Anticancer SAR Models for MCF-7 and MDA-MB-231 Breast Cell Lines

SHAHID QAMAR[1], C. ALEX CARRASQUER[1], SUZANNE L. CUNNINGHAM[1] and ALBERT R. CUNNINGHAM[1,2,3]

[1]*James Graham Brown Cancer Center, and Departments of* [2]*Medicine and* [3]*Pharmacology and Toxicology, University of Louisville, KY, U.S.A.*

Abstract. *The National Cancer Institute's Developmental Therapeutics Program (DTP) maintains the screening results obtained in 60 standardized cancer cell lines for ~43,000 compounds. Here the application of the categorical structure–activity relationship (cat-SAR) program for the identification of the structural attributes of identified compounds that display differential cytostatic or cytotoxic activity to one breast cancer cell line and not another is reported. The goal of this approach is to separate features associated with antiproliferative activity towards many cell lines from those that affect only a specific cell type. To assess this approach, SAR models were developed for cytostatic and cytotoxic activity against the human breast cancer cell lines MCF-7 and MDA-MB-231 and three differential activity models for compounds that were potent cytostatic and cytotoxic agents in MCF-7 cells, but relatively inactive against MDA-MB-231 cells. The MCF-7 and MDA-MB-231 models comprised the most potent 200 active and least potent 200 inactive compounds found in the DTP database and the differential activity models comprised 200 compounds potent in one cell line and not the other and 200 compounds equally potent between the cell lines. Leave-one-out validations of the individual MCF-7 and MDA-MB-231 models returned values between 83 and 85% concordance, with values obtained between 66 and 76% concordance for the differential activity models. The cat-SAR approach identified the chemical attributes associated with cytostatic and cytotoxic activity for the MCF-7 and MDA-MB-231 breast cancer cell lines included in the DTP and furthermore, were able to differentiate the selective activity of compounds between the two breast cancer lines. Thus it is conceivable that such cell line-specific mechanisms could be exploited for the discovery of highly specific anti-breast cancer agents and could also potentially facilitate the development of SAR models with sufficient resolution and clarity to identify chemical moieties associated with antiproliferative activity towards selective individual cancer types while being innocuous to other cell types.*

Correspondence to: Albert R. Cunningham, James Graham Brown Cancer Center, 529 South Jackson Street, University of Louisville, Louisville, KY 40202, U.S.A. Tel: +502 8523346, Fax: +502 8524311, e-mail: al.cunningham@louisville.edu

Key Words: cat-SAR, anticancer models, DTP, breast cancer, computational toxicology.

The National Cancer Institute's (1) Developmental Therapeutics Program (DTP) (1) contains screening results which have been obtained *in vitro* from ~43,000 compounds (1-2). These compounds have generally been tested in 60 cell lines representing leukemia, non-small cell lung, colon, central nervous system, melanoma, ovarian, renal and breast cancer to determine their cytostatic and cytotoxic potency. Cytostatic potency is reported as 50% growth inhibition (GI_{50}) and total growth inhibition (TGI) values and cytotoxic potency is reported as 50% lethal concentration (LC_{50}) (3-5). One of the major applications for the DTP is screening of specific compounds described by Shi *et al.* (6) as developing a profile or fingerprint of anticancer activity for each tested compound. The fingerprints of interesting anticancer agents can be used to search for other compounds with a similar spectrum of activity across the cell lines.

The information-rich data presented by DTP is being used for a variety of purposes by *in vitro* and *in silico* methods (7, 8). Computational methods can be used to more fully appreciate the information-rich dataset with the application of structure-activity relationship (SAR) modeling for the identification of pharmacophores associated with activity against specific cell lines. The SAR and quantitative SAR Q(SAR) expert systems are considered valid methods for drug discovery and are commercially available (9-11).

The DTP has been used to study chemical features associated with toxicity towards particular or aggregate cell types and many successful SAR models have been developed (12, 13). Fingerprint analysis (6, 14) and the use of seed structures (15) have also been useful for drug and mechanism discovery. Since the DTP contains chemical-by-chemical toxicity data across 60 cell lines, it is also possible,

by selecting one obvious cellular difference between two related cell lines and then selecting compounds that elicit contrasting responses in each of the lines, to investigate phenomena associated with cell type-specific toxicity and the identification of chemical features specifically associated with toxicity to only one cell type. Clearly, the higher the degree of cell type specificity, the greater the likelihood a potential drug will have a high therapeutic index.

Previously we developed SAR models for growth inhibition or lethality using the Multi Computer Automated Structure Evaluation (MCASE) expert system (16) which yielded concordance values between experimental and SAR-predicted results of 81% to 84% in individual cell models, and 72% for the differential activity model. Recently, using MCASE Chakravarti and Klopman obtained an average concordance value of 77% for differential cytotoxicity across multiple cell lines, although the sensitivity and specificity values were quite skewed at 58% and 92%, respectively (16).

In one of our earlier studies, the categorical-SAR (cat-SAR) expert system was used to distinguish rat mammary carcinogens from non-carcinogens (MC-NC) and mammary carcinogens from non-mammary carcinogens (MC-NMC) (17), achieving concordance between experimental and predicted values of 84%, sensitivity of 79%, and specificity of 89% for MC-NC and concordance of 78%, sensitivity of 82%, and specificity of 74% for MC-NMC. The cat-SAR expert system has also achieved concordance between experimental and predicted values between 79-81% for toxicity of developmental antithyroid drugs (18), between 80-90% for a set of environmental estrogen mimics (19), and 92% for a set of chemicals assessed for their ability to induce respiratory hypersensitivity (20).

Herein, are reported several new SAR models using the cat-SAR expert system to analyze the structural attributes of compounds that are selectively active against one breast cancer cell line and inactive towards another. Two sets of models were based on the cytotoxic and cytostatic activity towards estrogen receptor-positive (ER^+) MCF-7 and -negative (ER^-) MDA-MB-231 and two sets of models were based on differential activity between the two cell lines.

Materials and Methods

Learning set development. Four models were developed for each potency endpoint (*i.e.* GI_{50}, TGI and LC_{50}), one each for the individual MCF-7 and MDB-MB-231 cell lines and two differential activity models for compounds that were more potent toward MCF-7 than to MDA-MB-231 cells and *vice versa*. The reporting of potency values in the DTP required the development of several selection rules in order to sample the data accurately. Only compounds with GI_{50}, TGI and LC_{50} results reported in molar units were considered. For compounds tested at multiple concentration ranges, the lowest concentration was used. Potency values were transformed to their negative log values (hereafter simply referred to as GI_{50}, TGI and LC_{50}). Moreover, many of the compounds in the DTP are inactive at their maximum test concentration of (*e.g.*, 10^{-4} M). In these instances, the compounds are reported with a default potency value equal to the maximum concentration tested. Since a true potency value is unavailable for these compounds they were not included for modeling. Moreover, in the differential activity models only compounds with true potency values for both cell lines (*i.e.*, not default values) were used. The remaining compounds were then ranked according to their potency values. The top 50% were designated active and the bottom as inactive. For the developed models, the 200 most potent compounds were selected as active compounds and the 200 least potent as inactive ones. The potency ranges for the individual models is provided in Table I.

Table I. *Potency range for GI_{50}, LC_{50} and TGI values for the models.*

End point	MCF-7	MDA-MB-231	MCF-7 Diff	MDA-MB-231 Diff
GI_{50}	7.87-12.54	7.67-12.02	0.00-3.7	0.00-3.4
LC_{50}	5.36-9.60	5.39-9.04	0.00-1.84	0.00-2.7
TGI	6.52-9.99	6.67-10.46	0.00-2.34	0.00-3.83

GI_{50}: 50% Growth inhibition; LC_{50}: 50% lethal concentration; TGI: total growth inhibition; MCF-7, MDA-MB-231: breast cancer cell lines; Diff: differential activity.

For the differential activity model, first only those compounds that met the above criteria regarding activity in both cell lines were selected. 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. The GI_{50} values across the cell lines were compared only for the compounds tested in the same range. MDA-MB-231 GI_{50} values were then subtracted from the MCF-7 value, sorted by difference and the 200 with the greatest difference were selected as active and the 200 with the least difference as inactive compounds. The potency ranges for these models can be seen in Table I.

In silico chemical fragmentation and the compound-fragment data matrix. Using the Tripos Sybyl Hologram Quantitative Structure Activity Relationship (HQSAR) module (21), each chemical was fragmented *in silico* into all possible fragments meeting user-specified criteria. HQSAR allows the user to select attributes for fragment determination including atom count (*i.e.*, size of the fragment), bond types, atomic connections (*i.e.,* the arrangement of atoms in the fragment), hydrogen atoms, chirality and hydrogen bond donor and acceptor groups. The fragments can be linear, branched or cyclic moieties. The models developed herein contained fragments between two and seven atoms in size and considered atoms, bond types and atomic connections and explicate hydrogen atoms.

Upon completion of the fragmentation routine, a Sybyl HQSAR add-on procedure produces the compound-fragment data matrix as a text file. In the matrix, the rows are intact chemicals and the columns are molecular fragments. Thus for each chemical, a tabulation of all its fragments is recorded across the table rows and for each fragment all chemicals that contain it are tabulated in each column.

The HQSAR module is not used for statistical analysis or model development. The compound-fragment matrix is then analyzed, using the cat-SAR program to identify structural features associated with active and inactive compounds, validate the model, and predict the activity of untested compounds. The cat-SAR program, learning sets and the compound-fragment matrix are available through the corresponding author.

i) Identifying 'important' fragments of activity and inactivity: To ascertain any association between each fragment and biological activity (or inactivity), a set of rules is established to choose 'important' active and inactive fragments. The first selection rule is the number of times a fragment is identified in the learning set which, in this exercise, was set at between two and five compounds. It was reasoned that by looking at fragments that came from between two and five compounds in the learning set, models derived in the two to three range would be more inclusive (*i.e.*, higher coverage), while those in the four to five range would be more accurate (*i.e.*, higher concordance). In previous cat-SAR analyses, the fragment number was arbitrarily set to three.

The second rule relates to the proportion of active or inactive compounds that contribute to each fragment and in this investigation ranged from between 50% to 95%. Even if a particular fragment is associated with activity, there may be other reasons (*i.e.*, fragments) for its being inactive, thus it would not be expected to be found in 100% of the active compounds. A similar argument can be made for inactive fragments. Thus, by considering fragments toward the lower high end of the proportion scale (*e.g.*, derived from 60% active and 40% inactive) model would be expected to again be more inclusive (*i.e.*, higher coverage) while those derived from the higher end of the proportion scale (*e.g.*, 90% active and 10% inactive) would be more accurate (*i.e.,* higher concordance).

ii) Rule optimization: As in previous cat-SAR models, a relatively arbitrary setting of parameters for selecting important fragments (fragment compound counts and fragment activity proportion values) was used. For these analyses, a rule optimization routine was employed. The optimization routine in this instance allowed the Number Rule to range between 1 and 9 and the Proportion Rule to range between 0.50 and 0.95. Leave-one-out (LOO) validations were then conducted for each model and the final models selected to be both highly accurate (*i.e.,* had a high concordance between experimental and predicted values) and highly predictive (*i.e.*, made predictions on >90% of the chemicals in the learning set).

Model validation. A self-fit (*i.e.,* leave-none-out (LNO)) and two cross-validation routines (*i.e.,* LOO and multiple leave-many-out (LMO)) were conducted for each model. For the LOO cross-validation, each chemical, one at a time, was removed from the total fragment set and the n–1 model was derived. Using the same criteria described above, the activity of the removed chemical was then predicted using the n–1 model. Predicted *vs*. experimental values for each chemical were then compared and the model's concordance, sensitivity, and specificity were determined.

For the LMO cross-validation, randomly selected sets of 2.5% of the chemicals (*i.e.* 10 chemicals) were removed from the total fragment set and the n–2.5% model was derived. Again, the activity of each of the removed chemicals was then predicted using the n–2.5% model. Predicted *vs*. experimental values for the chemicals in the left out sets were then compared and the model's concordance, sensitivity and specificity were determined.

The cat-SAR predictions are based on two separate fragment sets (*i.e.,* the active fragments and the inactive ones) and the predicted activity of a chemical is based on the average probability of all the active and inactive compounds contributing to its fragments. To best classify compounds to an active or inactive category, an optimal cut-off point is selected that best separates the probabilistic prediction of active and inactive compounds based on the LOO validations.

Predicting activity. Once a final model is selected, the resulting list of fragments can then be used for mechanistic analysis, or to predict the activity of an unknown compound. In the latter circumstance, the cat-SAR program determines which, if any, fragments from the model's pool of significant fragments the test compound contains. If none are present, no prediction of activity is made for the compound. If one or more fragments are present, the number of active and inactive compounds containing each fragment is determined. The probability of activity or inactivity is then calculated based on the total number of active and inactive compounds that went into deriving each of the fragments.

The probability of activity was calculated with the cat-SAR FragSum routine. This method calculates the average probability of the active and inactive fragments contained in each compound and is weighted to the number of active and inactive compounds that contribute to each fragment. For example, if a compound contains two fragments, one being found in 9/10 active compounds in the learning set (*i.e.,* 90% active) and the other being found in 3/3 inactive compounds (*i.e.,* 0% active), the unknown compound will be predicted to have a probability of activity of 69% (*i.e.,* 9/10 actives + 0/3 actives=9/13 actives or 69% chance of activity).

Results and Discussion

Model analysis and validation. All together, 12 cat-SAR models were produced. The final set of models was developed with the cat-SAR Rule Optimizer that explored values for the Number and Proportion Rules in order to develop models with the optimal concordance between experimental and predicted values (Table II).

Overall, between 48,699 and 57,683 (average 58,821) unique chemical fragments between 3-7 non-hydrogen atoms were derived for the various MCF-7 and MDA-MB-231 models and between 51,267 and 57,637 (average 53,825) for the MCF-7-MDA-MB-231 and MDA-MB-231-MCF-7 differential activity models. There was no significant difference between the average number of fragments for the individual MCF-7 and MDA-MB-231 models and the two differential activity models. This suggested that the individual models and the differential models cover approximately the same range of structural diversity. On the other hand, looking at the fragments that ultimately derived the model, there were significantly fewer fragments derived from the differential activity models than from the individual MCF-7 and MDA-MB-231 (GI_{50}) models. Generally, chemicals with divergent activity cover a smaller area of structural space than compounds that are potent against two cell lines (*i.e.,* compounds with differential activity are structurally more distinct than others).

Table II. *Parameter optimized fragment summary, self-fit and cross validation results for MCF-7, MDA-MB-231 and differential activity models.*

Model	Total	Model	Active	Inactive	LNO			LOO			LMO		
					Sensitivity	Specificity	Concordance	Sensitivity	Specificity	Concordance	Sensitivity	Specificity	Concordance
GI_{50}													
MCF-7	49958	40305	18982	21323	0.95	0.95	0.95	0.83	0.83	0.83	0.83	0.83	0.83
					(190/200)	(184/193)	(374/393)	(165/200)	(160/194)	(325/394)	(4.0/4.9)	(3.9/4.7)	(8.0/9.7)
MDA-MB-231	58821	51329	23809	27520	0.94	0.96	0.95	0.84	0.85	0.85	0.86	0.82	0.84
					(187/200)	(191/199)	(378/399)	(168/200)	(169/199)	(337/399)	(4.2/4.9)	(4.0/4.8)	(8.2/9.8)
MCF-7 Diff	52316	2178	1799	379	0.82	0.82	0.82	0.75	0.77	0.76	0.73	0.76	0.75
					(149/182)	(146/178)	(295/360)	(134/178)	(141/184)	(275/362)	(3.3/4.5)	(3.4/4.5)	(6.7/8.9)
MDA-MB-231 Diff	57637	2473	1495	978	0.73	0.73	0.73	0.67	0.66	0.66	0.67	0.65	0.65
					(141/193)	(145/198)	(286/391)	(127/190)	(132/200)	(259/390)	(3.1/4.6)	(3.1/4.9)	(6.2/9.6)
TGI													
MCF-7	50013	43878	25192	18686	0.90	0.92	0.91	0.84	0.79	0.81	0.82	0.79	0.81
					(180/200)	(183/199)	(363/399)	(167/200)	(157/199)	(324/399	(3.9/4.8)	(3.9/5.0)	(7.9/9.8)
MDA-MB-231	56716	47656	17615	30041	0.96	0.97	0.97	0.87	0.83	0.85	0.86	0.84	0.85
					(192/200)	(194/200)	(386/400)	(173/200)	(164/197)	(337/397)	(4.3/5.0)	(4.0/4.7)	(8.3/9.9)
MCF-7 Diff	53837	11018	7528	3490	0.96	0.96	0.96	0.76	0.76	0.76	0.74	0.75	0.75
					(185/192)	(172/180)	(357/372)	(145/192)	(139/183)	(284/375)	(3.5/4.7)	(3.5/4.6)	(7.0/9.3)
MDA-MB-231 Diff	51267	39482	22147	17335	0.96	0.96	0.96	0.75	0.74	0.74	0.75	0.74	0.741
					(190/199)	(189/198)	(379/397)	(147/197)	(144/195)	(291/392)	(3.7/4.9)	(3.5/4.7)	(7.1/9.7)
LC_{50}													
MCF-7	48699	38267	20420	17856	0.96	0.96	0.89	0.76	0.76	0.76	0.76	0.75	0.75
					(192/200)	(188/195)	(380/395)	(152/199)	(149/195)	(301/394)	(3.7/4.9)	(3.6/4.9)	(7.3/9.8)
MDA-MB-231	57674	51072	23416	27656	0.92	0.89	0.90	0.84	0.80	0.82	0.81	0.81	0.81
					(183/200)	(178/200)	(361/400)	(167/200)	(160/200)	(327/400)	(4.0/4.9)	(3.9/4.9)	(7.9/9.8)
MCF-7 Diff	55053	1393	696	697	0.81	0.79	0.80	0.74	0.76	0.75	0.74	0.75	0.74
					(150/185)	(142/179)	(292/364)	(136/183)	(137/181)	(273/364)	(3.3/4.6)	(3.2/4.3)	(6.6/9.0)
MDA-MB-231 Diff	51300	4498	2769	1729	0.93	0.92	0.93	0.75	0.76	0.76	0.75	0.75	0.75
					(178/192)	(170/184)	(348/376)	(144/191)	(138/182)	(282/373)	(3.5/4.7)	(3.3/4.4)	(6.9/9.2)

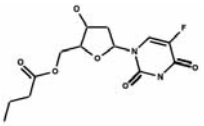
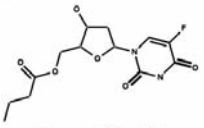
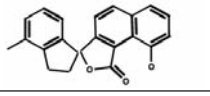
LNO: Leave-none-out; LOO: leave-one-out; LMO: leave-many-out; GI_{50}: 50% growth inhibition; TGI: total growth inhibition; LC_{50}: 50% lethal concentration; Total: number of fragments derived from learning set; Model: number of fragments meeting specified rules of the model; Active: number of fragments meeting specified rules to be considered as active; Inactive: number of fragments meeting specified rules to be considered as inactive; Sensitivity: number of correct positive predictions/total number of positive predictions; Specificity: number of correct negative predictions/total number of negative predictions; Concordance: number of correct predictions/total number of predictions.

The difference in the number of fragments between the individual models and the differential activity models was further defined when looking at the active and inactive fragments for the models. There were significantly fewer active fragments for the differential activity models when compared to the individual models (6,072 and 21,572). This again supported the notion that chemicals that have differential activity are structurally more distinct than chemicals with equal activity against both cell lines.

The LNO yielded concordance between observed and predicted results of 73% and 97% for the rule-optimized models (Table II). These high concordance values across all the models indicated that they were robust and that sufficient structural information was contained in the learning sets to distinguish between active and inactive compounds.

The LOO cross-validation concordance values ranged between 66% and 85% for the rule optimized models. The model derived for the single cell lines generally had a higher

Table III. *Fragment analysis of GI_{50} differential activity model.*

NSC	Fragment number	MCF-7 GI_{50} value	MDA-MB-231 GI_{50} value	Difference	Chemical structure
663791	1050	8.334	5.04	3.294	
663791	1275	8.33	5.04	3.29	
625587	9431	7.74	4.44	3.30	

NSC: Cancer Chemotherapy National Service Center; GI_{50}; 50% growth inhibition.

concordance value than those for the differential activity models which compared cell lines. For example, the MCF-7 and MDA-MB-231 GI_{50} models returned concordance values of 83% and 85%, respectively, while the MCF-7-MDA-MB-231 and MDA-MB-231-MCF-7 models had concordance values of 76% and 66% (GI_{50}), respectively. In all cases, the LMO cross-validations nearly equaled the LOO validation results, verifying that there was sufficient structural information contained in the learning sets to distinguish between active and inactive compounds.

In order to better judge how well these two models performed in general, the 'accuracy' or reproducibility of a standard *in vitro* toxicological test could be tested. For instance, the US National Toxicology Program's Salmonella mutagenicity database, which is derived from a standardized protocol, has been estimated to be about 85% reproducible as reported by Piegorsrch and Zeiger (22).

The overall results obtained from this exercise showed that the cat-SAR expert system could be a useful tool for analyzing DTP data. Firstly, the cat-SAR program, based on the LOO and LMO validation results, may be able to identify (*i.e., in silico* screening) agents that would be cytotoxic and/or cytostatic to MCF-7 and MDA-MB-231 cells, and other breast and non-breast cancer cell lines (data not shown). Furthermore, the cat-SAR program may identify agents, based on chemical structure that are uniquely potent against one breast cancer cell line and not another.

For example, three fragments Frag_1050, Frag_1275, and Frag_9431(Table III) were selected to illustrate how specific fragments can be used to identify compounds with differential activity between MCF-7 and MDA-MB-231 cells. Frag_1050 was identified in 12 compounds, all of which were more potent to MCF-7 cells than MDA-MB-231 cells. NSC 663791 (NSC refers to the Cancer Chemotherapy National Service Center) had the largest potency difference of 3.294 (MCF-7=8.334 and MDA-MB-231=5.04) and NSC 680418 had the smallest difference of 1.587 (MCF-7=6.308 and MDA-MB-231=4.721). Frag_1050 was a simple aromatic fluorine (F-C-C).

Frag_1275 was identified in 21 compounds, 19 of which were more potent to MCF-7 cells than to MDA-MB-231 cells and two of which were equipotent. Out of the 19 active compounds, again NSC 663791 had the largest potency difference of 3.294 (MCF-7=8.33 and MDA-MB-231=5.04), with NSC 668281 having the smallest difference of 1.58 (MCF-7=5.83 and MDA-MB-231=4.25), and both inactive chemicals NSC 684386 and 172112 had equipotent values of 4.60 in both MCF-7 and MDA-MB-231 cells.

Finally, Frag_9431 was identified in 10 compounds, nine of which were more potent to MCF-7 cells and one of which was equipotent. In this instance, NSC 625587 was the most potent towards MCF-7 cells, with a value of 3.30 (MCF-7=7.74 and MDA-MB-231=4.44), NSC 267469 had the smallest difference of the active compounds at 1.63 (MCF-7=8.27 and MDA-MB-231=6.63) and NSC 707850 was an inactive compound with a value of 4.75 for both cell lines. It is hypothesized that the structural features associated with differential activity therefore induce antiproliferative activity by mechanisms predominately found only in one of the two cell lines.

Recently, this analysis has been extended to include differential activity simultaneously in up to four other breast cancer cell lines or five other NCI DTP cell lines (data not shown). Hence, since one of the goals of the *in vitro* screening program is to indentify compounds that are specifically active against one cancer cell line and not another, differential activity cat-SAR models could be developed with a high success rate to virtually screen extremely large numbers of compounds in order to identify a few with a high likelihood of having specific activity against specific cell lines.

It is expected that the generated information could be used to identify the chemical moieties specific to the MCF-7 cell line. Thus the cat-SAR expert system produces models which are predictive and are based on mechanically sound attributes.

Acknowledgements

This research was supported by the Department of Defense Breast Cancer Research Program under award number W8XWH-05-1-0236 and the National Institutes of Health P20 RR018733. Views and opinions of, and endorsements by the author(s) do not reflect those of the US Army or the Department of Defense.

References

1 http://dtp.nci.nih.gov/docs/cancer/cancer_data.html.

2 Boyd MR: Status of the NCI preclinical antitumor drug discovery screen. *In*: Cancer. Principles and Practice of Oncology Update. DeVita VT, Hellman S and Rosenberg SA (eds.). J.B. Lippincott: Philadelphia, PA, pp. 1-12, 1989.

3 Boyd MR: The NCI *in vitro* anticancer drug discovery screen. *In*: Anticancer Drug Development Guide. Teicher BA (ed.). Humana Press: Totowa, NJ, pp. 23-42, 1997.

4 Monks A, Scudiero D, Skehan P, Shoemaker R, Paull K, Vistica D, Hose C, Langley J, Cronise P and Vaigro-Wolff A: Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. J Nat Cancer Inst *83(11)*: 757-7566, 1991.

5 Paull KB, Hamel E and Malspeis L: Prediction of biochemical mechanism of action from the *in vitro* antitumor screen of the National Cancer Institute. *In*: Cancer Chemotherapeutic Agents. Foye WO (ed.). American Chemical Society: Washington, DC, pp. 9-45,1995.

6 Shi LM, Fan Y, Myers TG, O'Connor PM, Paull KD, Friend SH and Weinstein JN: Mining the NCI anticancer drug discovery database: genetic function approximation for the QSAR of anticancer ellipticine analogues. J Chem Inf Comp Sci *38*: 189-199, 1998.

7 Yu L, Matias M, Scudiero DA, Hite KM, Monks A, Sausville EA and Waxman DJ: P450 enzyme expression patterns in NCI human tumor cell line panel. Drug Metab Dispos *29*: 304-321, 2001.

8 Cunningham AR, Cunningham SL and Day BW: Identification of structural components associated with cytostatic activity in MCF-7 but not in MDA-MB-231 cells. Bioorg Med Chem *11*: 5249-5258, 2003.

9 John NW, Timothy GM, Patrick MO, Stephen HF, Albert JF Jr., Kurt WK, Tito F, Susan EB, Lawrence VR, Leigh A, John KB, William WVO , Anne PM, Dominic AS, Edward AS, Daniel WZ, Barry B, Vellarkad NV, George SJ, Robert EW and Kenneth DP: An information-intensive approach to the molecular pharmacology of cancer. Science *275*: 343-349, 1997.

10 Snyder RD, Pearl GS, Mandakas G, Choy WN, Goodsaid F and Rosenblum IY: Assessment of the sensitivity of the computational programs DEREK, TOPKAT, and MCASE in the prediction of the genotoxicity of pharmaceutical molecules. Environ Mol Mutagen *43*: 143-158, 2004.

11 Klopman G: MULTICASE 1. A hierarchical computer-automated structure evaluation program. QSAR *11*: 176-184, 1992.

12 Fan Y, Shi LM, Kohn KW, Pommier Y and Weinstein JN: Quantitative structure–antitumor activity relationships of camptothecin analogues. Cluster analysis and genetic algorithm-based studies. J Med Chem *44*: 3254-3263, 2001.

13 Viswanadhan VN, Mueller GA, Basak SC and Weinstein JN: Comparison of neural net-based QSAR algorithm (PCANN) with hologram- and multiple linear regression-based QSAR approaches: application to 1,4-dihydropyridine-based calcium channel antagonists: J Chem Inf Comp Sci *41*: 505-511, 2001.

14 Shi LM, Yi F, Jae KL, Mark W, Darren TA, Uwe S, Kenneth DP and John NW: Mining and visualizing large anticancer drug discovery databases. J Chem Inf Comp Sci *40*: 367-379, 2000.

15 Zaharevitz DW, Holbeck SL, Bowerman C and Svetlik PA: COMPARE: a web accessible tool for investigating mechanisms of cell growth inhibition. J Mol Graph Mod *20*: 297-303, 2002.

16 Chakravarti SK and Klopman G: A structural analysis of the differential cytotoxicity of chemicals in the NCI-60 cancer cell lines. Bio Med Chem *16(7)*: 4052-4063, 2008.

17 Cunningham AR, Moss ST, Iype S, Qian G, Qamar S and Cunningham SL: Structure–activity relationship analysis of rat mammary carcinogens Chem Res Toxicol *21(10)*: 1970-1982, 2008.

18 Cunningham AR, Carrasquer CA and Mattison DR: A categorical structure-activity relationshiip analysis of the develpmental toxicity of antithyroid drug. Int J Pediatr Endocrinol, 2009.

19 Cunningham AR, Consoer DM, Iype SA and Cunningham SL: A structure–activity relationship analysis for the identification of environmental estrogens: The categorical-SAR (cat-SAR) approach. *In*: Endocrine Distrupting Modeling. Devillers J (ed.). CRC Press pp. 173-198, 2009.

20 Cunningham AR, Cunningham SL, Consoer DM, Moss ST and Karol MH: Development of an information-intensive structure–activity relationship model and its application to human respiratory chemical sensitizers. SAR QSAR Environ Res *16(3)*: 273-285, 2005.

21 Tripose, Ligand Based Design. St. Louis, MO: Tripose, Inc., 2002.

22 Piegorsrch WW and Zeiger E: Measuring intra-assay agreement for the ames salmonella assay. *In*: Statistical Methods in Toxicology. Hotham 1 (ed.). Springer, Heidberg *43*: 35-41, 1991.

Received May 21, 2011
Revised July 15, 2011
Accepted July 18, 2011