif for each compound class, the predicted and measured mean GI_{50} values and HF scores averaged over each compound class, and the SOM regions where these compounds are located. The compound classes clustered in the antimitotic (M) and nucleic acid synthesis and metabolism (S) regions show clearly above average GI_{50} and HF scores. Noteworthy is that these compound classes include several standard anticancer agents and known inhibitors of the protein tyrosine kinase p210bcr-abl and mitochondrial complex I.

These compounds are freely available upon request to our fellow cancer researchers who are interested in testing them.

Discussion

This study attempts to achieve a global assessment of compounds tested in the NCI_{60} in vitro anticancer screen in terms of molecular structure, biological activity, target specificity, and MOA and to determine the feasibility of deriving valid statistical models from structural/physicochemical descriptors to predict compound potency, as represented by mean GI_{50} and in vivo HF activity. Significant differences are observed in the distributions of common physicochemical descriptors between the active (mean $GI_{50} \geq 7$) and inactive (mean $GI_{50} \leq 4$) screened compounds. The active compound set has higher values in every structural descriptor than the inactive compound set. The structural descriptor LSUFC, as a measure of molecular complexity, shows the greatest difference between the active and inactive compound sets, followed by parent atom count and molecular weight, and the number of rotatable bonds shows the least difference. Molecular weight has been observed, within

the NCI's set of 166 standard anticancer agents, to contribute significantly to cancer cell growth inhibitory activities.[29] Molecular weight is directly proportional to the size of a molecule, and most of the other descriptors (e.g., PAC, HBA, HBD, RB, PSA) are also size-related. It is then not surprising that when one of these parameters is correlated with activity, the others will be more or less correlated as well. Our analysis, done from a global perspective, reveals that compounds with large structural/physicochemical descriptor values are more likely to be active or potent in cancer cell growth inhibition. This in turn indicates that it is not molecular weight per se, but rather molecule size and structural complexity in general, that contribute significantly to the activity or potency of a compound. The concept of molecular complexity has been shown to be valuable to achieve biological activity in pharmaceutical research.[40] However, complexity needs to be balanced with other molecular properties to avoid pharmacokinetic problems inherent in selecting large compounds. As we have observed early on in this study and will discuss next in the text, structure−activity

relationships are a collective consequence of specific compound features and MOA.

SOM clustering of GI_{50} response patterns segregates compounds into groups that share a similar putative MOA. As a way of relating molecular features to MOA, we have examined the structure−activity relationship for compounds within each SOM region. Consistent with earlier results, molecule size and complexity are found to contribute positively to compound potency in most cases, i.e., most physicochemical/structural descriptors are positively correlated with mean GI_{50}, while ALogP and RB show the largest variations in different SOM regions. However, compounds clustered in the Q-region, which are characterized by unique chemical features, exhibit structure−activity behaviors that are distinct from compounds in other regions, such that several size-related parameters (e.g., LSFC, HBA, HBD, PSA) are found to be negatively correlated with mean GI_{50}.[18] This exception to the general observation that large and more complex molecules are more potent is an indication that the manner in which physicochemical/structural descriptors determine compound activity is, at least in part, dependent on compound type and/or MOA.

Inspired by our observations that many molecular descriptors show good correlations with compound activity, specifically, mean GI_{50}, we have built linear regression models, which are shown to be statistically significant and valid, to predict compound activity (GI_{50} and HF activity) with suitable combinations of molecular descriptors. Efforts attempting to use molecular features as predictors of biological activity are not unprecedented and have shown success in small, focused data sets.[41] Our models built from large-scale data sets are proposed as prescreening tools to virtually filter compounds in large databases and select a smaller set with good predicted activity for subsequent testing and validation. The fact that compound activity can be predicted with reasonable accuracy, even on a global scale, reaffirms the concept of chemistry determining activity. One has to keep in mind, however, that the models, on average, can only explain about 20% of the variability in the data; thus, a large percentage of variance in activity still cannot be accounted for solely by the parameters used. Nevertheless, if the goal is to preselect the most potentially active compounds in a large database, the proposed models can be used effectively to eliminate most of the inactive compounds (other active compounds may be missed as a risk, since, when the model is highly selective, it is less sensitive).

The HF activity prediction model built with GI_{50} as an additional predictor is shown to be only slightly superior to the one built without, indicating that GI_{50} is a good, but not the sole, predictor of HF activity. Thus, molecular descriptors other than GI_{50} also contribute significantly to the HF activity of a compound, and GI_{50} may be replaced by other descriptors without a significant loss in the reliability and predictive power of the model. Another interesting observation is the repeated occurrences of structural feature count and ADME properties in the models as the most significant predictors of compound activity. Since these models are built on a grand scale using almost all available data, and general potency measures such as GI_{50} and HF activity are modeled without going into the details of specific target-oriented activity, the subtleties involving particular compound features will not be reflected but rather eliminated by the model, and only properties that are important for a compound's biological activity in general are thus revealed, which could be expected to include such properties as general molecular complexity and ADME properties.

Another property of common interest is target specificity. One way to measure target specificity is to determine whether a set of structurally similar compounds shares the same targets or MOA, as gauged by the strength of correlations between their differential GI_{50} response patterns. Compound potency (mean GI_{50}) is thus found to positively correlate with target specificity and is also the most significantly correlated parameter. Other physicochemical/structural descriptors such as molecular weight, feature count, parent atom count, rotatable bonds, hydrogen bond acceptors, and polar surface area also show some degree of correlation with target specificity, though much weaker. Reduction in molecular size and complexity is known to reduce the specificity and activity of natural products;[42] positive contribution from these molecular features to both specificity and potency is, therefore, expected. However, a direct correlation between potency and specificity, which is also exceedingly stronger than the correlations of specificity with other parameters, is very intriguing because this implies that it is specific targeting, rather than promiscuous poisoning, that gives rise to potency. In addition, a positive correlation, though not as strong, is also observed between compound specificity and in vivo HF activity. Compounds that are both potent and target specific are generally of great interest. Our observation that target specific compounds are more likely to be potent, or vice versa, has led us to the development of a strategy to exploit this relationship for mining novel anticancer agents. The combined filtering of our structural (DYFP) and activity (GI_{50}) SOMs reveals a set of compounds (the CSA set), the structurally similar subsets of which also show activity similarity (i.e., target specificity). As expected, these compounds show significantly higher potency and are predicted to be significantly more active in the HF assay than an average screened compound.

Finally, as a way to link compound activity to MOA via molecular features, we have compared the CSA compounds and the HF active and inactive compounds with compounds clustered in different SOM regions, i.e., compounds of different MOAs, in terms of physicochemical/structural descriptors and GI_{50}. Consistent with earlier findings, the HF active compounds appear to be larger, structurally more complex, and more potent, on average, than the HF inactive compounds. The CSA compounds exhibit characteristics very similar to the HF active compounds property-wise. Moreover, compounds clustered in SOM regions M, S, and V show higher "affinity" (stronger positive parameter correlations) for the HF active than the HF inactive compounds. Conversely, compounds in regions N, Q, and R seem to have selective affinity for HF inactive compounds. As a side note, even though our descriptor analysis identifies regions M and S as the most likely to contain potent and HF active compounds, this does not necessarily imply that these compounds will be absent in other SOM regions. To the contrary, the distribution of the set of known anticancer drugs collected by Leadscope on the SOM (Figure 7c, bottom histograms), plotted together with the average predicted HF activities for each SOM region (Figure 7c, top histograms), clearly shows that despite region S having the most known drugs, the other drugs are distributed fairly evenly across the other regions. The distribution of these drugs across the SOM regions actually has a good correlation with their predicted HF activity ($r = 0.61$), since M and S also have the highest predicted HF activity, whereas the rest of the regions show comparable activities.

Since SOM regions are indicative of compound MOAs, the parameter correlations between a compound set and each SOM region can then be used to predict the MOAs of that compound

set. That is, the putative MOA of a compound may be predicted based solely on molecular descriptors, with the optional use of GI_{50} as an additional parameter, via the SOM. The difference in the correlation patterns of the different compound sets with the SOM reveals differences in their MOAs. The primary MOA of a compound set is most likely to be similar to compounds in the SOM region with which it has the strongest physicochemical/structural parameter correlations.

Conclusions

In summary, this study presents a global analysis of compounds tested in the NCI_{60} in vitro anticancer screen in terms of molecular structure, biological activity (GI_{50} and in vivo HF activity), target specificity, and MOA, through the combined use of structural/physicochemical descriptors and the SOM. Our analysis finds that molecule size and structural complexity in general, contribute significantly to compound potency in cancer cell growth inhibition and HF activity. We have examined the structure−activity relationship for compounds within each GI_{50} SOM region to relate molecular features to MOA. Molecular size and complexity are found to positively contribute to compound potency in most cases, with exception occurring for compounds having distinct features and MOA. We have built linear regression models that are shown to be statistically significant and valid, which can be used as prescreening tools to filter in silico compounds in large databases to effectively eliminate most of the inactive compounds and select a smaller set with good predicted activity for subsequent testing and validation. The concept of chemistry determining activity is reaffirmed by the fact that compound activity is globally predictable with reasonable accuracy. Our results show that GI_{50} is a good, but not the sole, predictor of HF activity. A direct correlation is also found between potency and specificity, indicating that specific targeting, rather than promiscuous poisoning, gives rise to potency. We have proposed a strategy to exploit this relationship for mining novel anticancer candidates. Finally, we have shown that correlations between compounds and SOM regions via molecular descriptors can be used to predict compound MOAs.

Methods

Data Modeling. The molecular descriptor set is obtained by calculating the 26 families (except for Daylight search) of fast descriptors, which include 214 descriptors, in the Cerius^2 v. 4.9 molecular modeling software from Accelrys Inc. (San Diego, CA). We also included the Leadscope unique feature count (LSUFC), calculated by counting the number of unique structural features defined by Leadscope present in a molecule, as an additional parameter to cover the structural complexity aspect of the molecules, resulting in a total of 215 independent variables (descriptors) and one dependent variable (activity), to be modeled. One model is built with mean GI_{50} as the dependent variable to be predicted using about 27 000 compounds with assigned 2D structures and GI_{50} data. Using a set of 2847 out of the 3119 compounds tested in the HF assay with known structures, the HF activity data is modeled both with and without mean GI_{50} as an additional independent variable. To build each model, compounds are randomly divided into two sets, one set used for training and the other one for testing. The data are modeled using a multistep linear regression method available in the SAS statistical software (The SAS System V8, SAS Institute Inc., Cary, NC). Additional modeling algorithms such as PLS using either NIPALS or SVD to compute extracted PLS factors are also tested and the results do not seem to be superior.

Table 7. Model r^2 and Mean and SD of the r^2 values for 100 Random Models Obtained for the GI_{50} Prediction Model and the HF Activity Models with and without GI_{50}

model	r^2	mean random r^2	SD
GI_{50}	0.22	2.78×10^{-3}	6.50×10^{-4}
HF (with GI_{50})	0.24	0.025	5.77×10^{-3}
HF (without GI_{50})	0.19	0.024	5.92×10^{-3}

In the final model, each parameter estimate is associated with a *p*-value, which indicates the statistical significance of that parameter. A parameter estimate with a smaller *p*-value has less error and is therefore more accurate. Only variables that are significant at the $p < 0.15$ level are kept in the model.

Model Validation and Predictive Power. The model is built using only compounds in the training data set, where a set of descriptors that significantly contribute to activity is derived and their weights estimated. The activity value for each compound is then computed using the set of significant descriptor values and weights from the model, yielding the predicted values. The model is then cross-validated by predicting the activity values for the compounds in the testing data set using the set of descriptors derived from the training set. Correlations between the measured and predicted activity values are calculated for both the training and testing sets and compared. Compounds are sorted according to their predicted activity values and divided into 1000 groups for the GI_{50} model and 100 groups for the HF activity prediction models. The group averages are plotted against the measured activity values. This approach is used to reduce noise and observe global trends. For the GI_{50} prediction model (Figure 4a), the grouped training set yields a very good r^2 of 0.77 and the $r^2 = 0.67$ obtained for the testing set, i.e., the cross-validation r^2, is only slightly lower. For the HF activity prediction model with GI_{50} as an additional predictor (Figure 5a), the grouped training set yields an r^2 of 0.78, and the cross-validation r^2 of 0.61, obtained for the testing set, is lower but still reasonably good. Finally, for the GI_{50} free HF activity prediction model, the grouped training set r^2 is 0.68 and the cross-validation r^2 is 0.49.

Randomization tests are performed where the dependent variable, i.e., the activity data, is randomly permuted 100 times and a new model is built each time using the permuted data. The model correlation coefficient, r^2, which represents the amount of variability in activity that can be explained by the descriptors, is used as a measure of model quality. The model is considered statistically valid and predictive if the real model r^2 is significantly larger than those of the models generated from the randomly permutated data sets ($p < 0.05$). The model r^2, mean, and SD of the r^2 values for 100 random models obtained for the GI_{50} prediction model and the HF activity models with and without GI_{50} as an additional predictor are listed in Table 7. The r^2 values for all three models are larger than the maximum r^2 values generated from the random permutation tests; therefore, all three models are statistically significant at $p < 0.01$.

Another validation method used to evaluate the predictive power of a model is the receiver operating characteristic (ROC) curve, which is a plot of sensitivity versus (1 − specificity). Sensitivity, defined as the proportion of active compounds that are predicted as active [TP/(TP + FN)], describes how well a model identifies active compounds. Similarly, specificity is defined as the proportion of inactive compounds that are predicted as inactive [TN/(TN + FP)] and describes how well a model identifies inactive compounds. Predicted activities are computed for all compounds using the model, and the number of compounds that fall into each of the four categories (TP,

FP, TN, and FN) at different real activity levels are counted and the sensitivity and specificity values calculated. A good predictive model yields the greatest number of true positives with the least number of false positives, resulting in a ROC curve that tends upward while moving from right to left. The activity cutoff level that gives the maximum sensitivity and specificity, when the area under the ROC curve will be the maximum, will be the best cutoff to use when applying the model in compound selection.

Supporting Information Available: Parameter estimates for GI_{50} prediction model and hollow fiber activity prediction model built without GI_{50} as an additional predictor. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Hayflick, L. Recent advances in the cell biology of aging. *Mech. Ageing Dev.* **1980**, *14*, 59−79.

(2) Baguley, B. C.; Marshall, E. S. In vitro modelling of human tumour behaviour in drug discovery programmes. *Eur. J. Cancer* **2004**, *40*, 794−801.

(3) Boyd, M. R. The NCI human tumor cell line (60-Cell) screen: Concept, implementation, and applications. *Anticancer Drug Development Guide: Preclinical Screening, Clinical Trials, and Approval*, 2nd ed.; Humana Press: Totowa, NJ, 2004; pp 41−61.

(4) Suggitt, M.; Bibby, M. C. 50 years of preclinical anticancer drug screening: Empirical to target-driven approaches. *Clin. Cancer Res.* **2005**, *11*, 971−981.

(5) Gonzalez-Nicolini, V.; Fussenegger, M. In vitro assays for anticancer drug discovery−A novel approach based on engineered mammalian cell lines. *Anticancer Drugs* **2005**, *16*, 223−228.

(6) Tanaka, M.; Bateman, R.; Rauh, D.; Vaisberg, E.; Ramachandani, S.; et al. An unbiased cell morphology-based screen for new, biologically active small molecules. *PLoS Biol.* **2005**, *3*, e128.

(7) Vande Woude, G. F.; Kelloff, G. J.; Ruddon, R. W.; Koo, H. M.; Sigman, C. C.; et al. Reanalysis of cancer drugs: Old drugs, new tricks. *Clin. Cancer Res.* **2004**, *10*, 3897−3907.

(8) Novotny, J.; Szekeres, T. Recent developments in cancer chemotherapy oriented towards new targets. *Expert Opin. Ther. Targets* **2005**, *9*, 343−357.

(9) Shoemaker, R. H.; Monks, A.; Alley, M. C.; Scudiero, D. A.; Fine, D. L.; et al. Development of human tumor cell line panels for use in disease-oriented drug screening. *Prog. Clin. Biol. Res.* **1988**, *276*, 265−286.

(10) Monks, A.; Scudiero, D.; Skehan, P.; Shoemaker, R.; Paull, K.; et al. Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. *J. Natl. Cancer Inst.* **1991**, *83*, 757−766.

(11) Voigt, J. H.; Bienfait, B.; Wang, S.; Nicklaus, M. C. Comparison of the NCI open database with seven large chemical structural databases. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 702−712.

(12) Wallqvist, A.; Huang, R.; Thanki, N.; Covell, D. G. Evaluating Chemical Structure Similarity as an Indicator of Cellular Growth Inhibition. *J. Chem. Inf. Model.* **2006**, *46*, 430−437.

(13) Paull, K. D.; Shoemaker, R. H.; Hodes, L.; Monks, A.; Scudiero, D. A.; et al. Display and analysis of patterns of differential activity of drugs against human tumor cell lines: Development of mean graph and COMPARE algorithm. *J. Natl. Cancer Inst.* **1989**, *81*, 1088−1092.

(14) Boyd, M. R.; Paull, K. D. Some practical considerations and applications of the National Cancer Institute in vitro anticancer drug discovery screen. *Drug Dev. Res.* **1995**, *34*, 91−109.

(15) Weinstein, J. N.; Myers, T. G.; O'Connor, P. M.; Friend, S. H.; Fornace, A. J., Jr.; et al. An information-intensive approach to the molecular pharmacology of cancer. *Science* **1997**, *275*, 343−349.

(16) Shi, L. M.; Fan, Y.; Lee, J. K.; Waltham, M.; Andrews, D. T.; et al. Mining and visualizing large anticancer drug discovery databases. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 367−379.

(17) Rabow, A. A.; Shoemaker, R. H.; Sausville, E. A.; Covell, D. G. Mining the National Cancer Institute's tumor-screening database: Identification of compounds with similar cellular activities. *J. Med. Chem.* **2002**, *45*, 818−840.

(18) Covell, D. G.; Wallqvist, A.; Huang, R.; Thanki, N.; Rabow, A. A.; et al. Linking tumor cell cytotoxicity to mechanism of drug action: An integrated analysis of gene expression, small-molecule screening and structural databases. *Proteins* **2005**, *59*, 403−433.

(19) Huang, R.; Wallqvist, A.; Covell, D. G. Anticancer metal compounds in NCI's tumor-screening database: Putative mode of action. *Biochem. Pharmacol.* **2005**, *69*, 1009−1039.

(20) Voskoglou-Nomikos, T.; Pater, J. L.; Seymour, L. Clinical predictive value of the in vitro cell line, human xenograft, and mouse allograft preclinical cancer models. *Clin. Cancer Res.* **2003**, *9*, 4227−4239.

(21) Hollingshead, M. G.; Alley, M. C.; Camalier, R. F.; Abbott, B. J.; Mayo, J. G.; et al. In vivo cultivation of tumor cells in hollow fibers. *Life Sci.* **1995**, *57*, 131−141.

(22) Decker, S.; Hollingshead, M.; Bonomi, C. A.; Carter, J. P.; Sausville, E. A. The hollow fibre model in cancer drug screening: The NCI experience. *Eur. J. Cancer* **2004**, *40*, 821−826.

(23) Hall, L. A.; Krauthauser, C. M.; Wexler, R. S.; Hollingshead, M. G.; Slee, A. M.; et al. The hollow fiber assay: Continued characterization with novel approaches. *Anticancer Res.* **2000**, *20*, 903−911.

(24) Johnson, J. I.; Decker, S.; Zaharevitz, D.; Rubinstein, L. V.; Venditti, J. M.; et al. Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials. *Br. J. Cancer* **2001**, *84*, 1424−1431.

(25) Kohonen, T. *Self-Organizing Maps*; Springer-Verlag: Berlin, 1995.

(26) Kohonen, T. The self-organizing map. *Neurocomputing* **1998**, *21*, 1−6.

(27) Fan, Y.; Shi, L. M.; Kohn, K. W.; Pommier, Y.; Weinstein, J. N. Quantitative structure−antitumor activity relationships of camptothecin analogues: Cluster analysis and genetic algorithm-based studies. *J. Med. Chem.* **2001**, *44*, 3254−3263.

(28) Shi, L. M.; Fan, Y.; Myers, T. G.; O'Connor, P. M.; Paull, K. D.; et al. Mining the NCI anticancer drug discovery databases: Genetic function approximation for the QSAR study of anticancer ellipticine analogues. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 189−199.

(29) Ren, S. S.; Lien, E. J. Anticancer agents: Tumor cell growth inhibitory activity and binary QSAR analysis. *Curr. Pharm. Des.* **2004**, *10*, 1399−1415.

(30) Hansch, C.; Leo, A.; Hoekman, D. Exploring QSAR fundamentals and applications in chemistry and biology, volume 1. Hydrophobic, electronic and steric constants, volume 2 [Erratum for *J. Am. Chem. Soc.* **1995**, *117*, 9782]. *J. Am. Chem. Soc.* **1996**, *118*, 10678.

(31) Hansch, C.; Hoekman, D.; Leo, A.; Weininger, D.; Selassie, C. D. Chem-bioinformatics: Comparative QSAR at the interface between chemistry and biology. *Chem. Rev.* **2002**, *102*, 783−812.

(32) Hansch, C.; Leo, A.; Mekapati, S. B.; Kurup, A. QSAR and ADME. *Bioorg. Med. Chem.* **2004**, *12*, 3391−3400.

(33) Blower, P. E.; Cross, K. P.; Myatt, G. J.; Yang, C. Building predictive models from high-throughput screening data. *Abstracts of Papers, 228th ACS National Meeting, Philadelphia, PA, United States, August 22−26, 2004;* American Chemical Society: Washington, DC, 2004; CINF-090.

(34) Yang, C.; Cross, K.; Myatt, G. J.; Blower, P. E.; Rathman, J. F. Building predictive models for protein tyrosine phosphatase 1B inhibitors based on discriminating structural features by reassembling medicinal chemistry building blocks. *J. Med. Chem.* **2004**, *47*, 5984−5994.

(35) Dasmahapatra, G.; Rahmani, M.; Dent, P.; Grant, S. The tyrphostin adaphostin interacts synergistically with proteasome inhibitors to induce apoptosis in human leukemia cells through a reactive oxygen species (ROS)-dependent mechanism. *Blood* **2006**, *107*, 232−240.

(36) Hose, C.; Kaur, G.; Sausville, E. A.; Monks, A. Transcriptional profiling identifies altered intracellular labile iron homeostasis as a contributing factor to the toxicity of adaphostin: Decreased vascular endothelial growth factor secretion is independent of hypoxia-inducible factor-1 regulation. *Clin. Cancer Res.* **2005**, *11*, 6370−6381.

(37) Yu, C.; Rahmani, M.; Almenara, J.; Sausville, E. A.; Dent, P.; et al. Induction of apoptosis in human leukemia cells by the tyrosine kinase

inhibitor adaphostin proceeds through a RAF-1/MEK/ERK- and AKT-dependent process. *Oncogene* **2004**, *23*, 1364−1376.

(38) Tormo, J. R.; Gallardo, T.; Peris, E.; Bermejo, A.; Cabedo, N.; et al. Inhibitory effects on mitochondrial complex I of semisynthetic mono-tetrahydrofuran acetogenin derivatives. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4101−4105.

(39) Tormo, J. R.; Royo, I.; Gallardo, T.; Zafra-Polo, M. C.; Hernandez, P.; et al. In vitro antitumor structure−activity relationships of threo/trans/threo mono-tetrahydrofuranic acetogenins: Correlations with their inhibition of mitochondrial complex I. *Oncol. Res.* **2003**, *14*, 147−154.

(40) Selzer, P.; Roth, H. J.; Ertl, P.; Schuffenhauer, A. Complex molecules: Do they add value? *Curr. Opin. Chem. Biol.* **2005**, *9*, 310−316.

(41) Hodes, L. Selection of descriptors according to discrimination and redundancy. Application to chemical structure searching. *J. Chem. Inf. Comput. Sci.* **1976**, *16*, 88−93.

(42) Hall, D. G.; Manku, S.; Wang, F. Solution- and solid-phase strategies for the design, synthesis, and screening of libraries based on natural product templates: A comprehensive survey. *J. Comb. Chem.* **2001**, *3*, 125−150.