Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 16 (2008) 4052–4063
Bioorganic & Medicinal Chemistry

A structural analysis of the differential cytotoxicity of chemicals in the NCI-60 cancer cell lines

Suman K. Chakravarti[a,*] and Gilles Klopman[a,b]

[a]*Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA*
[b]*MultiCASE Inc., Suite 305, 23811 Chagrin Boulevard, Beachwood, OH 44122, USA*

Received 27 October 2007; revised 8 January 2008; accepted 14 January 2008
Available online 19 January 2008

Abstract—The primary functions of cancer chemotherapeutic agents are not only to inhibit the growth or kill the cancer cells, but to do so without eliciting unreasonable cytotoxic effects on the healthy cells and to withstand the ability of the cancer cells to develop resistance against it. This has unfortunately been proven so far to be a very difficult objective. In this perspective, the ability of small molecules (anti- tumor agents) to 'see' different cell types differently can be a key attribute. Thus the term 'differential cytotoxicity' is normally used to describe the drug's specificity. In the present paper, we have quantified differential cytotoxicity from a study of the chemicals tested in the National Cancer Institute's Developmental Therapeutics Program. The MULTICASE (Multiple Computer Automated Structure Evaluation) methodology was used to discover statistically significant structural fragments (biophores) related to the differential cytotoxicity of the compounds. We found that even small structural features often become important in this regard which is evident from the biophores that were found in structurally diverse chemicals. By utilizing the difference between the raw and normalized differential cytotoxicity indices, we found that the α,β-unsaturated carbonyl group (O=C–C=CH_2) is the major biophore associated with compounds with essentially parallel concentration profiles in the cell lines in question. These compounds have high non-normalized differential cytotoxicity but considerably low normalized differential cytotoxocity. The models developed were cross validated for their predictive ability.

1. Introduction

The National Cancer Institute's (NCI) Developmental Therapeutics Program (DTP) has screened thousands of compounds against 60 cancer cell lines derived from tumors from a variety of human tissues and organs.[1–4] These cell lines are known as NCI-60 cell lines and currently they include cells from eight melanomas, six leukemia, eight breast cancers, two prostate, nine lung, seven colon, six ovary, eight kidney, and six central nervous system cancers.[5,6] NCI screens approximately 10,000 compounds for anti-cancer activity every year against these cell lines. More than 70,000 different chemical compounds, including all common chemotherapeutics, have been screened so far and are available for download from the NCI web site.[†] Besides drug screening data, NCI has also published gene expression profiles for the untreated cell lines for thousands of genes. Inaccessibility of human tumors and normal tissue for anti-cancer compound screening has made such cell lines invaluable for anti-cancer drug discovery and early evaluation.

Free online availability of the NCI-60 screening results has made it a rich data source for cheminformatics, bioinformatics, data mining, and drug designers and helped to provide new insights in the challenging field of cancer chemotherapy. Examples include identification of molecular mechanisms for cellular drug resistance,[7–12] study of the drug sensitivity-gene expression correlations,[13] systematic variation in gene expression,[14] correlating molecular substructures with gene expression[15,16], and so on.

Drug activity profiles acquired from the growth inhibition assay performed on the panel of 60 cell lines provide important information on the mechanism of action of various compounds. Past studies have used the activity profile of a compound typically as a vector of 60 GI_{50} (50% growth inhibitory concentration)

Keywords: Differential cytotoxicity; NCI-60; MULTICASE; Anticancer.
* Corresponding author. Tel.: +1 216 310 4361; fax: +1 216 831 3742; e-mail: skc5@case.edu
† http://dtp.nci.nih.gov.

0968-0896/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2008.01.024

values.[17] Two compounds are considered acting through similar mechanisms if their profiles are similar.[18] Some studies also correlated the drug profiles with the gene expression profiles in the cell lines and discovered relevant genes.[19,20]

In the present work, we quantified the ability of anti-tumor compounds to elicit differential growth inhibitory profiles (differential cytotoxicity) among the panel cell lines and treated the resulting index as a biological property of the compounds. This index we believe is a function of the presence of certain specific structural features in the compounds. Indeed, after MULTICASE analysis[21] we identified several structural features which were significantly statistically related to this index.

Differential cytotoxicity of anti-tumor agents is of significant research interest from early on.[22–26] This attribute becomes particularly important since we do not have access to the data for the effects of these compounds on normal cell lines. If a compound does not have the ability to distinguish between the cancer cell lines while eliciting its cytotoxicity, there is a possibility that it will be toxic to normal cells as well. Although the reverse cannot be said with high confidence for the compounds with differential activity, at least they offer a possibility to narrow down the screening results. Another equally important issue is that of the cancer cells' ability to develop resistance towards chemotherapeutic agents. Cancer cells often develop unique solutions for survival by targeting key regulatory steps in the apoptotic process.[‡] Screening profiles of compounds with marked 'differential' growth inhibition and/or cytotoxicity have been of particular interest for studying mechanism of actions of compounds, as well as for the selection and prioritization of compounds for in vivo evaluation.[2]

In the past studies, a compound with differential cytotoxicity and known mechanism of action is treated as a 'seed' compound and has been used to search for other compounds in the NCI database which display similar differential cytotoxicity pattern.[22] Although this strategy helped find novel and unrelated leads acting through similar mechanisms, it does not help in explaining why the compounds exhibit differential cytotoxicity in the first place, which might enable us to discover interesting leads and design better compounds.

In order to compute an index for the differential cytotoxicity of the compounds, we have used three different concentrations (GI_{50}, TGI and, LC_{50}) for each compound as opposed to using a single inhibitory concentration (GI_{50}). Each of these three member vectors makes up a profile for a compound for one cell line. Figure 1 shows composite profiles for two different compounds. We categorize a compound as one with differential cytotoxicity if its profiles (total 60 for the cell lines) show considerable variation (**1**, Fig. 1). The index calculation methodology is a very simple one and is explained in the 'Data' section.

The MULTICASE methodology[21] was used to find the structural fragments or 'biophores' related to the differential cytotoxicity of the chemicals. The program is well suited for this purpose because it has the ability to generate very detailed structure–activity relationships from the input data. Although several biophores were found to be statistically significant, we only have listed and discussed some of the major ones in this paper. Among the identified biophores, some relate to well-known anticancer compounds. However, others were also found that were not as obvious because they appear in structurally very diverse chemicals.

2. Data

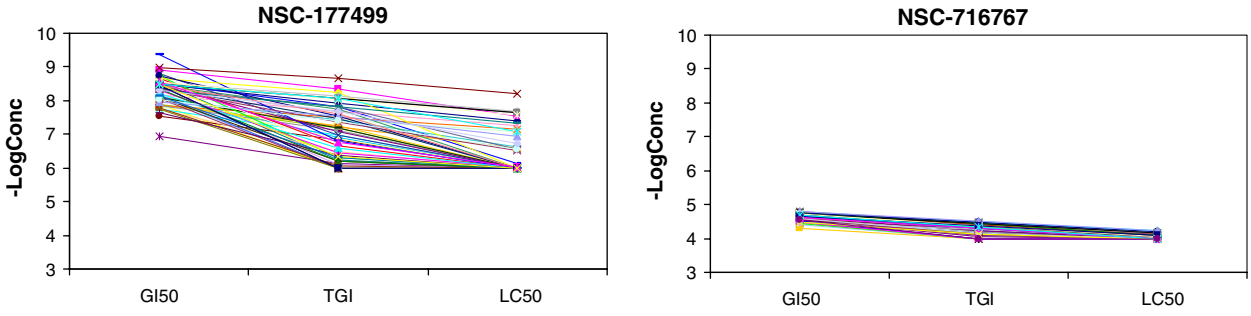
The data were obtained from the NCI web site for DTP Human Tumor Cell Line Screen[§] and downloaded in the form of three files representing three different growth inhibition concentrations; $-\log GI_{50}$ ($-\log$ of the concentration that causes 50% growth inhibition), $-\log TGI$ ($-\log$ of the concentration which causes total inhibition of cell growth/cytostatic effect), and $-\log LC_{50}$ ($-\log$ of the concentration required to kill 50% of the cells/cytotoxic effect). These files only contain the NCI's internal ID number (NSC No.) of each compound. We used these NSC numbers to cross search using the Enhanced NCI Database Browser[¶] and the SMILES codes for each of the chemicals were downloaded. The structures that could not be found with the enhanced browser were obtained from a SD file (September 2003 release), provided in the DTP web site. The raw activity and structural data were converted to a Microsoft Access® Database and, using appropriate queries, the data were finally converted to a format similar to that shown in Table 1. Compounds having elements other than the organic set (C, N, O, S, P, Cl, Br, I) and those which have been tested in less than 50 cell lines for all the three concentrations (GI_{50}, TGI, and LC_{50}) were removed. The process afforded us the required information for 26377 compounds.

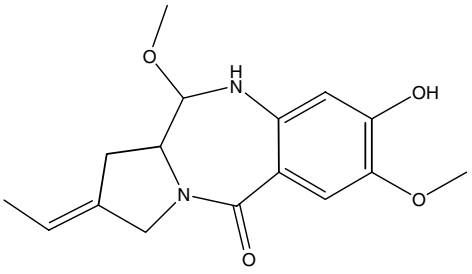
Concentration triplets ($-\log GI_{50}$, $-\log TGI$, and $-\log LC_{50}$ values) for each compound were used to calculate its differential cytotoxicity (DC) indices. For a particular compound, the purpose of the DC index is to quantify how different the concentration profiles are from one another. We have treated the three concentrations for each cell line as a point in the 3-dimensional space and calculated the Euclidean distances from each point to the others. For each compound, two slightly different versions of DCI were calculated; in the first one, the Euclidean distances over the entire cell line were averaged (raw DCI) whereas in the second one, the profiles were first normalized by subtracting the mean from each profile and then the Euclidean distances were calcu-

[‡] United States Patent 7125997, http://www.freepatentsonline.com/7125997.html.

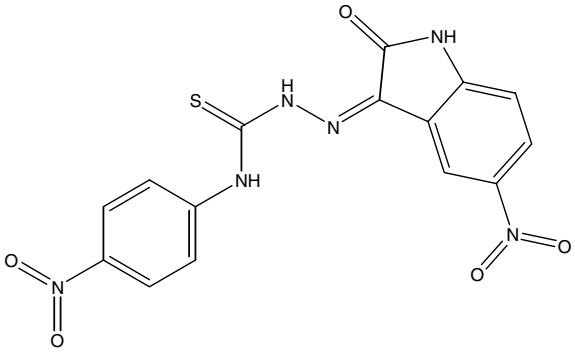
[§] http://dtp.nci.nih.gov/docs/cancer/cancer_data.html.

[¶] http://129.43.27.140/ncidb2/.





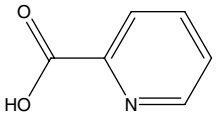
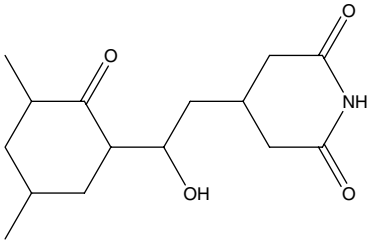
Compound **1** (**NSC-177499**)



Compound **2** (**NSC-716767**)

Figure 1. Examples of two compounds with markedly different inhibitory concentration profiles. Compound **1** has a significant differential cytotoxicity while compound **2** has low differential cytotoxicity.

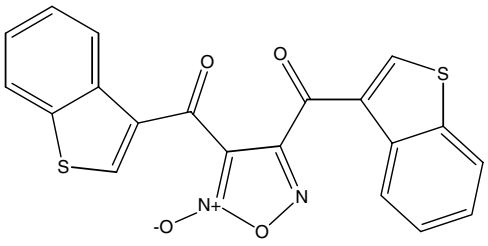
Table 1. A sample of the information content of the main database

NSC No.	Cell lines[a]	Structure	LCONC[b]	GI_{50}[c]	TGI[d]	LC_{50}[e]
89	NCI-H23		−4.00	4.209	4.131	4.053
	NCI-H522			6.159	5.977	4.978
	A549/ATCC			4.186	4.086	4.000
	.			.	.	.
	.			.	.	.
	.			.	.	.
171	NCI-H23		−4.00	4.000	4.000	4.000
	NCI-H522			4.000	4.000	4.000
	A549/ATCC			4.000	4.000	4.000
	.			.	.	.
	.			.	.	.
	.			.	.	.
.	.	.		.	.	.
.	.	.		.	.	.

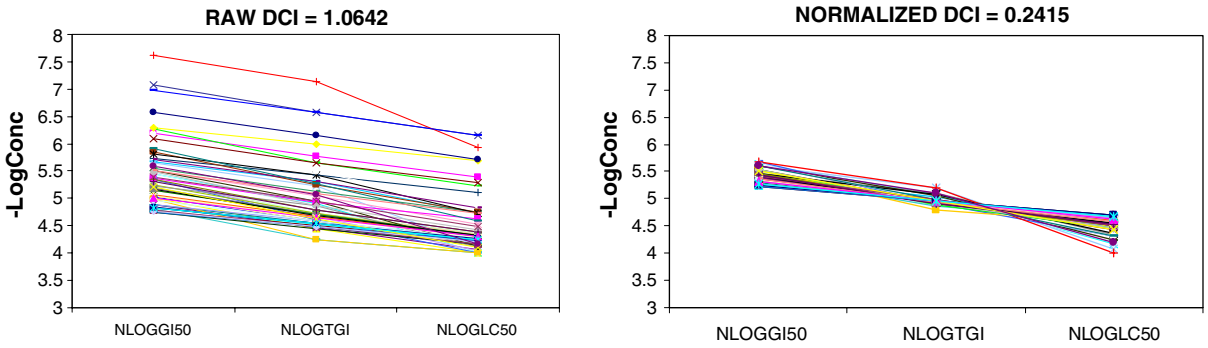
[a] Details of the cell lines can be found at the NCI web site: http://dtp.nci.nih.gov/docs/misc/common_files/cell_list.html.
[b] log of the highest molar concentration tested.
[c] −log of the concentration that causes 50% growth inhibition.
[d] −log of the concentration that causes total growth inhibition.
[e] −log of the concentration required to kill 50% of the cells.

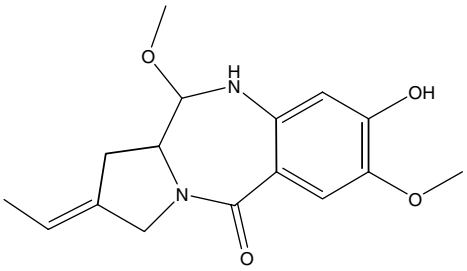
lated (Normalized DCI). The objective of the normalization is to avoid awarding high DCI values to those molecules whose profiles are parallel to each other and differ only slightly in shape and slope. This situation is documented by the examples of Figure 2. For the first compound (**1**, Fig. 2) a significant decrease in the DCI value can be seen after normalization whereas, the second one (**2**, Fig. 2) retained a high DCI value even after normalization. On the basis of the differences in the raw and normalized DCI values, the compounds in the database can be divided into three categories:

i. Compounds which retain a high DCI value after normalization,
ii. Compounds whose DCI values drops significantly after normalization and,



Compound **1** (**NSC-18877**)





Compound **2** (**NSC-177499**)

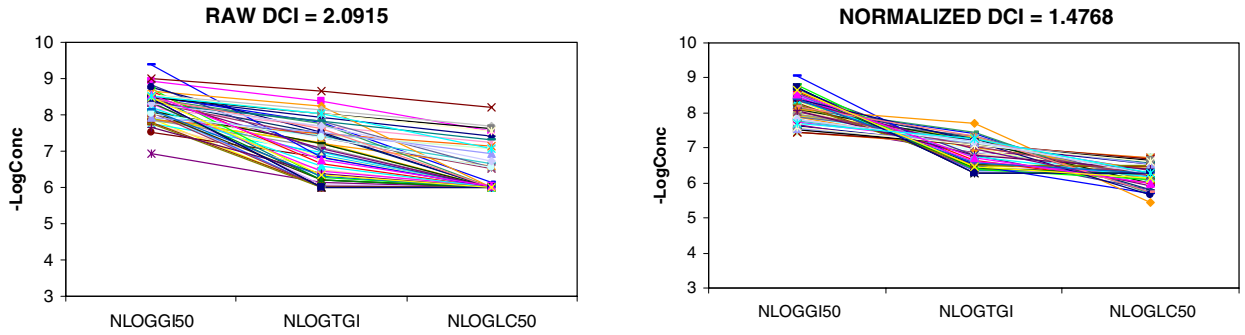


Figure 2. Effect of normalization on the differential cytotoxicity index of two compounds. Normalization results in a significant lowering of the DCI value for compound **1**, whereas compound **2** retains a high DCI even after normalization. The parallel nature of the concentration profiles is evident in the first compound.

iii. Compounds which have poor DCI value both before and after normalization.

The rationale behind the use of the second category of compounds is to find relevant structural features that are associated specifically with compounds that have essentially parallel concentration profiles, allowing us to draw a connection between mechanisms of action of cytotoxic compounds and their dose profiles. Exploiting this simple difference between the two versions of DCI might give us a crude discriminatory insight in the possible effects of the cytotoxic compounds on the complex cellular machinery.

3. Methods

3.1. MULTICASE methodology

The MULTICASE program[21,27] is based on a hierarchical statistical analysis of a database (called the training set) composed of a number of chemicals with their bio-

logical activity data. The program aims to discover biophores, that is, substructures that appear mostly in active molecules and may therefore be responsible for the observed activity. It starts by identifying the statistically most significant substructure existing within the learning set. This fragment, labeled the top biophore, is seen as responsible for the activity of the largest possible number of active molecules. The active molecules containing this biophore are then removed from the database and the remaining ones are submitted to a new analysis leading to the identification of the next biophore. This procedure is repeated until either the activity of all the molecules in the learning set have been accounted for or no additional statistically significant substructure can be found. For each set of molecules containing a specific biophore, MULTICASE identifies additional parameters, deemed modulators, which can be used in the construction of a quantitative structure–activity relationship within this reduced set of congeneric molecules. Modulators consist of the presence of certain substructures or the value of calculated parameters such as the highest occupied and lowest unoccupied orbital energies, octanol–water partition coefficient, and so on. The process is automated and proceeds with minimal human intervention. The knowledge that the program gains during the training process can then be used to predict the biological activity of new chemicals that were not included in the training set. The MULTICASE methodology and its applications have been extensively published in the scientific literature.[21,27–31]

3.2. MULTICASE database preparation

Data subsets were constructed from the full database of 26377 compounds with an objective to answer three questions:

i. What are the structural features associated with a high differential cytotoxicity?
ii. What are the structural features related to the differential cytotoxicity associated with compounds whose DCI value reduces significantly when raw DCI is transformed to normalized DCI?
iii. Key structural features that are associated with nonselective cell killing in compounds so that such structural features may be avoided in future drug design.

To answer the first question, databases were prepared with normalized DCI values treated as activity. The normalized DCI for the compounds varied from zero to 1.9644. We have categorized compounds as actives if their normalized DCI is above 0.468, marginally actives if this value is between 0.418 and 0.468 and inactive if it is below 0.418. We prepared three different databases of 6500 compounds each by randomly selecting compounds from the full set. Each of them contains ~2000 active, ~500 marginally active and ~4000 inactive compounds. After MULTICASE analysis we have considered only those biophores which appeared in two or more databases.

To answer the second question we sorted the compounds in descending order of reduction in their DCI values when transformed from raw to normalized states and labeled the top compounds (with normalized DCI equal or below 0.468) as actives. A single database of 6500 compounds was built with 433 actives and 6067 inactive compounds.

For the third query, compounds with 5.0 or more average cytotoxic concentration (negative log concentration) and with a very low normalized DCI (equal or less than 0.0166) were labeled as actives and rest of them were labeled as inactive. A database with 6500 compounds was built with 538 active molecules with high overall cytotoxicity but very low differential cytotoxicity and 5962 inactive molecules.

4. Results and discussion

We will first focus on the biophores obtained by MULTICASE when processing the three databases with normalized DCI values. For the sake of clarity and conciseness we have discussed only major biophores, that is, if they appear in two or more databases. The biophores are shown in Figure 3.

Biophore **1** (Fig. 3) appears in all three databases and statistically it is the most relevant biophore; 94 active compounds and only 2 inactive compounds carry this biophore. This biophore mainly consists of a –OH group and an –O– group separated by three aliphatic ring carbon atoms. The ring carbon next to the –OH group contains no hydrogens, the second carbon contains two hydrogens and the carbon next to the –O– group contains one hydrogen (OH–C–CH_2–CH–O–). Some minor variants of this biophore were also found which mainly differ in the number of hydrogens on the carbon atoms and sometimes nitrogen is found in place of oxygen. This biophore appears in structurally diverse molecules and some examples are shown in Figure 4. It is interesting to note that some of the well-known anti-cancer drugs happen to have this biophore, for example, paclitaxel, adriamycin, and daunorubicin. These drugs work by different mechanisms; paclitaxel is an antimitotic agent whereas adriamycin and daunorubicin interact with DNA by intercalation. These drugs are widely used in chemotherapy. At least one research group has used paclitaxel and adriamycin as reference compounds in assessing the differential cytotoxicity of fredericamycin A derivatives.[32] Another study[‖] points out that adriamycin has nearly 20 times selective inhibition towards tumor cell line MCF7 as compared to Normal Human Mammary Epithelial Cells (HMEC).

Biophore **2** was found to exist in 37 molecules with high differential cytotoxicity. The molecules that contain this biophore are mostly derivatives of antimitotic agent colchicine.

It should be mentioned here that antimitotic agents are known for their differential cytotoxicity. Paull et al.[22]

[‖] http://aiche.confex.com/aiche/2007/techprogram/P97795.HTM.

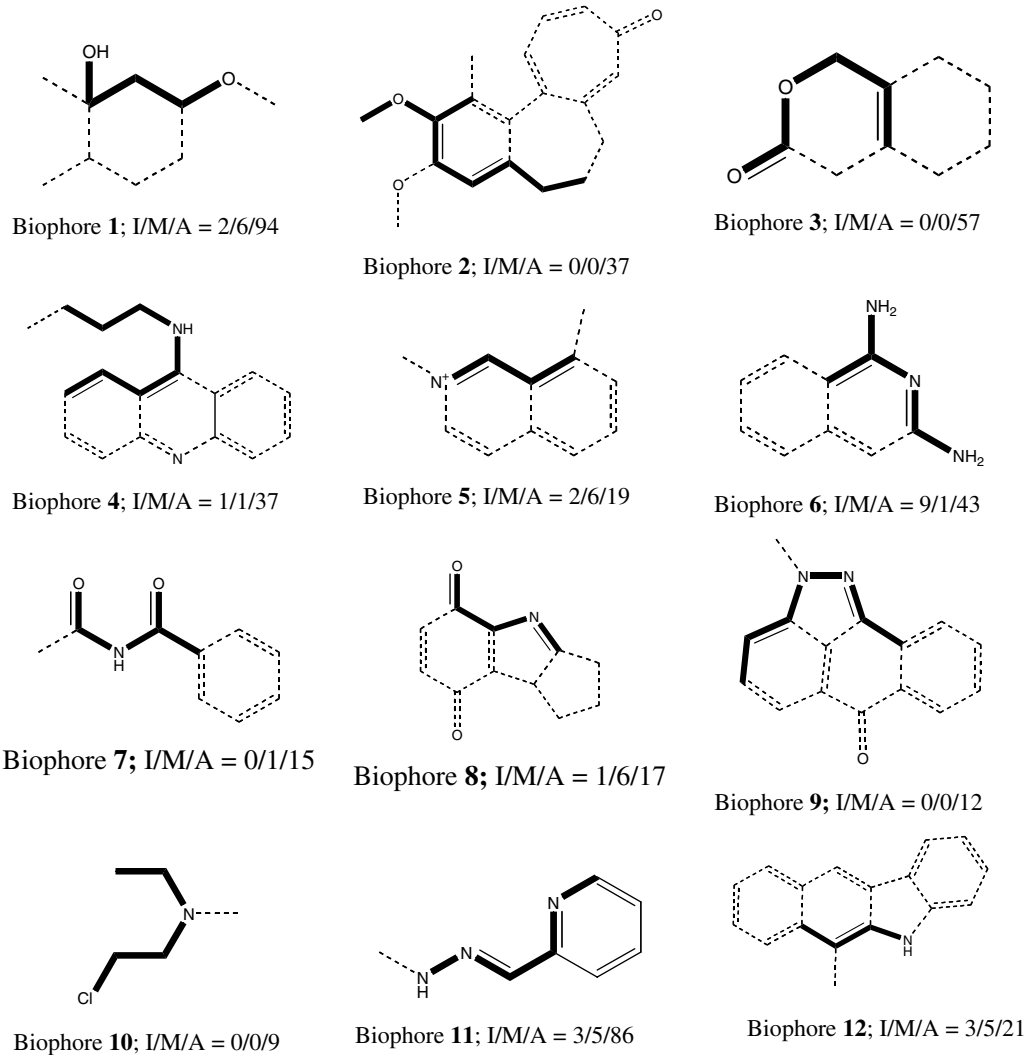


Figure 3. Major biophores associated with high normalized differential cytotoxicity (DC). I/M/A stands for distribution of the biophore in compounds with low DC, marginal DC and high DC, respectively.

mentioned in their paper that some of the possible explanations for greater sensitivity of some cell lines towards antimitotic agents may be relatively low concentrations of tubulin in certain cell lines or microtubules could have some specific function in structural integrity or metabolism of certain cell lines, rendering them hypersensitive to antitubulin agents. Also, certain cell lines could have increased levels of membrane tubulin, which could act as a specific transport mechanism for the uptake of antimitotic agents.

Biophore **3** can be a part of 5 or 6-member rings and is often found within the 2H-pyran-2-one-3,6-dihydro or 2(5H)-furanone ring system fused with another ring. Although in our databases, the majority of the compounds that contain this biophore are camptothecin derivatives, the results are certainly not limited to them. Indeed, heterocyclic benzodioxole lactones also contain this biophore. Some molecules with this biophore are shown in Figure 5. Camptothecin is an anti-cancer drug in use. Abel et al.[32] used camptothecin as a reference compound for assessing differential cytotoxicity of anti-tumor agents. Mechanism of action of camptothecin derivatives are thought to be the inhibition of the enzyme topoisomerase I.

Biophore **4** was another statistically significant biophore and was found almost exclusively within an acridine ring system. However, as it can be seen in Figure 3, the biophore spans only over two of the three rings of acridines and is therefore also found in ring systems like that of 4-quinolamine derivatives (e.g., NSC-709257).

MULTICASE identified several other statistically significant biophores; for example, Biophore **5** was found to be present mainly in two different types of molecules; some of them are shown in Figure 6; biophore **6** was found in compounds which are 2,4-pteridinediamine and 2,4-quinazolinediamine derivatives. Biophore **7** is found in benzoylurea derivatives. Biophore **8** was also found to be present in mainly two different chemical classes (Fig. 7). Biophore **10** is a part of 2-chloroethyl alkylamines and biophore **11** is found in pyridyl isonicotinoyl hydrazone derivatives.

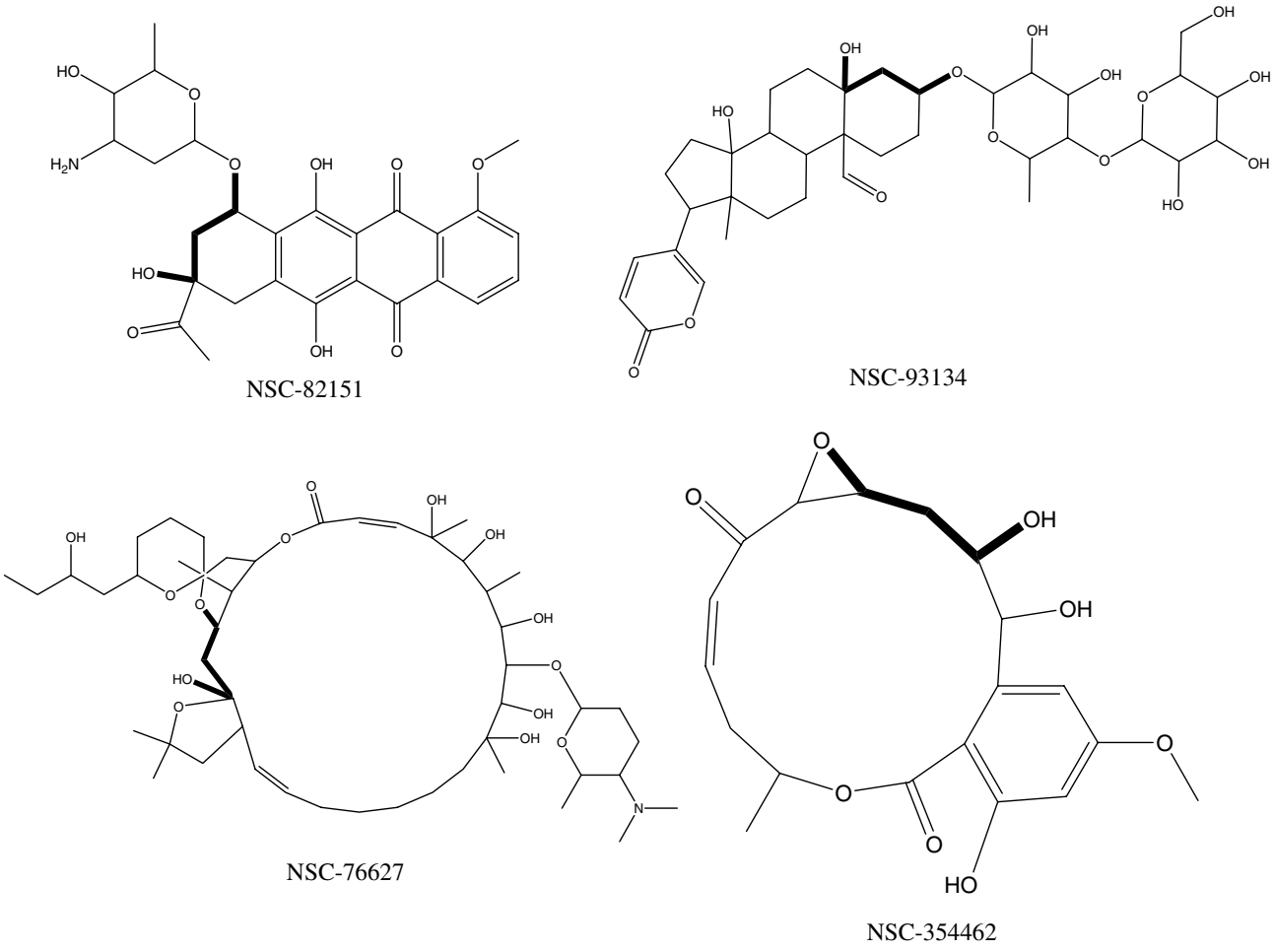


Figure 4. Some molecules containing Biophore **1** from Figure 3.

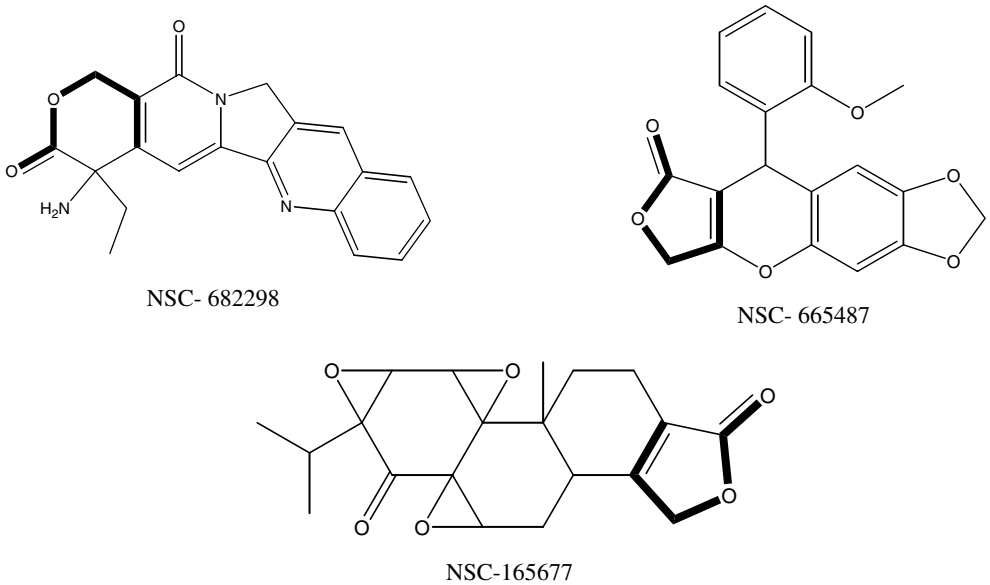


Figure 5. Some molecules containing biophore **3** from Figure 3.

As mentioned earlier, the second part of our study focuses on finding biophores that are associated specifically with compounds with DCI values that undergo marked reduction after normalization. Major biophores found by the MULTICASE analysis are shown in Figure 8. A commonality amongst the biophores in this list

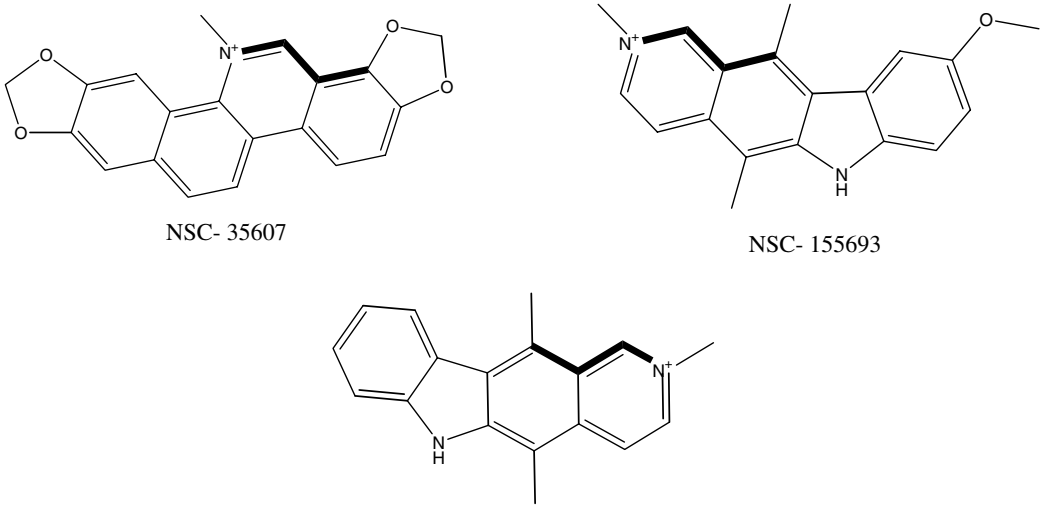


Figure 6. Some molecules containing Biophore **5** from Figure 3.

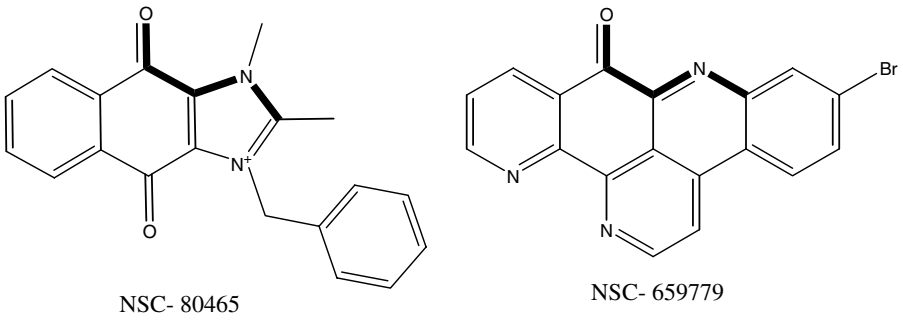


Figure 7. Some molecules containing biophore **8** from Figure 3.

is notable; majority of the biophores shown in Figure 8 contain an α,β-unsaturated carbonyl group, that is, O═C─C═CH_2 (Biophores **1**, **2**, **6**, **7**, and **8**). Biophores **4** and **5** seem to be different at first glance, but they also contain this feature. Cytotoxic activity of compounds containing an α,β-unsaturated carbonyl group is well known and many of these compounds have significant anti-tumor activity.[33,34] Cytotoxic sesquiterpene lactones are notable examples (e.g., germacranolides, elephantin, elephantopin, costunolide, and tulipinolide and many others). Lee et al.[35] showed that most immediate and direct factor responsible for cytotoxicity of these compounds is the introduction of the O═C─C═CH_2 system. It is believed that cytotoxicity of these compounds results from their chemical reactivity towards nucleophilic thiols via Michael-type conjugate addition resulting in alkylation and thereby inhibiting a variety of cellular functions and leads the cell to apoptosis.[33] They are reactive towards thiol-rich enzymes, including phosphofructokinase, glycogen synthetase, and amino acids such as cysteine. Some of the targets of this group of compounds are membrane based, for example, the reduced NADH: ubiquinone oxidoreductase in complex I, which is a membrane-bound protein of the mitochondrial electron transport system and the ubiquinone-linked NADH oxidase in the plasma membrane of cancerous cells.[36,37] Halmos et al.[38] reported that cytostatic unsaturated ketonucleosides strongly reduce membrane surface thiol levels and suggested that selective inhibition of certain key membrane thiols by these compounds might be an important event in their biological effect. α,β-unsaturated carbonyl compounds have also been shown to have inhibitory effect on membrane embedded multidrug resistance proteins (MRP1/MRP2) which belong to the ATP-binding transporter protein family. The parallel nature of the cytotoxic concentration profiles of the compounds containing this particular biophore might be the result of difference in the overall thiol levels of the NCI60 cell lines[39] or the effect of α,β-unsaturated carbonyl on the membrane transport mechanisms on the individual cell lines. However, more investigation is needed to bring certainty to this possibility.

Last part of our study involves finding structural features that are found in compounds with cytotoxicity but with negligible 'differential' cytotoxicity. These compounds kill all the cell lines non-selectively. Such features are not desirable in a chemotherapeutic agent and should be avoided in designing new drugs. A database was built to uncover such features containing 6500 compounds and 538 of them were labeled as active. Ac-

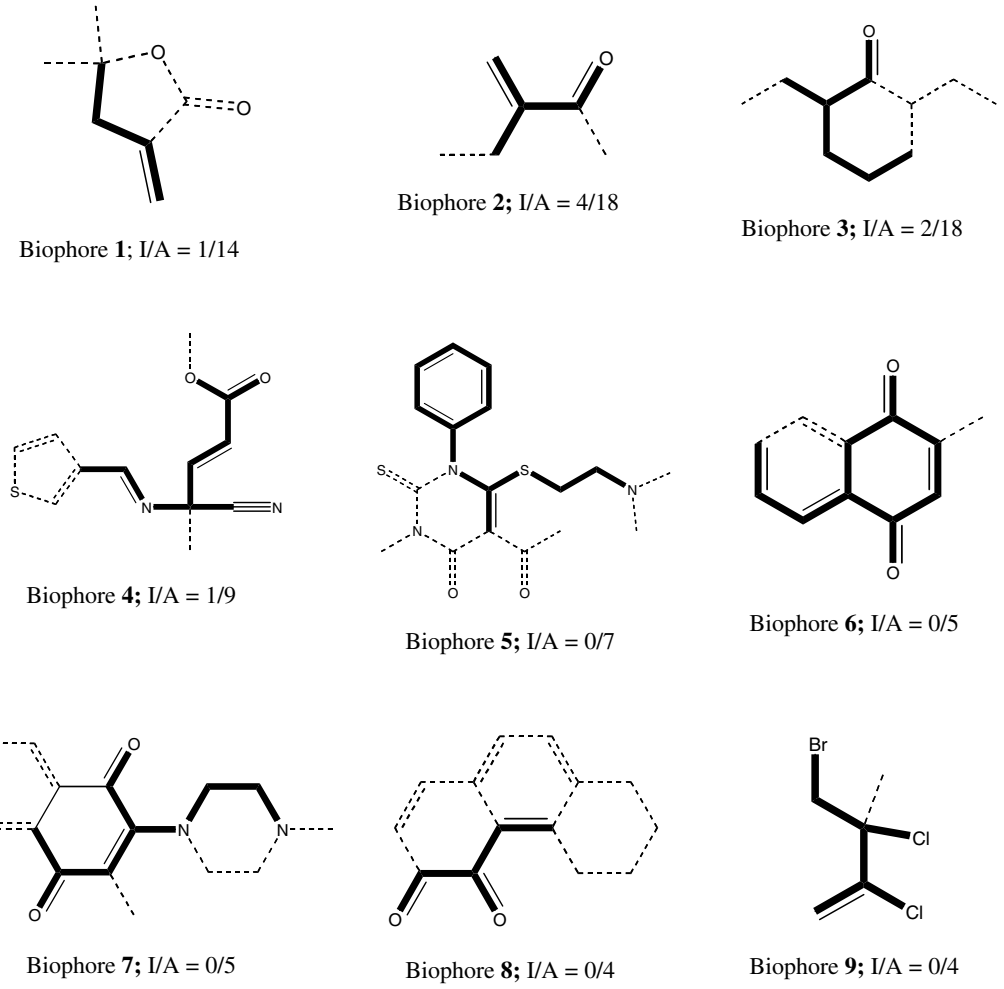


Figure 8. Major biophores which are associated only with non-normalized differential cytotoxicity (DC) index. After normalization, differential cytotoxicity of the compounds containing these biophores becomes negligible. I/A stands for distribution of the biophore in compounds that do not change their DC index after normalization and compounds whose DC index change significantly.

tive compounds are those with high cytotoxicity but with a very small differential component. The biophores found by MULTICASE analysis are shown in Figure 9. Only the top most statistically significant ones are shown. Biophore **1** occurs in diverse chemical structures and frequently is a part of side chains in those compounds. However, all the other biophores appear in restricted chemical classes represented by the biophore itself.

4.1. Validation and predictive ability

In order to test the ability of the models to predict the differential cytotoxicity potential of new molecules (molecules that are not present in the model), we have performed cross validation tests. The tests were carried out for one of the databases with normalized differential cytotoxicity index as the activity. In the cross validation tests, 3% of the compounds were taken out randomly from the total set of 6500 molecules and used as a test set. The remaining 97% of the molecules were used to rebuild the model and differential cytotoxicity potential of the 3% test molecules were predicted using this model. This procedure was repeated 3 times and the results are shown in Table 2.

It can be seen from Table 2 that an average of 10.4% test set molecules were not tested during cross validations. This is because those compounds were determined to be out of domain by the program and considered unsuitable for testing. There could be a number of reasons to classify a test compound to be out of domain. One of them is the presence of 'unknown' structural fragments in the test molecules. 'Unknown' structural fragments are those which were not found in the training set compounds. This is also a way to find structural domain that is not covered by the model.

For every 100 active compounds in the test set, an average of 57.9 compounds was predicted active by MULTICASE (sensitivity). On the other hand, for every 100 inactive compounds in the test set, averagely 92.0 compounds were predicted inactive by the program (specificity). Since our original databases were skewed towards inactive molecules with a 2:1 ratio between inactive and active molecules, a decrease in the sensitivity is expected. However, high specificity of the test results compensates for the mediocre sensitivity, that is, the program recovers 57.9% of the active molecules from the test set, nevertheless, out of this 57.9% fraction, 85.5% of compounds are true positives. These test re-

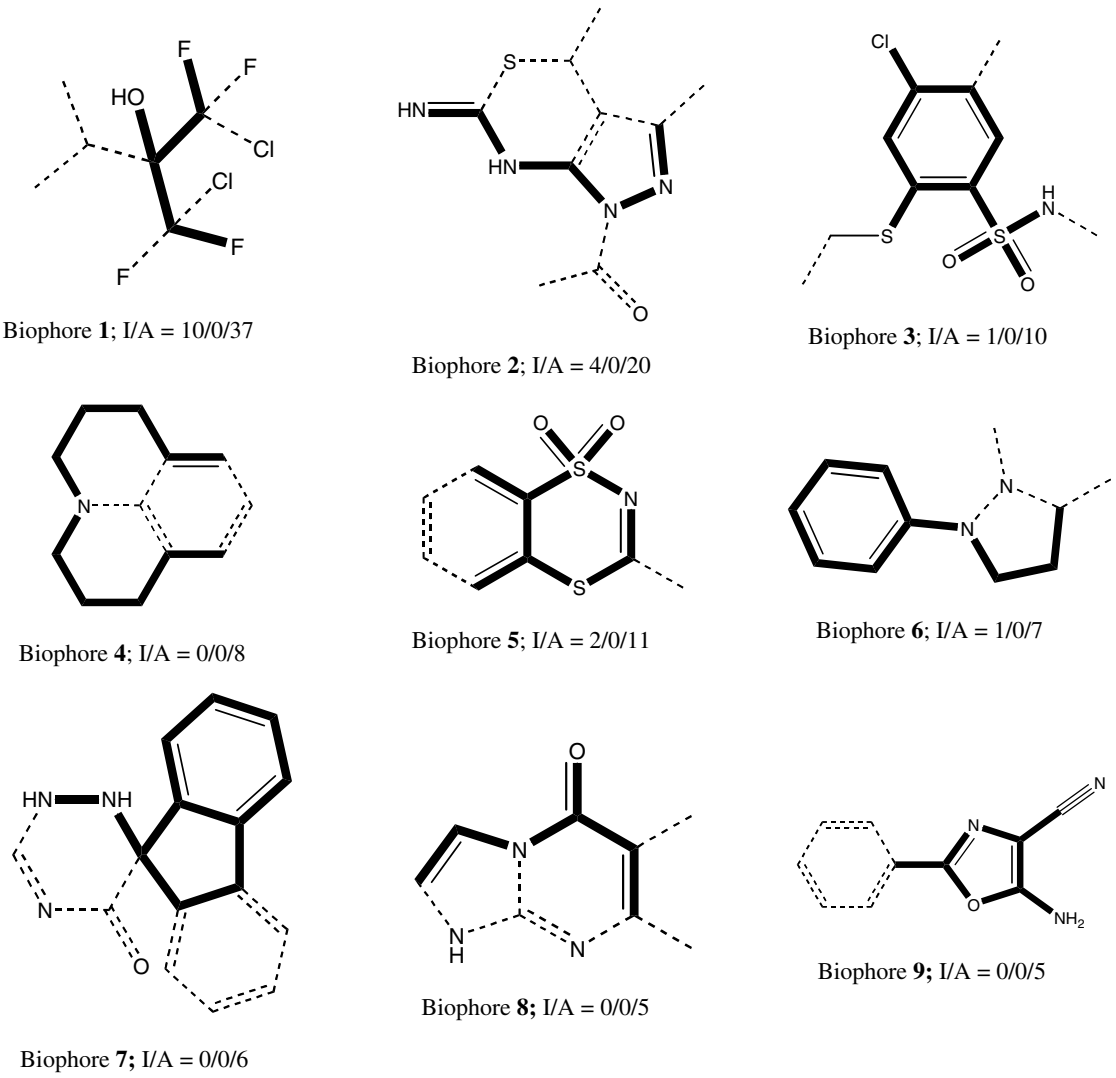


Figure 9. Major biophores associated with compounds that have non-selective cytotoxicity towards the cell lines, that is, those with negligible differential cytotoxicity. I/A stands for distribution of the biophore in compounds without non-selective cytotoxicity and with non-selective cytotoxicity, respectively.

Table 2. Results of the cross validation tests

Test no.	Concordance[a] (%)	Sensitivity[b](%)	Specificity[c] (%)	Number of test set molecules	No test[d] (%)
1	78.86	59.50	94.70	196	10.71
2	78.29	59.20	92.80	196	10.71
3	73.71	55.10	88.50	194	9.79
Average	76.95	57.93	92.00	195	10.40

[a] Concordance is a statistical measure of what is the overall correctness of the test in identifying the negative and positive cases.
[b] Sensitivity, or recall rate, is a statistical measure of how well the test correctly identifies the positive cases.
[c] The specificity is a statistical measure of how well the test correctly identifies the negative cases.
[d] Percentage of test set molecules that were not processed for prediction because the program considered them out of domain of the training set data.

sults indicate that the models developed can be used effectively to discover new molecules with desirable differential cytotoxicity profiles.

Activity prediction using a MULTICASE model for a particular compound contains a detailed account of how the prediction is reached including biophores/biophobes/modulators that were found in the test compound and a probability value which is the likelihood of the test compound to be active. If the prediction is correct (i.e., it matches with the actual experimental activity) then inactive compounds should receive low (close to zero) probability value whereas the active compounds receive high probability (close to one) values. In case of a poor prediction, this situation is reversed. A distribution of this probability value in the activity prediction of the test molecules in the three cross validation tests is shown in Figure 10. It can be

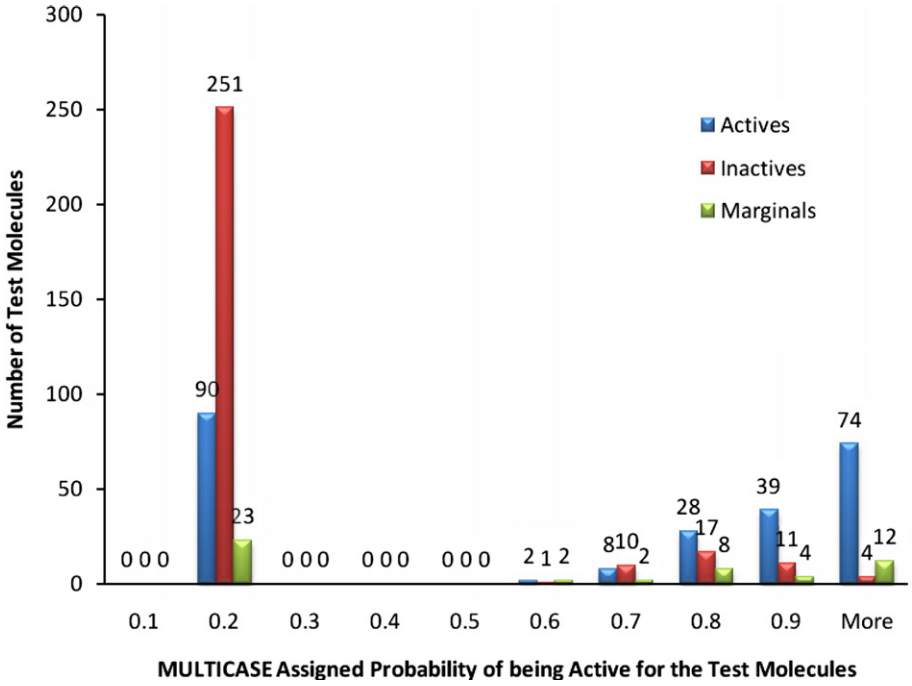


Figure 10. Distribution of MULTICASE predicted probabilities for a molecule to be active (with high differential cytotoxicity potential) for 586 test set compounds. The results are an accumulation of the three different test sets generated in the cross validation study. The colors of the bars represent the actual experimental activity category of the test compounds.

seen that majority of the inactive molecules were assigned a probability between (0.1 and 0.2) whereas majority of the active test molecules received a probability on the other end of the scale (0.7–1.0). Amongst the total of 586 test molecules, 90 of the active molecules were misclassified and therefore received low probability and grouped with inactive molecules. However, very few of the inactive molecules were misclassified and received higher probabilities. It can also be seen that there is a big gap in the distribution and only five molecules received probability more than 0.2 and less than 0.7. This gap offers a better discriminatory prediction between active and inactive molecules and a clear cutoff probability value to be used.

5. Conclusions

In this paper we have explored the idea that the ability of a chemical to exhibit differential cytotoxicity (against the panel of 60 NCI cancer cell lines) depends on certain molecular structural features. Using the MULTICASE program we have found biophores statistically related to the differential cytotoxicity index. Some biophores were found to be belonging to group of chemicals acting through more than one known mechanism of actions. In addition, we found the α,β-unsaturated carbonyl group ($O{=}C{-}C{=}CH_2$) as a major biophore related specifically to compounds which have essentially parallel concentration profiles. This biophore is well known to react with cellular targets rich in thiols and membrane proteins act as transporters. The findings of this study have the potential to aid in the design and discovery of novel anticancer agents.

Supplementary data

The list of compounds containing biophores shown in Figure 3 is provided as a supplementary file (*BphComps.txt*). The file contains NSC numbers and SMILES codes of the compounds. Also, the database of 6500 compounds (433 active compounds) which produced the biophores shown in Figure 8 is provided as a supplementary file (*Database433.txt*). This file contains NSC numbers, active/inactive flags and SMILES codes of the member compounds. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2008.01.024.

References and notes

1. Grever, M. R.; Schepartz, S. A.; Chabner, B. A. *Semin. Oncol.* **1992**, *19*, 622–638.
2. Boyd, Michael R. The NCI in vitro anti-cancer drug discovery screen: concept, implementation, and operation, 1985–1995; The NCI in vitro anti-cancer drug discovery screen: concept, implementation, and operation, 1985–1995, Div. Cancer Treatment, National Cancer Inst., Frederick Cancer Res. Devel. Center, Frederick, MD, USA.
3. Boyd, M. R.; Paull, K. D. *Drug Dev. Res.* **1995**, *34*, 91–109.
4. Monks, A.; Scudiero, D. A.; Johnson, G. S.; Paull, K. D.; Sausville, E. A. *Anti-cancer Drug Des.* **1997**, *12*, 533–541.
5. Stinson, S. F.; Alley, M. C.; Kopp, W. C.; Fiebig, H. H.; Mullendore, L. A.; Pittman, A. F.; Kenney, S.; Keller, J.; Boyd, M. R. *Anti-cancer Res.* **1992**, *12*, 1035–1053.
6. Myers, T. G.; Anderson, N. L.; Waltham, M.; Li, G.; Buolamwini, J. K.; Scudiero, D. A.; Paull, K. D.; Sausville, E. A.; Weinstein, J. N. *Electrophoresis* **1997**, *18*, 647–653.

7. Dai, Z.; Sadee, W.; Blower, P. *Curr. Pharmacogenomics* **2006**, *5*, 11–19.
8. Cao, X.; Fang, L.; Gibbs, S.; Huang, Y.; Dai, Z.; Wen, P.; Zheng, X.; Sadee, W.; Sun, D. *Cancer Chemother. Pharmacol.* **2007**, *59*, 495–505.
9. Szakacs, G.; Annereau, J. P.; Lababidi, S.; Shankavaram, U.; Arciello, A.; Bussey, K. J.; Reinhold, W.; Guo, Y.; Kruh, G. D.; Reimers, M.; Weinstein, J. N.; Gottesman, M. M. *Cancer Cell* **2004**, *6*, 129–137.
10. Gottesman, M. M. *Annu. Rev. Med.* **2002**, *53*, 615–627.
11. Fojo, T.; Bates, S. *Oncogene* **2003**, *22*, 7512–7523.
12. Staunton, J. E.; Slonim, D. K.; Coller, H. A.; Tamayo, P.; Angelo, M. J.; Park, J.; Scherf, U.; Lee, J. K.; Reinhold, W. O.; Weinstein, J. N.; Mesirov, J. P.; Lander, E. S.; Golub, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10787–10792.
13. Andersson, C. R.; Fryknaes, M.; Rickardson, L.; Larsson, R.; Isaksson, A.; Gustafsson, M. G. *J. Chem. Inf. Model* **2007**, *47*, 239–248.
14. Ross, D. T.; Scherf, U.; Eisen, M. B.; Perou, C. M.; Rees, C.; Spellman, P.; Iyer, V.; Jeffrey, S. S.; Van de Rijn, M.; Waltham, M.; Pergamenschikov, A.; Lee, J. C. F.; Lashkari, D.; Shalon, D.; Myers, T. G.; Weinstein, J. N.; Botstein, D.; Brown, P. O. *Nat. Genet.* **2000**, *24*, 227–235.
15. Huang, Y.; Blower, P. E.; Yang, C.; Barbacioru, C.; Dai, Z.; Zhang, Y.; Xiao, J. J.; Chan, K. K.; Sadee, W. *Pharmacogenomics J.* **2005**, *5*, 112–125.
16. Blower, P. E.; Yang, C.; Fligner, M. A.; Verducci, J. S.; Yu, L.; Richman, S.; Weinstein, J. N. *Pharmacogenomics J.* **2002**, *2*, 259–271.
17. Scherf, U.; Ross, D. T.; Waltham, M.; Smith, L. H.; Lee, J. K.; Tanabe, L.; Kohn, K. W.; Reinhold, W. C.; Myers, T. G.; Andrews, D. T.; Scudiero, D. A.; Eisen, M. B.; Sausville, E. A.; Pommier, Y.; Botstein, D.; Brown, P. O.; Weinstein, J. N. *Nat. Genet.* **2000**, *24*, 236–244.
18. Covell, D. G.; Huang, R.; Wallqvist, A. *Mol. Cancer Ther.* **2007**, *6*, 2261–2270.
19. Nakatsu, N.; Yoshida, Y.; Yamazaki, K.; Nakamura, T.; Dan, S.; Fukui, Y.; Yamori, T. *Mol. Cancer Ther.* **2005**, *4*, 399–412.
20. Rickardson, L.; Fryknaes, M.; Dhar, S.; Loevborg, H.; Gullbo, J.; Rydaker, M.; Nygren, P.; Gustafsson, M. G.; Larsson, R.; Isaksson, A. *Br. J. Cancer* **2005**, *93*, 483–492.
21. Klopman, G. *Quantitative Structure–Activity Relationships* **1992**, *11*, 176–184.
22. Paull, K. D.; Lin, C. M.; Malspeis, L.; Hamel, E. *Cancer Res.* **1992**, *52*, 3892–3900.
23. Sawicki, S. G.; Godman, G. C. *J. Cell. Biol.* **1971**, *50*, 746–761.
24. Lampidis, T. J.; Shi, Y.-F.; Calderon, C. L.; Kolonias, D.; Tapiero, H.; Savaraj, N. *Leukemia* **1997**, *11*, 1156–1159.
25. Hoki, Y.; Fujimori, A.; Pommier, Y. *Cancer Chemother. Pharmacol.* **1997**, *40*, 433–438.
26. Freije, J. M. P.; Lawrence, J. A.; Hollingshead, M. G.; de la Rosa, A.; Narayanan, V.; Grever, M.; Sausville, E. A.; Paull, K.; Steeg, P. S. *Nat. Med. (N.Y.)* **1997**, *3*, 395–401.
27. Klopman, G. *J. Am. Chem. Soc.* **1984**, *106*, 7315–7321.
28. Klopman, G.; Chakravarti, S. K.; Harris, N.; Ivanov, J.; Saiakhov, R. D. *SAR and QSAR in Environmental Research* **2003**, *14*, 165–180.
29. Klopman, G.; Chakravarti, S. K. *Chemosphere* **2003**, *51*, 445–459.
30. Klopman, G.; Chakravarti, S. K. *Chemosphere* **2003**, *51*, 461–468.
31. Klopman, G.; Tu, M. *J. Med. Chem.* **1999**, *42*, 992–998.
32. Abel, U.; Simon, W.; Eckard, P.; Hansske, F. G. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3292–3297.
33. Cassady, J. M.; Byrn, S. R.; Stamos, I. K.; Evans, S. M.; McKenzie, A. *J. Med. Chem.* **1978**, *21*, 815–819.
34. Kupchan, S. M.; Eakin, M. A.; Thomas, A. M. *J Med Chem 14*, 1147–1152.
35. Lee, K. H.; Huang, E. S.; Piantadosi, C.; Pagano, J. S.; Geissman, T. A. *Cancer Res.* **1971**, *31*, 1649–1654.
36. Shimada, H.; Grutzner, J. B.; Kozlowski, J. F.; McLaughlin, J. L. *Biochemistry* **1998**, *37*, 854–866.
37. Friedrich, T.; van Heek, P.; Leif, H.; Ohnishi, T.; Forche, E.; Kunze, B.; Jansen, R.; Trowitzsch-Kienast, W.; Hoefle, G. *Eur. J. Biochem.* **1994**, *219*, 691–698.
38. Halmos, T.; Cardon, A.; Antonakis, K. *Chem. -Biol. Interact.* **1983**, *46*, 11–29.
39. Pati, H. N.; Das, U.; Sharma, R. K.; Dimmock, J. R. *Mini-Rev. Med. Chem.* **2007**, *7*, 131–139.