Electrophoresis 1997, *18*, Artikel 2510

A protein expression database for the molecular pharmacology of cancer

Timothy G. Myers[1,4]
N. Leigh Anderson[2]
Mark Waltham[1]
Guang Li[1]
John K. Buolamwini[1]
Dominic A. Scudiero[3]
Kenneth D. Paull[4]
Edward A. Sausville[5]
John N. Weinstein[1]

[1]Laboratory of Molecular Pharmacology, Division of Basic Science, National Cancer Institute (NCI), Bethesda, MD, USA
[2]Large Scale Biology, Inc., Rockville, MD, USA
[3]SAIC/NCI-FCRDC, Frederick, MD, USA
[4]Information Technology Branch, Developmental Therapeutics Program (DTP), Division of Cancer Treatment, Diagnosis, and Centers (DCTDC), NCI
[5]Office of the Director, DTP, DCTDC, NCI, Bethesda, MD, USA

In the last six years, the Developmental Therapeutics Program (DTP) of the US National Cancer Institute (NCI) has screened over 60 000 chemical compounds and a larger number of natural product extracts for their ability to inhibit growth of 60 different cancer cell lines representing different organs of origin. Whereas inhibition of the growth of one cancer cell type gives no information on drug specificity, the relative growth inhibitory activities against 60 different cells constitute patterns that encode detailed information on mechanisms of action and resistance (as reviewed in Boyd and Paull, *Drug Devel. Res.* 1995, *34*, 19–109 and Weinstein *et al.*, *Science* 1997, *275*, 343–349). In order to correlate the patterns of activity with properties of the cells, we and other laboratories are characterizing the cells with respect to a large number of factors at the DNA, mRNA, and protein levels. As part of that effort, we have developed a two-dimensional gel electrophoresis (2-DE) protein expression database covering all 60 cell types (Buolamwini *et al.*, submitted). Here we present analyses of the correlations among protein spots (i) in terms of their patterns of expression and (ii) in terms of their apparent relationships to the pharmacology of a set of 3989 screened compounds. The correlations tend to be stronger for the latter than for the former, suggesting that the spots have more robust signatures in terms of the pharmacology than in terms of expression levels. Links to pertinent databases and tools of analysis will be updated progressively at http://www.nci.nih.gov/intra/lmp/jnwbio.htm and http://epnwsl.nciferf.gov.2345/dis3d/dtp.html.

1 Introduction

In September of 1996, in US National Cancer Institute (NCI) announced creation of the "Cancer Genome Anatomy Project", whose overall goal was stated as follows: "To achieve the comprehensive molecular characterization of normal, precancerous and malignant cells, with the aim of making it possible to characterize and recognize all major cellular mechanisms and steps of tumor development. Such molecular characterization will allow correlation of disease progression and outcome; improve the evaluation of treatments; stimulate new approaches to prevention, detection and therapy; and provide high-throughput diagnostic tools for clinical application". Quite appropriately, the first initiatives under this program emphasize characterization at the DNA and mRNA levels, but proteome research will presumably make major contributions in the long run. Although many aspects of protein biochemistry and cell physiology can be inferred from studies at the DNA and mRNA levels, there will continue to be great gaps, in part because of (i) the trivial but vexing problem of errors in the DNA sequence databases that can, for example, throw off reading frame assignments for proteins; (ii) post-translational modifications, including the phosphorylations important in cell cycle and signaling pathways; (iii) lack of diret correspondence between levels of mRNA and levels of protein in cells; and (iv) complex pathways and feedback mechanisms that control protein levels [1].

When molecular characteristics of cancer cells are considered in relation to therapy, two reciprocally related questions arise: (i) How can clinical treatment be individualized, given molecular characterization of a patient's tumor? (ii) How can agents selectively active against tumors with particular molecular characteristics be discovered or designed? The first question raises all of the complex opportunities and problems associated with clinical tumor marker studies; the second poses a challenge that this paper will address, albeit in a highly tentative and preliminary way. One way in which proteome research can further the development of new therapies is afforded by the NCI Developmental Therapeutics Program's (DTP) cancer drug discovery effort [2–8]. Since 1990, DTP has screened more than 60 000 compounds, plus a larger number of natural product extracts, for their ability to inhibit the growth of 60 different human cancer cell lines. Included currently are melanomas, leukemias, and cancers of breast, prostate, lung, colon, ovary, kidney, and central nervous system origin. This "disease-oriented" strategy for drug discovery was originally based on the hypothesis that selective activity *in vitro* against cancer cell lines from a particular organ would predict selective activity against corresponding tumors in humans. We find, however, that patterns of activity observed in the screen are predictive in a more powerful way at the molecular level: They provide incisive information on molecular targets and modulators of activity within the cancer cells [7, 9–18]. We continue to refer to this test system as a "screen", but is also serves as a way to profile, or fingerprint, candidate therapeutic agents.

Correspondence: Dr. Timothy G. Myers, NCI, EPN Rm. 811, 6130 Executive Blvd., Rockville, MD 20852, USA (Tel: +301-496-8747; Fax: +301-480-4808; E-mail: tgm@nih.gov); Dr. John N. Weinstein, NCI, Bldg 37, Rm 5C-25, 9000 Rockville Pike, Bethesda, MD 20892, USA (Tel: +301-496-9571; Fax: +301-402-0752; E-mail: weinstein@dtpax2.nciferf.gov)

Nonstandard abbreviations: A, activity database; P, protein expression database

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Protein database / Cancer / Pharmacology / Chemosensitivity / Cell culture / Cell line / Drug

0173-0835/97/0303-0000 $17.50+.50/0

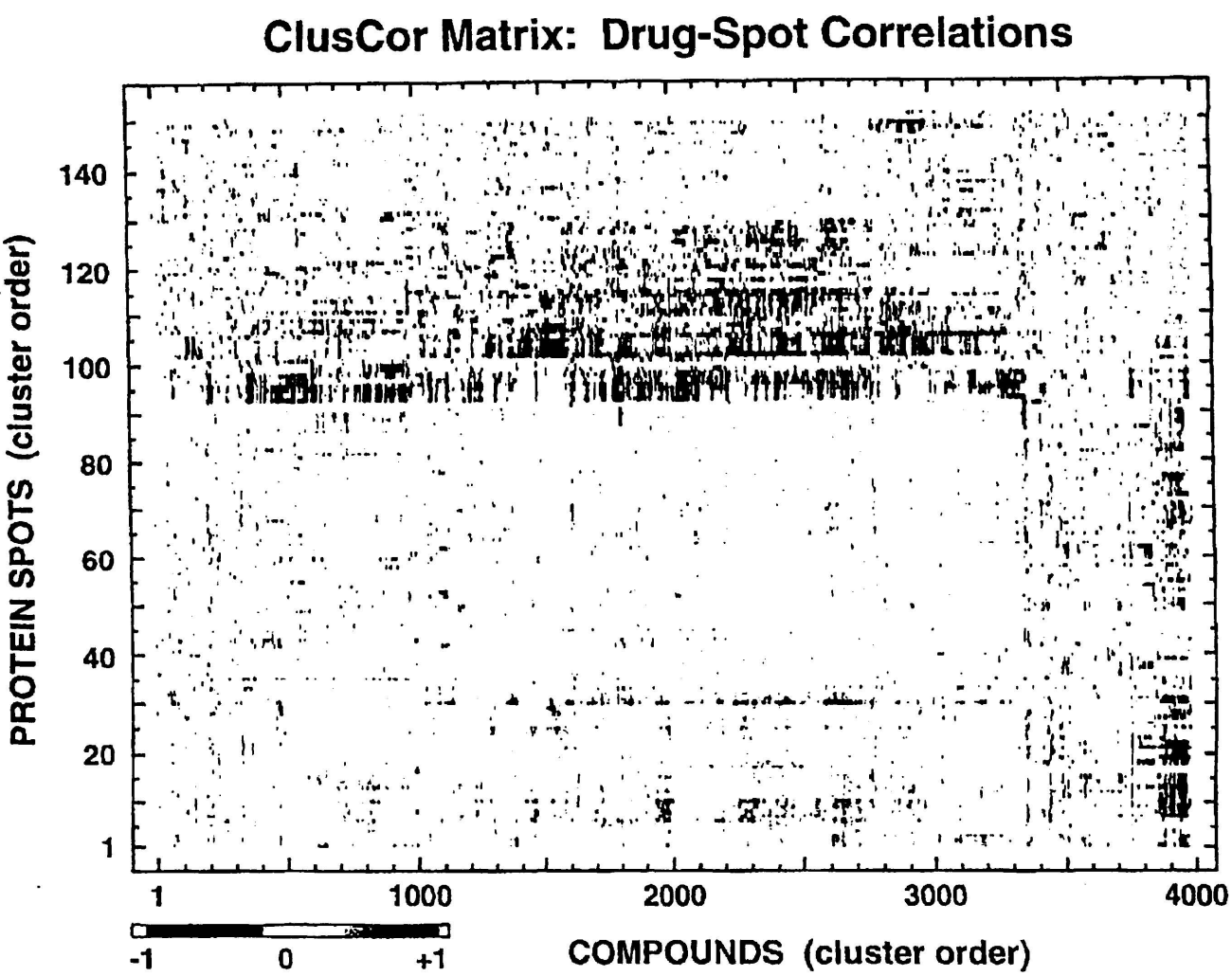

Figure 1. ClusCor matrix (PA) showing similarities and differences (in terms of the Pearson correlation coefficient) between the patterns of expression of 151 protein spots and the patterns of activity of 3989 compounds. Red, orange, and yellow indicate positive correlations; blue and black indicate negative correlation. Patterns of correlation between protein abundance and cell sensitivity appear as homogeneous rectangles because rows and columns of the matrix have been cluster ordered. The cluster tree used to order the compounds (columns) was based on similarity (Pearson correlation coefficient) between activity patterns in the 60 cell lines. The cluster tree used to order the protein spots (rows) was created by clustering the spot patterns present in rows of the PAAP matrix (represented in Fig. 2B) using Euclidean distance as the similarity metric. The latter tree is shown in Fig. 2.

An important component of the NCI drug discovery program is the "molecular targets" initiative. Investigators at the NCI and other institutions are characterizing the 60 cell lines in terms of a variety of molecular factors at the DNA, mRNA protein, and functional levels. Among the targets assessed to date by various research groups are oncogenes, tumor suppressor genes, drug resistance-mediating transporters, heat shock proteins, telomerase, cytokine receptors, molecules of the cell cycle and apoptotic pathways, DNA repair enzymes, components of cellular cytoarchitecture, intracellular signaling molecules, and enzymes of metabolism (as summarized in [18] and at DTP's World Wide Web site; see URL above). The guiding hypothesis of the molecular targets program is that patterns of chemosensitivity in the screen can be related to molecular characteristics of the cells. If so, the resulting correlations may be useful as guides to rationally directed drug development and, given data on clinical markers, to possible individualization of therapy. Limitations of this drug screen database are clear. The two most fundamental are that (i) cell lines *in vitro* are not fully representative of tumors in humans, and (ii) the data are correlative, since the cell lines do not represent isogenic cell sets. With respect to (i), we consider the cells simply as complex collections of molecular biological and cell physiological aspects of cancer. Empirically, their patterns of response appear to encode useful, interpretable information; with respect to (ii), statistical information arising from the analyses is treated as the basis for hypothesis generation, not as accomplished fact.

Most of the effort thus far has gone into characterizing individual cellular factors, one at a time. An alternative approach that we have initiated is to search for targets and modulators of activity in a more comprehensive way at the protein level by establishing a two-dimensional gel electrophoresis (2-DE) database for the 60 cell lines of the screen [19]. Here we briefly summarize how the

database was generated and then analyze relationships among the spots in terms of their patterns of expression across the 60 cell lines. We also present a preliminary look at the relationships between protein expression and cell growth inhibitory activity for a set of 3989 tested compounds.

2 Materials and methods

2.1 Cells and screening protocol

Procedures for growing cells and testing compounds in the NCI-DTP cancer drug discovery screen have been described previously [2–6, 8, 20]. Very briefly with respect to the assay, cells are plated on day 0 at a density individualized for each cell line so that they will generally be sub-confluent at the end of the assay period. On day 1, a compound is added in the format for a duplicate-well, 5-dose, ten-fold-interval dose response study. No-drug, no-cell, and no-growth controls are included. On day 3 the cells are processed for staining with sulforhodamine B (SRB), which reflects the amount of cell mass present at the end of a 48 h exposure to the test agent. From dose-response curves based on the SRB data, various parameters can be determined. The most useful, and the one used in the present study, is the GI_{50}, defined as that concentration of compound required to inhibit growth of the cell line by 50%. More precisely, the quantity used in the calculations to be described is the potency measure $-\log_{10}GI_{50}$.

2.2 Activity database

A table consisting of growth inhibition (GI_{50}) values for 60 cell lines and 3989 compounds was created from the DTP *in vitro* cancer screen database. These were the non-confidential compounds considered sufficiently interesting to have been tested more than once (as of the time when the table was compiled). Dose-response curves from the two or more experiments were used to calculate an average potency, $-\log_{10}(GI_{50})$, for each cell line for each compound. For compounds that had been tested at more than one dose range, an algorithm we term "BestVec" (T. G. Myers *et al.*, unpublished) was used to select the dose range that best represented the data. Fewer than 10% of the 60 × 3989 table entrices were missing. Each missing value was replaced by the mean potency of the compound over the remaining cell lines.

2.3 Cell harvesting for protein studies

Our experimental methods are described in detail elsewhere [19]. Briefly, the cells were grown in 96-well microtiter plates in parallel with those being used for drug screening. This procedure ensured that conditions were similar to those used in generating the data on chemosensitivity, but the cells used for analysis of proteins were not exposed to test agents. At the time of harvest, cells were examined under a microscope, quickly washed in ice-cold buffer, and lysed by vortex-mixing in a detergent-containing lysis buffer [19]. Gels were run in batches of 20 by Large Scale Biology Corp. (Rockville, MD) using the ISO-DALT 2-D gel system [21].

2.4 2-D PAGE

Samples were prepared for electrophoresis by precipitating the protein in 12% trichloroacetic acid, vortex-mixing, and then centrifuging for 60 s in a microfuge (Eppendorf, Hamburg). The resulting pellets were redissolved in 100 µL of lysis buffer in which the Pharmalyte 8–10.5 had been replaced by BDH 4–8 carrier ampholytes (Gallard-Schlesinger, Carleplace, NY). The protein samples were precipitated and then resolubilized in order to increase protein concentration. Practical constraints prevented successfully loading more than 15–20 µL on first-dimensional gels and this volume did not contain enough protein for optimal 2-D gels when attached cultured cells were harvested from microtiter plates. The method was tested on rat liver samples, asking whether proteins were lost during the precipitation and resolubilization steps. No losses were apparent. Samples of 10–20 µL were applied to tube gels and run for 33 000–34 500 Vh for first-dimensional isoelectric focusing, using a progressively increasing voltage. Second-dimensional sodium dodecyl sulfate slab gels, 20 × 25 cm, were run at 10°C in DALT tanks with buffer recirculation [22]. The top 5% of the gel was 11% total acrylamide; the remaining 95% varied linearly from 11% to 18%. The tube gels were loaded directly onto the slab gels without equilibration and held in place by polyester fabric wedges to avoid the use of hot agarose. After electrophoresis the gels were fixed overnight in 50% ethanol/2% phosphoric acid, washed three times with cold deionized water, transferred to 34% methanol/17% ammonium sulfate/2% phosphoric acid for 1 h, and stained with Coomassie Blue G-250 for 4 days [19]. After digitization with an Ektron 1412 scanner (Ektron, Bedford, MA) gel images were archived on optical disks and processed using the Kepler software to yield a spot list giving position, shape, and density information for each spot detected [22]. Two-dimensional least squares optimization was used to refine parameters for a bivariate Gaussian representation of each spot, and quantitation of spots was done in terms of their "volume", *i.e.*, the digitized staining intensity integrated over the area of the bivariate normal representation.

2.5 Protein expression database

The 151 spots for this study were chosen on the basis of two criteria: (i) presence on al gels for at least 50% of those cell lines in which the spot was detected at all, and (ii) p*I* between 4.6 and 5.4 (the region of the gel best resolved in the cell line patterns). Spot volumes from the 110 best gel images (50 cell lines represented in duplicate, 10 in singlicate) served as the source for the final expression table (P). The volumes were transformed by first replacing any value below 500 (including "0") with a threshold value of 500 (approximately the limit of detection). The thresholded volumes were converted to $\log_{10}$ form, and duplicate log(volume) values were averaged to create the 60 cell line x 151 spot table.

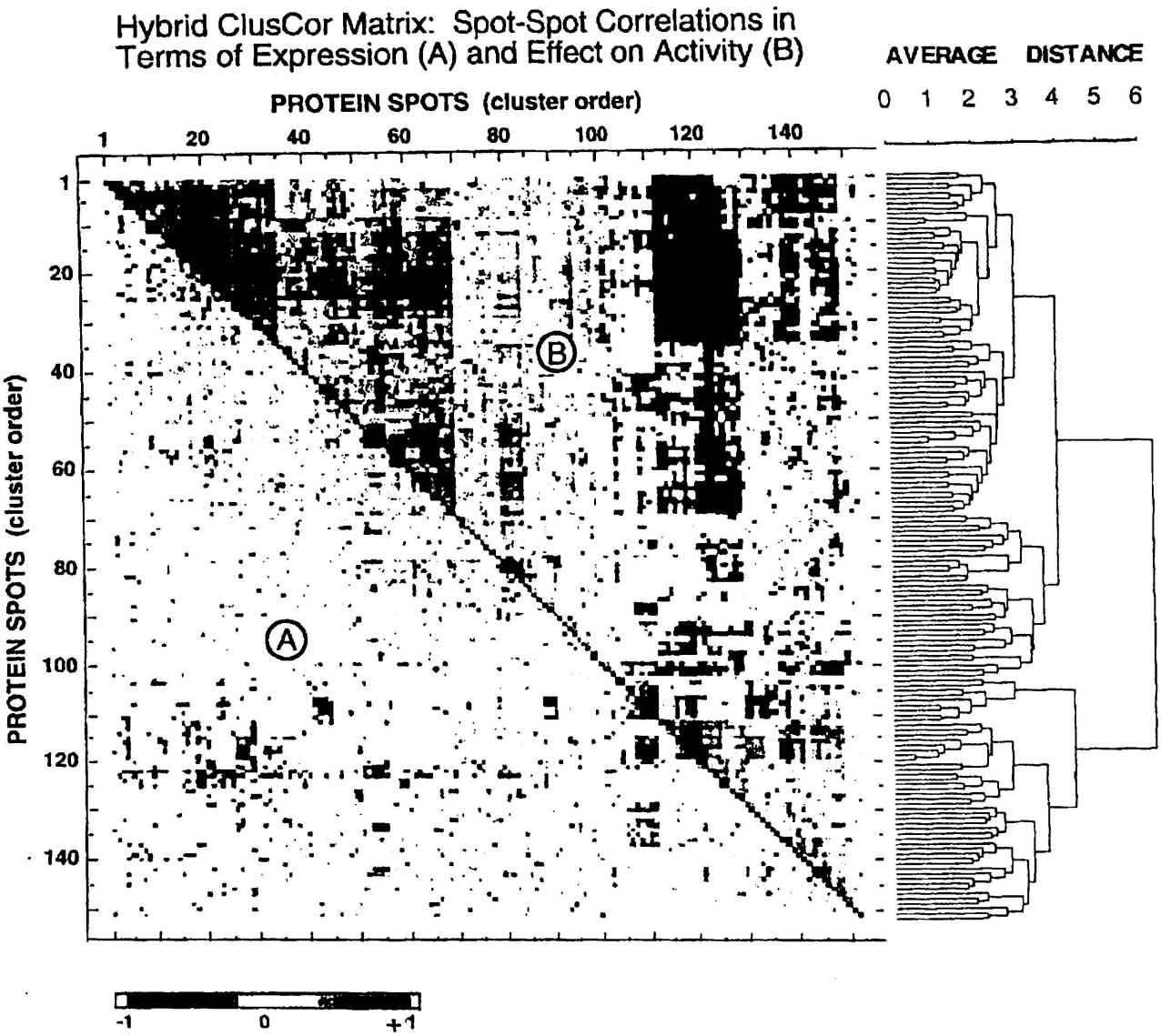

Figure 2. A hybrid ClusCor matrix showing the relationship of spot to spot for a quality-controlled database of 151 spots. (A) Spot-spot relationships in terms of pattern of expression across the 60 cell lines: lower-left half of PP (as described in Section 3). (B) Spot-spot relationships, not in terms of patterns of expression, but in terms of their correlations with the activity patterns of the 3989 compounds: upper-right half of PAAP (as described in Section ●●). Both of the original tables are symmetrical around the principal diagonal, so only half of each need be shown here. Because any pattern is, by definition, 100% correlated with itself, all values on the diagonal are color-coded red. Rows and columns of the hybrid matrix are ordered according to the cluster tree shown. Clustering was done by the average-linkage algorithm with Euclidean metric, based on spot patterns (rows) in the AP table (*i.e.*, on correlations of protein abundance with potency). The axis next to the tree indicates the average distance (inverse of similarity) between members of the two branches joined at each node. Because of the cluster ordering, families of spots correlated in expression appear as red patches near the principal diagonal, whereas negatively correlated families of spots are indicated by dark blue, off-diagonal patches. Most of the spots in this database have not yet been identified.

The mean difference between log-transformed spot volumes measured on duplicate gels (from cell samples harvested at least one month apart and, in most cases, with gels run at different times) was 0.13 log units (N = 7550).

2.6 Patterns of correlation

Patterns of correlation among spots, cell lines, and potency values were analyzed using the DISCOVERY programs [16–18], which map coherent patterns in the data in a variety of ways. The particular analyses shown here make use of what we term the "clustered correlation", or CluCor algorithm, as will be explained in Section 3. Correlation calculations and cluster analyses for this study made use of routines in SAS (SAS Institute, Cary, NC) and S-PLUS (MathSoft, Inc., Seattle, WA).

2.7 Identification of spots

Our procedure for identification of proteins by Western blotting of 2-D gels is described in detail elsewhere [19].

We have also developed a protocol for identification of spots by in-gel proteolytic digestion followed by matrix-assisted laser desorption ionization (MALDI) mass spectrometry, as described by Li *et al.* [23] in this volume. However, most of the spot identifications cited in the present study were obtained by cross-indexing highly conserved proteins from previous 2-DE analyses of rat liver and human lymphocytes (N. L. Anderson *et al.*, unpublished).

3 Results and discussion

As discussed elsewhere [19], statistical analysis indicates that the protein spot database is informationally coherent in the sense that correlation between different collections of the same cell line are higher than correlations between different cell lines. However, there remains room for improvement, in part by increasing the number of replicate collections for each cell line. Figure 1 shows the correlation of protein expression patterns with the activities of tested agents. For purposes of analysis, the protein expression table (P) can be considered as a matrix having the number of rows equal to the number of spots (151) and number of columns equal to the number of cell lines (60). Entries in the matrix are of the form $\log_{10}$(spot volume), with threshold of detection taken as 500 arbitrary units for each spot volume. Similarly, the activity table is a matrix (A) with number of rows equal to the number of compounds (3989 in this case) and number of columns equal to the number of cell lines (60). Entrices in the matrix are values of potency ($-\log GI_{50}$). Then, each row (spot) of P can be correlated with each row (compound) of A to form a new matrix AP having 151 rows and 3989 columns. Each entry in AP is a Pearson correlation coefficient (r) that indicates the degree of similarity between the pattern of a protein spot's expression across the 60 cell lines and the pattern of a compound's activity values across the same 60 cell lines. To bring out patterns in the matrix, a further step is necessary: cluster ordering. In Fig. 1, the compounds (columns) are cluster ordered along the abscissa by activity pattern. That is, compounds with the most similar patterns of activity in the screen appear side by side (*cf.* Fig. 1 in ref. [18]). The protein spots (rows) appear along the ordinate in cluster order based on the correlation with activities. That is, the spots most similar in their apparent effects on (or, more properly, correlation with) activity over the set of 3989 compounds appear side by side in the figure.

Table 1 lists index numbers assigned to the spots when the database was constructed using Kepler. The table also lists the spot positions in Fig. 1 and 2, as well as identities where assigned. The most striking observation in Fig. 1 is that, broadly speaking, there appear to be two major categories of spots: Numbers 1 to 93 and 94–151. This dichotomy is also reflected in the cluster tree in Fig. 2 (which was used to order the spot in Figs. 1 and 2). Similarly, the compounds in this figure appear to fall into two broad categories, with a boundary at approximately compound number 3340. This picture and its implications for various families of agents will be addressed in detail elsewhere. Another instructive view

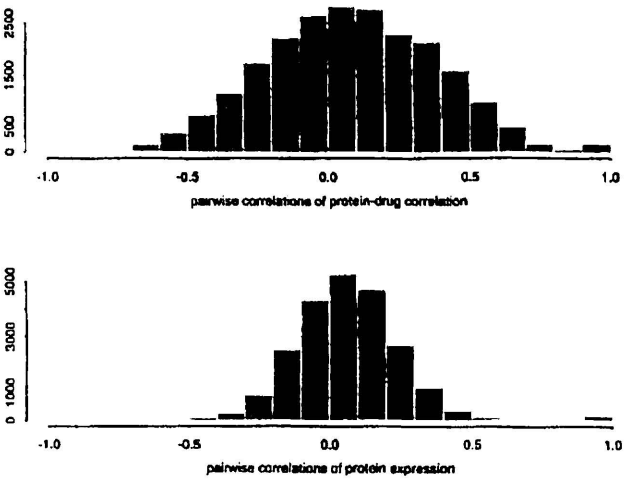

Figure 3. Histograms of Pearson correlation coefficients from Fig. 2. (A) Correlation coefficients from (PP), reflecting relationships in terms of levels of protein expression. (B) Correlation coefficients from (PAAP), reflecting spot-spot relationships in terms of association with the activities of the 3989 compounds.

of the data is given by the hybrid ClusCor matrix in Fig. 2. The lower left part (A) represents one half of PP, a matrix of correlations between protein expression patterns. That is, PP gives the relationship of spot to spot in terms of expression patterns. The upper right part (B) represents one half of PAAP. PAAP, like PP, compares spots with spots. However, the patterns compared are the patterns of a spot's correlations with compounds (rows of the AP table). That is, correlation values in the PAAP table give the relationship of spot to spot in terms of association with the activities of the 3989-compound A table. Since both PP and PAAP are symmetrical around the principal diagonal, only one half of each need be shown. Both rows and columns in this hybrid matrix (Fig. 2) are ordered according to the cluster tree, which reflects similarities between spot patterns (rows) in the AP table.

Many features stand out in Fig. 2, and only a few illustrative relationships will be indicated here: (i) Spots 18–24 are highly correlated (mean $r = 0.76$) in terms of their relationship to activity but not very highly in terms of their expression levels (mean $r = 0.38$). They fall within a larger group (1–34) that shows similar internal correlations. (ii) This large group (1–34) is highly inversely correlated (mean $r = -0.46$) with spots 112–128 in terms of activity level, as indicated by the large blue-purple patch. (iii) Spots 51–55 are highly positive in their internal correlations, both with respect to expression levels (mean $r = 0.52$) and associations with activity (mean $r = 0.65$). Spots 53, 54, and 55 are, respectively, Grp75, Hsp60, and Hsc70. Hence, this cluster represents a set of heat shock proteins that are correlated in their basal levels of expression. (iv) Spots 103–111, particularly 107–111, are highly correlated in terms of both expression (mean $r = 0.63$) and association with activity (mean $r = 0.75$). Spots 108, 110, and 111 are three cytokeratins. (v) This last group is negatively correlated in expression pattern with spots 115–118 (mean $r = -0.26$), which are positively correlated with each other in terms of both expression

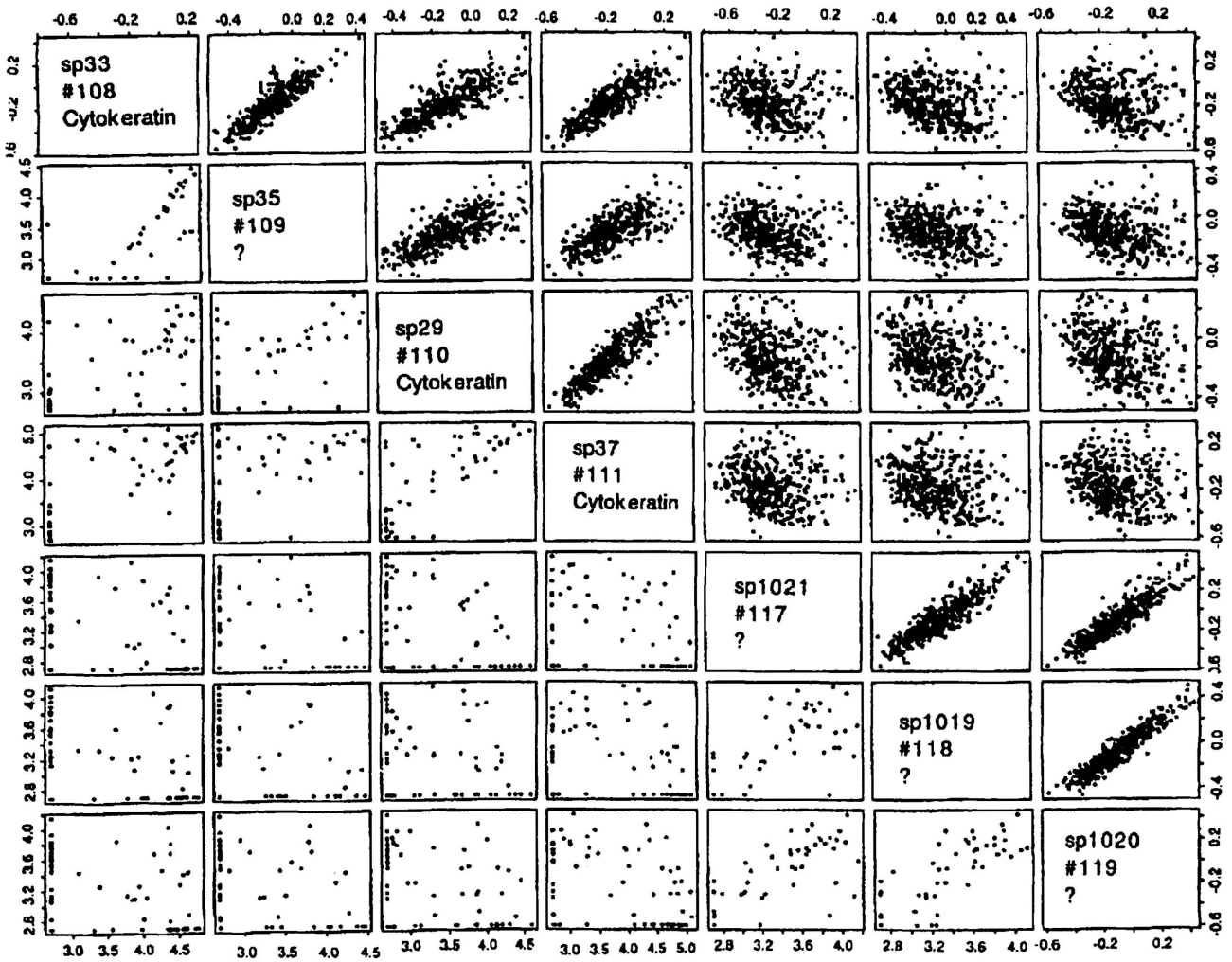

Figure 4. Scattergrams indicating the relationships among seven protein spots in terms of protein expression levels (panels at lower left) and association with activity of the 3989 compounds (panels at upper right). The lower left panels have 60 points, 1 for each cell line; the upper right panels have 400 data points, representing a set of compounds randomly selected from the 3989 (the random selection having been made to avoid the problem of showing 3989 points on each panel).

(mean $r = 0.55$) and association with activity (mean $r = 0.79$). These few examples indicate that it is possible to see high positive or negative correlations in pattern of expression, association with activity, or both. The histograms in Fig. 3 show that the correlations in terms of associations with activity (*i.e.*, in PAAP) tend to be larger in absolute value than those in terms of expression levels (*i.e.*, in PP). That difference is shown for a set of seven individual spots in Fig. 4. The three cytokeratins are better correlated with relation to activity of the tested compounds (mean $r = 0.76$) than with respect to expression levels (mean $r = 0.38$). Overall, these findings suggest that the protein spots are more robustly represented in terms of their pharmacological signatures than in terms of their expression levels.

4 Concluding remarks

In developing this 2-DE database for cancer cell lines of the NCI drug discovery program, our principal aim was to establish a useful link between proteome science and the molecular pharmacology of cancer. The cells of the NCI drug discovery program represent a disparate set of cell types from nine different organs of origin, hence this database includes a broad range of the properties of cultured cancer cells, although not necessarily those of clinical tumors. As we are able to identify increasing numbers of spots by mass spectrometry [23], the utility of this database can be expected to increase dramatically. Its utility, particularly for hypothesis testing, should increase further as we add additional cancer cell types, including transfected lines and cell isolates from microdissected clinical tumors.

We thank William W. van Osdol, Barry Bunow, and Lawrence V. Rubinstein for helpful discussions and advice. We also thank Karen Hite and the staff of SAIC/NCI-FCRDC for assistance in setting up the cells, Ricardo Esquer-Blasco and Abita Sachdev for running the 2-D gels, and Norman G. Anderson for advice on the gel systems. We are grateful to Michael R. Boyd, Robert E. Wittes, Michael R. Grover, Bruce A. Chabner, Robert Shoemaker, George S. Johnson, George VandeWoude, and Saul A. Schepartz for

their efforts in support of the DTP cell target characterization program.

5 References

[1] Anderson, N. G., Anderson, N. L., *Electrophoresis* 1996, *17*, 443–453.

[2] Boyd, M. R., in: DeVita, V. T., Hellman, S., Rosenberg, S. A. (Eds.), *Cancer: Principles and Practice of Oncology Update*, J. B. Lippicott, Philadelphia 1989, pp. 1–12.

[3] Boyd, M. R., in Neiderhuber, J. E. (Ed.), *Current Therapy in Oncology*, Decker, Philadelphia, 1992, pp. 11–22.

[4] Alley, M. C., Scudiero, D. A., Monks, A., Hursey, M. L., Czerwinski, M. J., Fine, D. L., Abbot, B. J., Mayo, J. G., Shoemaker, R. H., Boyd, M. R., *Cancer Res.* 1988, *48*, 589–601.

[5] Grever, M. R., Schepartz, S. A., Chabner, B. A., *Seminars Oncol.* 1992, *19*, 622–638.

[6] Monks, A., Scudiero, D., Skehan, P., Shoemaker, R., Paull, K., Vistica, D., Hose, C., Langely, J., Cronise, P., Vaigro-Wolff, A., Grey-Goodrich, M., Campbell, H., Mayo, J., Boyd, M. R., *J. Natl. Cancer Inst.* 1991, *83*, 757–766.

[7] Paull, K. D., Shoemaker, R. H., Hodes, L., Monks, A., Scudiero, D. A., Rubinstein, L., Plowman, J., Boyd, M. R., *J. Natl. Cancer Inst.* 1989, *81*, 1088–1092.

[8] Scudiero, D., Shoemaker, R. H., Paull, K. D., Monks, A., Tierney, S., Nofziger, T. H., Currens, M. J., Seniff, D., Boyd, M. R., *Cancer Res.* 1988, *48*, 4827–4833.

[9] Weinstein, J. N., Kohn, K. W., Grever, M. R., Viswanadhan, V. N., Rubinstein, L. V., Monks, A. P., Scudiero, D. A., Welch, L., Koutsoukos, A. D., Chiausa, A. J., Paull, K. D., *Science* 1992, *258*, 447–51.

[10] Paull, K. D., Hamel, E., Malspeis, L., in: Foye, W. E. (Ed.), *Cancer Chemotherapeutic Agents*, American Chemical Soc. Books, •• location? 1993, pp. 1574–1581.

[11] Weinstein, J. N., Myers, T. G., Buolamwini, J. K., Raghavan, K., van Osdol, W., Licht, J., Viswanadhan, V. N., Kohn, K. W., Rubinstein, L. V., Koutsoukos, A. D., Monks, A. P., Scudiero, D. A., Anderson, N. L., Zaharevitz, D., Chabner, B. A., Grever, M. R., Paull, K. D., *Stem Cells* 1994, *12*, 13–22.

[12] Weinstein, J. N., *Handbook Intl. Forum on Advances in Screening Technol. and Data Management*, •• publisher, location? 1994, 79–83.

[13] van Osdol, W. W., Myers, T. G., Paull, K. D., Kohn, K. W., Weinstein, J. N., *J. Natl. Cancer Inst.* 1994, *86*, 1853–1859.

[14] Weinstein, J. N., Myers, T. G., *J. Liposome Res.* 1995, *5*, 837–844.

[15] Boyd, M. R., Paull, K. D., *Drug Develop. Res.* 1995, *34*, 91–109.

[16] Myers, T. G., Weinstein, J. N., Raghavan, K., Buolamwini, J. K., Anderson, N. L., O'Connor, P., Kohn, K. W., Scudiero, D. A., Monks, A. P., Friend, S., Zaharevitz, D., Rubinstein, L. V., Paull, K. D., *Proc. Am. Assoc. Cancer Res.* 1995, *36*, 305.

[17] Myers, T. G., *Proc. Am. Assoc. Cancer Res.* 1996, *37*, 299.

[18] Weinstein, J. N., Myers, T. G., O'Connor, P. M., Friend, S. H., Fornace, A. J., Kohn, K. W., Fojo, 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* 1997, *275*, 343–349.

[19] Buolamwini, J. K., Myers, T. G., Anderson, N. L., Waltham, M., Li, G., Scudiero, D. A., Paull, K. D., Peterfy, M., Sausville, E. A., Kohn, K. W., Weinstein, J. N., submitted. •• accepted?

[20] Stinson, S. F., Alley, M. C., Kopp, W. C., Fiebig, H. H., Mullendore, L. A., Pittman, A. F., Kenney, S., Keller, J., Boyd, M. R., *Anticancer Res.* 1992, *12*, 1035–1053.

[21] Anderson, N. L., *Two-dimensional Electrophoresis: Operation of the ISO-DALT System*, Large Scale Biology Press, Washington, DC 1991.

[22] Anderson, N. g., Anderson, N. L., *Anal. Biochem.* 1978, *85*, 331–340.

[23] Li, G., Waltham, M., Anderson, N. L., Unsworth, E., Treston, A., Weinstein, J. N., *Electrophoresis* 1997, *18*, ••–••.

Table 1. Spot index numbers in the database, locations Figs. 1 and 2, and tentative indentity (when available). Identifications are from [19]

Order # in Fig. 1	Order # in Fig. 2	Database Index No.	Tentative Identity	Order # in Fig. 1	Order # in Fig. 2	Database Index No.	Tentative Identity	Order # in Fig. 1	Order # in Fig. 2	Database Index No.	Tentative Identity
60	1	sp284		91	52	sp12		58	102	sp225	
54	2	sp38		88	53	sp11	GRP75	99	103	sp149	
31	3	sp25	beta-F1ATPase	89	54	sp15	HSP60	100	104	sp405	
30	4	sp507		90	55	sp7	HSC70	97	105	sp144	
56	5	sp112		53	56	sp850		98	106	sp52	
55	6	sp661		49	57	sp40		101	107	sp81	
17	7	sp258		50	58	sp194		95	108	sp33	Cytokeratin
32	8	sp379		51	59	sp407		94	109	sp35	
1	9	sp261		74	60	sp289		93	110	sp29	Cytokeratin
2	10	sp100		83	61	sp121		96	111	sp37	Cytokeratin
3	11	sp520		69	62	sp692		106	112	sp98	
4	12	sp730		84	63	sp19		114	113	sp243	
5	13	sp475		23	64	sp206		115	114	sp385	
6	14	sp491		75	65	sp1024		108	115	sp120	
11	15	sp207		78	66	sp115		102	116	sp105	
12	16	sp280		73	67	sp811		103	117	sp1021	
20	17	sp82		85	68	sp30		104	118	sp1019	
9	18	sp226		82	69	sp424		105	119	sp1020	
10	19	sp321		139	70	sp148		112	120	sp51	Gamma-Actin
7	20	sp57		135	71	sp75		113	121	sp9	
8	21	sp61		44	72	sp129		109	122	sp34	Beta-Actin
14	22	sp213		138	73	sp169		107	123	sp789	
15	23	sp268		151	74	sp97		110	124	sp685	
19	24	sp232		146	75	sp244		111	125	sp792	
13	25	sp448		133	76	sp396		149	126	sp13	Protein disulfide isomerase
28	26	sp177		137	77	sp872		119	127	sp114	
21	27	sp39		33	78	sp205		121	128	sp613	
22	28	sp481		92	79	sp47		147	129	sp178	
18	29	sp89		68	80	sp480		145	130	sp1062	
24	30	sp1002		71	81	sp67		134	131	sp24	
25	31	sp50		66	82	sp10	HSP90	136	132	sp250	
39	32	sp65		70	83	sp540		148	133	sp589	
29	33	sp578		142	84	sp306		36	134	sp140	
27	34	sp71		131	85	sp126		150	135	sp4	GRP78
86	35	sp55		132	86	sp545		144	136	sp62	
64	36	sp124	Lamin B	47	87	sp271		143	137	sp292	
65	37	sp341		140	88	sp295		122	138	sp332	
57	38	sp196		48	89	sp703		124	139	sp1	Endoplasmin
61	39	sp146		130	90	sp165		123	140	sp845	
63	40	sp386		46	91	sp99		126	141	sp302	
41	41	sp163		35	92	sp352		128	142	sp16	
78	42	sp335		42	93	sp134		129	143	sp72	
81	43	sp1026		34	94	sp316		125	144	sp246	
72	44	sp356		43	95	sp383		127	145	sp53	
26	45	sp135		52	96	sp1054		45	146	sp173	
77	46	sp249		59	97	sp186		120	147	sp339	
79	47	sp358		37	98	sp18		117	148	sp428	
80	48	sp490		38	99	sp8		116	149	sp360	
62	49	sp345		40	100	sp14	Beta-Tubulin	141	150	sp189	
67	50	sp496		16	101	sp138		118	151	sp229	
87	51	sp220									