Pergamon

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 11 (2003) 5249–5258

BIOORGANIC & MEDICINAL CHEMISTRY

Identification of Structural Components Associated with Cytostatic Activity in MCF-7 but not in MDA-MB-231 Cells

Albert R. Cunningham,[a,*] Suzanne L. Cunningham[a,c] and Billy W. Day[b,c]

[a]*Department of Environmental Studies, Louisiana State University, Baton Rouge, LA 70803, USA*
[b]*Departments of Pharmaceutical Sciences and of Chemistry, University of Pittsburgh, Pittsburgh, PA 15261, USA*
[c]*Department of Environmental and Occupational Health, University of Pittsburgh, Pittsburgh, PA 15260, USA*

Received 7 January 2003; revised 29 July 2003; accepted 7 August 2003

Abstract—The National Cancer Institute's Developmental Therapeutics Program maintains the screening results obtained in 60 standardized cancer cell lines and contained 37,836 compounds for this study. This dataset has shown to be an outstanding resource for the development of structure–activity relationship (SAR) models describing anticancer activity. We report here a novel SAR modeling approach based on a subtractive protocol to develop models that describe cell type-specific molecular descriptors of cytotoxicity. The goal of this approach is to separate features associated with antiproliferative activity to many cell lines from those that effect only a specific cell type. To assess this approach, we developed SAR models for cytostatic activity against the human breast cancer cell lines MCF-7 and MDA-MB-231 and one differential activity model for compounds that were potent cytostatic agents in MCF-7 cells but relatively inactive against MDA-MB-231 cells. The models were between 72 and 84% accurate when challenged with compounds not in the learning sets. Structural features associated with the differential activity model highlighted how the use of this approach can selectively identify chemical moieties associated with potent cytostatic action to MCF-7 but not to MDA-MB-231 cells. We surmise that outgrowth of this method can facilitate the development of SAR models with sufficient resolution and clarity to identify chemical moieties associated with antiproliferative activity to selective individual cancer types while being innocuous to other cell types.

Introduction

At the start of this study in the Fall 2001, the National Cancer Institute's (NCI) Developmental Therapeutics Program (DTP) database of in vitro screening results consisted of 37,836 compounds. The database is continually updated and compounds are typically assayed against up to 60 standardized cell lines to determine their antiproliferative (GI_{50}, TGI values) and cytotoxic (LC_{50}) potencies.[1–3] The current database, supporting information, and tools can be accessed at dtp.nci.nih.gov. The cell lines are derived from a variety of human cancer types including colorectal, renal, ovarian, breast, prostate, lung, CNS, leukemias and melanomas. Shi et al.[4] have referred to the overall process of screening the compounds in the 60 cell lines for the three endpoints as developing a profile or 'fingerprint' of anticancer activity for each tested compound. The COMPARE algorithm has been developed to assist in searching through the DTP to identify compounds with similar biological activities.[3,5] The program identifies compounds with similar spectrum of activity across the 60 cell lines based on a 'seed' compound's activity pattern. This methodology has successfully been applied to finding novel agents operating through a number of important anticancer mechanisms including anti-tubulin[6] and topoisomerase II active agents.[7] By using the fingerprint of a compound with a known desirable therapeutic activity as a guide, it is possible to search for other similarly acting compounds in the database. For example, cluster analysis reveals that a group of camptothecin analogues display activity patterns similar to topoisomerase inhibitors.[8] Since two of these analogues have also demonstrated topoisomerase activity, this method demonstrates a feasible technique to identify others. The general conclusion from this type of study is that compounds which exhibit similar in vitro biological activity patterns to a known drug will be likely to act via a similar biochemical mechanism of action and thus potentially possess its therapeutic activity.

*Corresponding author. Tel.: +1-225-578-9422; fax: +1-225-578-4286; e-mail: arc@lsu.edu

0968-0896/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2003.08.018

Results from previously screened compounds in the DTP also serve as a vast resource for structure–activity relationship (SAR) modeling with one of the goals being the identification of structural attributes associated with antitumor activity. Historically, SAR and quantitative SAR (QSAR) analyses have been conducted using a single empirical value for the dependent (activity) variable. However, given that the DTP contains observations across 60 cell lines and for three endpoints, Weinstein and colleagues have examined the ability of a number of SAR techniques to extract meaningful information about the structural requirements for antitumor activity when using multiple observations. Again, using the set of camptothecin analogues as an example, Fan et al. used mean GI_{50} value across the 60 cells lines and a series of molecular descriptors as the independent variables to identify structural attributes associated with antitumor activity.[8] Likewise, cluster analysis helped to identify different ellipticine analogues based on the activity patterns of the individual compounds used in the analysis.[4]

Overall, fingerprinting entails an additive method of simultaneous comparison of multiple assay results (e.g., mean GI_{50} values or cluster analysis). We report here a novel approach based on a subtractive protocol to develop SAR models that describe molecular descriptors associated with cell type specific activity. The goal of this type of approach is to identify and remove structural moieties associated with general antiproliferative or cytotoxic activity to certain cell lines and by so doing identify those attributes that are influential only to a specific cell type. In other words, we seek to identify structural moieties that potentially interact with cellular components present only in the targeted cell type.

To assess the feasibility of this approach, we selected DTP results from MCF-7 and MDA-MB-231 human breast cancer cell lines. These two lines have a well known difference, estrogen receptor alpha (ER) status or responsiveness. We selected ER status as the differing point in the present study because it is well characterized and understood. However, we note ER status is not the only difference in these two cell lines. For example, Yu and colleagues demonstrated that these two cell lines differentially express various P450 isozymes[9] and, likewise, PCR analyses have shown differential expression of a variety of RNA messages in these cell lines.[10]

In this study, we developed 'control' SAR models for GI_{50} in MCF-7 and MDA-MB-231 cells. For the experimental model, we used differential activity values obtained by subtracting the GI_{50} values for compounds tested in MDA-MB-231 cells from those of the same compounds tested in MCF-7 cells. Therefore, compounds with a high differential activity value were more cytostatic to MCF-7 cells, while low differences indicated equipotency in the two cell lines. The results from this study indicate the ability to identify structural attributes of antitumor activity that can be focused on particular types of cancer cells.

Methods

SAR modeling

We used the MCASE suite of algorithms[11–13] to develop models from our growth inhibition models. The models are based on the occurrence of identified significant molecular features in each learning set. Briefly, the MCASE modeling process begins with the compilation of a learning set of chemical structures (i.e., MOL files from the DTP) and their experimentally derived biological activity value. Each compound in the learning set is broken down, in silico, to all possible fragments from 2 to 10 heavy (i.e., non-hydrogen) atoms. Each fragment is labeled with the name and activity of its parent chemical. Upon completion of this process, the program organizes the list of fragments and tabulates the number of chemicals containing each of them. The program then identifies those fragments that were identified predominately in active chemicals and refers to these fragments as biophores. The selection of biophores is based on the binary experimental results of each chemical. For example, here we used compounds that showed no growth inhibitory potential as inactives and compounds with growth inhibitory activity as active. An MCASE module then selects from the list of biophores the most statistically significant one based on its occurrence in the largest number of chemicals in the learning set. At this point in the MCASE routine, a congeneric series of chemicals has been identified with the biophore being the unifying feature. The program then performs a series of defined chemical substitutions of the atoms in the first biophore (e.g., one halogen for another halogen or an sp^2 nitrogen for a carbon in aromatic systems) and then searches for these expanded definitions of the biophore in the library of previously identified significant chemicals. All chemicals containing the biophore and the expanded definitions are then grouped together. Thus, a biophore may consist of a single feature or a family of chemically similar features.

Using the molecules contained in this family of chemicals as a new learning set, MCASE identifies modulators of their activity. These modulators may be chemical, physicochemical or quantum mechanical parameters. Modulators augment or decrease the activity of the chemicals containing the biophore. Some values and coefficients are localized to particular atoms of a chemical (e.g., a charge or HOMO coefficient on an individual atom as calculated by a modified Hückel method). The biophore and identified modulators are then used to derive local QSAR equations for chemicals within this subset. If the entire learning set is congeneric, then the single biophore and associated modulators may explain the activity of the entire set; this usually does not occur with a large ensemble of chemicals and there is typically a group of molecules not explained by the single biophore and associated modulators. When this happens, the program will remove from consideration the molecules already explained and will search for the next biophore. The process is iterated until all of the active molecules in the learning set have been explained or until no significant fragments can be found to explain them.

The resulting list of biophores can then be used in mechanistic studies or to predict the activity of yet untested molecules.[14] For example, upon submission for evaluation, MCASE will determine if an unknown molecule contains a biophore. If the molecule does not contain a biophore, it will be predicted, by default, to be inactive. When the molecule contains a biophore, the program will make a qualitative prediction that the chemical is biologically active with an associated probability that this prediction is correct. Moreover, MCASE will inspect the molecule for the presence of modulators associated with this biophore. The program then incorporates the parameters for the identified modulators into the QSAR equation and produces a quantitative prediction for the potency of the chemical. In essence, while biophores are the determining structures for an activity call, the modulators will determine whether and to what extent the biological potential of the chemical containing the biophore is expressed.

Learning sets

We developed three models from DTP data for analysis: one each based on $-\log(GI_{50})$ values for MCF-7 and MDB-MB-231 cells and one differential activity model for compounds that were more potent growth inhibitors to MCF-7 than to MDA-MB-231 cells. In actuality, these were compounds that showed high potency to MCF-7 and minimal activity against MDA-MB-231 cells. The reporting of GI_{50} values in the DTP required the development of several selection rules in order to sample the data accurately. Only compounds with GI_{50} results reported in molar units were considered. For compounds tested at more than one concentration range, the lowest concentration range was used. GI_{50} values were transformed to their negative log values (hereafter simply referred to as GI_{50}). Many compounds in the DTP are inactive at their maximum test concentration of (e.g., 10^{-4} M). In such instances, the compounds are reported with a default GI_{50} value equal to the maximum concentration tested. Since a true GI_{50} value is unavailable for these compounds they were not considered for inclusion in the models. The rationale for not considering these compounds further is similar to the rationale utilized by the DTP for not testing these compounds at higher concentration. Since our goal is to identify attributes useful for drug discovery we focused our analyses on compounds that showed measurable cytostatic activity in the standardized test ranges chosen by the DTP. Moreover, since one of the goals of this project is the development of models based on comparing the activity of compounds in two separate assays, we only could include compounds with true GI_{50} values for both cell lines, not default values. We then ordered the remaining compounds and ranked them according to their GI_{50} values. The top 50% were designated active and the bottom as inactive. For this specific exercise, we selected the 200 most potent as active compounds and the 200 least potent as inactive compounds. The GI_{50} range for the active compounds was between 8.069 and 12.55 for the MCF-7 model and 7.756 to 12.03 for the MDA-MB-231 model. Future work will entail multiple random samples from the active and inactive pools, including analysis of the utility of using compounds with no activity at the maximum test concentration (i.e., not active at the default value).

For the differential activity model, we first selected only compounds that met the above criteria in *both* cell lines regarding activity. Moreover, during the course of DTP analyses, compounds are often tested multiple times and at different concentration ranges. Typically, for each assessment, compounds are tested within a concentration range of four log units and the highest of these is reported to indicate the range used. We compared the GI_{50} values only for compounds tested in the same range (i.e., we did not mix values from different concentration ranges since we observed that even when the same chemical was tested in different concentration ranges the GI_{50} values sometimes differed significantly). We then subtracted the MDA-MB-231 GI_{50} value from the MCF-7 value, sorted the compounds on the difference and selected the 200 with the greatest difference as actives and the 200 with minimal difference as inactives. The range of differences for the active category was between 3.777 and 1.474 units. The inactive category was composed of compounds with virtually the same potency to both cell lines.

Additionally, MCASE requires single organic molecules for consideration. Prior to analysis, we removed entries that were mixtures or contained metals. Organic salts were analyzed as the free base or acid.

Results and Discussion

A 10-fold cross-validation of the MCF-7 and MDA-MB-231 models yielded correct classification rates of 81 and 84%, respectively (Table 1). The differential activity model achieved a correct classification rate of 72% (Table 1). Overall, these values indicated that our procedure was capable of generating mechanistically sound models for growth inhibition in general and for antiproliferative activity specific to MCF-7 cells. We have consistently observed that MCASE models demonstrating good predictivity are based on mechanistically sound attributes. For example, a model of *Salmonella* mutagens that correctly classified 76% of mutagens and non-mutagens not in the model is based predominately on molecular fragments depicting electrophilic or proelectrophilic moieties.[15] Likewise, models of rat and mouse carcinogens with good predictivity are predominately based on structural moieties similar to the

Table 1. Predictive performance summary for MCF-7, MDA-MB-231, and differential activity models

Model	Concordance	Sensitivity	Specificity
MCF-7	0.81	0.81	0.81
MDA-MB-231	0.84	0.81	0.87
Differential toxicity	0.72	0.63	0.82

Concordance, number of correct predictions/total number of predictions; Sensitivity, number of correct positive predictions/total number of positives; Specificity, number of correct negative predictions/total number of negatives.

'structural alerts' described by Ashby.[16,17] Predictivity is also an accepted measure for assessing the meaningfulness of models and their descriptors developed from the DTP.[4]

To judge the predictive performance of our models, we compared them to two previously developed MCASE models. One model is based on the National Toxicology Program's *Salmonella* mutagenicity database. The *Salmonella* database is derived from a standardized protocol and, more importantly, has been analyzed for reproducibility and accuracy by replicate analyses of chemicals.[18] The interlaboratory reproducibility of the *Salmonella* mutagenicity assay is 85%.[18] The other was a mouse carcinogenicity model based on the Carcinogenic Potency Database.[19] Likewise, in analyses investigating the reproducibility of rodent cancer bioassays, the reproducibility of mouse assays in the Carcinogenic Potency Database has been estimated to be 80%.[20,21] MCASE models of these two datasets yielded a correct prediction rate of 76% for the *Salmonella* mutagenicity model[22] and 70% for the mouse carcinogenicity model.[23] The MCF-7 and MDA-MB-231 models outperformed these predictive performance standards. This high rate of predictivity could be attributed to the high quality of the data, a result of the simplicity of the growth inhibition assay and the lack of interlaboratory variability. Overall, the data suggest that the predictive performance of these models is comparable to the accuracy of the assay itself. This high degree of predictivity further substantiates the view that the models are based on and are capable of describing the underlying mechanistic attributes of cytostatic activity.

The MCF-7 and MDA-MB-231 models were more accurate than the differential activity model, while achieving this accuracy with fewer structural features (i.e., model parameters). The biophores making up each of the models are listed in Tables 2–4 and illustrated in Figs 1–3. For consideration here, we only included biophores derived from more than two active compounds. The MCF-7 model consisted of 11 biophores (made from 24 fragments or expanded definitions of biophores) (Table 2 and Fig. 1) and one biophobe. Likewise, the MDA-MB-231 model consisted of 12 biophores (from 33 fragments) and 3 biophobes (Table 3 and Fig. 2). The differential activity model, on the other hand, consisted of 13 biophores (from 58 fragments) (Table 4 and Fig. 3). Thus, the differential activity model was composed of approximately twice as many features, with a large number of single-occurrence fragments not shown in Table 4 and was less accurate.

The observation that the differential activity model was less predictive may be explained by the mechanics of the validation exercises and the nature of the endpoint. The 10-fold cross-validation process consisted of the random removal of 10% of the compounds from the learning sets, generation of a model based on the remaining 90% of the compounds, and use of that model to predict the activity of the left-out 10%. The ideal situation is that removal of only 10% of the compounds for validation does not drastically alter the model. In practice, however, there are typically always a few active compounds that do not fall into any of the biophore structural classes. They are therefore explained by a unique feature, and accurate prediction of the activity of a compound with a unique biophore is difficult since by removing the compound from the learning set, the biophore is not 'discovered' for the model. This phenomena has been recently been observed and discussed by Grant et al. when developing an MCASE model for the mouse lymphoma forward mutation assay.[24]

Given the fact that the data were all gathered from the DTP and our analysis techniques were identical, we surmise that the MCF-7 and MDA-MB-231 models are describing robust routes or mechanisms to cytostatic activity where significant (i.e., large) numbers of structurally similar compounds are inducing cytostatic activ-

Table 2. Biophores associated with cytostatic activity (GI_{50}) in MCF-7 cells and the proportion of cytostatic compounds in the differential activity and MDA-MB-231 models that contained the MCF-7 biophores

Biophore	Differential activity		MCF-7		MDA-MB-231	
	A/*n*	(%, *p*- value)	*A*/*n*	(%, *p*- value)	*A*/*n*	(%, *p*- value)
1. [N –] <–11.7Å–> [N–]	31/68	(46, <0.001)	68/76	(89)	66/81	(81, 0.157)
2a. O –CH –CH_2–CH –	20/32	(63, 0.001)	39/42	(93)	39/39	(100, 0.089)
2b. O –CH –CH_2–CH_2–	3/15	(20)	9/9	(100)	13/13	(100)
2c. O^ –C –CH –CH –	2/2	(100)	5/5	(100)	4/4	(100)
3. c. =n –c. =	32/39	(82, 0.662)	19/22	(86)	13/20	(65, 0.104)
4. cH =c –c =c <– <3-O>	9/12	(75)	20/22	(91)	24/27	(89, 0.816)
5. CS –NH –N =	6/7	(86)	6/6	(100)	7/10	(70)
6. c.″ –CO –c. =c <–	13/13	(100)	13/13	(100)	12/14	(86)
7. OH –C –CH_3	4/7	(57)	9/10	(90)	6/8	(75)
8. N –CH_2 –CH–CH_2–	7/11	(64)	12/12	(100)	11/12	(92)
9. O –C –CH–C –	7/10	(70)	14/14	(100)	11/11	(100)
10. CH –CH –O –CH–N –CH =C –	14/14	(100)	4/5	(80)	1/2	(50)
11. c. =c –n =c <– <2–NH_2>	3/3	(100)	7/8	(88)	1/3	(33)
B1. Cl –c =cH –cH =c –	3/12	(25)	0/7	(0)	3/7	(43)
	151/233	(65, <0.001)	225/244	(93)	208/244	(86%, 0.015)

A, total number of active compounds with biophore; *n*, total number of compounds deriving biophore. Biophore interpretation: c, aromatic carbon; < >, attachment of electron withdrawing or electron donating group; ^, epoxide; <#-atom>, substitution on biophores at atom # with substituent. See Figure 1 for illustration of biophores.

Table 3. Biophores associated with cytostatic activity (GI_{50}) in MDA-MB-231 cells and the proportion of cytostatic compounds in the differential activity and MCF-7 models that contained the MDA-MB-231 biophores

Biophore	Differential activity		MCF-7		MDA-MB-231	
	A/*n*	(%, *p*- value)	*A*/*n*	(%, *p*- value)	*A*/*n*	(%, *p*- value)
1. O –CH –CH_2 –CH –	20/32	(63, <0.001)	39/42	(93, 0.089)	39/39	(100)
1b. N –CH –CH –CH_3	4/5	(80)	7/8	(88)	7/7	(100)
1c. O –CH_2 –C –CH –	7/11	(64)	9/10	(90)	10/10	(100)
2. O –CH –CH –	48/78	(62, <0.001)	65/75	(87, 0.766)	68/80	(85)
2b. N –CH_2 –CH –	8/13	(62)	13/15	(87, 0.411)	12/16	(75)
3. [O –] <– 6.5Å–> [OH–]	18/36	(50, <0.001)	44/51	(87, 0.230)	44/47	(94)
4. [N –] <– 6.5Å–> [N=]	6/11	(55)	10/15	(67)	13/14	(93)
5. CH_3 –O –c =c <–cH= <3–c=>	9/11	(82)	20/22	(91, 0.894)	23/25	(92)
6. n =c –C =N –	4/5	(80)	7/8	(88)	7/7	(100)
7. CH =C –NH –c. =	5/8	(63)	9/10	(90)	9/9	(100)
8. CH_2 –CO –CH –	0/2	(0)	5/5	(100)	6/6	(100)
9. CO –C –OH =	26/32	(81)	16/18	(89)	6/6	(100)
10. CH –CH –CH –C –	1/2	(50)	6/7	(86)	6/6	(100)
11. CH_3 –O –C =CH –	2/2	(100)	4/4	(100)	6/6	(100)
12. CO –N –c. =	8/11	(73)	10/12	(83)	8/8	(100)
B1. NH =C –	7/15	(47)	0/5	(0)	0/6	(0)
B2. n =c –c. = <2–NH >	1/2	(50)	1/2	(50)	0/5	(0)
B3. CH″ –N –CH – <2–CO >	17/17	(100)	3/11	(27)	0/5	(0)
	166/257	(65%, <0.001)	264/300	(88%, 0.081)	264/286	(92)

A, total number of active compounds with biophore; *n*, total number of compounds deriving biophore. Biophore interpretation: c, aromatic carbon; <>, attachment of electron withdrawing or electron donating group; ^, epoxide; <#-atom>, substitution on biophores at atom # with substituent. See Figure 2 for illustration of biophores.

Table 4. Biophores associated with differential cytostatic activity (GI_{50}) in between MCF-7 and MDA-MB-231 cells and the proportion of cytostatic compounds in the MCF-7 and MDA-MB-231 models that contained differential activity biophores

Biophore	Differential activity		MCF-7		MDA-MB-231	
	A/*n*	(%, *p*- value)	*A*/*n*	(%, *p*- value)	*A*/*n*	(%, *p*- value)
1. [N –] <– 4.0Å–> [N=]	35/37	(95)	16/20	(80, 0.087)	4/15	(27, <0.001)
2. [O –] <– 7.9Å–> [NH–]	40/44	(91)	24/32	(75, 0.060)	23/33	(67, 0.017)
3. N –CH –O –CH –CH_2 –	21/21	(100)	4/13	(31)	1/8	(13)
4a. CO –C =C –CH_3	8/8	(100)	4/4	(100)	3/6	(50)
4b. CO –c =c –OH	5/5	(100)	2/3	(67)	3/3	(100)
5. NH_2–c =cH –cH =	9/9	(100)	1/4	(25)	1/5	(20)
6. c.″ –CO –c. =c –	13/13	(100)	13/14	(93)	12/15	(80)
7. CH =C –C =CH –	6/6	(100)	2/4	(50)	2/8	(25)
8. OH –CH –CH –c=	8/9	(89)	7/7	(100)	5/5	(100)
9. cH =c –c <=c <– <2–O>	14/16	(88)	24/36	(67, 0.118)	24/27	(89, 0.891)
10. NH –c =cH –c <=	4/5	(80)	1/1	(100)	0/1	(0)
11. O^ –CH_2 –	5/5	(100)	10/10	(100)	8/9	(89)
12. CH_2 –CH –CH –CH_2–CH –	11/11	(100)	4/8	(50)	3/5	(60)
13. CO –C =C. –	6/6	(100)	5/5	(100)	4/6	(67)
	185/195	(95)	117/161	(73, <0.001)	95/145	(66, <0.001)

A, total number of active compounds with biophore; *n*, total number of compounds deriving biophore. Biophore interpretation: c, aromatic carbon; <>, attachment of electron withdrawing or electron donating group; ^, epoxide; <#-atom>, substitution on biophores at atom # with substituent. See Figure 3 for illustration of biophores.

ity through common mechanisms. On the other hand, the differential activity model is likely describing more elusive (or unique) and cell line-specific routes where a significant number of structurally diverse compounds are inducing cytostatic activity above and in excess to the more general routes. Although these minor biophores may in fact be associated with specific antiproliferative activity only in MCF-7 cells, they are at this juncture difficult to assess. Therefore, we chose to only consider biophores found in five or more compounds per learning set for the remainder of this exercise.

In order to assess model and biophore similarity, we compared the number of active compounds associated with biophores for each of the models and learning sets. Comparison of the biophores between these sets provided a measure of model similarity and, ultimately, biological similarity. It is expected that biophores developed for a particular learning set describe a high proportion of active compounds for that endpoint. On the other hand, it would not be expected that these same biophores would be associated with a significant proportion of active compounds for another endpoint unless the two endpoints were related.

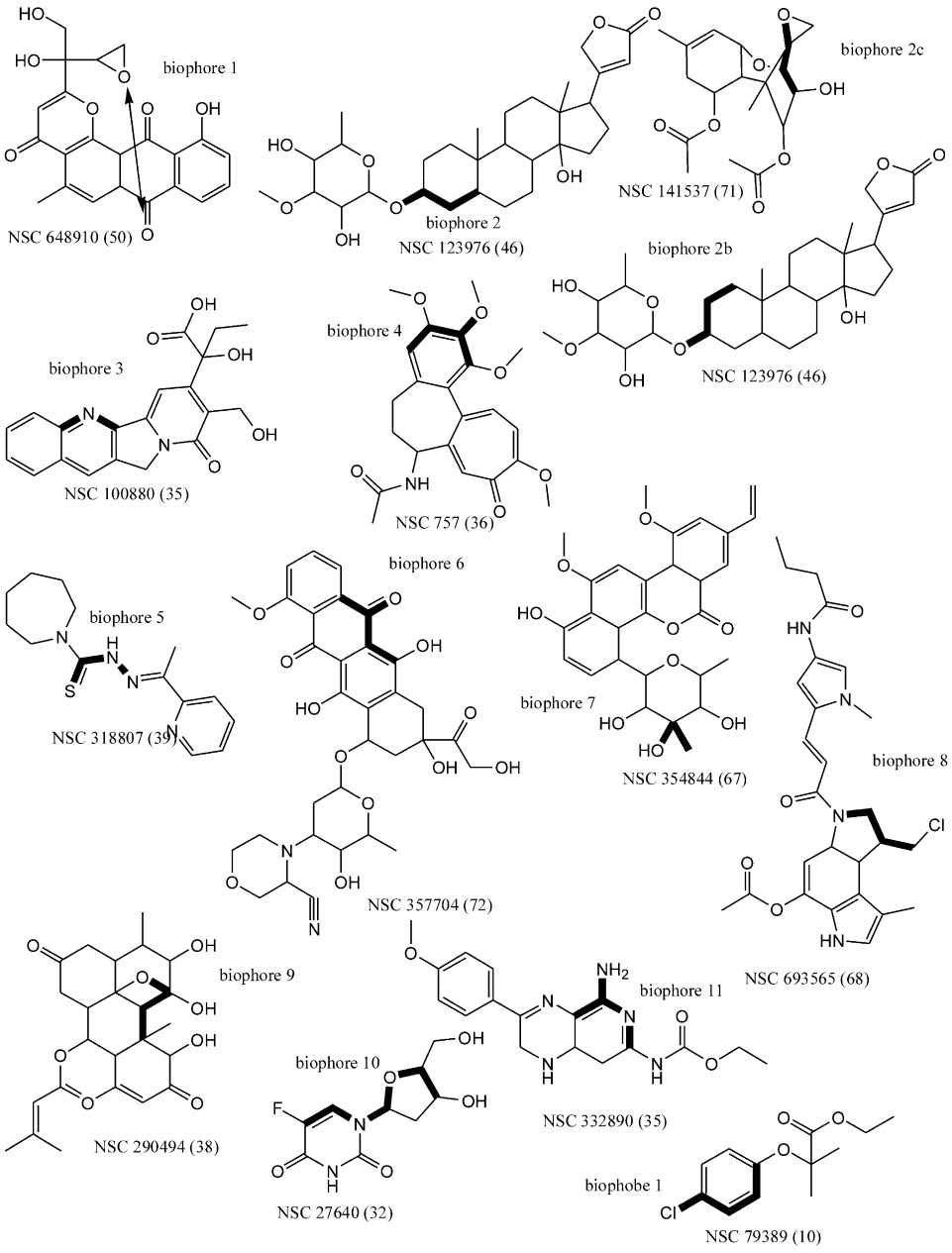

Figure 1. Illustration of biophores associated with cytostatic activity in MCF-7 cells.

Considering the sum total of biophores developed from greater than five compounds per learning set, 93 and 92% of the compounds in the MCF-7 and MDA-MB-231 learning sets that contained them were active (Tables 2 and 3). Of the MCF-7 compounds identified as having MDA-MB-231 biophores, 88% were active (Table 3). Likewise, of the MDA-MB-231 compounds identified as having MCF-7 biophores, 86% were active (Table 2). Overall, in terms of the major biophores, the structural components of the MCF-7 model were not significantly different ($p = 0.001$) from those of the MDA-MB-231 model ($p = 0.015$) (Table 2), nor vice versa ($p = 0.081$) (Table 3).

On the other hand, of the MCF-7 compounds identified as having differential activity biophores, 73% were active (Table 4) and only 66% of the corresponding MDA-MB-231 compounds were active (Table 4). Therefore, in terms of the major biophores, the MCF-7 and MDA-MB-231 models were significantly different as gauged by the differential activity model ($p < 0.001$) (Tables 2 and 3) and vice versa ($p < 0.001$) (Table 4). In other words, the structural attributes associated with antiproliferative activity toward both cell lines were different from those that contribute to additional cytostatic activity only to MCF-7 cells.

It is noteworthy from a model validity standpoint that the differential activity model was more closely related to the MCF-7 model than to the MDA-MB-231 model. This was consistent with our compound selection procedure in that the pool of compounds used to select for the residual activity model are by definition a subset of the MCF-7 pool.

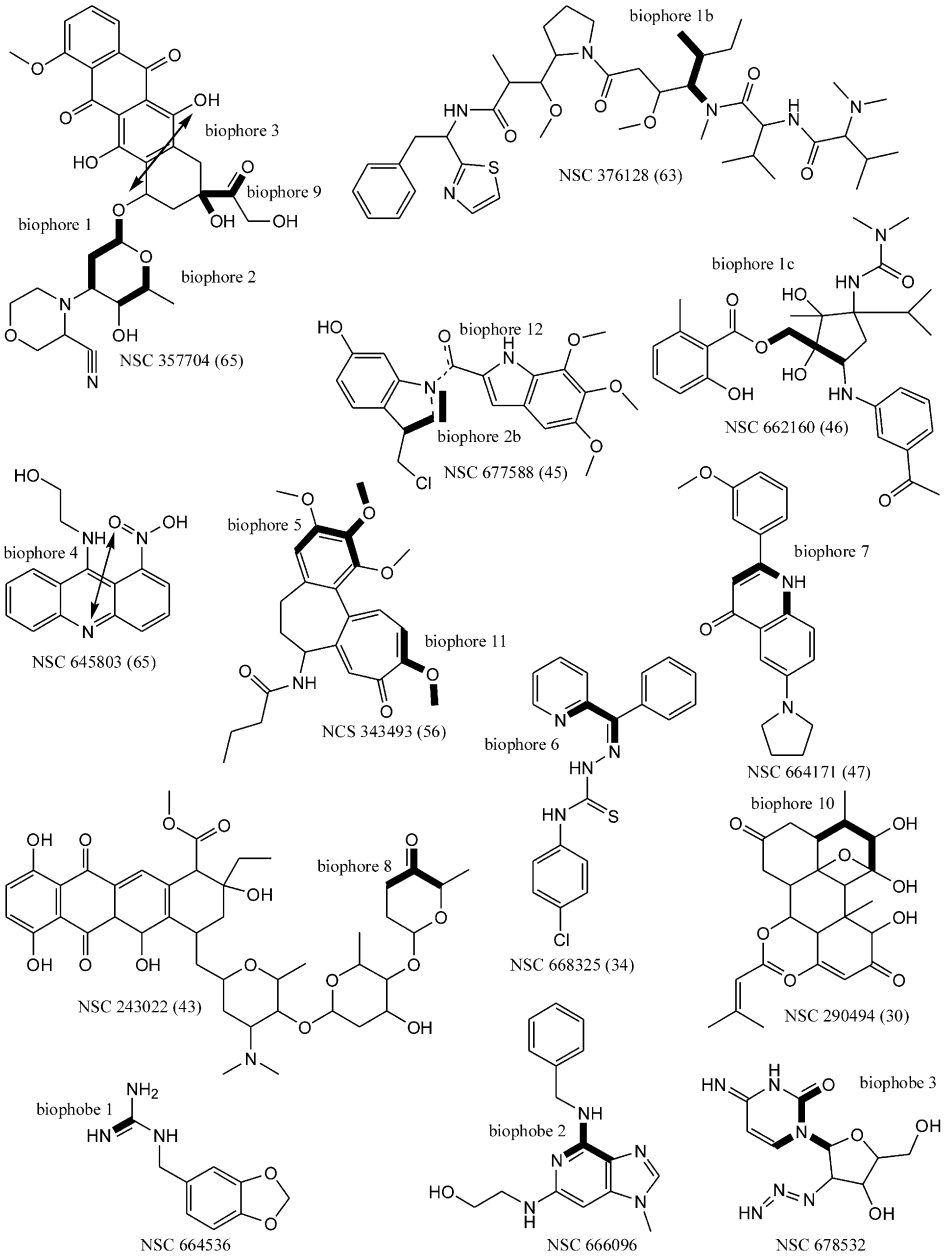

Figure 2. Illustration of biophores associated with cytostatic activity in MDA-MB-231 cells.

Analysis of individual biophores gives an indication as to which of the structural features are associated with general cytostatic activity (i.e., activity in both cell lines), as opposed to features associated only with excess antiproliferative activity in MCF-7 cells. It deserves mentioning, but perhaps goes without stating, that our modeling approach is not perfect. Consider the case of MCF-7 biophore 4 (Fig. 1), MDA-MB-231 biophore 5 (Fig. 2) and the differential activity biophore 9 (Fig. 4). They all approximately depict 2,3,4-hydroxy or-methoxy substituted aromatic rings. There was no significant difference in the percentage of active compounds containing this biophore in the three learning sets for the differential activity model compared to MCF-7 ($p = 0.118$) and for the differential activity model compared to MDA-MB-231 ($p = 0.891$) (Table 4). This indicated a requirement to compare the differential activity model to 'control-type' models for verification of attributes associated with excess antiproliferative activity in MCF-7 cells.

A number of biophores associated with either MCF-7 or MDA-MB-231 cytostatic activity were not associated with a comparable percentage of active differential activity compounds containing the biophores. This indicated that on a structure-by-structure basis the differential activity model was separable from both the MCF-7 and MDA-MB-231 models. Moreover, biophores could be identified that were not cell line-specific but rather were describing general cytostatic activity (i.e., common activity in both cell lines). For example, MCF-7 biophore 2a (Fig. 1) was identical to MDA-MB-231 biophore 1 (Fig. 2) and there was no significant difference between the number of active compounds

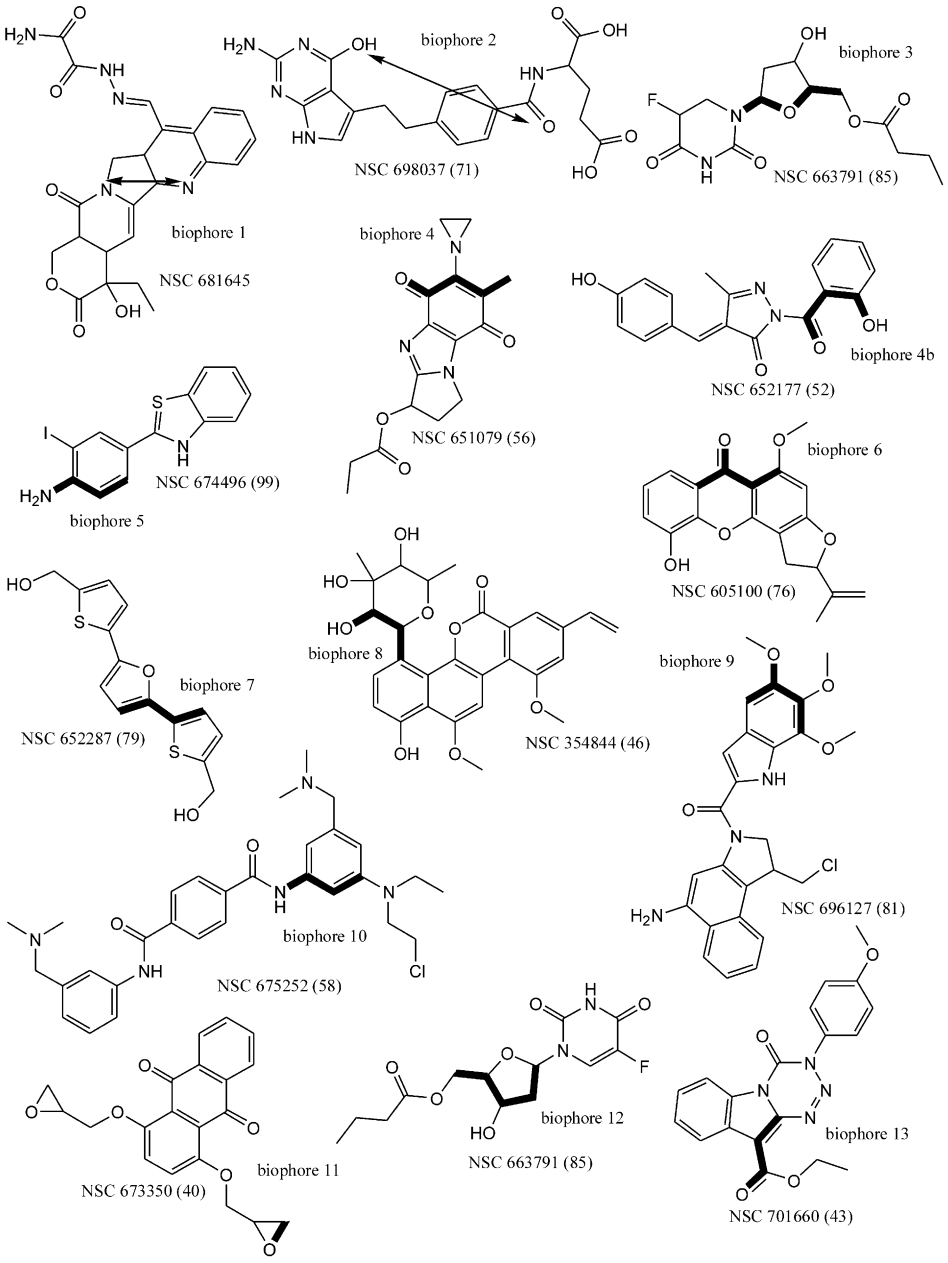

Figure 3. Illustration of biophores associated with differential cytostatic activity between MCF-7 and MDA-MB-231 cells.

containing the feature in either learning set ($p = 0.089$) (Tables 2 and 3). However, only 20/32 (63%) of the differential activity compounds with this biophore were active, which represented a significant difference ($p < 0.001$) (Tables 2 and 3). This comparison indicated that structural features associated with general cytostatic activity in MCF-7 cells were similar to those inducing similar activity in MDA-MB-231 cells.

Of most interest is that the differential activity model identified structural attributes highly associated with potent antiproliferative activity in MCF-7 cells that were *not* present in the model derived from MDA-MB-231 cells. These MCF-7-specific biophores therefore were associated with the targeting of only MCF-7 cells and seemed to be innocuous to MDA-MB-231 cells. The most significantly different biophore was differential activity biophore 1 (Fig. 3). This biophore described a calculated 2-dimensional distance of 4 Å between two nitrogen atoms. This biophore was associated with a 95% chance of inducing excess cytostatic activity in MCF-7 cells with a significantly lower chance, only 27%, of inducing cytostatic activity in MDA-MB-231 cells ($p < 0.001$) (Table 4). Likewise, the differential activity biophore 2 had a 91% chance of targeting only MCF-7 cells with a significantly lesser chance of 67% of inducing cytostatic activity in MDA-MB-231 cells (Table 2).

Other differential activity biophores were also found to be highly associated with cytostatic activity against MCF-7 cells but with minimal activity toward MDA-MB-231 cells (see Table 4). However, since less than 15 members of the MCF-7 and MDA-MB-231 learning

sets contained these biophores, the level of statistical significance was not calculated. For example, these included biophore 3, which was found in 21 potent MCF-7 cytostatic agents and no inactive ones (i.e., 100%). Yet, only 31% and 13% of compounds active in the MCF-7 or MDA-MB-231 learning sets that had this biophore are potent cytostatic agents. The differential activity biophore 5 was another example of a structural attribute associated only with potent cytostatic activity in MCF-7 cells (i.e., 100%) and minimally associated with general cytostatic activity in MCF-7 (i.e., 25%) and MDA-MB-231 (i.e., 20%).

To further our understanding of the differential activity model we selected three of its biophores for further study. These were biophores 1, 3, and 5 (Table 4). We selected these because they had a large difference between the percentage of differentially active to solely active in MCF-7 and MDA-MB-231.

The differential activity biophore 1 had significant differences ($p < 0.001$) in percent active compounds of 95% and 27% between compounds in the differential activity model and those in the MDA-MB-231 model, respectively (Table 4). The most striking characteristic of the compounds that went into the derivation of this biophore is the large number of camptothecin analogues. Recently, Fan et al. have published models around the topoisomerase I inhibitory potential of this family of agents. These were a mean (i.e., across the 60 cell lines) GI_{50} QSAR model of 167 camptothecin analogues[8] and a DNA cleavage potency model of 18 camptothecin analogues.[25] These models by Fan compared either the mean potency of the agents across the 60 cells lines[8] or very specific receptor binding site values.[25]

We searched the DTP database for the key word 'camptothecin' and identified 57 test results for these compounds. When considering only compounds tested in both MCF-7 and MDA-MB-231 cells and only the lowest concentrations of those tested, results from 36 compounds remained. Of these, all were more potent growth inhibitors of MCF-7 than MDA-MB-231 cells except for one with equal activity.

The differential activity biophore 3 also had a large difference in percent active compounds between compounds in the differential activity model and those in the MDA-MB-231 model [i.e., 100 and 13%, respectively (Table 4)]. As with biophore 1, we could readily identify an underlying common mechanism for members of the biophore group. These are antimetabolites and some have been used or tested as breast cancer therapies. Administration of tamoxifen enhanced the cytotoxicity of 5-fluorouridine in both MCF-7 and 47-DN cells.[26] More recently, Yamamoto and colleagues have suggested the use of fluoropyrimidines with docetaxel as a means to produce both antitumor and anticachectic effects through lower IL-6 levels.[27]

Unlike other breast cancer cell lines, differential activity biophore 5 (Table 3) showed the same pattern of potency to MCF-7 cells as did differential activity biophores 1 and 3. However, unlike biophores 1 and 3, we were unable to identify any studies using them as cancer treatments.

Conclusions

This series of exercises demonstrated that it is possible to build meaningful SAR models that describe selective biological activity to one cell line and not to another closely related one. Since the models developed for antiproliferative activity in MCF-7 and MDA-MB-231 cells were very similar, we suspect that they describe general routes to antiproliferative activity. However, the fact that they are not 100% identical suggests that each also possesses attributes of mechanisms to cytostasis not present in the other. We suspect that the structural features associated with differential activity are therefore inducing antiproliferative activity by a mechanism found only in MCF-7 cells and not MDA-MB-231 cells.

It did not escape us that none of the compounds of biophores 1, 3, nor 5 had an obvious association with the ER although they have been employed as antibreast cancer agents. These observations do lay open the question as to why these families of compound with antitumor (and antibreast tumor) significance are more potent against ER positive cell lines than ER negative ones. Further study of these compounds using more detailed QSAR methods may give clues to their possible influence on the estrogen receptor, its associates, and their utility in ER positive breast cancers.

Also of significance is that although we dichotomized on only one obvious trait (i.e., ER sensitivity) it might be possible to isolate other characteristics by comparing multiple cell lines simultaneously. The preliminary observations that the mean graphs had a consistent pattern across several ER positive and ER negative cell lines leaves open the possibility that a sufficient number of compounds may be obtained for modeling when divergent activity in multiple cells lines is considered (e.g., potent in lines A and B and weak in lines B, C, and D).

The outgrowth of the method we report here may facilitate the development of SAR and QSAR models with sufficient resolution to identify chemical moieties associated with antiproliferative activity that are highly selective to individual cancer types. This could be accomplished by selecting agents with antiproliferative activity in one cell line that are conversely innocuous to multiple cell lines (instead of just the one cell line used as an example in this exercise). These structural features would then have the potential to serve as pharmacophores wherein compounds that possess them would target only one cell type. Therefore, these basic units could be used to develop new and highly selective cancer chemotherapies.

References and Notes

1. Monks, A.; Scudiero, D. A.; Skehan, P.; Shoemaker, R.; Paull, K. B.; Vistica, D.; Hose, C.; Langley, J.; Cronise, P.; Viagro-Wolff, A.; Gray-Goodrich, M.; Campbell, H.; Mayo, J.; Boyd, M. R. *J. Natl. Cancer Inst.* **1991**, *83*, 757.
2. Boyd, M. R. In *Anticancer Drug Development Guide*; Teicher, B. A., Ed.; Humana: Totowa, NJ, 1997; p 23.
3. Paull, K. B.; Hamel, E.; Malspeis, L. In *Cancer Chemotherapeutic Agents*; Foye, W. O., Ed.; American Chemical Society: Washington, DC, 1995.
4. Shi, L. M.; Fan, Y.; Myers, T. G.; O'Connor, P. M.; Paull, K. D.; Friend, S. H.; Weinstein, J. N. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 189.
5. Paull, K. B.; Shoemaker, R. H.; Hodes, L.; Monks, A.; Scudiero, D. A.; Rubinstein, L.; Plowman, J.; Boyd, M. R. *J Natl. Cancer Inst.* **1989**, *81*, 1088.
6. Paull, K. B.; Lin, C. M.; Malspeis, L.; Hamel, E. *Cancer Res.* **1992**, *52*, 3892.
7. Paull, K. B.; Hamel, E.; Malspeis, L. *National Cancer Institute Developmental Therapeutics Program* **2002**.
8. Fan, Y.; Shi, L. M.; Kohn, K. W.; Pommier, Y.; Weinstein, J. N. *J. Med. Chem.* **2001**, *44*, 3254.
9. Yu, L. J.; Matias, M.; Scudiero, D. A.; Hite, K. M.; Monks, A.; Sausville, E. A.; Waxman, D. J. *Drug Metab. Dispos.* **2001**, *29*, 304.
10. Kirschmann, D. A.; Seftor, E. A.; Nieva, D. R.; Marino, E. A.; Hendrix, M. J. *Breast Cancer Res Treat.* **1999**, *55*, 127.
11. Klopman, G. *J. Am. Chem. Soc.* **1984**, *106*, 7315.
12. Klopman, G. *Quant. Struct. Act. Relat.* **1992**, *11*, 176.
13. Klopman, G.; Rosenkranz, H. S. *Mutat. Res.* **1994**, *305*, 33.
14. Rosenkranz, H. S.; Cunningham, A. R.; Zhang, Y. P.; Claycamp, H. G.; Macina, O. T.; Sussman, N. B.; Grant, S. G.; Klopman, G. *SAR QSAR Environ. Res.* **1999**, *10*, 277.
15. Rosenkranz, H. S.; Klopman, G. *Mutat. Res.* **1990**, *228*, 1.
16. Ashby, J.; Tennant, R. W. *Mutat. Res.* **1991**, *257*, 229.
17. Ashby, J.; Paton, D. *Mutat. Res.* **1993**, *286*, 3.
18. Piegorsrch, W. W.; Zeiger, E. Statistical Methods in Toxicology. In *Statistical Methods in Toxicology*, Hotham, L., Ed.; Springer: Heidlberg, 1991; Vol. 43; p 35.
19. Gold, L. S.; Slone, T. H.; Manley, N. B.; Backman, G. M.; Garfinkle, G. B.; Rohrbach, L.; Ames, B. N. In *Handbook of Carcinogenic Potency and Genotoxicity Databases*; Gold, L. S., Zeiger, E., Eds.; CRC: Boca Raton, 1997; p 1.
20. Gold, L. S.; Wright, C.; Bernstein, L.; deVeciana, M. *JNCI* **1987**, 78.
21. Gold, L. S.; Sloan, T. H.; Ames, B. N. In *Handbook of Carcinogenic Potency and Genotoxicity Databases*; Gold, L. S., Zeiger, E., Eds.; CRC: New York, 1997; p 661.
22. Zeiger, E.; Ashby, J.; Bakale, G.; Enslein, K.; Klopman, G.; Rosenkranz, H. S. *Mutagenesis* **1996**, *11*, 471.
23. Cunningham, A. R.; Rosenkranz, H. S.; Zhang, Y. P.; Klopman, G. *Mutat. Res.* **1998**, *398*, 1.
24. Grant, S. G.; Zhang, Y. P.; Klopman, G.; Rosenkranz, H. S. *Mutat. Res.* **2000**, *465*, 201.
25. Fan, Y.; Weinstein, J. N.; Kohn, K. W.; Shi, L. M.; Pommier, Y. *J. Med. Chem.* **1998**, *41*, 2216.
26. Benz, C. C.; Cadman, E.; Gwin, J.; Wu, T.; Amara, J.; Eisenfeld, A.; Dannies, P. *Cancer Res.* **1983**, *43*, 5298.
27. Yamamoto, S.; Keurebayashi, J.; Kurosumi, M.; Kunisue, H.; Otsuki, T.; Tanaka, K.; Sonoo, H. *Cancer Chemother. Pharmacol.* **2001**, *48*, 283.