0026-895X/98/020241-11$3.00/0

MOLECULAR PHARMACOLOGY, **53:**241–251 (1998).

Mining the National Cancer Institute Anticancer Drug Discovery Database: Cluster Analysis of Ellipticine Analogs with p53-Inverse and Central Nervous System-Selective Patterns of Activity

LEMING M. SHI,[1] TIMOTHY G. MYERS, YI FAN, PATRICK M. O'CONNOR, KENNETH D. PAULL, STEPHEN H. FRIEND, and JOHN N. WEINSTEIN

Laboratory of Molecular Pharmacology (L.M.S., Y.F., P.M.O., J.N.W.), Division of Basic Sciences, and Information Technology Branch (T.G.M., K.D.P.), Division of Cancer Treatment, Diagnosis, and Centers, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892; and Fred Hutchinson Cancer Research Center/National Cancer Institute (S.H.F.), Seattle, Washington 98104

Received May 9, 1997; Accepted October 20, 1997

This paper is available online at http://www.molpharm.org

ABSTRACT

The United States National Cancer Institute conducts an anticancer drug discovery program in which ~10,000 compounds are screened every year *in vitro* against a panel of 60 human cancer cell lines from different organs. To date, ~62,000 compounds have been tested in the program, and a large amount of information on their activity patterns has been accumulated. For the current study, anticancer activity patterns of 112 ellipticine analogs were analyzed with the use of a hierarchical clustering algorithm. A dramatic coherence between molecular structures and their activity patterns could be seen from the cluster tree: the first subgroup (compounds 1–66) consisted principally of normal ellipticines, whereas the second subgroup (compounds 67–112) consisted principally of N^2-alkyl-substituted ellipticiniums. Almost all apparent discrepancies in this clustering were explainable on the basis of chemical transformation to active forms under cell culture conditions. Correlations of activity with p53 status and selective activity against cells of central nervous system origin made this data set of special interest to us. The ellipticiniums, but not the ellipticines, were more potent on average against p53 mutant cells than against p53 wild-type ones (i.e., they seemed to be "p53-inverse") in this short term assay. This study strongly supports the hypothesis that "fingerprint" patterns of activity in the National Cancer Institute *in vitro* cell screening program encode incisive information on the mechanisms of action and other biological behaviors of tested compounds. Insights gained by mining the activity patterns could contribute to our understanding of anticancer drugs and the molecular pharmacology of cancer.

The NCI is conduc[1]ting an anticancer drug discovery program in which ~10,000 compounds are screened every year *in vitro* against a panel of 60 different human cancer cell lines (Monks *et al.*, 1991; Boyd, 1997). Currently included in the screen are cells from eight melanomas, six leukemias, and eight cancers of breast; two of prostate; nine of lung; seven of colon; six of ovary; eight of kidney; and six of CNS. The purpose of the screen is to provide the initial evaluation of compounds for cytotoxic or growth-inhibitory activity against a diverse panel of cancer cell types. Compounds that show interesting activity patterns are selected for subsequent *in vivo* testing. This screen profiles, or "fingerprints," the tested compounds in terms of their anticancer activity patterns.

To date, ~62,000 synthetic compounds, plus a larger number of natural product extracts, have been tested. Similarity in activity patterns very often indicates similarity in mechanism of action, mode of drug resistance, and molecular structure of the tested compounds. Several different algorithms have been introduced to use the information in the discovery of anticancer drugs and for understanding the molecular pharmacology of cancer (Weinstein *et al.*, 1997). The COMPARE program (Paull *et al.*, 1989, 1995; Koo *et al.*, 1996; Boyd, 1997) has proved useful in finding agents with activity patterns similar to that of a "seed" compound and in finding compounds with activity patterns that correlate well (positively or negatively) across the 60 cell lines with the expression levels of particular cellular targets. Back-propagation neural networks (Weinstein *et al.*, 1992), Kohonen self-organizing maps (van Osdol *et al.*, 1994), principal component analysis (Koutsoukos *et al.*, 1994; Shi *et al.*, 1997a), hierarchical cluster analysis (Shi *et al.*, 1997a; Weinstein *et al.*, 1997), and multidimensional scaling (Shi *et al.*, 1997a)

[1] Current affiliation: R.O.W. Sciences, Inc., National Center For Toxicological Research, Jefferson, AR 72079.

ABBREVIATIONS: NCI, United States National Cancer Institute; CNS, central nervous system.

have been used to predict mechanism of action or organize compounds into families based on activity patterns. This "information-intensive" approach to the molecular pharmacology of cancer and anticancer drug discovery (Weinstein *et al.*, 1992, 1994, 1997; Shi *et al.*, 1997a) has proved useful in identifying subgroups of compounds related to particular biological targets. Growth-inhibitory activity for a single cell line is not very informative, but activity patterns across the 60 cell lines provide incisive information on the mechanism of action of screened compounds and on molecular targets and modulators of activity within the cancer cells.

Our approach to the discovery of anticancer drugs and the molecular pharmacology of cancer involves three kinds of databases (Weinstein *et al.*, 1994, 1997): 1) anticancer activity data (**A**) for compounds across the 60 human tumor cell lines, 2) chemical structural information (**S**) for the tested compounds, and 3) information on possible targets or modulators (**T**) of activity in the 60 cell lines. Currently, the size of database **A** is ~62,000 × 60. The number of targets or modulators characterized in the cell lines is well beyond 100 and growing quickly (Bates *et al.*, 1995; Li *et al.*, 1997; Myers *et al.*, 1997). The **S** database can be encoded in terms of any set of two- or three-dimensional molecular structural descriptors or experimentally measured or theoretically calculated physicochemical properties. The NCI Drug Information System (Milne *et al.*, 1994), a major resource for drug discovery (Klopman *et al.*, 1997), contains structural information for nearly 500,000 molecules, including the 62,000 tested to date. For the analysis and display of these large databases, we developed the DISCOVERY program set, which maps coherent patterns in the data rather than treating the compounds and targets one pair at a time (Myers *et al.*, 1997; Weinstein *et al.*, 1994, 1997).

One particular example of the application of our approach was the identification of a series of p53-inverse compounds (Weinstein *et al.*, 1997). The *p53* tumor suppressor gene is mutated in >50% of human tumors, more than any other gene examined to date (Hollstein *et al.*, 1991; Harris, 1996). p53 functions as a transcriptional regulator with the ability to both transactivate and suppress gene transcription. It is activated in response to DNA damage and can orchestrate a number of cellular responses to genotoxic stress, including G1 arrest and apoptosis (Harris, 1996). The p53 sequences in the NCI 60-cell line screen have been determined (O'Connor *et al.*, 1997). Nineteen of the 60 lines are p53 wild-type, and 41 are p53 mutant. Our analyses showed that the majority of clinical agents for cancer treatment seemed more active on average in this screen against the p53 wild-type cell lines than against the p53 mutant ones (O'Connor *et al.*, 1997; Weinstein *et al.*, 1997). We termed these compounds p53-direct. It seemed desirable to search for compounds (p53-inverse) that were more active against p53 mutant cell lines, at least in this assay. To search for such compounds, we analyzed the database of activity patterns using the COMPARE and DISCOVERY program sets. As part of that process, cluster analysis led to the identification of 1057 agents belonging to 37 cluster families that seemed predominantly p53-inverse. One particularly interesting group of such compounds was a subset of the ellipticine family, which forms the focus of the current study.

Ellipticine (5,11-dimethyl-6*H*-pyrido[4,3-*b*]carbazole; Fig. 1) is one of the simplest naturally occurring alkaloids with a

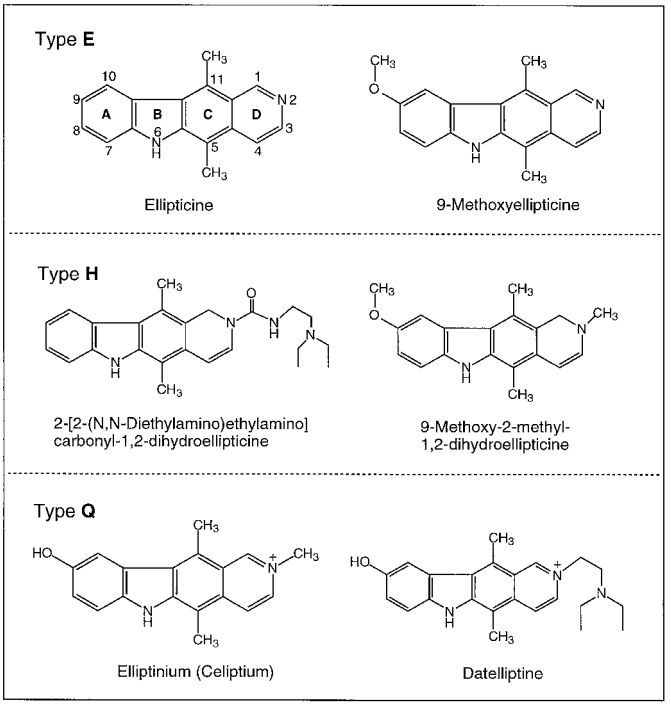

Fig. 1. The ellipticine analogs can be classified on the basis of chemical structure into three subgroups: the normal ellipticines (*E*), N^2-substituted ellipticiniums (*Q*), and 1,2-dihydrogenated ellipticines (*H*).

planar structure. It was first isolated in 1959 from the leaves of the evergreen tree *Ochrosia elliptica* Labill (Apocynaceae), which grows wild in Oceania (Goodwin *et al.*, 1959), but its biological activities were not recognized at that time. In 1967, the synthesis and anticancer activity of ellipticine, 9-methoxyellipticine, and other derivatives were reported (Dalton *et al.*, 1967). Since then, extensive attention has been paid to the design, synthesis, and structure-activity relationships of this class of compounds (Gribble, 1990; Acton *et al.*, 1994; Anderson *et al.*, 1994; Jurayi *et al.*, 1994; Devraj *et al.*, 1996a, 1996b; Shimamoto *et al.*, 1996). Several groups have reported that some ellipticine analogs, specifically ellipticiniums, are selectively active against cancer cell lines of CNS origin (Acton *et al.*, 1994; Anderson *et al.*, 1994; Jurayi *et al.*, 1994), and Boyd and colleagues (Vistica *et al.*, 1994; Kenney *et al.*, 1995) found that cellular uptake was a determinant of the CNS selectivity. Studies on the mechanisms of cytotoxicity and anticancer activity of the ellipticine analogs indicate a complex set of effects (Kohn *et al.*, 1975; Gribble, 1990; Sainsbury, 1990), including 1) DNA intercalation, 2) inhibition of topoisomerase 2, 3) covalent alkylation of macromolecules, and 4) generation of cytotoxic free radicals.

Because cardiovascular toxicity and hemolysis were observed during preclinical studies, development of the parent compound, ellipticine, was halted. Interest shifted to the 9-substituted derivatives, including 9-hydroxyellipticine, 9-methoxyellipticine, and Elliptinium (Celiptium). Only limited activity was observed in clinical trials with 9-methoxyellipticine and 9-hydroxyellipticine. Although phase II clinical trials of Elliptinium yielded moderately promising results, none of the ellipticine derivatives have reached clinical practice in the United States. However, members of the ellipticine family may still yield clinically useful anticancer drugs if their mechanisms of action and relative activities in

tumors with particular molecular characteristics are better understood (Gribble, 1990; Sainsbury, 1990).

For this study, we collected information on 112 ellipticine derivatives that have been tested in the NCI anticancer drug discovery program. We were interested in them partly because of their potent anticancer activity and partly because some of them seemed to be p53-inverse (Weinstein, 1997) (i.e., they seemed more active against p53 mutant cell lines than against p53 wild-type cell lines in the NCI screen).

We initially classified these ellipticine analogs into three subtypes (Fig. 1) according to their chemical structures. A molecule was classified as type E if it was a normal Ellipticine (the D ring was a noncharged pyridyl ring); type H if the D ring was 1,2-dihydrogenated; and type Q if the D ring was an N^2-substituted pyridyl ring with a permanent positive charge (i.e., quaternized). In our data set, 30 compounds were type E, 28 were type H (including three 1,2,3,4-tetrahydroellipticine analogs), and 54 were type Q, according to the above classification criteria. As shown in Discussion, there are pathways by which some of the ellipticine analogs can be transformed from one type to another under cell culture conditions.

Materials and Methods

Cell screen and activity data. Details of the NCI cell screening protocols and reporting procedures have been described previously (Paull *et al.*, 1989, 1995; Monks *et al.*, 1991; Jurayi *et al.*, 1994; Boyd, 1997). Briefly, dimethylsulfoxide solutions of the compounds are diluted routinely by a factor of 500 with aqueous medium. Aliquots of the resulting aqueous solutions or suspensions are tested against the cells in microtiter plates for 48 hr of exposure, and cell growth then is assayed spectrophotometrically by staining for total cellular protein with sulforhodamine B. The growth-inhibitory activity of a tested compound is expressed in terms of the quantity $-\log(GI_{50})$, where GI_{50} is the concentration required to inhibit growth by 50% in comparison with the untreated control. For each compound, 60 activity values (1 for each cell line) make up the activity pattern, or fingerprint, of the compound. Overall, our database included 14% missing values, each of which was replaced by the mean value for the drug in question before further data analyses. It should be noted that sulforhodamine B staining is not the only method for obtaining an index of cell growth, but it is a standard one. The screening capacity, reproducibility, and quality control all seemed favorable when the sulforhodamine B staining method was used (Boyd, 1997).

Cluster analysis. We used the agglomerative hierarchical clustering ("hclust") function in the S-Plus statistical package (StatSci Division, MathSoft, Seattle, WA) to cluster compounds in terms of their *in vitro* activity patterns across the 60 cell lines. At each step in the clustering process, the two clusters (or individuals) nearest to each other by some chosen criterion are combined to form one larger cluster, a process called "merge." The procedure continues to aggregate clusters together until there is only one. Compounds similar in activity pattern appear together, and dissimilar ones appear to be widely separated in the resulting cluster tree.

A convention is needed at each merge of the cluster tree to specify which cluster should go on the top and which should go on on the bottom (i.e., in terms of the orientation in Figs. 2 and 4). For a data set of n individuals, there are $n - 1$ merges. Therefore, there are 2^{n-1} possible orderings of the leaves in a tree. The convention used in hclust orders the clusters so that the "tighter" (more homogeneous) one is always on the top. Individuals are considered the "tightest" clusters possible and therefore always are put on the top when merged with clusters. If a merge involves two individuals, they are placed in order by observation number in the data set.

The shape of the cluster tree formed by a particular data set is determined by a combination of the distance metric and clustering method. The distance metric can be euclidean, maximum, Manhattan, binary, or some other user-selected measure. Clustering methods implemented in S-Plus include average linkage, single linkage, and complete linkage. For average linkage clustering, the distance between two clusters is the average of the distances between the data points in one cluster and the data points in the other. For single linkage clustering, the distance between two clusters is the minimum distance between data points in the first cluster and data points in the second. For complete linkage clustering, the distance is the largest distance between data points in one cluster and data points in the other.

For this study, we investigated different clustering algorithms but settled on average linkage clustering. The distance metric used was $1 - r$, where r is the Pearson correlation coefficient between the activity patterns of two compounds. When we compared different clustering algorithms and distance metrics, it seemed that the combination of average linkage clustering and $1 - r$ or euclidean distance gave the most coherent results, but all of the methods except single linkage yielded essentially the same regularities as those presented here.

Results

Cluster analysis based on cell screen activity patterns. Fig. 2 is a dendrogram showing the hierarchical clustering of 112 ellipticine analogs based on their anticancer activity patterns. The cluster tree indicates a distinct separation of the compounds into two subgroups, with a remarkable separation between compounds **66** and **67**. By examining the chemical structures, we found that the first subgroup (compounds **1–66,** denoted as EE) consisted principally of normal ellipticines (E), whereas the second subgroup (compounds **67–112**, denoted as QQ) consisted principally of N^2-alkyl-substituted ellipticiniums (Q). Fig. 3, a clustered correlation (ClusCor) matrix (Myers *et al.*, 1997; Weinstein *et al.*, 1997), also shows the distinct separation of compounds into two subgroups. The average pairwise Pearson correlation coefficients were 0.401 for compounds within the EE subgroup and 0.464 for those within the QQ subgroup. The average correlation coefficient between compounds of the EE and QQ subgroups was only 0.103.

Most of the H-type compounds clustered in the EE subgroup, with a few exceptions that fell in the QQ subgroup. Some of the Q-type compounds clustered in the EE subgroup. We initially thought that these exceptions might reflect random experimental noise in the screen data. After examination of the chemical structures of the "outlier" compounds, however, we were able to explain the clustering results almost completely, as described in Discussion.

To determine how well the *in vitro* cell screen activity patterns reflect the chemical structures and mechanisms of action of tested compounds, we clustered an expanded data set formed by adding 167 camptothecin analogs to the 112 ellipticines. The cluster tree for the 279 compounds is shown in Fig. 4 (distance metric, $1 - r$; clustering method, average linkage).

The camptothecins clustered side by side in one major branch of the tree, with two exceptions (**1** and **274**). The activity patterns of compounds **1** and **274** were different from those of the remainder of the data set; they merged last in the cluster tree. Further examination revealed, however, that this apparent dissimilarity was an artifact. In the NCI screen, if a compound is not sufficiently active to inhibit cell

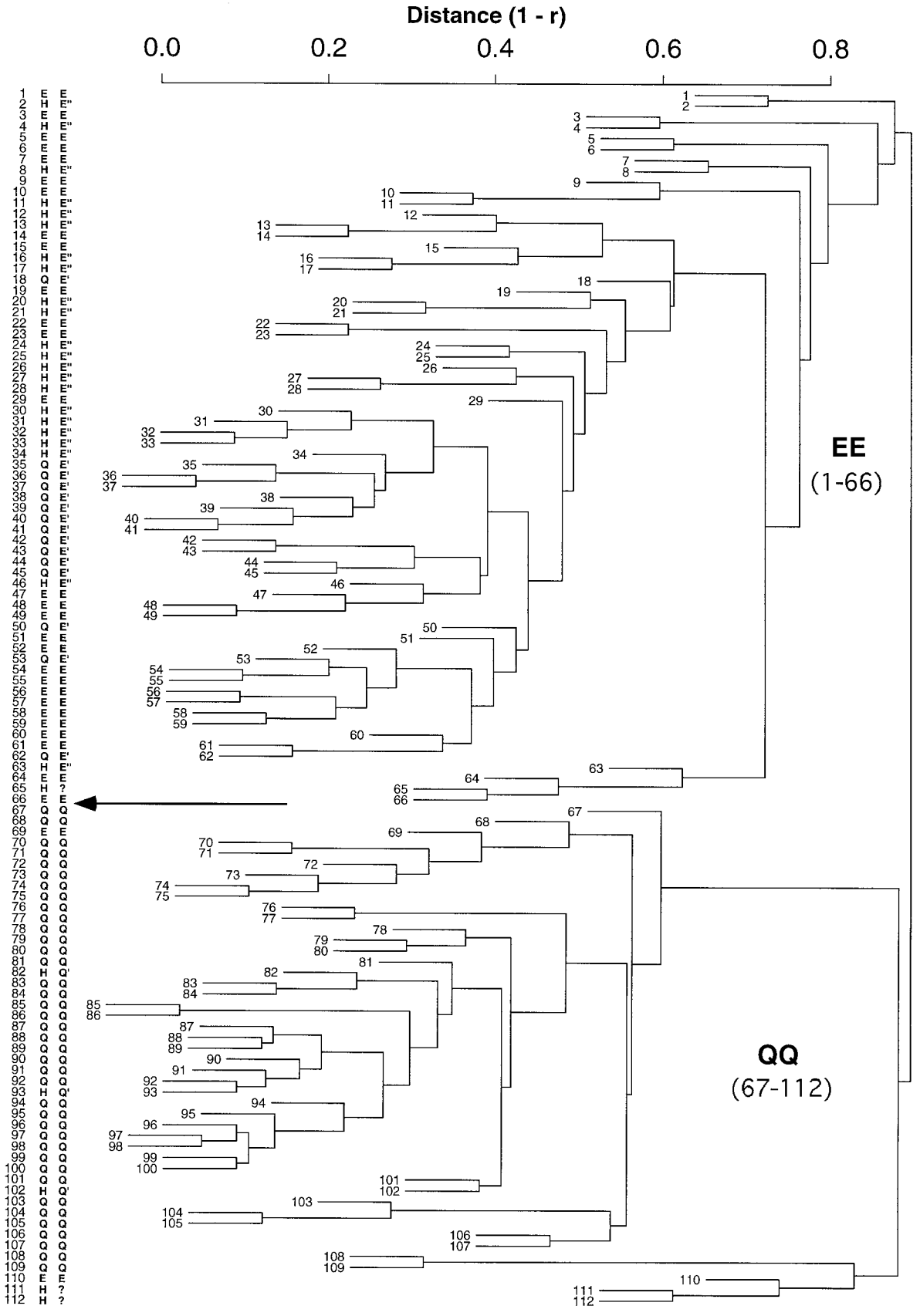

Fig. 2. Dendrogram showing the hierarchical clustering of 112 ellipticine analogs based on their activity patterns across 60 human tumor cell lines. First column, compound numbers in cluster order. Second column (*E*, *H*, or *Q*), original structural type of the compound. Third column (*E*, *E′*, *E″*, *Q*, or *Q′*), active structural type to which the compound probably was transformed in the cell culture medium. ?, Compounds with active forms of which we are unsure. The distance between two clusters is measured by $1 - r$, where r is the Pearson correlation coefficient. There is a major gap between compounds **66** and **67**, dividing the data set into two subgroups (*EE* and *QQ*).

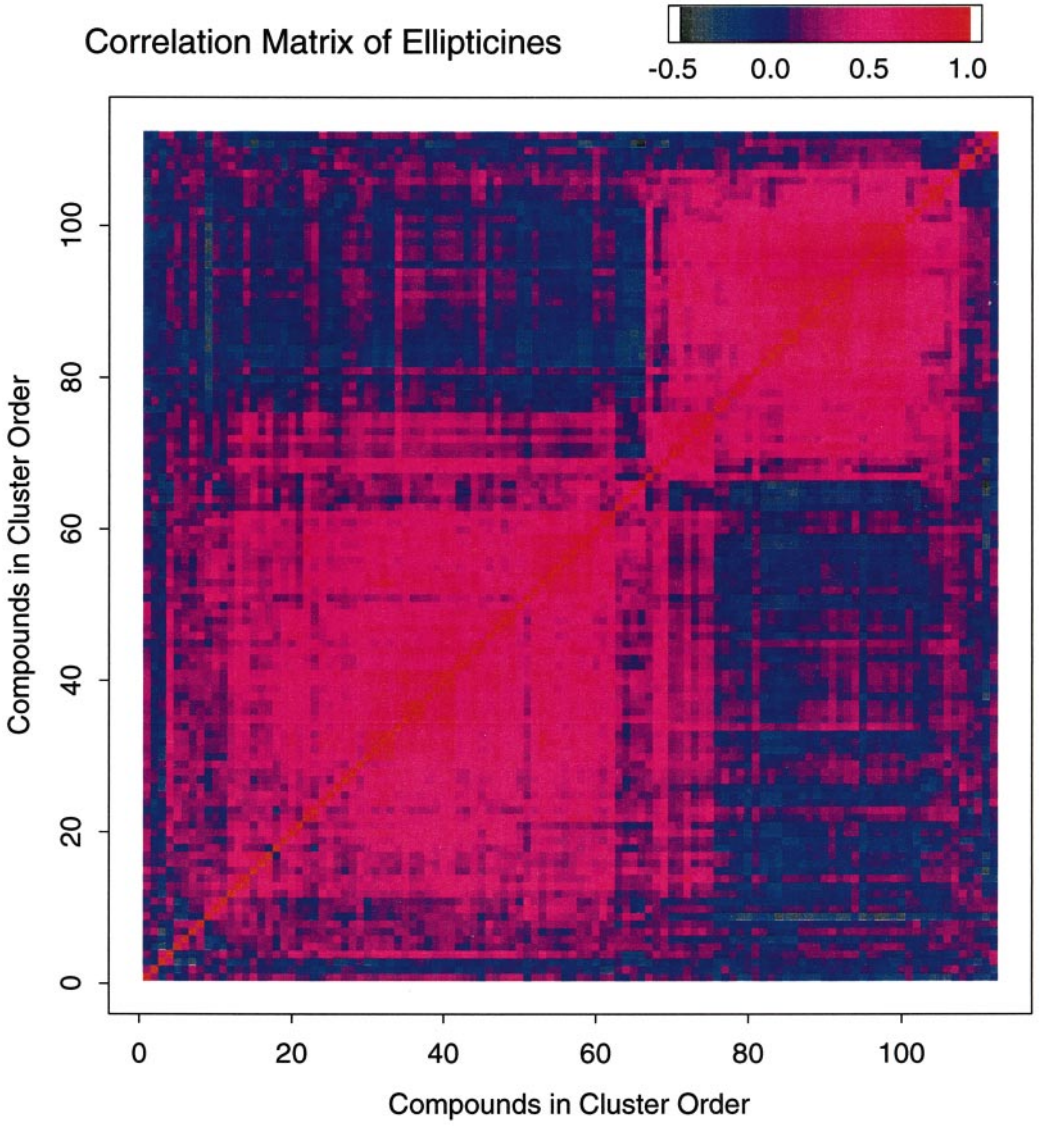

Fig. 3. A clustered correlation (ClusCor) matrix (Myers *et al.*, 1997; Weinstein *et al.*, 1997) for the 112 ellipticine analogs in terms of their activity patterns across 60 human cancer cell lines. Each data entry represents a Pearson correlation coefficient between the activity patterns of two compounds. This figure clearly shows the separation between the EE and QQ subgroups.

growth by 50% at the highest concentration tested (usually 10^{-4} M), the GI_{50} value is reported as the highest concentration tested. These two compounds, 12-nitro-17-hydroxy-20(*S*)-camptothecin (**1**) and 9-amino-20(*R*)-camptothecin (**274**), were almost completely inactive in the screen; only 1 or 2 of the 60 cell types were sufficiently sensitive for 50% growth inhibition below the highest concentration tested (10^{-4} M). Therefore, there was not enough information encoded in the activity patterns to characterize the biological behavior of these two compounds. If it were possible to test them at higher concentrations, their activity patterns might emerge, and they might cluster with the other camptothecins.

As shown in Figs. 4 and 5, the camptothecins are more similar to each other in activity pattern than are the ellipticines. The average pairwise Pearson correlation coefficient (r) between activity patterns for the camptothecin set was 0.696 (standard deviation = 0.207), whereas that for the ellipticine analogs (including both EE and QQ compounds) was only 0.267 (standard deviation = 0.256). In fact, the EE-type ellipticines were somewhat closer in activity pattern to the camptothecins (average r = 0.248, standard deviation = 0.296) than they were to the QQ-type ellipticiniums (average r = 0.103, standard deviation = 0.162). These observations may reflect the current view that camptothecins act by a single, specific mechanism of action (i.e., by inhibition of topoisomerase 1), whereas ellipticines and ellipticiniums act in a variety of ways, including intercalation in DNA, inhibition of topoisomerase 2, and alkylation of DNA.

p53 status and CNS-related indices. To explore the p53- and CNS cell-related properties of the ellipticine analogs, we defined two activity indices: 1) CNS.sel (mean. CNS − mean.0.60), where mean.CNS for a particular compound is the mean activity for the 6 CNS cell lines, and mean.60 is the mean activity across all 60 cell lines. A positive CNS.sel value indicates greater potency in cells of CNS origin. 2) p53.MW (mean.p53M − mean.p53W), where mean.p53M is the mean activity for the 41 p53 mutant cell lines, and mean.p53W is the mean activity for the 19 p53 wild-type cell lines. A positive p53.MW value indicates a p53-inverse agent.

Values of CNS.sel and p53.MW were calculated for the 112 ellipticine analogs, as shown in Fig. 6. The compounds in Fig. 6 were arranged in the same order as in the cluster tree (Fig. 2). Clearly, most compounds in the second subgroup (QQ) of

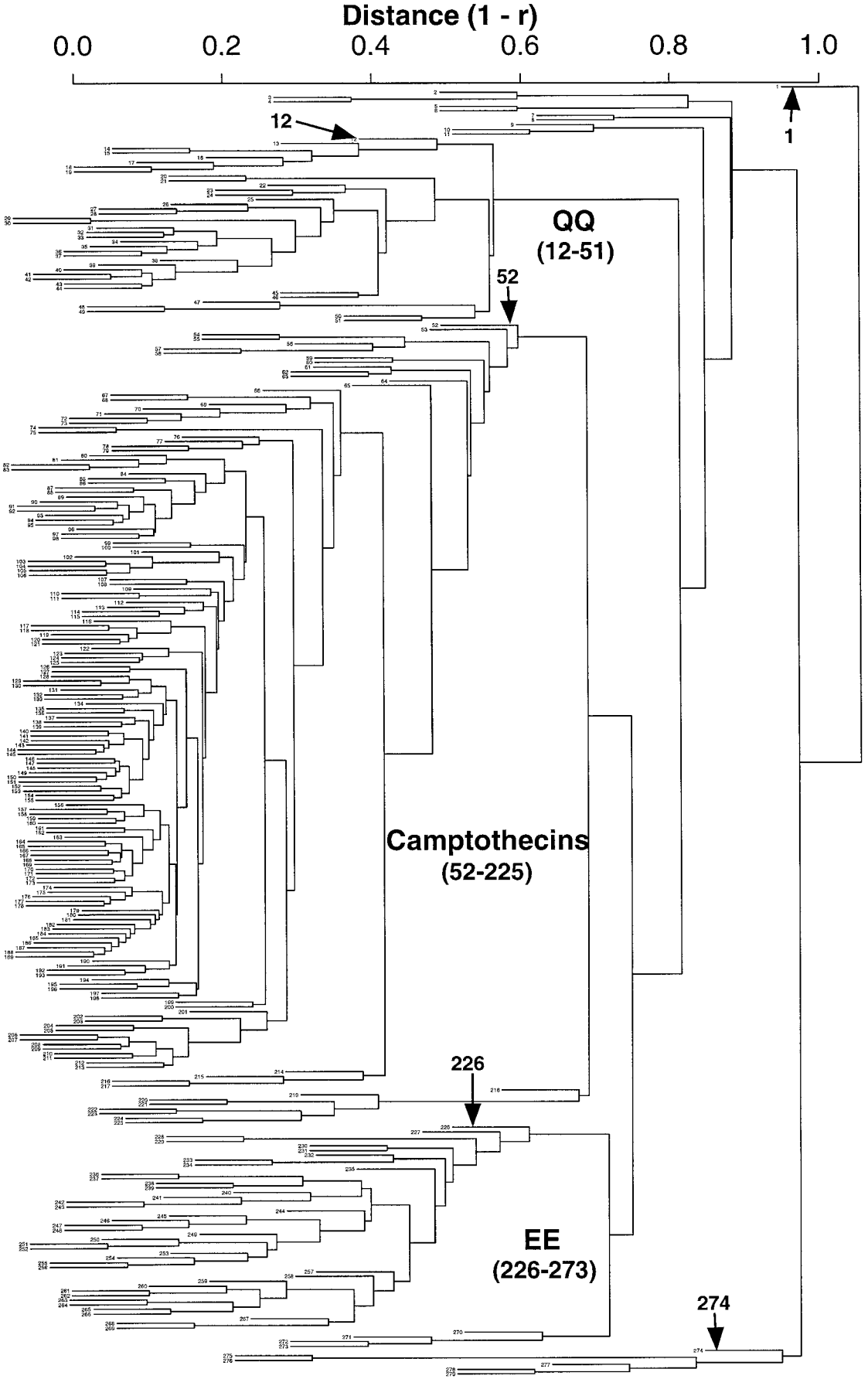

Fig. 4. Dendrogram showing the hierarchical clustering of 112 ellipticine analogs (*EE* or *QQ*) and 167 camptothecin derivatives based on their activity patterns across 60 human tumor cell lines. The distance between two clusters is measured by $1 - r$, where r is the Pearson correlation coefficient. The activity patterns of camptothecins seem much more homogeneous than do those of the ellipticines.

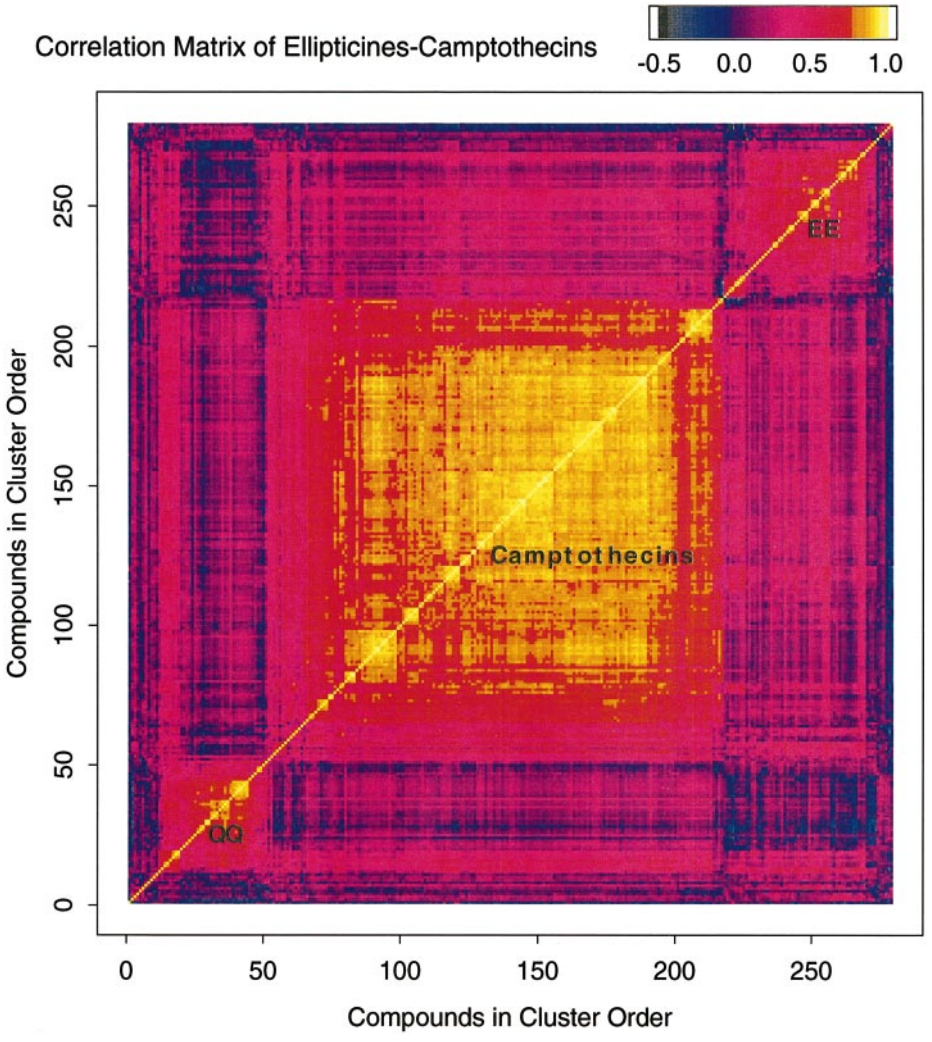

Fig. 5. A clustered correlation (ClusCor) matrix (Myers *et al.*, 1997; Weinstein *et al.*, 1997) for the 279 ellipticine and camptothecin analogs in terms of their activity patterns across 60 human cancer cell lines. Each entry represents a Pearson correlation coefficient between the activity patterns of two compounds.

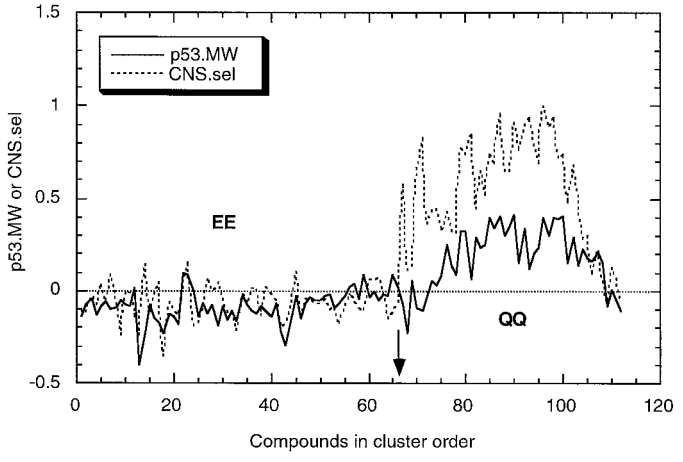

Fig. 6. Chemosensitivity differences p53.MW and CNS.sel for 112 ellipticines. Compounds are listed in cluster order from Fig. 2. There is a high correlation (r = 0.800) between the two indices.

the cluster tree (**67–112**) seemed both p53-inverse and CNS selective, whereas most compounds in the first subgroup (EE, **1–66**) were neither p53-inverse nor CNS selective. There was a high correlation between p53.MW and CNS.sel (r = 0.800). Because only 1 (SF-539) of the 6 CNS cell lines is p53 wild-type, we initially suspected that the p53-inverse property of the QQ subgroup simply might be a result of the superior activity of these compounds against the CNS cell lines. We therefore excluded the 6 CNS cell lines and calculated the p53.MW(no CNS) index. Fig. 7 indicates that there was no major difference between p53.MW and p53.MW(no CNS).

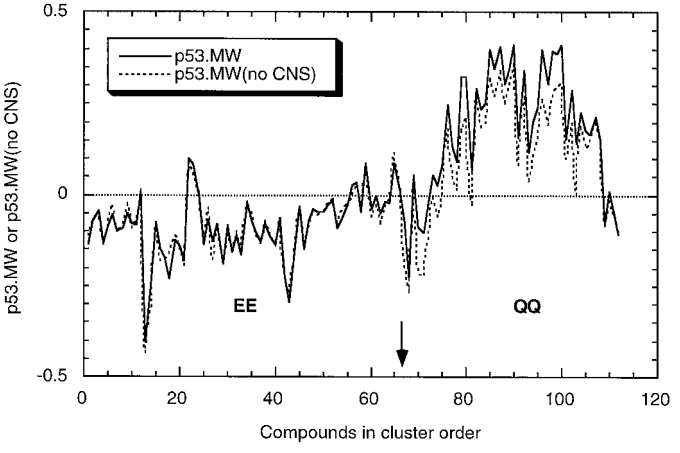

Fig. 7. Chemosensitivity differences p53.MW and p53.MW(no CNS) for 112 ellipticines. Compounds are listed in cluster order from Fig. 2. There is a high correlation (r = 0.975) between the two indices, indicating that the apparent p53-inverse status of ellipticinium compounds was not attributable to the CNS cell lines alone.

The general trends of the indices essentially were unchanged. The relationship between p53.MW and p53.MW(no CNS) could be described by the following statistically highly significant linear equation: p53.MW = 0.0289 + 1.114*p53.MW(no CNS), with $r = 0.975$ and $n = 112$.

This relationship indicated that the apparent p53-inverse status of the ellipticiniums (QQ) was not an artifact of the CNS selectivity. It seemed possible that the p53-related activity index, p53.MW, might indeed reflect some kind of inherent difference between the EE and QQ subgroups in terms of their inhibitory activity against the p53 wild-type and mutant cell lines.

We then decided to calculate the p53.MW index within each organ-of-origin subpanel of cell lines; the results are shown in Fig. 8. The numbers of cell lines per subpanel were insufficient for statistical analysis, but p53-inverse status seemed most prominent in the colon, CNS, and kidney subpanels. They were the main contributors to the overall p53-inverse status of the QQ subgroup. Differences between indices for EE and QQ subgroups were not seen for the leukemia, melanoma, ovarian, or breast subpanels. There were many missing values for the breast subpanel because the breast cell lines were not included in the cell screen program during its early stages. Because both prostate cell lines in the screen were p53 mutant, a p53.MW index for the prostate subpanel could not be calculated. For the non-small cell lung cancer lines (Fig. 8B), the situation was different: the difference in p53.MW index between the EE and QQ subgroups was considerable, but the mean value for the QQ subgroup was −0.122. Therefore, this subpanel of cell lines did not contribute to the overall p53-inverse status of the QQ subgroup.

The overall p53.MW index (Fig. 8I), although not very large in magnitude, did remain characteristic and significant with respect to the activity patterns and chemical structural issues used to generate the QQ and EE subgroups.

Discussion

We noticed that in Fig. 2, 15 Q-type compounds (**18**, **35–45**, **50**, **53**, and **62**) clustered in the EE subgroup rather than in the QQ subgroup as one might have expected. By examining their chemical structures, we found that 12 of the 15 compounds were ellipticiniums substituted at the N^2-position with —CH_2—O—CO—R, where R is an alkyl group of two to six carbon atoms in 10 cases (**18** and **35–43**) and a phenyl ring in 2 cases (**44** and **45**). The other three "misclustered" compounds (**50**, **53**, and **62**) were ellipticiniums substituted at the N^2-position with —$CH(CH_3)$—O—CO—OCH_2CH_3. We suspected that all 15 of these Q-type compounds could be transformed to the normal ellipticine form by the metabolic pathway shown in Fig. 9. We have not investigated whether the conversions depicted in Fig. 9 actually did occur under cell culture conditions; however, by analogy with what has

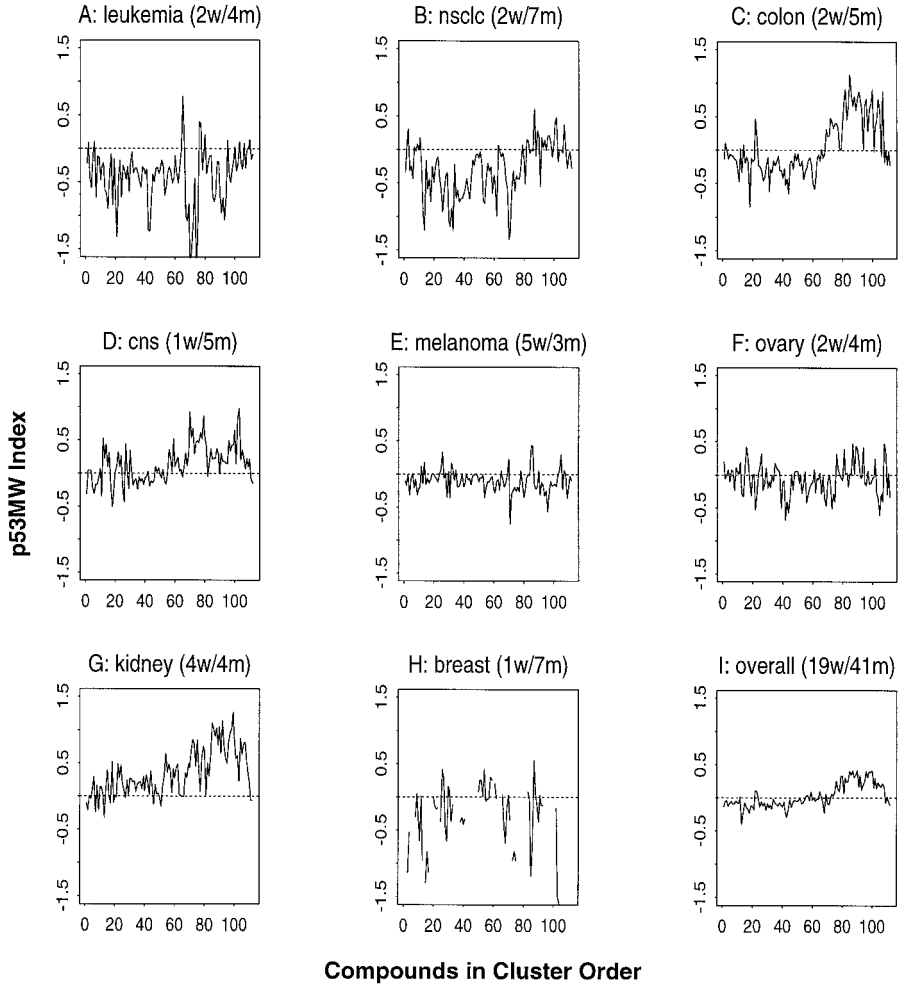

Fig. 8. Chemosensitivity difference (p53.MW) between p53 wild-type and mutant cell lines within each organ-of-origin subpanel (compare with Figs. 6 and 7). A–H, p53. MW index within each subpanel. I, Overall p53. MW index across all 60 cell lines. There were many missing values for the breast subpanel because the breast cell lines were not included in the NCI cell screening program until 1992; compounds tested before that date simply were not tested for their anticancer activity against these breast cell lines.

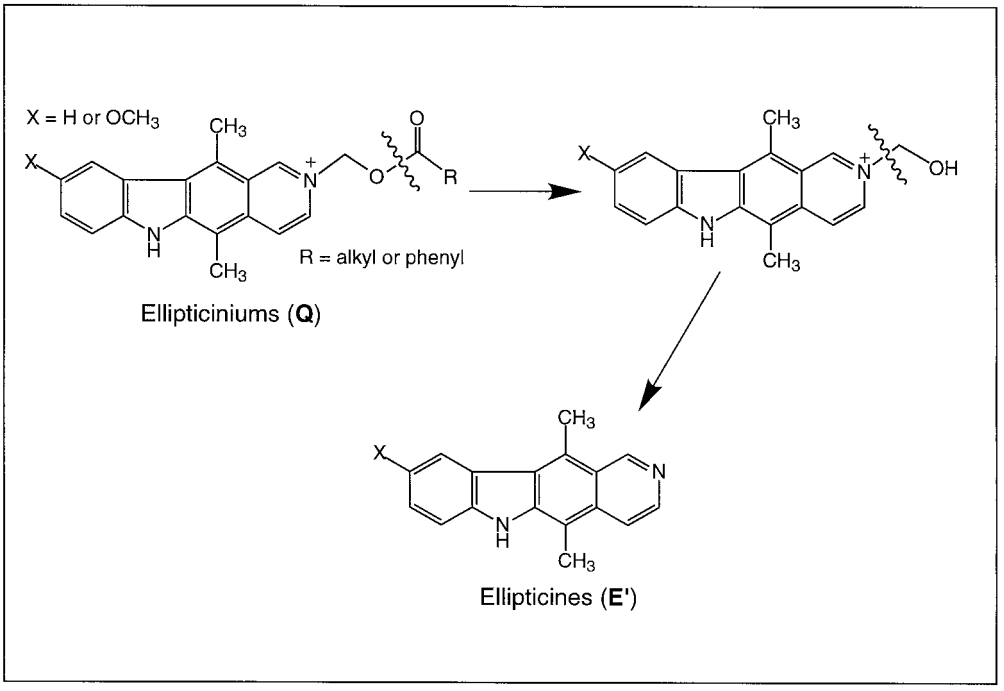

Fig. 9. Possible pathway for conversion of a subgroup of ellipticiniums (Q) to active ellipticine form (E′).

been reported in the literature (Varia *et al.*, 1984; Bundgaard, 1991), these conversions are very likely. The first step in Fig. 9 is a simple hydrolysis of the ester. It could happen easily in the presence of esterases (Bundgaard, 1991). The intermediate, *N*-hydroxymethyl derivative, that would result from the first step has been shown for similar types of compounds to be very unstable (Varia *et al.*, 1984). This intermediate can readily be broken down to formaldehyde and the parent compound (Varia *et al.*, 1984; Bundgaard, 1991), as shown by the second step in Fig. 9.

The simplest interpretation of the clustering is that these transformations took place sufficiently quickly under cell culture conditions that the activity patterns of these compounds seemed closer to those of the EE subgroup than to those of the remainder of the QQ subgroup. As discussed, these 15 compounds did not show CNS selectivity or p53-inverse status demonstrated by other Q-type compounds (**67**, **68**, **70–81**, **83–92**, **94–101**, and **103–109**). To distinguish them from the E-type and other Q-type compounds, these 15 compounds were reclassified as type E′ (see Figs. 2 and 9).

Although most of the H-type compounds clustered within the EE subgroup, a few of them (**82**, **93**, and **102**) clustered in the QQ subgroup. The H subset, in fact, consisted of prodrugs that might convert to different active forms under cell culture conditions. Fig. 10 indicates pathways by which these compounds might be converted.

First, most of the H subset (22 of 28) were N^2-acyl-1,2-dihydroellipticines. They could be converted to the normal ellipticine structures by oxidation and hydrolysis (Fig. 10, *pathway A*) and then would be expected to exhibit biological effects similar to those of normal ellipticines. Indeed, these H-type compounds were intermingled with the EE subset in the cluster tree. These 22 compounds were reclassified as type E″ in Fig. 2 to indicate that they could be converted from H-type to E-type.

Second, three other H-type compounds (**82**, **93**, and **102**) were N^2-methyl-1,2-dihydroellipticines, which could be oxidized to the corresponding ellipticinium derivatives by pathway B in Fig. 10. As might be expected, these three H-type compounds clustered with the QQ subgroup (Fig. 2). They were reclassified as type Q′, indicating that they were probably converted from H- to Q-type. In fact, Jurayi *et al.* (1994) observed the conversion of N^2-methyl-1,2-dihydroellipticines to their corresponding ellipticiniums through oxidization in air. The activity patterns of these three compounds seemed to reflect this type of conversion under cell culture conditions. Compounds **65**, **111**, and **112** are 1,2,3,4-tetrahydroellipticines. Their activity patterns was distinct from those of other ellipticine analogs, which is consistent with their unique chemical structural features.

Cluster analyses were performed on the *in vitro* cell screen activity patterns of 112 ellipticine analogs. The results were in good agreement with previous observations that activity patterns generated from the screening program could distinguish diverse compounds by structure and mechanism of action (Paull *et al.*, 1989, 1995; Weinstein *et al.*, 1992, 1997). In this study, we have demonstrated, at a "microlevel" (for a more homogenous data set), that the NCI *in vitro* cell screen program generates incisive information about the mechanisms of action and selective anticancer activity of tested compounds. After taking into account the biochemical transformations likely to have taken place under cell culture conditions, we found an essentially perfect match between chemical structures and their positions (determined by their activity patterns) in the cluster trees of Figs. 2 and 4.

The striking coherence between chemical structures and activity patterns suggested a causal relationship. We suspected that each subset (EE or QQ) might behave in a unique way in inhibiting tumor cell growth. As mentioned, the EE subset consists of the normal ellipticines without a permanently charged D-ring, whereas the QQ subset consists of compounds permanently positively charged on the D-ring. This structural difference may play a major role in determining their difference in mechanism of action or the way or

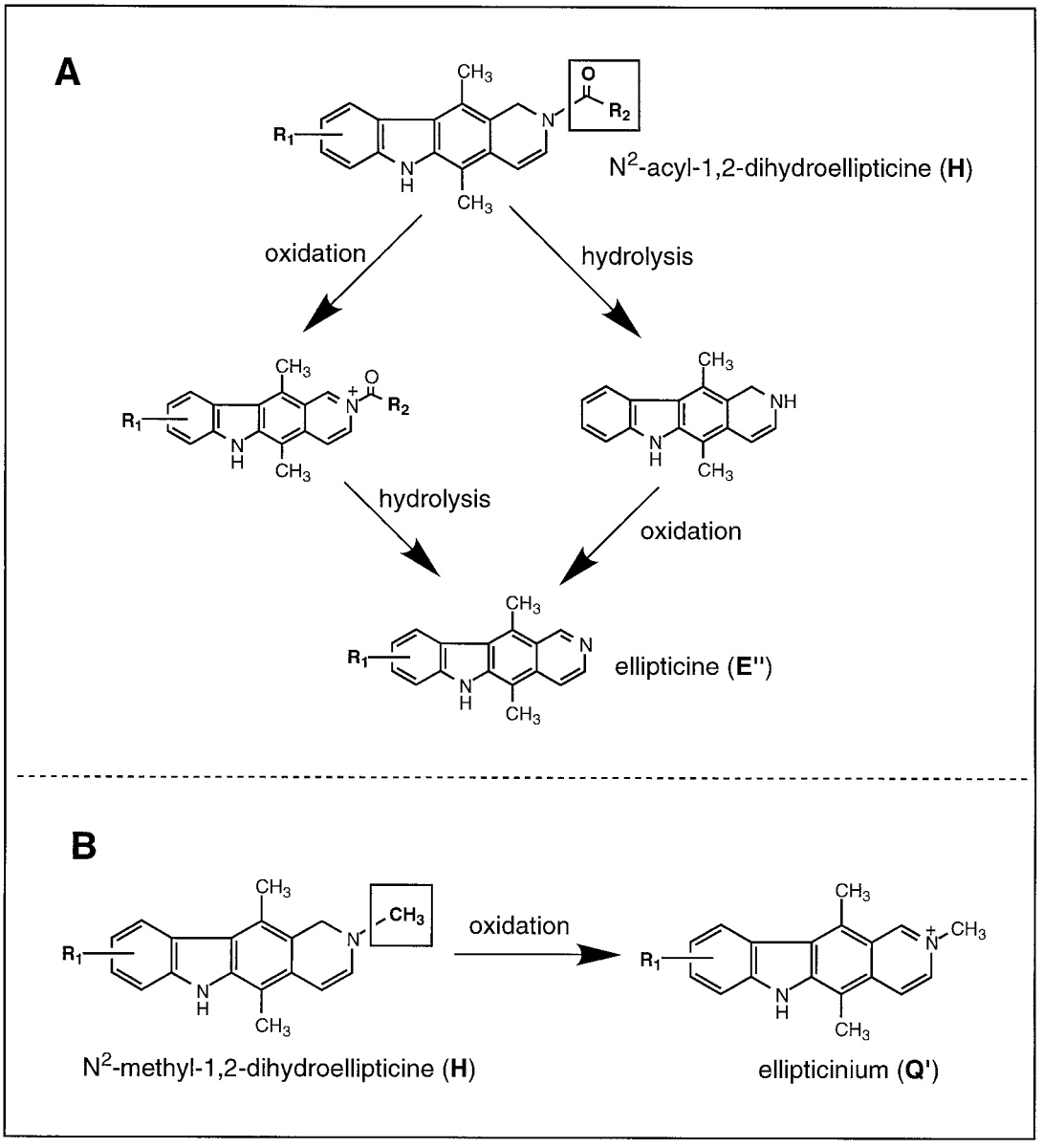

Fig. 10. Possible pathways for conversion of 1,2-dihydroellipticines (H) to the active form, ellipticines (E″) or ellipticiniums (Q′).

ways in which these compounds are transported and metabolized by tumor cells.

The most immediate reasons for our interest in the ellipticine analogs were the apparent p53-inverse and CNS cell-selective properties of part of the data set (i.e., the QQ subset). Our previous analyses showed the majority of clinical anticancer agents to be more active on average against the p53 wild-type cell lines than against the p53 mutant ones in the NCI 2-day growth-inhibition assay (O'Connor *et al.*, 1997; Weinstein *et al.*, 1997). These findings were consistent with those for a series of Burkitt's lymphomas and lymphoblastoid cells (Fan *et al.*, 1994). However, the impact of p53 on chemosensitivity is complex and multifactorial; it depends on the genotypic/phenotypic context of the cell type and on the nature of the assay (Lowe and Jacks, 1997). In particular, the two-day sulforhodamine B assay emphasizes rapid growth inhibition and cytotoxicity, including some apoptosis.

Our analyses (Fig. 8) suggest a context dependence of the effect of p53 on chemosensitivity. The QQ subgroup seems to be p53-inverse for the subpanels of colon, CNS, and kidney origin but not for the subpanels of leukemia, melanoma, ovary, or breast origin. However, the numbers of cell lines in each category are small, and other lines should be studied prospectively to pin down the point.

One might infer that p53 is, directly or indirectly, involved in the processes that determine the anticancer activity of the compounds. However, we do not have experimental evidence for that proposition; the conclusions are correlative not causal. From the statistical point of view, the differences are significant. These observations already have triggered more insightful thinking about the impact of p53 status on chemosensitivity to ellipticines and ellipticiniums.

In a companion study (Shi *et al.*, 1998), we used the recently developed genetic function approximation (Rogers and Hopfinger, 1994; Shi *et al.*, 1997b) method to model structure-activity relationships of the ellipticine data set. Molecular descriptors calculated with use of the Cerius^2 molecular modeling package (Molecular Simulations, San Diego, CA)

were used to predict a number of activity indices, including p53.MW and CNS.sel. Good points of agreement between the activity indices and molecular structural descriptors were found. Exemplar compounds selected in part on the basis of the cluster analysis and genetic function approximation studies will be tested in the p53-isogenic cell sets (Fan *et al.*, 1995, 1997).

Acknowledgments

We are grateful to Drs. Kurt Kohn, Yves Pommier, and Mark Waltham for helpful discussions. We also are grateful to members of the Developmental Therapeutics Program, NCI, for the chemosensitivity data used in this report.

References

Acton EM, Narayanan VL, Risbood PA, Shoemaker RH, Vistica DT, and Boyd MR (1994) Anticancer specificity of some ellipticinium salts against human tumors in vitro. *J Med Chem* **37:**2185–2189.

Anderson WK, Gopalsamy A, and Reddy PS (1994) Design, synthesis and study of 9-substituted ellipticine and 2-methylellipticinium analogues as potential CNS-selective antitumor agents. *J Med Chem* **37:**1955–1963.

Bates SE, Fojo AT, Weinstein JN, Myers TG, Alvarez M, Paull KD, and Chabner BA (1995) Molecular targets in the National Cancer Institute drug screen. *J Cancer Res Clin Oncol* **121:**495–500.

Boyd MR (1997) The NCI in vitro anticancer drug discovery screen: concept, implementation, and operation, 1985–1995, in *Anticancer Drug Development Guide: Preclinical Screening, Clinical Trials, and Approval* (Teicher BA, ed) pp 23–42, Humana Press, Totowa, NJ.

Bundgaard H (1991) Novel chemical approaches in prodrug design. *Drugs Future* **16:**443–458.

Dalton LK, Demerac S, Elmes BC, Lorder JW, Swan JM, and Teitel T (1967) Synthesis of the tumor-inhibitory alkaloids, ellipticine, 9-methoxyellipticine, and related pyrido[4,3-b]carbazoles. *Aust J Chem* **20:**2715–2727.

Devraj R, Barrett JF, Fernandez JA, Katzenellenbogen JA, and Cushman M (1996b) Design, synthesis, and biological evaluation of ellipticine-estradiol conjugates. *J Med Chem* **39:**3367–3374.

Devraj R, Jurayi J, Fernandez JA, Barrett JF, and Cushman M (1996a) Synthesis of a series of cytotoxic 2-acyl-1,2-dihydroellipticines which inhibit topoisomerase II. *Anti-Cancer Drug Design* **11:**311–324.

Fan S, Chang JK, Smith ML, Duba D, Fornace AJ Jr, and O'Connor PM (1997) Cells lacking CIP1/WAF1 genes exhibit preferential sensitivity to cisplatin and nitrogen mustard. *Oncogene* **14:**2127–2136.

Fan S, el-Deiry WS, Bae I, Freeman J, Jondle D, Bhatia K, Fornace AJ Jr, Magrath I, Kohn KW, and O'Connor PM (1994) p53 gene mutations are associated with decreased sensitivity of human lymphoma cells to DNA damaging agents. *Cancer Res* **54:**5824–5830.

Fan S, Smith ML, Rivet DJ, Duba D, Zhan Q, Kohn KW, Fornace AJ Jr, and O'Connor PM (1995) Disruption of p53 function sensitizes breast cancer MCF-7 cells to cisplatin and pentoxifylline. *Cancer Res* **55:**1649–1654.

Goodwin S, Smith AF, and Horning EC (1959) Alkaloids of Ochrosia elliptica Labill. *J Am Chem Soc* **81:**1903–1908.

Gribble GW (1990) Synthesis and antitumor activity of ellipticine alkaloids and related compounds, in *The Alkaloids: Chemistry and Pharmacology* (Brossi A, ed) pp 239–352, Academic Press, San Diego.

Harris CC (1996) p53 tumor suppressor gene: from the basic research laboratory to the clinic: an abridged historical perspective. *Carcinogenesis* **17:**1187–1198.

Hollstein M, Sidransky D, Vogelstein B, and Harris CC (1991) p53 mutations in human cancers. *Science (Washington DC)* **253:**49–53.

Jurayi J, Haugwitz D, Varma RK, Paull KD, Barrett JF, and Cushman M (1994) Design and synthesis ellipticinium salts and 1,2-dihydroellipticines with high selectivities against human CNS cancers in vitro. *J Med Chem* **37:**2190–2197.

Kenney S, Vistica DT, Linden H, and Boyd MR (1995) Uptake and cytotoxicity of 9-methoxy-N^2-methylellipticinium acetate in human brain and non-brain tumor cell lines. *Biochem Pharmacol* **49:**23–32.

Klopman G, Shi LM, and Ramu A (1997) Quantitative structure-activity relationship of multidrug resistance reversal agents. *Mol Pharmacol* **52:**323–334.

Kohn KW, Waring MJ, Glaubiger D, and Friedman CA (1975) Intercalative binding of ellipticine to DNA. *Cancer Res* **35:**71–76.

Koo H-M, Monks A, Mikheev A, Rubinstein LV, Gray-Goodrich M, McWilliams MJ, Alvord WG, Oie HK, Gazdar AF, Paull KD, Zarbl H, and Vande Woude GF (1996) Enhanced sensitivity to 1-*beta*-D-arabinofuranosylcytosine and topoisomerase II inhibitors in tumor cell lines harboring activated ras oncogenes. *Cancer Res* **56:**5211–5216.

Koutsoukos AD, Rubinstein LV, Faraggi D, Simon RM, Kalyandrug S, Weinstein JN, Kohn KW, and Paull KD (1994) Discrimination techniques applied to the NCI in vitro anti-tumour drug screen: predicting biochemical mechanism of action. *Stat Med* **13:**719–730.

Li G, Waltham M, Unsworth E, Treston A, Mushine J, Anderson NL, Kohn KW, and Weinstein JN (1997) Rapid protein identification from two-dimensional polyacrylamide gels by MALDI mass spectrometry. *Electrophoresis* **18**:391–402.

Lowe SW and Jacks T (1997) (Scientific correspondence). *Nature (Lond)* **385:**124–125.

Milne GWA, Nicklaus MC, Driscoll JS, Wang S, and Zaharevitz D (1994) National Cancer Institute drug information system 3D database. *J Chem Inf Comput Sci* **34:**1219–1224.

Monks A, Scudiero DA, Shoemaker RH, Paull KD, Vistica D, Hose C, Langley J, Cronise P, Vaigro-Wolff A, Gray-Goodrich M, Campell H, Mayo J, and Boyd MR (1991) Feasibility of a high-flux anticancer screen using a diverse panel of cultured human tumor lines. *J Natl Cancer Inst* **83:**757–766.

Myers TG, Waltham M, Li G, Buolamwini JK, Scudiero DA, Rubinstein LV, Paull KD, Sausville EA, Anderson NL, and Weinstein JN (1997) A protein expression database for the molecular pharmacology of cancer. *Electrophoresis* **18:**647–653.

O'Connor PM, Jackman J, Bae I, Myers TG, Fan S, Mutoh M, Scudiero DA, Monks A, Sausville EA, Weinstein JN, Friend S, Fornace AJ Jr, and Kohn KW (1997) Characterization of the p53-tumor suppressor pathway in cell lines of the National Cancer Institute anticancer drug screen and correlations with the growth-inhibitory potency of 123 anticancer agents. *Cancer Res* **57:**4285–4300.

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

Paull KD, Shoemaker RH, Hodes L, Monks A, Scudiero DA, Rubinstein L, Plowman J, and Boyd MR (1989) Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. *J Natl Cancer Inst* **81:**1088–1092.

Rogers D and Hopfinger AJ (1994) Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J Chem Inf Comput Sci* **34:**854–866.

Sainsbury M (1990) Ellipticines, in *The Chemistry of Antitumor Agents* (Wilman D, ed) pp 410–435, Chapman and Hall, New York.

Shi LM, Fan Y, Myers TG, Paull KD, and Weinstein JN (1997a) Mining the anticancer activity database generated by the NCI anticancer drug discovery program using statistical and artificial intelligence techniques. *Mathematic Model Sci Comput* **8**, in press.

Shi LM, Fan Y, Myers TG, and Weinstein JN (1997b) Genetic function approximation in the molecular pharmacology of cancer. Proceedings of the 1997 International Conference on Neural Networks (ICNN '97), Houston, TX, pp 2490–2493. Institute of Electrical and Electronics Engineers, Piscataway, NJ.

Shi LM, Fan Y, Myers TG, O'Conner PM, Paull KD, Friend SH, and Weinstein JN (1998) Mining the NCI anticancer drug discovery databases: genetic function approximation for the QSAR study of anticancer ellipticine analogs. *J Chem Inf Comput Sci* **38,** in press.

Shimamoto T, Imajo S, Honda T, Yoshimura S, and Ishiguro M (1996) Structure-activity relationship study on N-glycosyl moieties through model building of DNA and ellipticine N-glycoside complex. *Bioorg Med Chem Lett* **6:**1331–1334.

van Osdol WW, Myers TG, Paull KD, Kohn KW, and Weinstein JN (1994) Use of the Kohonen self-organizing map to study the mechanisms of action of chemotherapeutic agents. *J Natl Cancer Inst* **86:**1853–1859.

Varia SA, Schuller S, Sloan KB, and Stella VJ (1984) Phenytoin prodrugs. III: water-soluble prodrugs for oral and/or parenteral use. *J Pharm Sci* **73:**1068–1073.

Vistica DT, Kenney S, Hursey ML, and Boyd MR (1994) Cellular uptake as a determinant of cytotoxicity of quaternized ellipticines to human brain tumor cells. *Biochem Biophys Res Commun* **200:**1762–1768.

Weinstein JN, Kohn KW, Grever MR, Viswanadhan VN, Rubinstein LV, Monks AP, Scudiero DA, Welch L, Koutsoukos AD, Chiausa AJ, and Paull KD (1992) Neural computing in cancer drug development: predicting mechanism of action. *Science (Washington DC)* **258:**447–451.

Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJ, Jr., Kohn KW, Fojo T, Bates SE, Rubinstein LV, Anderson NL, Buolamwini JK, van Osdol WW, Monks AP, Scudiero DA, Sausville EA, Zaharevitz DW, Bunow B, Viswanadhan VN, Johnson GS, Wittes RE, and Paull KD (1997) An information-intensive approach to the molecular pharmacology of cancer. *Science (Washington DC)* **275:**343–349.

Weinstein JN, Myers T, Buolamwini J, Raghavan K, van Osdol W, Licht J, Viswanadhan VN, Kohn KW, Rubinstein LV, Koutsoukos AD, Monks A, Scudiero DA, Anderson NL, Zaharevitz D, Chabner BA, Grever MR, and Paull KD (1994) Predictive statistics and artificial intelligence in the U. S. National Cancer Institute's Drug Discovery Program for Cancer and AIDS. *Stem Cells* **12:**13–22.

Send reprint requests to: John N. Weinstein, M.D., Ph.D., NCI, NIH, Building 37, Room 5D05, 9000 Rockville Pike, Bethesda, MD 20892. E-mail: weinstein@dtpax2.ncifcrf.gov