Journal of Computer-Aided Molecular Design, **15:** 219–234, 2001.
KLUWER/ESCOM

A multivariate insight into the *in vitro* antitumour screen database of the National Cancer Institute: classification of compounds, similarities among cell lines and the influence of molecular targets

Giuseppe Musumarra*, Daniele F. Condorelli*, Alessandro S. Costa & Maria Fichera
Dipartimento di Scienze Chimiche, Università di Catania, Viale A. Doria 6, 95125 Catania, Italy

Received 27 June 2000; accepted 12 October 2000

Key words: antitumour agents, chemoinformatics, biometrics, molecular targets, NCI antitumour screen database

Summary

A multivariate insight into the in vitro antitumour screen database of the NCI by means of the SIMCA package allows to propose hypotheses on the mechanism of action of novel anticancer compounds. As an example, the application of multivariate analysis to the NCI standard database provided clues to the classification of drugs whose mechanism is either unknown or controversial. Moreover, the influence of intrinsic biochemical cell line properties (molecular targets) on the sensitivity to drug treatment could be evaluated simultaneously for classes of compounds which act by the same mechanism. Interestingly, the present approach can also provide a correlation between the molecular targets and the therapeutical fingerprint of novel active compounds thus suggesting specific biochemical studies for the investigation of new mechanisms of drug action and resistance. The statistical approach reported here represents a valuable tool for handling the enormous data sets deriving from recent genome-wide investigations of gene expression in the NCI cell lines.

Introduction

Prodigious information has been accumulated in the last decade by the National Cancer Institute (NCI) and the *in vitro* anticancer screening program employs several human tumour cell lines that have been grouped in disease subpanels including leukemia, non-small-cell lung, small-cell lung, central nervous system, colon, melanoma, ovarian, and renal tumours cell lines. The standard database includes now 175 compounds for which log GI_{50}, log TGI and log LC_{50} have been determined [1].

By means of COMPARE [2, 3] a probe or 'seed' compound, specified by using the compound's NCI accession number (the NSC number), can be compared with all compounds in the entire database, which are ranked in order of the similarity of the responses of 60 selected cell lines for the compounds in the database with respect to the responses of the same cell lines for the seed compound. Whereas the results are based on COMPARE, pairwise correlations indicate that compounds high in this ranking may act through a mechanism of action similar to that of the seed compound. The COMPARE methodology is also able to provide an estimate of the similarity between a single compound and a single molecular target (or between two molecular targets) by correlating their 'responses' with respect to a given number of cell lines. COMPARE is widely used and has been applied in the preliminary screening of new drugs, helping us to form an approximate idea on their mechanism of action.

In the past two decades, multivariate data analysis by means of the SIMCA (Soft Independent Modelling of Class Analogy) method [4–6] has been applied in many fields of science demonstrating to be most suitable in handling complex data sets represented in the form of matrices where a number of objects are characterized by a number of variables. The main advantage of multivariate analysis is that it investigates the relationships among all objects and all variables

*To whom correspondence should be addressed. E-mail: gmusumarra@dipchi.unict.it or condorda@mbox.unict.it

simultaneously, providing a statistically reliable overall picture of the database under study. The application of the above methodology allows to have an insight into existing databases from a different viewpoint and results in a significant time saving for the collection and for the preliminary examination of new data.

We here report the first multivariate insight into the *in vitro* antitumour screen database of the NCI, with the aim to illustrate the advantages of looking from a new perspective at this database and to point out the potentialities of the multivariate approach in future data pattern recognition.

The NCI *in vitro* antitumour screen database is a precious information source which can be represented in the form of a matrix, where the standard compounds (e.g. objects) may be represented as characterized by a multivariate biological 'fingerprint', given by the activity patterns with respect to more than one hundred cell lines (e.g. variables). Such a matrix is suitable for multivariate data analysis by means of the SIMCA method [4–6] including Principal Component Analysis (PCA), which is able to detect similarities among variables (cell lines), among objects (compounds) and to provide a statistically reliable criterion for compounds classification based on their *in vitro* potential therapeutic activity fingerprint *vs.* different cancer cells, which is expected to be related to the mechanisms of drug action and drug resistance. PCA, which investigates the relationships among all objects and all variables in a single context, can be applied to the NCI database in an alternative way by considering the cell lines as 'objects' characterized by the compounds activities ('variables'). In the latter case the SIMCA method, aimed at objects classification, could evidence a biometric cell line grouping based on the *in vitro* antitumour activity of the standard database compounds. In the above alternative matrix arrangement (i.e. cell lines as objects) the so-called molecular targets (i.e. intrinsic biochemical cell line properties that can influence sensitivity to drug treatment) can be also inserted as 'variables' and their overall similarities evidenced by SIMCA, by considering all variables simultaneously, rather than by pairwise correlations as in COMPARE.

PCA is able to classify objects and to evaluate the variable information content, but cannot establish cause-effect relationships between a set of descriptor variables (the X matrix) and a dependent variable (the response). The latter objective can be achieved by partial least squares in latent variables (PLS), another multivariate method included in the SIMCA package [5], aimed at finding the relationships between one or more 'dependent' variable(s) and a group of explanatory variables (descriptors). Furthermore the block of molecular targets (called the cell lines 'descriptor' variables in the PLS procedure) can be related to the therapeutical 'fingerprint' of a given compound (called the 'response' in the PLS procedure) *vs.* the same cell lines and the influence of each molecular target in determining the response that can be evaluated by means of a PLS statistical parameter called "descriptor variables modelling power". The application of the above methodology to compounds whose molecular targets are well known will enable to test the consistency of the PLS results with current knowledge.

In this context, the present multivariate insight into the *in vitro* antitumour screen database of the National Cancer Institute is aimed at highlighting the following points:

i. compounds classification.
   PCA will be used to gain information about compounds with unclassified mechanism of action, either by postulating the biometrically most probable mechanism of action for a given compound or alternatively by ruling out the possibility that certain mechanisms may be operating. Moreover, it will be possible to verify if the known mechanism of action is unique or alternative mechanisms may also be postulated for a given compound;
ii. cell line classification.
   A biometric cell line classification will be attempted by evaluating if cell lines belonging to the same organ or with similar biochemical characteristics exhibit similar information content.
iii. influence of molecular targets on the *in vitro* drug sensitivity.
   The PLS algorithm allows to relate the molecular targets to the activity of a single compound or of a class (group) of compounds.

Multivariate Statistical Methods

The data sets used for the PCA were tables (matrices), in which a number of biological cell line responses (variables) were given for each compound (object) in the cases of models 1–3, and *vice versa* a number of biological responses of compounds or molecular targets (variables) provided for each cell line (object) in models 4–8.

The results of PCA depend upon the weighting of the data; in SIMCA the variables are generally

autoscaled by multiplying the variables by appropriate weights (the reciprocal of the variable standard deviation) to give them unit variance (i.e. the same importance).

In models 1–3 the SIMCA method [4–6] carried out PCA on a data matrix containing elements x_{ik}, where index k is used for the experimental measurements (variables) and index i for the samples (objects). Autoscaled matrix elements, are then fitted into model given by Equation (1), where the number A of significant cross terms (components), and the parameters p_{ak} and t_{ia} are calculated by minimising the residuals e_{ik}, after subtracting $\overline{x}_k$ (the mean value of the i experimental quantities x_k).

$$x_{ik} = \overline{x}_k + \sum_{a=1}^{a=A} t_{ia}\, p_{ak} + e_{ik} \qquad (1)$$

In models 1–3, parameters $\overline{x}_k$ and p_{ak} (the loadings) depend only on the cell lines (variables), and the t_{ia} (scores) only on the compounds. In models 4–8, where the cell lines were considered as objects and the compounds or molecular targets as variables, the scores refer to the cell lines and the loadings to the drugs or to the molecular targets.

The deviations from the model are expressed by the residuals e_{ik}. The number of significant components (A) is determined using the cross-validation technique [7].

The relevance of each variable in describing the mathematical model is given by its modelling power Ψ_k (Equation 2) where s_k is the residual standard deviation for each variable after A dimensions and after dimension zero.

$$\Psi_k = \frac{1 - s_k(A = A)}{s_k(A = 0)} \qquad (2)$$

However, PCA is not aimed at finding out cause-effect relationships. A correct statistical approach, able to cope both with the interpretation of results and the prediction of unmeasured data, is provided by a method called PLS (partial least squares or projection to latent structures) analysis [8–10].

When there is more than a single dependent variable, we have two blocks of variables, and it is possible to define a 'dependent' matrix Y and an 'independent' matrix X [9–11]. The question under investigation is whether or not the response or the members of the Y matrix can be described as a function of the members of the X matrix. Rather than computing principal components models for each of the two matrices and looking for a linear relationship between the principal components of these two blocks by a two step procedure, the PLS algorithm achieves these two steps simultaneously.

This method, denoted PLS if there is a single dependent variable and PLS2 with more than one dependent variable, is computationally much faster than PCA followed by multivariate regression analysis (MRA), and leads to a better prediction of the members of the Y matrix [11]. The PLS method gives a description of the X matrix by the one principal component-like model (Equation 1) and a description of the *y* vector as a predictive relation with the latent variables *t* (Equation 3), where b_a is a proportionality coefficient for each dimension.

$$y_{ia} = \sum b_a t_{ia} + h_{ia} \qquad (3)$$

The algorithm used in the SIMCA/Macup method, presented in detail elsewhere [9] is iterative for each dimension as in PCA. It consists in finding the latent variables of the X matrix t_{ia} in such a way that the relationship between y_i and t_i is maximized.

The statistical results obtained by the PLS method may be used to detect what variables in the X block are relevant to determine the dependent variable(s) and to predict the response (*y* variables) for unmeasured samples.

Details on the objects and on the variables for the data sets used in all PCA and PLS calculations are available on the WWW under http://www.unict.it/dipchi/ or from the corresponding authors.

Results and Discussion

Classification of compounds according to their mechanism of action

NCI reports the *in vitro* antitumour activity of 175 compounds, expressed as log GI_{50} and log LC_{50}, for 106 cell lines. The NCI renamed the IC_{50} value, the concentration that causes 50% growth inhibition (cytostatic effect), as the GI_{50} value to emphasize the correction for the cell count at time zero; thus, GI_{50} is the concentration of test drug where $100 \cdot (T - T_0)/(C - T_0) = 50$. The LC_{50}, which indicates a cytotoxic effect, is the concentration of drug where $100 \cdot (T - T_0)/T_0 = -50$. For some compounds the mechanism of action is known and is coded by the first letter in the second column of Table A provided in the WWW [12]: **A** = alkykating agents, **R**

Table 1. PCA for models 1–3.

		N. of Objects	N. of Variables	N. of PC	% variance explained
Model 1		171[a]	71[b]	6	97.4(94+1.5+0.7+0.6+0.3+0.3)
Class	A	36	71	6	98.8 (95.9+1+0.7+0.5+0.4+0.3)
	R	13	71	5	95.4 (79.5+8.5+3.4+2.3+1.7)
	D	16	71	6	98.0 (84.2+7.9+2.2+1.5+1.2+1.0)
	M	6	71	3	98.1 (88.5+7.4+2.2)
	T	16	71	6	98.1 (87.9+4.9+3.0+1.2+0.6+0.5)
Model 2		62[a]	71[b]	4	93.7 (86.3+5.1+1.4+0.9)
Class	A	10	71	4	99.1 (97.1+1.1+0.5+0.5+0.4)
	D	5	71	2	99.1 (96.9+2.2)
	M	5	71	2	97.2 (92.7+4.5)
	T	12	71	4	87.5 (59.1+16.3+7.1+5.0)
Model 3		62[a]	142[b]	5	93.8 (79.8+9.9+2.4+0.9+0.8)
Class	A	10	142	4	99.0 (96.0+2.0+0.5+0.5)
	D	5	142	2	82.0 (59.9+22.1)
	M	5	142	2	94.7 (87.1+7.6)
	T	12	142	5	92.6 (64.8+14.6+6.2+4.8+2.2)

[a] See supporting information: Table A, column 3.
[b] See supporting information: Table B, column 3.

= RNA/DNA antimetabolites,**D** = DNA antimetabolites, **M** = antimitotic agents, **T** = topoisomerase II inhibitors, while for many others (84 compounds) it is unclassified and designated by letter **U** = unclassified.

It is noteworthy that for many compounds designated by **U**, a possible mechanism has already been postulated, but it is not reported in the database probably because it is not unique or has not been yet confirmed. A multivariate biometric classification of the above unclassified compounds could therefore be conveniently compared with available literature data on their mechanism of action.

GI_{50} Data (Model 1)

After selecting 71 of the 106 variables (cell lines) of the log GI_{50} data set, excluding those exhibiting many missing values and including 171 objects (compounds listed in table A and cell lines listed in table B, both provided as supporting information [12]), a data matrix with 12141 elements (data) was analysed by principal component analysis (PCA) using the SIMCA program. The analysis provided 6 Principal components (PC) explaining 97.4% total variance (see Table 1), with the first PC explaining already 94% variance ordering the compounds in order of average activity. The number of significant PC was determined according to the cross validation procedure [7] adopted by the SIMCA package [5].

In the present work, compounds with known mechanism of action were divided into five separate classes each including drugs with the same mechanism, (**A**: alkylating agents, **R**: RNA/DNA antimetabolites, **D**: DNA antimetabolites, **M**: antimitotic agents, **T**: topoisomerase II inhibitors), as proposed in the NCI standard database [1] and already used as a training set for neural network analysis of drug mechanism of action [2]. Of course alternative and more detailed classification criteria might be adopted according to the specific aim of the analysis.

The SIMCA classification procedure implies, in the present case, calculation of separate PCA class models for each set of compounds with the same mechanism of action (classes **A**, **R**, **D**, **M**, **T**), fitting of compounds with unclassified mechanism of action (e.g. those whose code in the second column of table A starts with letter **U**) into each class, and their classification by means of their residual standard deviation (RSD) with respect to each class.

The above approach can easily evidence outliers belonging to neither class [4–5, 11] as well as objects

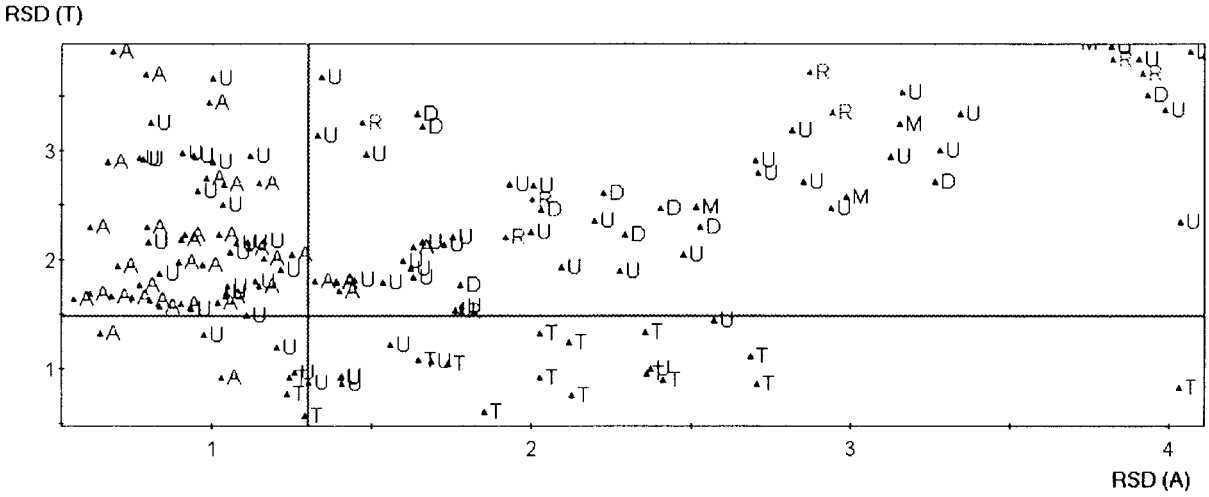

Figure 1. Cooman's plot for drug classification into classes **A** and **T**.

which belong to more than one class by inspection of the RSD for each class. The present method, providing a tool to classify **U** compounds, represents therefore an advancement with respect to previous classification of compounds 'with known mechanism of action' by LDA, KNN, neural networks and PCA scores plots [3c].

RSD values relative to unclassified compounds are recorded in Table 2, from which it is possible to evaluate the 'goodness' of fit for each compound into each class: the lower the RSD, the better the fit.

A graphical comparison of compounds fitting into two classes is shown by the so called Cooman's plot [5], where compounds in the upper right rectangle belong to neither class and those in the lower left rectangle belong to both classes. Figure 1 points out a good separation of classes A and T, while Figure 2 evidences that RNA/DNA antimetabolites (**R**) are well separated from alkylating agents (**A**), differently than alkylating agents which are located in the lower left rectangle, and therefore can be also classified as RNA/DNA antimetabolites. This finding is not surprising taking into account that classification as alkylating agent is based on a chemical criterion rather than on a biological one. Indeed the effects of some RNA/DNA antimetabolites and alkylating compounds are mainly mediated by a block in DNA based processes such as transcription and replication.

In conclusion, model 1, based on GI_{50} data only, indicates that not always the compounds belong to a single class. Moreover, it was possible to assign to a single class (either to RNA/DNA antimetabolites or to topoisomerase II inhibitors) 19 of the 84 compounds with unclassified mechanism of action, in particular:

RNA-DNA antimetabolites: 15200 (gallium nitrate), 24559 (mitramycin), 126849 (3-deazauridine), 127755 (triazinate), 129943 (ICRF-159), 169780 (ICRF-1), 284751(8Cl-cyc-AMP), 339004 (Chloroquinoxaline sulfonamide), 339555 (Bryostatin 1), 339638 (fostriecin), 375575 (cyclopentenylcytosine).

Topoisomerase II inhibitors : 3053 (actinomycin D), 58514 (Chromomycin A3), 165563 (bruceantin), 349156 (pancratiastatin), 526417 (echinomycin).

LC_{50} Data (Model 2)

A similar analysis was carried out with the log LC_{50} data set in order to obtain better separated classes, as that these data contain better information about the selectivity of the compounds. After selecting 62 of the 171 compounds, excluding those with no selectivity, and 71 of the 106 variables, not including those with many missing values, the new matrix (elements) was analysed by principal component analysis whose results are summarised in Table 1.

Only four of the previous five classes could be defined, as class R would contain only a few compounds (objects). Table 2 shows that model 2 (LC_{50} data only) allows classification of six of the 29 unclassified compounds into a single class, either **A**: alkylating agent or **T**: topoisomerase 2 inhibitor), in particular:

Alkylating agents: 3053 (actinomycin D), 253272 (caracemide).

Topoisomerase 2 inhibitors: 58514 (chromomycin A3), 141633 (homoharringtonine), 325014 (bactobolin), 609699 (topotecan).

Table 2. SIMCA classification of compounds with unknown mechanism of action: RSD values for fitting in classes from models 1–3[a].

		Classes from model 1					Classes from model 2[b]				Classes from model 3[b]			
N.	Code	A	R	D	M	T	A	D	M	T	A	D	M	T
	RSD[c]	1.30	1.49	1.46	1.77	1.47	1.55	1.71	1.73	1.48	1.48	1.62	1.64	1.46
7	U3051	**0.96**	1.67	**1.01**	6.85	5.73	-	-	-	-	-	-	-	-
8	U3053	2.57	3.08	4.26	2.09	**1.44**	**1.02**	3.66	2.44	5.22	**1.46**	9.75	2.84	4.21
10	U4728	1.60	**1.26**	**1.33**	5.43	4.25	-	-	-	-	-	-	-	-
15	U13875	0.80	0.93	1.01	3.53	2.16	-	-	-	-	-	-	-	-
16	U15200	2.80	**1.48**	2.01	5.90	5.17	2.28	2.53	2.81	4.64	2.47	13.0	7.70	6.66
18	U21548	2.12	**1.29**	**1.37**	5.71	4.38	-	-	-	-	-	-	-	-
19	U23759	**0.94**	**1.01**	**0.93**	4.35	2.95	-	-	-	-	-	-	-	-
20	U24559	2.28	**1.31**	1.93	2.26	1.89	4.13	5.01	4.83	3.19	3.54	2.69	2.90	3.84
22	U26271	**0.96**	**0.98**	**0.95**	3.86	2.63	-	-	-	-	-	-	-	-
26	U32946	3.35	1.94	2.31	3.28	3.34	2.90	2.89	3.34	4.93	2.86	5.81	3.98	4.40
28	U37364	**1.12**	**1.06**	**1.17**	4.34	2.96	-	-	-	-	-	-	-	-
29	U38721	2.00	**1.16**	**1.37**	2.34	2.25	-	-	-	-	-	-	-	-
30	U45388	1.63	**1.12**	**1.19**	2.86	1.92	-	-	-	-	-	-	-	-
34	U58514	2.37	2.59	3.27	2.23	**1.00**	3.29	5.01	5.03	**1.36**	2.84	7.33	3.08	**1.39**
37	U68075	1.45	**1.15**	**1.45**	1.96	1.81	-	-	-	-	-	-	-	-
41	U77037	1.54	**1.11**	**1.23**	1.84	1.79	-	-	-	-	-	-	-	-
42	U77213	**0.91**	**1.01**	**1.04**	4.38	2.98	-	-	-	-	-	-	-	-
46	U85998	0.81	**1.03**	**1.00**	4.58	3.25	-	-	-	-	-	-	-	-
49	U95580	**1.20**	**1.40**	**0.99**	6.55	5.31	-	-	-	-	-	-	-	-
53	U104801	1.66	**1.09**	**1.39**	2.25	2.16	1.88	2.48	2.64	3.14	1.76	5.64	4.20	3.05
55	U109229	4.99	5.57	4.75	11.9	11.5	-	-	-	-	-	-	-	-
56	U109724	**0.79**	**0.99**	**0.98**	4.24	2.92	-	-	-	-	-	-	-	-
57	U118742	1.63	**1.25**	**1.24**	5.20	4.36	-	-	-	-	-	-	-	-
62	U125066	3.90	2.57	2.76	3.62	3.84	-	-	-	-	-	-	-	-
65	U126849	3.86	**1.39**	2.56	4.51	4.05	-	-	-	-	-	-	-	-
67	U127755	2.82	**1.15**	2.16	2.24	3.19	1.74	1.80	1.88	1.91	2.13	2.06	2.91	2.24
68	U129943	1.93	**1.22**	1.54	3.93	2.70	-	-	-	-	-	-	-	-
70	U132319	**0.90**	1.63	**0.92**	6.82	5.62	-	-	-	-	-	-	-	-
71	U133100	1.60	**1.16**	**1.09**	2.49	1.98	**1.46**	**1.42**	1.79	**1.32**	**1.46**	5.60	4.67	2.06
74	U139490	3.61	1.56	2.59	5.03	4.25	-	-	-	-	-	-	-	-
75	U141537	**1.26**	1.90	2.79	2.25	**0.96**	3.96	5.33	5.18	3.10	2.91	6.61	3.33	2.59
77	U141633	3.82	2.38	2.94	2.52	3.96	1.71	2.92	2.67	**1.38**	2.28	4.31	2.96	1.94
82	U150014	1.48	**0.87**	**1.10**	4.29	2.97	-	-	-	-	-	-	-	-
85	U157365	5.95	7.85	7.22	15.0	14.3	-	-	-	-	-	-	-	-
88	U165563	1.41	1.68	2.44	2.12	**0.86**	4.90	6.54	6.40	3.18	3.56	5.83	3.82	2.74
90	U169780	2.01	**1.31**	1.52	3.90	2.68	-	-	-	-	-	-	-	-
93	U180973	0.98	**0.92**	**1.12**	**1.62**	**1.31**	**1.22**	1.96	**1.29**	**0.89**	**1.44**	2.62	4.84	2.01
95	U192965	**1.21**	**0.96**	**1.24**	**1.55**	**1.19**	3.19	3.21	3.25	2.65	2.38	1.88	4.70	2.90
96	U208734	**1.30**	**1.49**	2.26	2.11	**0.87**	2.64	3.55	3.86	1.83	2.24	5.36	2.73	1.48
97	U218321	**0.77**	**0.98**	**0.94**	4.35	2.94	-	-	-	-	-	-	-	-
99	U226080	5.48	2.95	3.87	3.98	5.35	**1.31**	**1.62**	**1.50**	1.81	3.16	3.45	3.39	2.87
100	U237020	2.23	2.95	1.70	8.73	7.59	6.30	7.47	10.1	23.0	3.47	19.7	10.4	21.8
103	U253272	1.76	**1.13**	**1.04**	2.96	2.21	**1.44**	1.75	2.16	2.81	**1.33**	6.65	4.78	3.09
105	U261726	2.09	**1.14**	**1.24**	2.72	1.93	-	-	-	-	-	-	-	-
107	U267213	**1.00**	**1.06**	**0.97**	4.40	2.90	-	-	-	-	-	-	-	-

Table 2. Continued.

		Classes from model 1					Classes from model 2[b]				Classes from model 3[b]			
N.	Code	A	R	D	M	T	A	D	M	T	A	D	M	T
	RSD[c]	1.30	1.49	1.46	1.77	1.47	1.55	1.71	1.73	1.48	1.48	1.62	1.64	1.46
112	U278214	**1.14**	**1.00**	**1.20**	2.93	1.80	-	-	-	-	-	-	-	-
113	U280594	6.25	2.60	3.68	4.57	5.31	-	-	-	-	-	-	-	-
114	U281272	3.28	1.80	1.67	4.07	3.00	-	-	-	-	-	-	-	-
115	U283162	**1.05**	**0.93**	**1.12**	2.09	1.69	**1.37**	1.91	**1.30**	**1.24**	2.32	5.26	6.01	2.86
116	U284356	**1.08**	**0.93**	**1.07**	3.68	2.14	-	-	-	-	-	-	-	-
117	U284751	2.94	**1.47**	1.64	2.79	2.47	-	-	-	-	-	-	-	-
118	U286193	3.78	2.01	2.80	3.79	4.10	-	-	-	-	-	-	-	-
119	U291643	0.84	**0.92**	**0.76**	3.02	1.87	-	-	-	-	-	-	-	-
120	U293015	**1.12**	**1.12**	**1.25**	3.49	2.16	-	-	-	-	-	-	-	-
124	U303861	**1.06**	**0.85**	**0.84**	3.20	2.07	**1.01**	**1.03**	**1.72**	3.41	**0.94**	7.79	5.04	3.78
125	U305884	1.56	**1.16**	**1.39**	1.98	**1.22**	**1.16**	**1.56**	1.78	**0.93**	**1.25**	3.79	4.98	**1.20**
127	U312887	3.12	1.82	1.85	3.74	2.94	-	-	-	-	-	-	-	-
128	U314055	1.35	**1.31**	**1.28**	5.18	3.68	-	-	-	-	-	-	-	-
129	U320846	1.63	**1.18**	**1.27**	2.72	1.83	-	-	-	-	-	-	-	-
130	U321803	**0.94**	**0.94**	**1.14**	2.29	1.54	-	-	-	-	-	-	-	-
131	U322921	3.16	1.63	1.95	3.77	3.54	-	-	-	-	-	-	-	-
132	U325014	1.69	**1.27**	1.65	**1.72**	**1.06**	2.00	2.80	2.69	**1.30**	1.83	2.59	2.29	**1.42**
133	U325319	1.73	2.47	3.19	2.42	2.14	4.07	5.40	5.43	3.16	3.16	7.20	3.81	3.19
134	U326231	2.16	1.74	1.67	6.20	5.28	-	-	-	-	-	-	-	-
135	U328426	4.03	2.32	3.31	2.37	2.34	1.70	3.35	2.54	2.69	2.26	6.78	3.12	2.55
139	U333856	1.76	**1.17**	1.71	**1.63**	1.53	**1.30**	2.25	1.89	**1.03**	**1.40**	**1.37**	4.01	1.82
140	U336628	**1.11**	**1.10**	**1.15**	2.40	1.48	-	-	-	-	-	-	-	-
142	U338720	**1.05**	**1.07**	**1.15**	2.58	1.76	-	-	-	-	-	-	-	-
144	U339004	2.71	**1.43**	1.88	3.58	2.80	-	-	-	-	-	-	-	-
145	U339555	2.48	**1.46**	1.81	2.44	2.04	-	-	-	-	-	-	-	-
146	U339638	2.20	**1.27**	1.55	2.19	2.35	2.35	2.52	2.51	2.18	2.06	3.26	3.77	2.14
147	U343513	**1.22**	**1.12**	**1.22**	2.93	1.91	-	-	-	-	-	-	-	-
149	U347512	**1.04**	**1.19**	**1.16**	3.91	2.51	-	-	-	-	-	-	-	-
151	U349156	1.41	1.52	2.08	2.04	**0.92**	2.80	3.91	3.73	2.03	2.40	4.69	2.47	1.94
153	U349438	**1.25**	**1.47**	**1.10**	6.51	5.39	-	-	-	-	-	-	-	-
158	U356894	1.34	**1.09**	**1.13**	4.66	3.14	-	-	-	-	-	-	-	-
160	U361456	**1.01**	**1.12**	**0.93**	4.98	3.67	-	-	-	-	-	-	-	-
163	U366241	3.99	1.90	2.54	2.70	3.37	-	-	-	-	-	-	-	-
165	U375575	2.85	**1.37**	2.03	2.58	2.72	-	-	-	-	-	-	-	-
166	U406021	1.02	1.18	0.89	5.57	4.37	-	-	-	-	-	-	-	-
168	U526417	1.41	2.33	3.10	2.44	**0.92**	4.47	5.98	5.57	1.98	3.51	7.33	3.27	1.96
169	U602668	1.17	**1.11**	**1.09**	3.42	2.16	-	-	-	-	-	-	-	-
170	U609699	2.70	2.10	2.14	2.94	2.91	1.65	2.37	2.37	**1.34**	2.27	5.08	2.31	2.29
171	U619003	1.78	1.58	1.61	2.16	1.56	**1.40**	2.37	2.22	**0.97**	**1.44**	4.22	2.80	**1.02**

[a] Values in bold indicate that the compound belongs to the class.
[b] See supporting information: Table A, column 3.
[c] RSD class limit value.

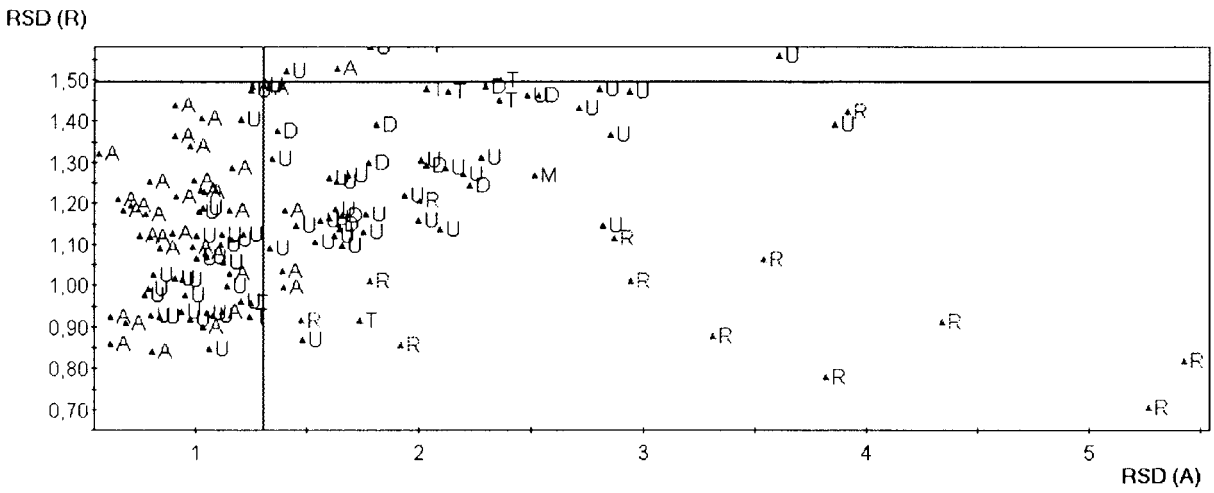

Figure 2. Cooman's plot for drug classification into classes **A** and **R**.

GI_{50} + LC_{50} Data (Model 3)

A new matrix containing the same 62 compounds (objects) and 71 log GI_{50} plus 71 log LC_{50} values, with respect to the same cell lines (142 variables), was analysed. The results of this PCA including both GI_{50} and LC_{50} data are recorded in Table 1, show that the p_1 loadings separate well log GI_{50} values from log LC_{50} and allow a new classification for eight of the 29 unclassified compounds:

Alkylating agents: 3053 (actinomycin D), 133100 (rifamycin SV), 180973 (tamoxifen), 253272 (caracemide), 303861 (L-cysteine analogue).

Topoisomerase 2 inhibitors: 58514 (chromomycin A3), 325014 (bactobolin), 208734 (aclacinomycin A).

In conclusion, as far as the compounds classification is concerned, the SIMCA classification results can be summarised as follows:

1. Comparison of the results for all the three PCA models shows that a unique consistent classification as topoisomerase 2 inhibitor can be achieved only for 58514 (chromomycin A3), while 349156 (pancratiastatin) and 526417 (echinomycin) are fitted in the same class only by model 1. Indeed, echinomycin and chromomycin A3 are nucleic-acid binding ligands. The antibiotic echinomycin is known to bisintercalate into DNA and chromomycin is thought to bind in the minor groove as a dimer to runs of GC base pairs [14, 15]. These DNA interactions may result both in a block of DNA transcription and in DNA damage. Topoisomerase 2 inhibitors form a ternary complex with topoisomerase 2 and DNA, leading to an accumulation of DNA breaks and cell death [16]. Therefore the inclusion of echinomycin and chromomycin A3 in the **T** class is consistent with previous information on the mechanism of action of these drugs and stimulates further studies on their effects in topoisomerase 2 activity. To this regard, it is worth mentioning that actinomycin D (3053, compound 8 in Table 2), another classical intercalating agent, is classified as topoisomerase 2 inhibitor using model 1 (GI_{50}). This result is consistent with the reported ability of actinomycin D to inhibit the activity of topoisomerase 2 [17]. However, when LC_{50} data are taken into consideration (models 2 and 3), actinomycin D shows a response profile more similar to that of alkylating agents. It is possible that at the higher doses necessary to achieve a lethal effect (LC_{50}) the ultimate mechanism causing cell death related to DNA damage is similar for alkylating agents and actinomycin D.
2. Compound 325014 (bactobolin) was classified only in class **T** by the second and third models, although it belongs to more classes in the first classification, i.e. **R** and **T**. Therefore this compound might share some crucial pathways of the cytotoxic effect with topoisomerase 2 inhibitors.
3. Compound 303861 (L-cysteine ethyl ester, S-(N-methylcarbamate monohydrochloride) belongs to more classes (**A**, **R** and **D**) in the first and second models and only to class **A** in the third model. It has been reported [18] that this compound behaves as an L-glutamine antagonist, selectively acting on the enzymes of purine nucleotide biosynthesis in accordance with its classification as **R** and **D**.

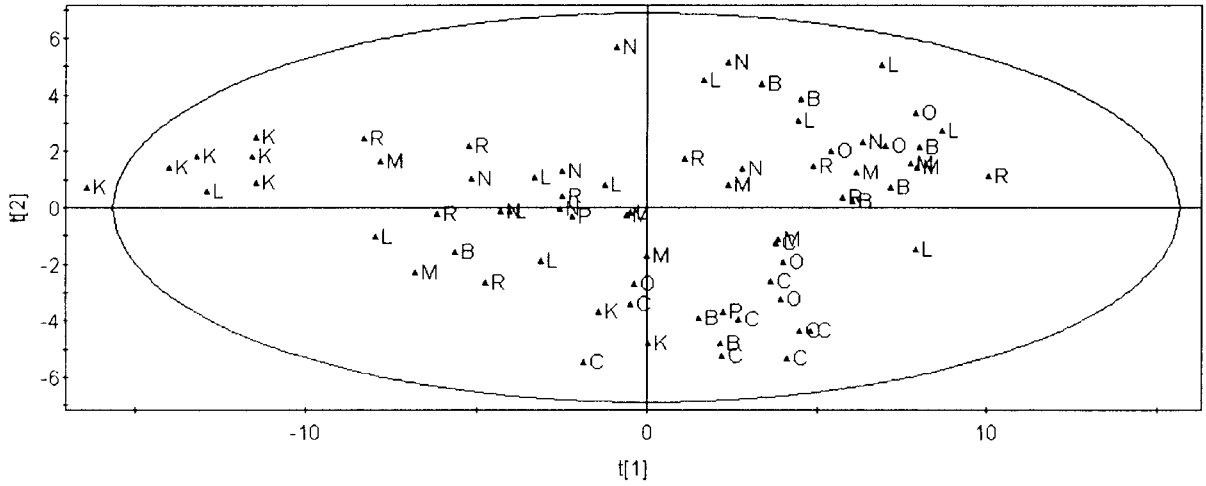

Figure 3. Scores plot from PCA (log GI_{50} data) using 71 cell lines as objects and 87 drugs as variables. Cell lines codes according to the tissue of origin are: **K**: leukaemia; **L**: non small cell lung; **C**: colon; **N**: CNS; **M**: melanoma; **O**: ovarian; **R**: renal; **P**: prostate; **B** breast.

4. The eleven unclassified compounds belonging to class **R** according to first model, for which no log LC_{50} data were available, could not be classified in any of the four defined classes in the second and third models. However, this classification, not yet supported by other models based on independent data, restricts the range of possible mechanisms, thus providing a firm basis for further studies.

Classification of cell lines

In order to evidence biometric cell line grouping based on the *in vitro* drug sensitivity profile (or on the so called molecular targets) PCA was carried out on a matrix arranged with 71 cell lines as objects (see Table B in supporting information) and their sensitivity (expressed as logGI_{50}) to 87 standard database drugs with known mechanism of action as variables (see Table D in supporting information). The above alternative arrangement, as already mentioned in the introduction, could exploit better the SIMCA object classification capability. Aim of such analysis was to establish if the cell line tissue of origin exerts any influence on the drug sensitivity pattern. Figure 3, reporting the $PC_1$and PC_2 scores accounting for 54% explained variance, points out grouping of leukaemia cell lines in the left part of the plot (i.e. negative t_1 values) with RPMI-8226 and K-562 exhibiting t_1 close to zero. This finding is consistent with the results of recent cell line cluster analysis showing that the above-cited two cell lines cluster separately from the other leukaemia cell lines [19]. Further PCA performed separately with each class of drugs with the same known mechanism of action (classes **A**, **R**, **D**, **M**, **T**) evidenced the same grouping of leukaemia cell lines only for alkylating agents and DNA antimetabolites, but not for other classes.

PCA using as variables the molecular targets, intrinsic biochemical properties which could influence the mechanism of drug action and drug resistance, did not reveal cell line grouping according to the tissue of origin. The novel techniques DNA microarrays have been recently used to analyse the expression of approximately 8000 unique genes among the 60 cell lines used in the NCI screen for anticancer drugs [19]. A consistent relationship between the gene expression patterns and the tissue of origin of the tumours was observed. The 33 molecular targets used in the present study were selected because of their possible involvement in the mechanism of drug resistance or drug activity. Therefore it is not surprising that the expression of this group of selected genes is not related to the histologic origin of the cell lines. Interestingly, Scherf et al. [20] reported a certain degree of cell line clustering according to the organ of origin on the basis of their drug responses, although it is not as strong as it was on the basis of gene expression. The lower degree of clustering according to tissue of origin, when classification is based on drug responses, is probably due to variations of expression of small group of genes which are responsible for drug sensitivity and resistance. Accordingly, our multivariate analysis, using selected molecular targets as variables, provides no significant grouping depending on the histologic origin of the cell lines.

Table 3. PLS for models 4–8.

c		N. of variables	N. of objects[a]	N. of PC	% of variance explained
Model 4 (**A**)	X	33[b]	60	2	17.3 (9.1+8.2)
	Y	36[c]			31.2 (24.8+6.4)
Model 5 (**R**)	X	33[b]	60	1[d]	9.7
	Y	13[c]			18.4
Model 6 (**D**)	X	33[b]	60	2	20.3 (8.9+11.4)
	Y	16[c]			27.2 (20.6+6.6)
Model 7 (**M**)	X	33[b]	60	1[d]	8.8
	Y	6[c]			26.7
Model 8 (**T**)	X	33[b]	60	1	9.9
	Y	16[c]			24.7

[a]See supporting information, Table C, column 3.
[b]See legend in Figure 4.
[c]See supporting information, Table D, column 3.
[d]The first PC was found to be hardly significant by cross-validation.

Molecular targets

A number of molecular and cellular features (called 'molecular targets'), suggested to be involved in the sensitivity to different therapeutical agents, were measured in the 60 cell line panel of the NCI. Indeed, several papers already reported the presence or absence of correlation between the levels of specific molecular targets and the *in vitro* drug sensitivity [21–25]. However, the PLS procedure of the SIMCA method (Table 3, models 4–8) allows a simultaneous analysis of all the molecular targets analysed up to now, providing a statistical parameter (the modelling power, W*) that indicates their correlation with the sensitivity to a class of compounds or to a single compound.

In Figures 4–8, W* values of 33 molecular targets are reported for each of the five classes of compounds of the NCI standard database. As already mentioned, the compounds of the standard database have been distinguished in five classes based on their mechanism of action (classes **A**, **R**, **D**, **M**, **T**). The bar plots of Figures 4–8 provide an immediate picture of the correlations which may be revealed by means of such an analysis. For example, the inverse correlation between the level of the P53 protein and the *in vitro* sensitivity of **T**, **D**, **A**, **R** compounds is in agreement with the previous analysis showing that the majority of clinical agents for cancer treatment is more active against the p53 wild-type cell lines than against the p53 mutant ones [22, 23]. The p53 tumor suppressor gene is mutated in 50% of human tumors and it is known that mutated p53 protein accumulates at higher intracellular levels [26–28].

This interpretation is supported by the observation that the level of X-ray induction of the mdm2, CIP-1 and gadd45 genes, and the intensity of G1 arrest by gamma irradiation are inversely correlated to **T**, **D**, **A**, **R** drug sensitivity. Indeed it is well-known that the wild type p53 plays a role both in mdm2, CIP-1 and gadd45 gene transactivation and in the cell cycle arrest after gamma irradiation [28].

A notable exception to the influence of the p53 activity on the sensitivity to anticancer drugs is represented by the **M** group: in this case no correlation with the presence of mutated p53 and its levels, with the level of X-ray induction of the mdm2, CIP-1 and gadd45 genes, and with the intensity of G1 arrest, was detected (Figure 7). A simple explanation is that the p53 gene participate to the antiproliferative and apoptotic effect of several anticancer drugs acting on DNA, while the anti-tubulin drugs (**M** group) are acting downstream and therefore are not influenced by alterations of the p53 system. Indeed, these observations are fully confirmatory of the results obtained in a previous analysis of the NCI anticancer drug screen database with a different method [22].

Another well-established result [29] confirmed by the PLS2 analysis is the inverse correlation between the expression of the multidrug resistant protein (MDR) and the sensitivity to some drug classes (**M** and **T** in Figures 7 and 8, respectively).

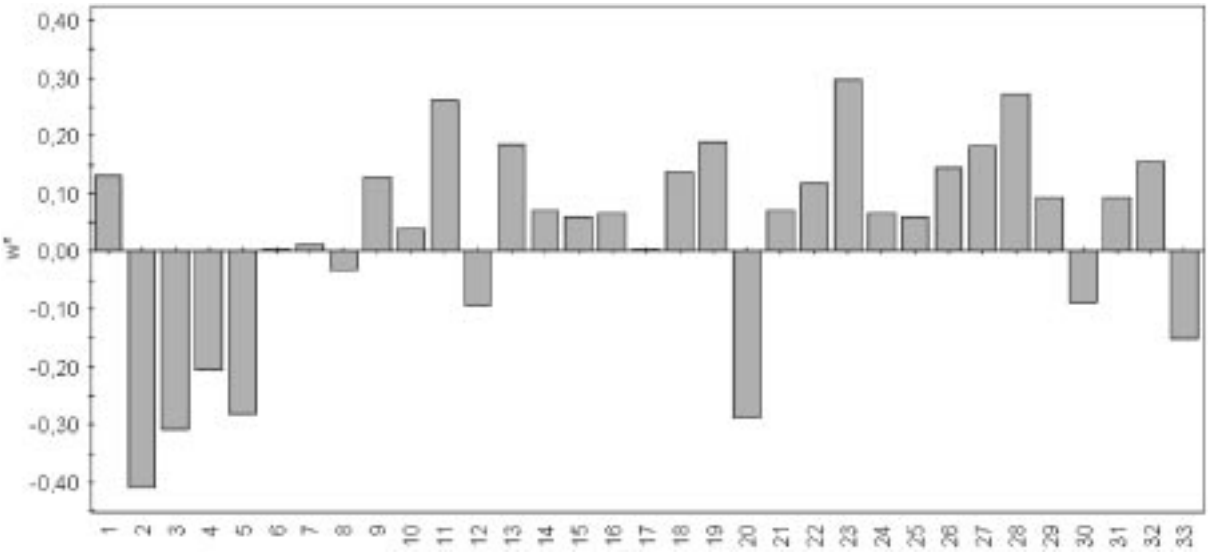

Figure 4. Modelling powers for molecular targets 1–33 in model 4, alkylating agents. Number codes are as follows: **1:** p53 protein level; **2**: X-ray induction of msm2; **3:** X-ray induction of CIP1/WAF1; **4**: X-ray induction of gadd45; **5**: G1 arrest (6.3 Gy of gamma); **6**: topoisomerase II alpha mRNA; **7**: topoisomerase II beta mRNA; **8**:Raf-1; **9**: MDR-1 by PCR; **10**: MDR-1 by Rhodamine efflux; **11**: DT-Diaphorase Activity; **12**: NADH: Cytocrome-b5 Reductase; **13**: Thioredoxin mRNA Level; **14**: Thioredoxin Reductase mRNA Level; **15**: Class 1 Aldehyde Dehydrogenase Activity; **16**: Class 3 Aldehyde Dehydrogenase Activity; **17**: EGF receptor expression; **18**: c-erbB2 expression; **19**: TGF-alpha expression; **20**: metallothionein content; **21**: metallothionein nuclear/cytoplasm ratio; **22**: MRP (Rat Antibody); **23**: LRP; **24**: Nm23 expression; **25**: Glutathione S-transferase Pi; **26**: Glutathione S-transferase M3; **27**: Glutathione S-transferase A1; **28**: Glutathione S-transferase π **29**: Glutathione S-transferase μ **30**: &gamma-glutamyl Transpeptidase; **31**: Glyoxalase-I; **32** Dihydrodiol Dehydrogenase; **33**: Glutathione.

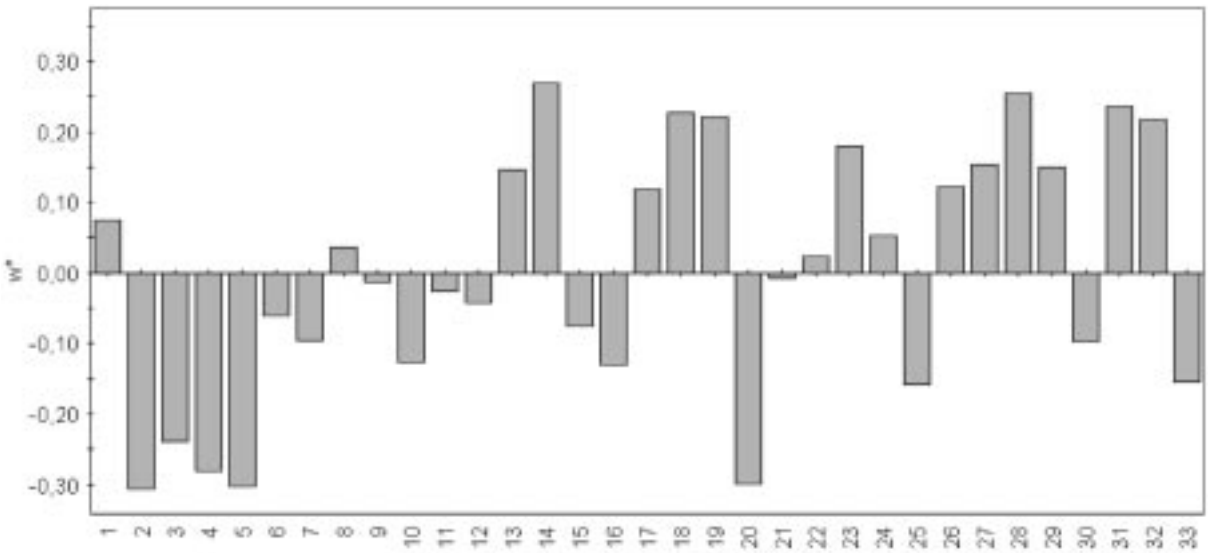

Figure 5. Modelling powers for molecular targets 1–33 (numbered as Figure 4) in model 5, RNA antimetabolites.

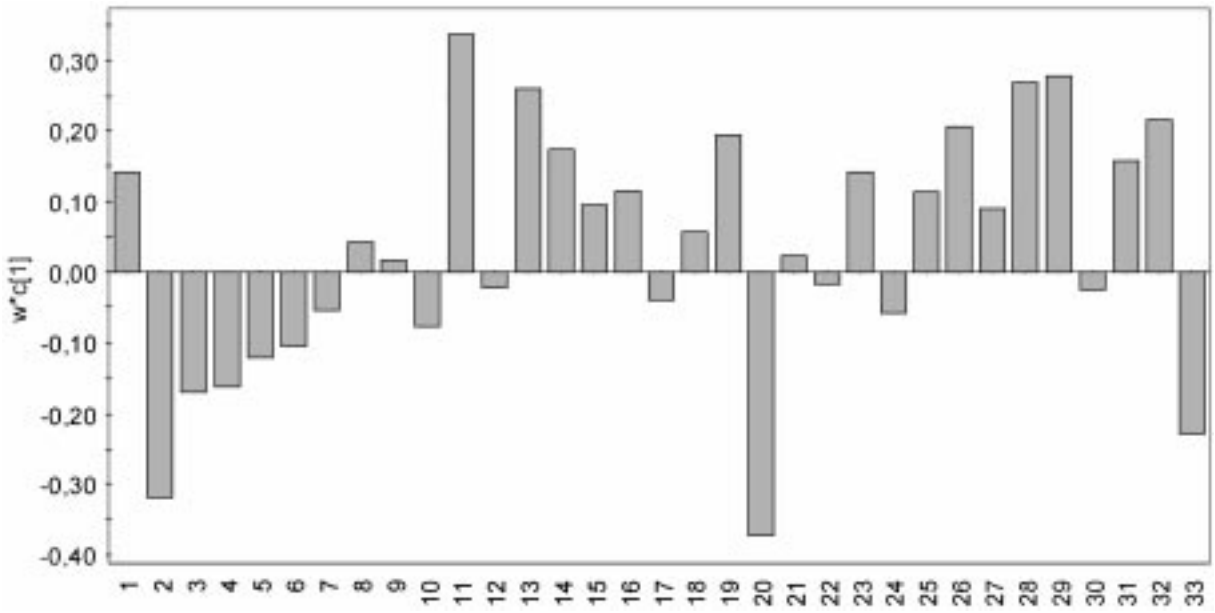

Figure 6. Modelling powers for molecular targets 1–33 (numbered as Figure 4) in model 6, DNA/RNA antimetabolites.

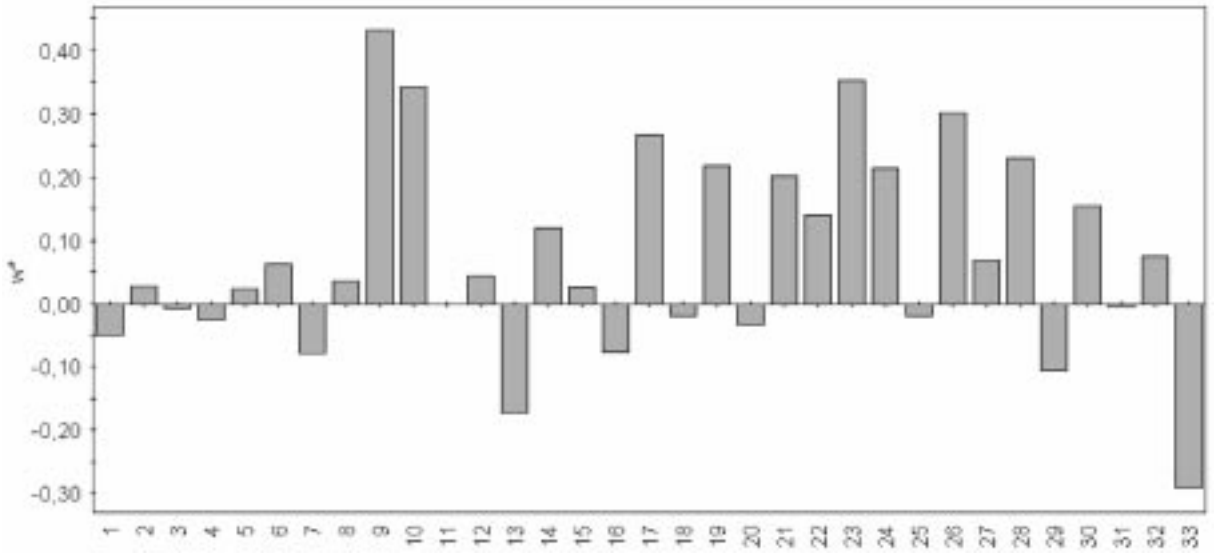

Figure 7. Modelling powers for molecular targets 1–33 (numbered as Figure 4) in model 7, antimitotic agents.

Moreover, other interesting correlations not fully described in previous studies may be revealed by the PLS2 analysis. In addition to MDR, several other proteins related to multidrug resistance have been described [30]. In the NCI molecular targets database the data related to two of these proteins are available. The Multidrug-Resistance-associated Protein (MRP) is one of the ATP-binding-cassette (ABC) transporters. The Lung-Resistance Protein (LRP) is the major component of human vaults, organelles involved in intracellular transport processes. Both proteins were previously analysed in the NCI cell panels [31, 32]. Interestingly, our analysis reveals that both MRP and LRP expression are inversely correlated to sensitivity to the **T** class, while LRP expression has a predominant contribution in the sensitivity to **A**, **D**, and **R** classes. It is noteworthy that drugs of the

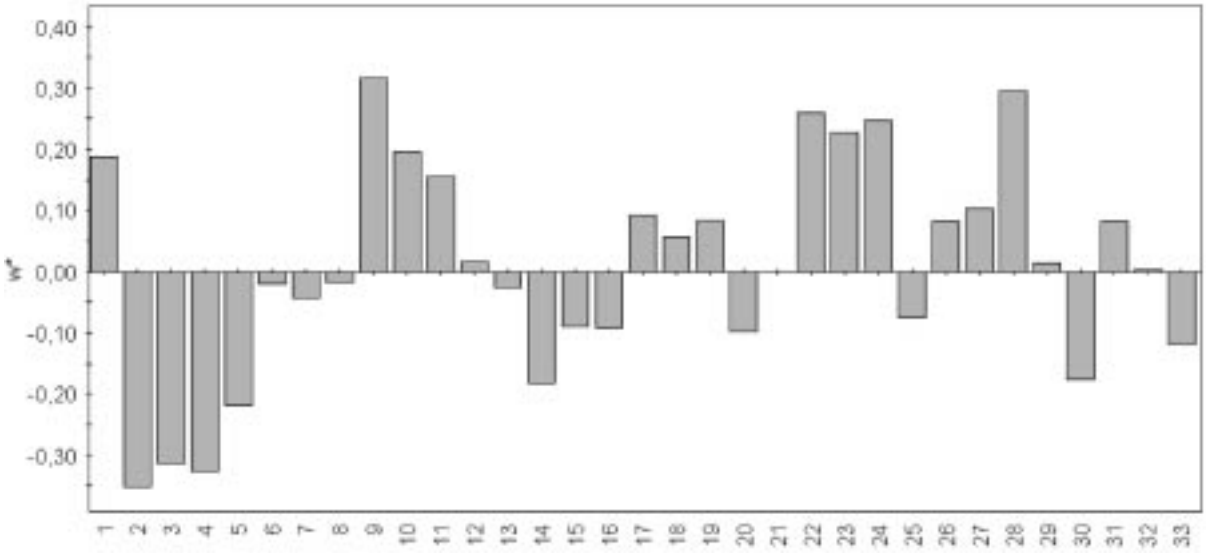

Figure 8. Modelling powers for molecular targets 1–33 (numbered as Figure 4) in model 8, topoisomerase II inhibitors.

D and **R** classes, not influenced by MDR and MPR expression, are influenced by LPR expression.

The expression of thioredoxin and thioredoxin reductase is inversely correlated to the sensitivity to **A**, **D** and **R** classes. Thioredoxin and thioredoxin reductase expression is increased several fold in human solid tumors and thioredoxin is released in the medium by cultured cancer cells. It has been suggested [33] that secretion of thioredoxin, which is known to have a direct growth stimulating activity, might lead to the stimulation of cancer cell growth. This autocrine stimulatory loop may be responsible for the activation of biochemical pathways that contribute to an intrinsic resistance of cancer cells to several anticancer drugs, with the notable exception of topoisomerase II inhibitors (**T** class, see Figure 8) or antimitotic drugs (**M** class) exhibiting an increased sensitivity with increasing level of cellular thioredoxin.

The sensitivities of the **A**, **D** and **R** classes of compounds show a direct correlation with the metallothionein content. Metallothioneins are major thiol-containing intracellular proteins that bind metals, are induced by stress, and are involved in resistance to some metal-containing anticancer drugs [34–36]. Interestingly, the present results show that the sensitivity of the **A**, **R** and **D** classes of compounds is directly correlated with the total metallothionein content, but not with its nucleo/cytoplasmic distribution (Figures 4, 5, 6). To our knowledge this is the first indication that total metallothionein content may increase sensitivity to specific classes of anticancer drugs.

An inverse correlation was found between DT-diaphorase activity and the sensitivity to **A**, **D** and **T** classes (see Figures 4, 6, 8). In previous studies only a direct correlation between DT-diaphorase activity and chemosensitivities to mitomycin C and EO9 was reported and explained suggesting the involvement of this enzyme in the metabolic activation of antitumor drugs.

Of course, the above correlations do not necessarily imply a cause-effect relationship as they might represent only an epiphenomenon. However they provide a clue to design direct experiments aimed to test the presented hypotheses and to confirm a causal link between the molecular property and the mechanism of action or of resistance for a specific group of anticancer drugs.

The ellipticine analogs case

The PLS approach can also be applied to relate the molecular targets to the therapeutical fingerprints of novel active compounds with the aim to evaluate the influence of each molecular target on their mechanism of action. When the resulting overall pattern is different from those of Figures 4–8 there is a clear indication that a different (i.e. new) mechanism may occur.

As an example, we decided to examine by PLS2 the 'fingerprint' activity patterns of ellipticine analogs, recently analysed by cluster analysis [24]. The interesting feature of this class of compounds (so called p53-inverse) is their higher activity against p53 mu-

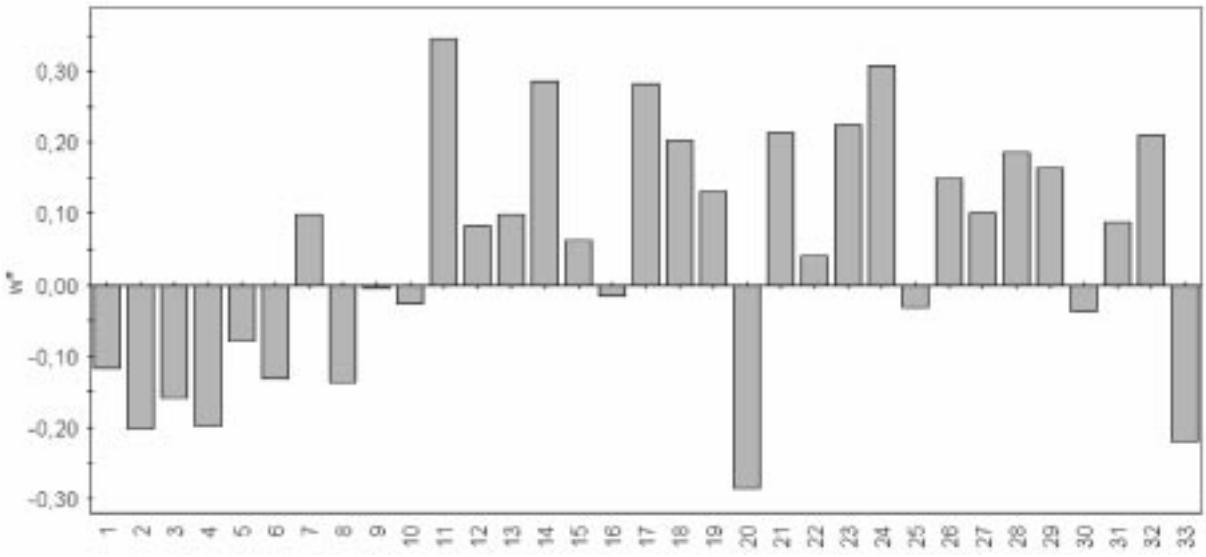

Figure 9. Modelling powers for molecular targets 1–33 (numbered as Figure 4) in modelling 9 hellipticine derivatives (NCI: 100594, 162907, 176326, 176327, 176328, 237068, 338258, 376451, 37645).

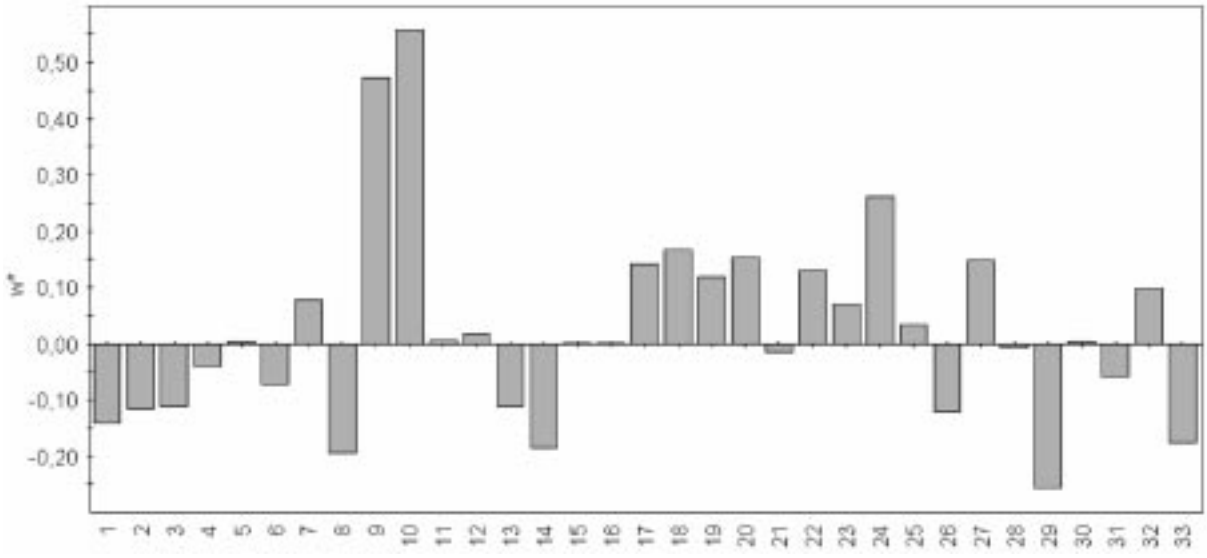

Figure 10. Modelling powers for molecular targets 1–33 (numbered as Figure 4) in modelling 3 hellipticine derivatives (NCI: 359449, 604574 637651).

tant cell lines. In contrast the majority of agents for cancer treatment are more active against the p53 wild-type cell lines (p53-direct compounds) [22–23]. PLS2 modelling was performed using the 33 molecular targets as descriptors and the available activities, expressed as log GI_{50}, for 14 ellipticines (NCI: 98949, 100594, 162907, 176326, 176327, 176328, 237068, 338258, 359449, 376451, 376452, 604574, 637651, 657149) as dependent variables. A preliminary PLS2 analysis pointed out that the activities of compounds 98949 and 657149 were not well described by the descriptors and therefore the above compounds were excluded from further analyses. Further PLS2 provided a 2PC (principal components) significant model, with the second component needed to explain the activity of only three derivatives (359449, 604574 and 637651). Therefore ellipticines were divided into two sub-groups. The resulting PLS2 model for 9 compounds (NCI: 100594, 162907, 176326, 176327, 176328, 237068, 338258, 376451, 376452) exhibits

only one statistically significant PLS2 component explaining 23% of Y variance (i. e. the 33 molecular targets account for 23% of the information embedded in the activity data) and provides the modelling powers shown in Figure 9. As expected a clear inverse correlation between p53 levels and drug sensitivity revealed by this analysis (molecular target 1, Figure 9), is consistent with the reported p53-inverse activity of ellipticine analogs [19].

A second PLS2 model for the remaining 3 hellipticine derivatives (NCI: 359449, 604574 and 637651) exhibits again only one statistically significant PLS2 component explaining up to 55% of Y variance (i.e. the 33 molecular targets account for 55% of the information embedded in the activity data) and provides the modelling powers reported in Figure 10. It is evident that in this case MDR-1 has by far the most relevant effect for these ellipticines which exhibit moderate anticancer activities, thus suggesting that these compounds are probably substrates for the MDR protein.

The above results support the hypothesis that the activity 'fingerprint' patterns of ellipticines (and in general of other drugs reported in the NCI database) encode information on their mechanism of action [24]. In addition, the present study provides an easy criterion to select compounds whose fingerprint is determined by the same molecular targets pattern and allows an easy identification of the most relevant targets responsible for the $\log GI_{50}$ of drugs acting by the same novel mechanism.

Conclusions and outlook

A multivariate insight into the *in vitro* antitumour screen database of the NCI by means of the SIMCA package allowed to propose hypotheses on the mechanism of action of the NCI standard database compounds whose mechanism is either unknown or controversial. In addition, the effect of intrinsic biochemical cell line properties such as the molecular targets that can influence the sensitivity to drug treatment, could be evaluated simultaneously for classes of compounds which act by the same mechanism. More interestingly, the PLS correlation between the molecular targets and the therapeutical fingerprint of a single novel active compound could allow an immediate identification of the molecular targets which are expected to be involved in its mechanism of action, providing clues which allow significant reduction of the time and of the cost of complex investigations aimed at the identification of the mechanism. It is noteworthy that such correlations are able to point out molecular target patterns different from those characteristic of well established mechanisms of action, thus suggesting specific biochemical studies for the investigation of new mechanisms.

In conclusion, the present paper points out the advantages of looking at information rich biological databases by means of multivariate methods which emphasise problem-solving applications. The lack of emphasis on application has recently been indicated [37] as the main reason for decrease of interest in biometrics, a ' lively subfield of biology ... 50 years ago", a discipline which due to the lack of ideas from biology has moved ' from science to become part of statistics". With the present work we provide a contribution to renew excitement on biometric applications which, in our opinion, are susceptible of great expansion in the near future, where biochemistry and pharmacology will need the help of statistics to understand complex biological processes on a molecular basis.

Acknowledgements

Financial support of the University of Catania, of MURST and CNR (Roma) is gratefully acknowledged.

References

1. Database resources currently available on the World Wide Web: http://www.dtp.nci.nih.gov/.
2. Weinstein, J. N., et al. Science, 258 (1992) 447.
3. a. Weinstein, J. N., Myers, T., Buolamwini, J., Raghavan, K., van Osdol, W., Licht, J., Viswanadhan, V. N., Kohn, K. W., Rubinstein, L. V., Koutsoukos, A. D. et al., Stem Cells, 12 (1994) 13; b. Chabner, B. A., Weinstein, J. N., Paull, K. D., Grever, M. R., (Eds.: Banzet, P., Holland, J. F., Khayat, D., Weil, M.), in Cancer Treatment, an Update, Springer-Verlag, Paris, 1994, pp. 10-16; c. Koutsoukos, A. D., Rubinstein, L. V., Faraggi, D., Simon, R. M., Kalyandrug, S., Weinstein, J. N., Kohn, K. W., Paull, K. D., Stat. Med., 13 (1994) 719; d. Bates, S. E. et al., J. Cancer Res. Clin. Oncol., 121 (1995) 495.
4. Wold, S., & Sjöström, M., American Chemical Society Symposium Series 52, 1977, pp. 243–282.
5. SIMCA 7.0 Manual Umetri AB, Umeå, Sweden.
6. Wold, S., Chemometr. Intell. Lab. Syst., 2 (1987) 37.
7. Wold, S., Technometrics, 20 (1978) 397.
8. Mardia, K. V., Kent, J. T., Biddy, J. M., in Multivariate Analysis, Academic Press, London, 1979.

9. Wold, S., Albano, C., Dunn, W. J., Esbensen, K., Hellberg, S., Johannsson, E., Sjöström, M. (Eds.: Martens, H. and Russwurm, H.), in Food Research and Data Analysis, Applied Science, London, 1983, pp. 147–162.
10. Wold, S., Albano, C., Dunn, W. J., Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Lindenberg, W., Sjöström, M. (Ed. B. R. Kowalski NATO ASI Series C.), Chemical Mathematics and Statistics in Chemistry, D. Reidel, Dordrecht, 1984, pp. 17–95.
11. Clementi, S., Cruciani, G., Curti, G., Anal. Chimica Acta, 191 (1986) 149.
12. WWW under http://www.unict.it/dipchi/.
13. Koutsoukos, A. D., Rubinstein, L. V., Faraggi, D., Simon, R. M., Kalyandrug, S., Weinstein, J. N., Kohn, K. W., Paull, K. D., Stat. Med., 13 (1994) 719.
14. Chen, H., Patel D. J., J. Mol. Biol., 246 (1995) 164.
15. Ren, J., Chaires, J. B., Biochemistry, 38 (1999) 16067.
16. Pommier, Y., Fesen, M. R., Goldwasser, F., (Eds. B. A. Chabner and D. L. Longo), Cancer Chemoterapy: Principles and Practice 2nd ed., J. B. Lippincott Co., Philadelphia, 1995.
17. Malone, H., Atassi, G., Anticancer Drugs, 8 (1997) 811.
18. Jayaran, H. N., Lui, M. S., Plowman, J., Pillwein, K., Reardon, M. A., Elliot, W. L., Weber, G., Cancer Chemother. Pharmacol., 26 (1990) 88.
19. 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., Lashkari, D., Shalon, D., Myers, T. G., Weinstein, J. N., Botstein, D., Brown, P. O., Nat Genet. 24 (2000) 227.
20. Scherf, U., Ross, D., Waltham, M., Smith, L., Lee, J., Tanabe, L., Kohn, K., Reinhold, W., Myers, T., Andrews, D., Scudiero, D., Eisen, M., Sausville, E., Pommier, Y., Botstein, D., Brown, P., Weinstein, J., Nat Genet., 24 (2000) 236.
21. Peters, A. C., Smythe, A. M., Wu, L., Monks, A., Boyd, M. R. and Shoemaker, R. H., Oncology Reports, 1 (1994) 907.
22. O'Connor, P. M., Jackman, J., Bae, I., Myers, T. G., Fan, S., Mutoh, M., Scudiero, D. A., Monks, A., Sausville, E. A., Weinstein, J. N., Friend, S., Fornace, A. J. J., Kohn, K. W., CancerRes., 57 (1997) 4285.
23. Weinstein, J. N., Myers, T. G., O'Connor, P. M., Friend, S. H., Fornace, A. J. J., Kohn, K. W., Foyo, T., Bates, S. E., Rubinstein, L. V., Anderson, N. L., Buolamwini, J. K., Van Osdol, W. W., Monks, A. P., Scudiero, D. A., Sausville, E. A., Zaharevitz, D. W., Bunow, B., Viswanadhan, V. N., Johnson, G. S., Wittes, R. E., Paull, K. D., Science, 275 (1997) 343.
24. Shi, L. M., Myers, T. G., Fan, Y., O'Connor, P. M., Paull, K. D., Friend, S. H., Weinstein, J. N., Mol. Pharmacol., 53 (1998) 241.
25. Shi, L. M., Fan, Y., Myers, T. G., O' Connor, P. M., Paull, K. D., Friend, S. H., Weinstein, J. N., J. Chem. Inf. Comput. Sci., 38 (1998) 189.
26. Hollstein, M., Sidransky, D., Vogelstein, B., Harris, C. C., Science, 253 (1991) 49.
27. Hollstein, M., Herenhahn, M., Yang, Q., Bartsch, H., Wang, Z. Q., Hainaut, P., Mutat. Res., 431 (1999) 199.
28. Harris, C. C., J. Natl. Cancer Inst., 88 (1996) 1442.
29. Lee, J.S., Paull, K., Alvarez, M., Hose, C., Monks, A., Grever, M., Fojo, A. T. and Bates, S., Mol. Pharm., 46 (1994) 628.
30. Flens, M. J., Scheffer, G. L., Van der Valk, P., Broxterman, H. J., Eijdems, E. W., Huysmans, A. C., Izquierdo, M. A., Scheper, R. J., Int. J. Cancer, 73 (1997) 249.
31. Alvarez, M., Robey, R., Sandor, V., Nishiyama, K., Matsumoto, Y., Paull, K., Bates, S., Fojo, T., Mol. Pharmacol., 54 (1998) 802.
32. Izquierdo, M. A., Shoemaker, R. H., Flens, M. J., Scheffer, G. L., Wu, L., Prather, T. R., Scheper, R. J., Int J. Cancer, 65 (1996) 230.
33. Berggren, M., Gallegos, A., Gasdaska, J. R., Gasdaska, P. Y., Warneke, J. and Powis, G., Anticancer Res., 16 (1996) 3459.
34. Lazo, J. S., Kuo, S. M., Woo, E. S., Pitt, B. R., Chem. Biol. Interact., 111–112 (1998) 255.
35. Woo, E. S., Monks, A., Watkins, S. C., Wang, A. S., Lazo, J. S., Cancer Chem. Pharmacol., 41 (1997) 61.
36. Kasahara, K., Fujiwara, Y., Nishio, K., Ohmori, T., Sugimoto, Y., Komiya, K., Matsuda, T., Saijo, N., Cancer. Res., 51 (1991) 3237.
37. Wold, S., Sjöström, M., Chemometr. Intell. Lab. Syst., 44 (1998) 3.