Proteomic Maps−Toxicity Relationship of Halocarbons Studied with Similarity Index and Genetic Algorithm

Marjan Vračko,*,† Subhash C. Basak,‡ Kevin Geiss,§ and Frank Witzmann||

National Institute of Chemistry, Hajdrihova 19, Ljubljana, Slovenia, Natural Resources Research Institute, Duluth, Minnesota 55811, Human Effectiveness Directorate, Air Force Research Laboratory, Dayton, Ohio 45433, and Indiana University School of Medicine, Indianapolis, Indiana 46202

Received June 23, 2005

In this work we analyzed proteomic maps obtained from hepatocytes, which were treated with 14 halocarbons. A similarity index was introduced as a robust measure of similarity between two maps or between two selections of spots within the maps. A searching algorithm was used to identify the spots that may play an important role in toxicity mechanism. The highest correlation coefficients obtained between the similarity index and biological parameter were larger than 0.9.

INTRODUCTION

Halocarbons are used widely for a variety of purposes which include their use as solvents, starting materials in synthetic processes, and are therefore present throughout our environment. Since the earliest observations of the cell toxicity associated with carbon tetrachloride and chloroform, there has been an interest in understanding the molecular basis of halocarbon toxicity in better understanding which halocarbons are more toxic and which are less toxic.[1−5]

In the postgenomic era the technologies of genomics and proteomics are being increasingly used in elucidating the detailed biochemical basis of the mechanism of action of toxicants. Two-dimensional electrophoresis (2DE) is a powerful tool for the analysis of protein expression.[6,7] It involves the orthogonal separation of complex protein mixtures based on protein isoelectric point (pI) via first-dimension isoelectric focusing, and molecular mass via second dimension sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Despite its limitations, 2DE generates high resolution 2D proteomics maps where each protein (or a few proteins of nearly identical pI and mass) occupies distinct x, y coordinates. If cells are exposed to xenobiotics their biological response is in changing of protein expression, which can be recorded as changes in intensity of protein spots. Usually we do not understand completely the events in a cell, but we can recognize the changes in 2D proteomics maps. Several attempts have been done to characterize the 2D proteomics maps or submaps (selection of spots within the map) numerically. The basic theme of such attempts is to represent a protein map or particular submaps with a single number (or a set of numbers) that enables algorithmic (computational) treatment of a large number of protein maps. A series of approaches was done considering the 2D protein maps as pattern in three-dimensional space where x and y coordinates determinate the position of a spot in the 2D map and the z coordinate represents the intensity of spot. Such patterns of spots from different proteomics maps can be compared used robust similarity index.[8,9] Alternatively, they can be associated to curves in space, which can be further described with different invariants.[10−12] A similar strategy has been applied when representing DNA sequences on graphical or numerical ways.[13−18]

Trohalaki et al. studied the toxic effects of a set of 20 halocarbons in the rat hepatocytes as measured by different cellular toxicity indicators.[19] The exposed hepatocytes were analyzed using 2-D gel electrophoresis technology in the lab of Frank Witzmann.[20,21] As shown in previous studies, rat hepatocytes have proven useful in vitro to investigate mechanisms of toxicity that would be manifested in the whole organ system.[1,2,19] The perturbation in the 2-D gel pattern can be looked upon as the toxic effects. But the principal problem is that there are data on over 1400 spots. A visual inspection of these proteomics patterns does not lead to any discernible pattern easily. Therefore, our research team has resorted to the development of compact biodescriptors and individual biomarkers using methods of discrete mathematics and statistics, respectively, as outlined below:

1. Invariants of graphs/matrices associated with proteomics maps[12]
2. Information theoretical biodescriptors[22]
3. Biodescriptors from spectrum-like representation of proteomics maps[8]
4. Statistical approaches to discover critical protein biomarkers.

We report the study of 2DE data recorded from rat hepatocytes, which were treated with 14 halocarbons. For all 14 halocarbons six biological parameters were determined. Our goal was to find the reliable relationship between a pattern of spots and biological parameters. The strategy and methods are described in the section below. Briefly, we used the similarity index to express the relationship between the selections of spots in untreated (reference map) and treated cells. A searching algorithm, which mimics a natural selection, was used to search over thousands of possible

* Corresponding author phone: +386 1 4760315; fax: +386 1 4760300; e-mail: marjan.vracko@ki.si.
† National Institute of Chemistry.
‡ Natural Resources Research Institute.
§ Air Force Research Laboratory.
|| Indiana University School of Medicine.

10.1021/ci0502597 CCC: $33.50
Published on Web 12/18/2005

Table 1. Chemical Dose Range

chemical	dose range (mM)
CCl_4	0.043−1.380
CBr_4	0.075−0.200
$CHBrCl_2$	0.500−10.000
$CHBr_2Cl$	0.290−2.350
CBr_2Cl_2	0.100−1.000
$CBrCl_3$	0.310−2.500
CH_2Br_2	2.500−25.000
$CHCl_3$	1.000−10.000
C_2Cl_4	0.100−2.000
C_2HCl_3	0.277−2.660
1,2-$C_2H_4Cl_2$	3.160−38.000
1,1,2-$C_2H_3Cl_3$	1.080−6.480
1,1,1-$C_2H_3Cl_3$	0.063−0.501
1,1,2-$C_2H_3Br_3$	0.310−2.460
1,1,2,2-$C_2H_2Cl_4$	1.188−4.750
CH_2Cl_2	10.000−100.000
$CHBr_3$	0.143−4.580
CH_2BrCl	10.000−100.00
1,2-C_2H_4BrCl	0.420−3.364
1,2-$C_2H_4Br_2$	0.070−8.630

patterns (combinations of spots). At the end we analyzed the patterns with correlation coefficients to a biological parameter larger than 0.75.

MATERIALS AND METHODS

Chemicals. Chemical and biological reagents were purchased from Sigma Chemical Co. (St. Louis, MO) unless otherwise stated. This includes most of the halogenated test chemicals (111-$C_2Cl_3H_3$, 112-$C_2Br_3H_3$, 112-$C_2Cl_3H_3$,12-$C_2Br_2H_4$, 12-C_2BrClH_4, 12-$C_2Cl_2H_4$, C_2Cl_3H, $C_2Cl_4H_2$, CBr_2H_2, $CBrClH_2$, CCl_2H2, $CHBr_3$). C_2Cl_4 and $CHCl_3$ were purchased from Fisher Scientific (Pittsburgh, PA). Chemicals were used as provided by the manufacturer, without further purification. The halogenated chemicals are summarized in Table 1 and shown their names and their experimental dose ranges. Doses were selected for each compound that corresponded to a level eliciting a biologically significant response in the six endpoints selected for this study. Normalized amounts of protein were used for expression analysis, and equivalent loadings were used for each gel sample. Ketamine (Injectable, U.S.P. grade) was purchased from Parke-Davis (Moris Plains, NJ). Xylazine (Injectable, U.S.P. grade) was purchased from Mobay Corporation (Shawnee, KS). Collagenase (Type D) and Protein Assay ELS reagents were purchased from Roche (formerly Boehringer-Mannheim) Biochemicals (Indianapolis, IN). Type I rat tail collagen was purchased from Upstate Biotechnology (Lake Placid, NY). CHEE's modified Eagle Medium (Formula No. 88-5046EA) and Hank's blanched salt solution (HBSS) were purchased from GibcoBRL/Life Technologies (Rockville, MD).

Animals. Fischer $CD^{+2}F^{R}$(F344)/CrlBR (F-344) male rats (225−300 g) were obtained from Charles River Laboratories. All animals used in this study were handled in accordance with the principles stated in the "*Guide for the Care and Use of Laboratory Animals*", National Research Council, 1996, and the Animal Welfare Act of 1988, as amended. Rats were anesthetized with 1 mL/kg of a mixture of ketamine (70 mg/mL) and xylazine (6 mg/mL) prior to undergoing in situ liver perfusion.

Hepatocyte Isolation and Culture. Rat livers were digested by perfusion using the two-step method of Seglen.[23]

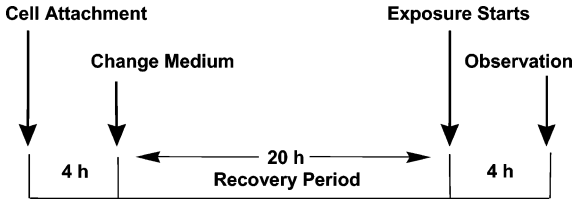

Figure 1. The cell culturing exposure protocol. The time for each segment is indicated.

In the first perfusion step, the liver was perfused via the hepatic portal vein with perfusion buffer (37 °C) consisting of Hank's balanced salt solution (HBSS; pH 7.2) lacking calcium and magnesium and supplemented with 15 mM 4-(2-hydroxyethyl)-1-piperazineethane-sulfonic acid (HEPES), heparin (2.0 U/mL), and ethylene-bis(oxyethylenenitrilo)-tetraacetic acid (EGTA; 0.5 mM). This first step of the antigrade two-step perfusion method was accomplished by cannulation of the portal vein followed by clamping of the posterior vena cava anterior to the diaphragm. Flow of perfusion buffer was then activated (20 mL/min), immediately followed by the cutting of the posterior vena cava (anterior to the renal vein) to allow drainage of the perfusion buffer. Following complete removal of blood from the liver, the second step of the liver perfusion was initiated by continuous perfusion with digestion buffer. In this second perfusion step, digestion buffer (37 °C) consisting of complete HBSS (pH 7.2) supplemented with collagenase (0.26 Wunsche Units/mL) was perfused through the liver to digest interstitial connective tissue. Perfusion was continued until the hepatocytes were completely disaggregated (approximately 20 min). Viable primary rat hepatocytes were washed (37 °C) with complete HBSS (pH 7.2) and enriched three times by low speed centrifugation at 50 g for 3 min. Typical viabilities of isolated hepatocytes ranged from 80 to 95% with yields of 250 to 400 million cells as determined by trypan blue dye exclusion. For the cell culture studies, freshly isolated hepatocytes in suspension were adjusted to a cell density of 1.0×10^6 cell/mL in cell attachment medium consisting of CHEEs modified culture medium (CHEE; pH 7.2) supplemented with HEPES (10mM), insulin/transferring/sodium selenite solution (final concentration of 5 μg/mL, 5 μg/mL, and 5 ng/mL, respectively), gentamycin (0.1 mg/mL), and dexamethasone (0.4 mg/mL). Cells were seeded in 6-well (1.0×10^6 cells/well) culture plates. Plates were previously coated with Type I rat tail collagen (25.0 μg/mL stock) at 2.6 μg/cm^2. After 4 h of incubation in a 95% air/5% CO_2 incubator at 37 °C, cell attachment medium was removed, and the rat hepatocytes were incubated in fresh CHEEs culture medium lacking dexamethasone. Hepatocytes were incubated for an additional 20 h in order to recover from the stress incurred during the isolation procedure. Exposures to halogenated chemicals were initiated at this point. The overall cell culture scheme is shown in Figure 1.

Cell Culture and Chemical Exposure. All exposure to the test chemicals were accomplished in the VITROBOX exposure system.[24,25] Chemicals were equilibrated in both air and medium, to ensure consistent dosage and preventing loss to the atmosphere, which overcomes a shortcoming of traditional studies in vitro. The system was basically a rectangular glass chamber, approximately $8 \times 25 \times 25$ cm interior dimensions (~5 L volume). For each chemical dose,

there was a separate VITROBOX chamber and Tedlar dosing bag. Chemical dose ranges are also shown in Table 1.

Prior to exposures (~30 min) the approximate chemical concentration was added to the Tedlar dosing bags and placed back into a 37 °C incubator to allow the chemicals to volatilize. Also, immediately prior to cell culture dosing, the chemicals were diluted into the appropriate medium. No other solvents or vehicles (e.g. ethanol) were used in preparation of the dosing solutions. Media in the cell culture plates was replaced with the chemical dosing media. Plates were then immediately placed into the chamber without plate lids. The chamber was then closed, and the dosing and capture bags were attached to the chamber. The dosing of the chamber was accomplished by manually compressing the dosing bag, forcing the dosing atmosphere into the chamber, while it exhausted into the capture bag. Following the dosing process, all ports were sealed, and the chamber was placed into a 37 °C incubator. The cells were exposed in the chamber for 4 h and then removed from the chambers for analysis.

Two-Dimensional Electrophoreses. Hepatocytes were prepared for 2DE essentially as described elsewhere.[7,21,24] Samples were analyzed for the control and $EC20_{SH}$ samples. $EC20_{SH}$ was selected as a key indicator of biological response, where the cells should clearly be in an oxidatively challenged state.[26] Briefly, the cultured cells were solubilized directly in situ after removal of medium. Four hundred microliters of lysis buffer containing 9 M urea, 4% Igepal CA-630 ([octylphenoxy] polyethoxyethanol), 1% DTT, and 2% ampholytes (pH 8−10.5) were added directly to each well. The culture plates were then placed in a 37 °C incubator for 1 h with intermittent manual agitation. Following the 1 h solubilization, the entire volume was removed from each well and placed in 2 mL Eppendorf tubes. Each sample was then sonicated with a Fisher Sonic Dismembranator using 3 × 2 s bursts at instrument setting #3. Sonication was carried out every 15 min for 1 h after which the fully solubilized samples were transferred to a cryotube for storage at −80 °C until thawed for analysis.

Proteins were resolved by 2D electrophoresis (2DE) using the ISO-DALT System in which up to 20 first- and second-dimension gels can be run simultaneously.[21]

After staining, the 2DE protein patterns were scanned under visible light at 200 μm/pixel resolution using the Fluor-S MAX MultiImager System (Bio-Rad, Hercules CA). Image data were analyzed using PDQuest software, (Bio-Rad)Background was subtracted, and peaks for the protein spots were located and counted. Because total spot counts and the total optical density are directly related to the total protein concentration, individual protein quantities were thus expressed as parts-per-million (PPM) of the total integrated optical density. A reference pattern was constructed, and gel matching was performed with image analysis software matching each gel in the match set to the reference gel. Numerous proteins that were uniformly expressed in all patterns were used as landmarks to facilitate rapid gel matching until the majority of protein spots were matched across all gel patterns. Protein abundances were normalized based on the total image density of all sample gels in the match set, and the reference spot number, x and y coordinates, and integrated densities were exported to spreadsheet software for further analysis.

Table 2. Distribution of Spots Respecting the Relative Error (SD/Mean)

SD/mean	no. of spots
0.1	14
0.2	69
0.3	105
0.4	97
0.5	90
0.6	81
0.7	74
0.8	59
0.9	54
1.0	54
>1.0	703

The considered data set consists of 92 2D proteomic maps; 27 of them were recorded from untreated cells (control), and another 65 maps were recorded from treated cells. In each of the maps 1401 spots were assigned. Twenty-seven control samples formed a basis for the first statistical analysis. For each spot's intensity the average value (mean) and the standard deviation (SD) over all 27 samples were calculated. The distribution of spots regarding the relative error (SD/mean) is given in Table 2 and in Figure 2. For further analysis we selected a set of 188 spots with the relative error smaller than 0.39.

Biological Parameters. The cells have been exposed to 14 halocarbons: 1,1,1-trichloroethane (111-$C_2Cl_3H_3$), 1,1,2-tribromoethane (112-$C_2Br_3H_3$), 1,1,2-trichloroethane (112-$C_2Cl_3H_3$), 1,2-dibromoethane (12-$C_2Br_2H_4$), 1-bromo-2-chloroethane (12-C_2BrClH_4), 1,2-dichloroethane (12-$C_2Cl_2H_4$), trichloroethylene (C_2Cl_3H), tetrachloroethylene (C_2Cl_4), tetrachloroethane ($C_2Cl_4H_2$), dibromomethane (CBr_2H_2), chlorobromomethane ($CBrClH_2$), dichloromethane (CCl_2H2), trichloromethane (CCl_3H), tribromomethane ($CHBr_3$). The biological parameters are expressed as a logarithm of nanomolar concentration causing a defined biological effect. Six parameters were considered: $EC50_{MTT}$ = effective concentration which causes 50% decrease in the measured response for the MTT endpoint, $EC50_{LDH}$ expresses the 50% decrease in response in the LDH endpoint, $EC20_{SH}$ = effective concentration causing a 20% decrease in the zhiol levels, LEC_{LP} = lowest effective concentration causing lipid peroxidation, LEC_{ROS} = lowest effective concentration causing reactive oxygen species production, LEC_{CAT} =

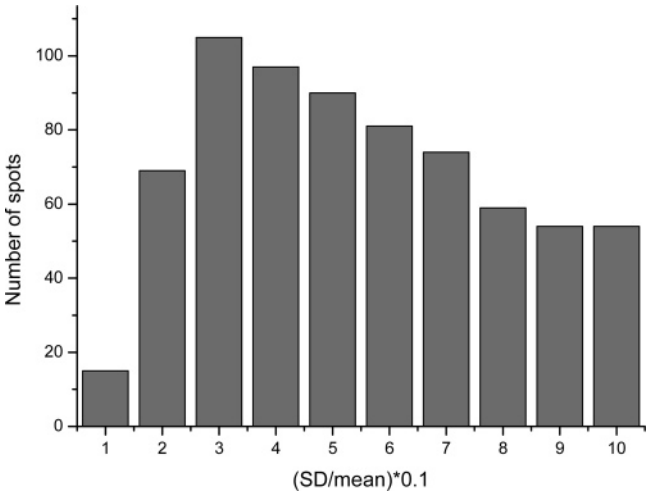

Figure 2. Distribution of spots respecting the relative error (SD/mean).

lowest effective concentration causing inhibition of catalase enzyme.

COMPUTATIONAL METHODS

Similarity Index. One defines a similarity index as a weighted projection between two maps *a* and *b*

$$s^{a,b} = \frac{1}{N}\sum_{i=1}^{N}\frac{z_i^a z_i^b}{\max\,(z_i^a, z_i^b)^2} \qquad (1)$$

where z^a and z^b are intensities of a spot located at the same position in maps. Some further details on the similarity index can be found elsewhere.[8,9] In the present study we defined the similarity index between the treated sample and the control sample. If the sum in eq 1 runs over all spots in proteomic maps the similarity index measures a global similarity between maps. It is to emphasize that the information of entire map is compressed into a single number. In the hierarchical view of descriptors and QSAR models it is a global biodescriptor.[27,28] If the sum runs over a selection of spots the index expresses the local similarity (local biodescriptor). In the presented study we created thousands of different selections of spots and calculated the similarity indices between reference maps and treated maps. In a further step we calculated correlation coefficients between indices and biological parameters. The first generation of selections were generated randomly, and other generations were created via algorithm described below.

Searching Algorithm. An algorithm, which mimics a natural evolution, was used for searching of spots, which correlate to a biological parameter. Such algorithms, which are in a more general form known as genetic algorithms, are often used to search for the most important variables.[29−31] The applied algorithm follows the idea of stepwise selection as follows.

1. Five thousand different combinations of spots (patterns) were created each containing 10 to 20 spots. The combinations were set randomly in such a way that all spots were considered equivocally.

2. For all patterns the similarity indices between the control sample and the treated one were calculated according to the eq 1.

3. The correlation coefficients between similarity indices and biological parameters were calculated. The correlation coefficient is the criterion to evaluate the patterns.

4. The patterns were arranged accordingly to the descendent of correlation coefficient.

5. The neighboring patterns exchanged their spots to build the new generation of patterns. The steps 2−5 are repeated 50 times.

At the end we count for each spot how often it occurs in patterns with a correlation coefficient larger than 0.75.

RESULTS AND DISCUSSION

General Consideration. Figure 3 shows the highest correlation coefficient between the similarity index and the parameter $EC50_{MTT}$ for each generation. One recognizes from the figure that after 10 generations the correlation coefficients reach the plateau and oscillate between values 0.9 and 0.95. Figure 4 shows the distributions of correlation coefficients for the parameter $EC50_{MTT}$ for all considered selections. The sharp peak in the first generation means that most of the correlation coefficients are centered at 0.5, while in the next generations the peak becomes broader and covers regions to lower and higher values. Figure 5 shows the distribution of correlation coefficients after 50 generations for six different biological parameters. There are differences in distributions, as for example, the difference in distributions of parameter $EC50_{MTT}$ and parameters $EC20_{SH}$ and LEC_{CAT}.

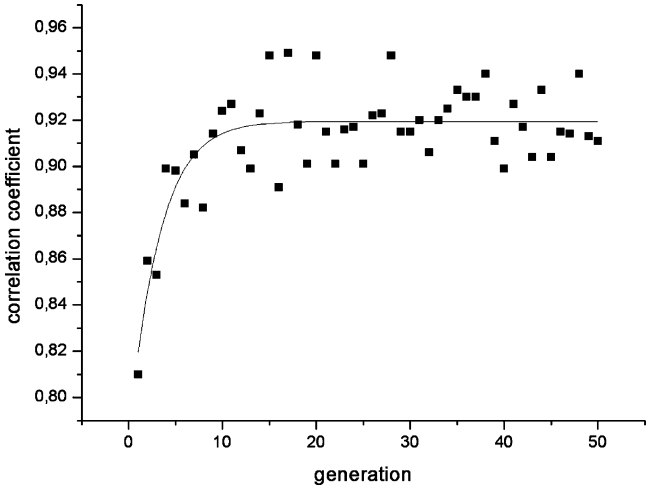

Figure 3. Highest correlation coefficients over 50 generations.

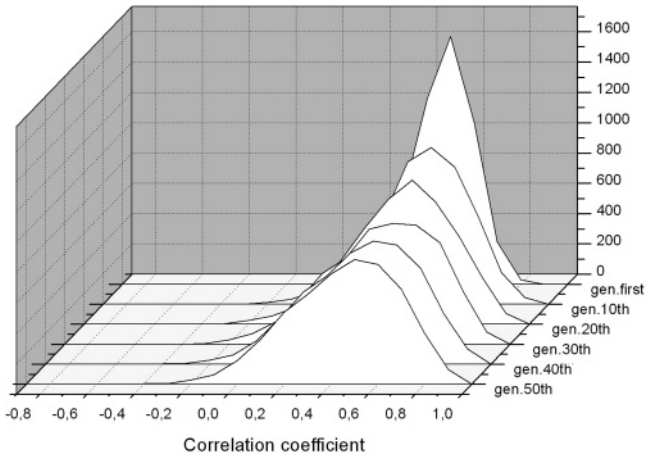

Figure 4. Distribution of correlation coefficients over 50 generations.

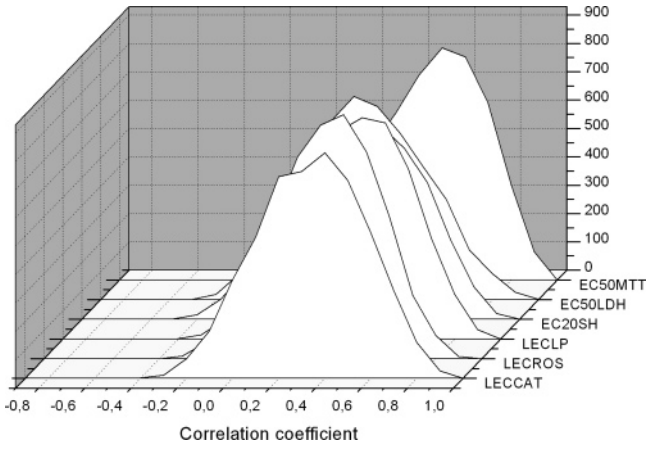

Figure 5. Distribution of correlation coefficients for the 50th generation for six different biological parameters.

Table 3. Selections of the Most Important Spots for Six Biological Parameters[a]

	r_{max}	r_{avr}	N
$EC50_{MTT}$			
15 most important spots:			
70 108 27 73 64 161 15 55 77 45 143 51 86 7 24	0.9272	0.8970	446
removed spots: 70 108 27			
15 most important spots:			
77 64 73 15 45 55 161 7 143 175 32 51 146 14 111	0.9428	0.8751	182
removed spots: 70 108 27 73 64 161			
15 most important spots:			
77 143 15 55 175 45 7 17 51 86 144 14 32 105 59	0.9096	0.8685	34
removed spots: 70 108 27 73 64 161 15 55 77			
15 most important spots:			
74 14 156 143 50 24 49 101 89 175 81 7 17 105 86	0.8184	0.8043	42
removed spots: 70 108 27 73 64 161 15 55 77 45 143 51			
15 most important spots			
110 86 17 7 175 81 50 169 24 59 160 118 89 116 65	0.8029	0.7778	46
$EC50_{LDH}$			
15 most important spots:			
108 55 45 15 161 70 175 101 77 51 17 169 89 137 160	0.9139	0.8488	116
removed spots: 108 55 45			
15 most important spots:			
15 161 70 17 175 77 101 169 123 125 137 102 19 183 65	0.8952	0.8223	135
removed spots: 108 55 45 15 161 70			
15 most important spots:			
17 175 77 126 102 89 101 82 160 124 169 137 183 61 156	0.8409	0.7937	76
removed spots: 108 55 45 15 161 70 175 101 77			
15 most important spots:		0.6363	0
no correlation with $r > 0.75$			
$EC20_{SH}$			
15 most important spots:			
70 108 15 64 161 73 27 45 79 51 32 143 77 126 24	0.9167	0.8500	124
removed spots: 70 108 15			
15 most important spots:			
64 27 77 161 45 175 73 126 32 143 89 55 79 158 160	0.8747	0.8000	119
removed spots: 70 108 15 64 161 73			
15 most important spots	0.8579	0.7978	98
143 27 45 175 77 55 89 7 32 106 156 24 79 126 61			
removed spots: 70 108 15 64 161 73 27 45 79			
15 most important spots:	/	0.7090	1
no correlation with $r > 0.75$			
LEC_{LP}			
15 most important spots:			
108 160 83 15 70 55 114 45 163 89 161 61 102 64 118	0.9216	0.8539	301
removed spots: 108 160 83			
15 most important spots:			
15 55 70 102 61 45 89 183 118 17 161 126 7 114 29	0.8883	0.8449	181
removed spots: 108 160 83 15 70 55			
15 most important spots:			
102 114 160 61 17 45 89 156 175 118 64 126 29 8	0.8820	0.8247	92
removed spots: 108 160 83 15 70 55 114 45 163			
15 most important spots:			
126 102 183 137 89 69 17 49 152 156 117 159 110 61 16	0.7986	0.7646	6
removed spots: 108 160 83 15 70 55 114 45 163 89 161 61			
15 most important spots:			
95 7 147 115 48 59 128 134 85 168 171 60 101 72 30	0.7677	0.7330	2
LEC_{ROS}			
15 most important spots:			
70 108 77 15 64 161 27 32 156 24 45 73 169 107 158	0.8728	0.8265	84
removed spots: 70 108 77			
15 most important spots:			
161 143 66 15 27 175 45 32 158 24 169 7 73 17 79	0.8566	0.7972	96
removed spots: 70 108 77 15 64 161			
15 most important spots:			
143 45 27 32 175 158 17 126 169 55 156 24 153 7 144	0.8259	0.7863	68
removed spots: 70 108 77 15 64 161 27 32 156			
15 most important spots:			
no correlation with $r > 0.75$	/	0.6806	0
LEC_{CAT}			
15 most important spots:			
45 70 55 77 15 108 82 101 17 137 175 161 183 160 114	0.8946	0.8535	153
removed spots: 45 70 55			
15 most important spots:			
77 15 108 101 137 160 175 17 82 19 161 156 124 183 127	0.8929	0.8329	177
removed spots: 45 70 55 77 15 108			
15 most important spots:			
16 81 175 161 101 160 144 19 72 143 69 156 137 61 85	0.8700	0.8162	107
removed spots: 45 70 55 77 15 108 82 101 17			
15 most important spots:			
51 161 169 68 69 79 81 92 111 18 14 175 136 149	0.7684	0.7209	1
removed spots: 45 70 55 77 15 108 82 101 17 137 175 161			
15 most important spots:			
50 160 76 77 81 84 103 107 143 139 156 158 8 178	0.7547	0.6926	2

[a] r_{max} is the maximal correlation coefficient, r_{avr} is the average of maximal correlation coefficients over 49 generations, and N is the number of correlations larger than 0.7 found in the last, i.e., 50th generation.

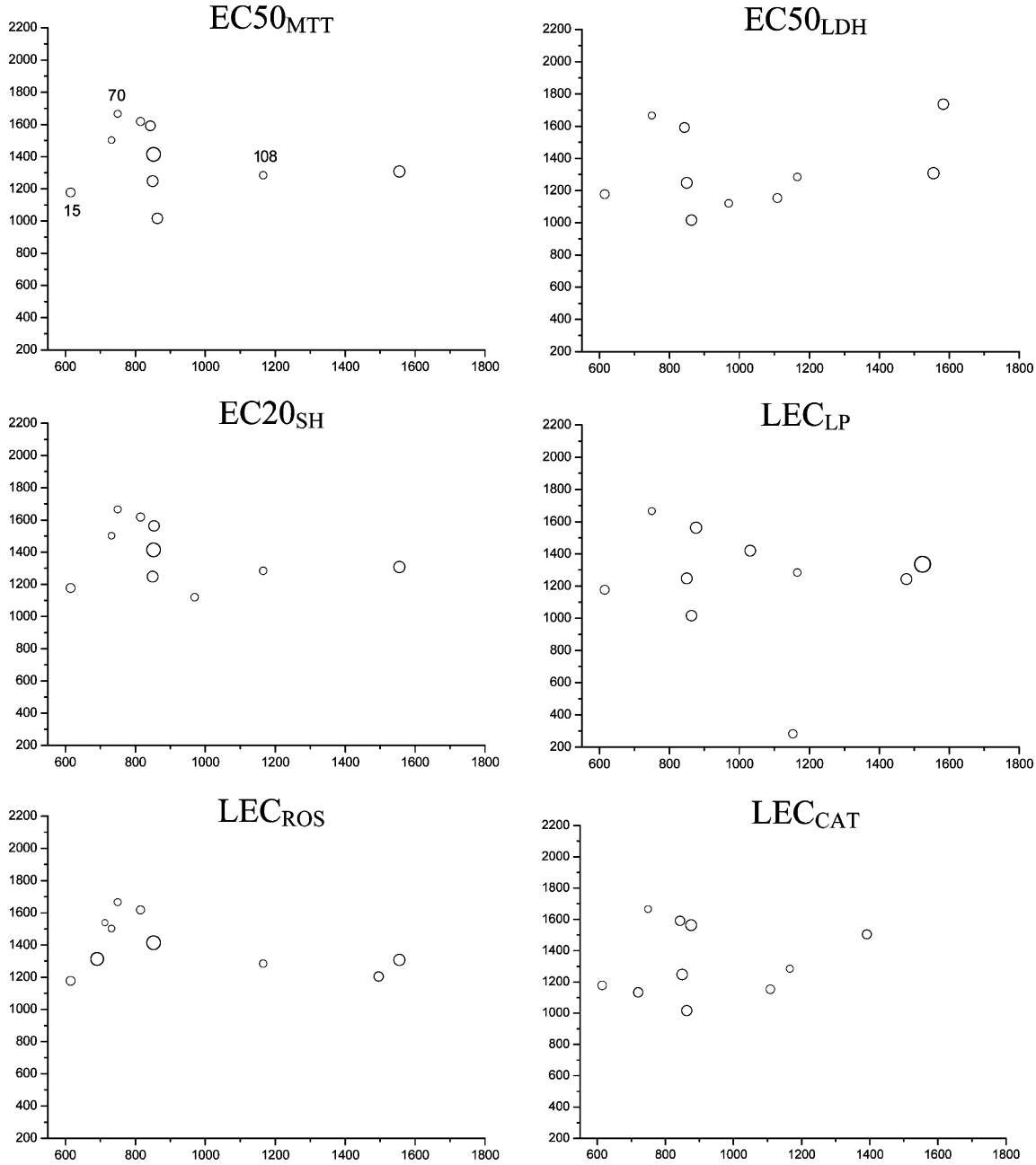

Figure 6. Ten mostly occurring spots for six different biological parameters as they appear in the 2D gel. The size of the bubbles is proportional to the logarithm of intensity. (The integrated intensity of protein spots was calculated by the image analysis software, PDQuest.) In the first picture ($EC50_{MTT}$) the spots 15, 70, and 108 are labeled.

Selecting of Spots. We analyzed all relationships between similarity indices and parameters with correlation coefficients larger than 0.75. In this analysis we determined for each spot the number that expresses how often it occurs in those relationships. In this way we selected for each parameter 10 spots, which are shown in Table 3. The positions as they appear in 2D gels are shown in Figure 6. To investigate the importance of spots and to check if eventual chance correlations were found several leave-out-spots calculations were performed. In these runs the algorithm was applied on reduced data set where three, six, nine, or twelve 'important' spots were removed from the input data. Results are shown in Table 3 where we report the following: the largest correlation coefficients found in an individual run, the average of largest correlation coefficients over all generations, and as additional information, the number of correlations larger than 0.7 found in the last generation (*N*). A general trend is obvious from the table, when the important spots are removed the correlation coefficients and the *N* drop. However, there are differences between different parameters. The parameter $EC50_{MTT}$ shows the best correlation to the spots. After the deleting of 12 spots the average correlation coefficient drops from 0.8970 to 0.7778. In the parameter $EC50_{LDH}$ the average *r* drops from 0.8488 to 0.6363 when nine spots are removed. No correlation with *r* larger than 0.75 was found. Similar trends can be observed in LEC_{ROS} and $EC20_{SH}$ parameters. The fundamental question is if the same spots occur as important in all six biological parameters. Looking to Table 3 different conclusions can be made, as for example, the spots 15, 70, and 108 appear as important in all six biological endpoints (see Figure 6). Spots 27 and 73 are important in endpoints $EC50_{MTT}$, LEC_{ROS}, and $EC20_{SH}$

but not in the endpoints $EC50_{LDH}$, LEC_{LP}, and LEC_{CAT}. Spot 77 appears in five endpoints with the exemption of LEC_{LP}. It is to emphasize that our conclusions ground only on numerical analysis of data.

CONCLUSIONS

In the presented report we studied proteomic maps derived from hepatocytes, which were treated with 14 halocarbons. On the other side, six different biological parameters were known for the same set of compounds measured on the same in vitro testing system. Using the similarity index and searching algorithm we selected sets of spots associated with proteins, which may play key roles in biological processes. The criterion to locate the spots was the correlation coefficient between the similarity index, which is a robust measure of similarity between reference and treated samples, and the biological parameters, which were determined in in vitro studies. In several test runs the algorithm was applied on reduced sets of spots where "the important spots" were leaved-out. In all cases the highest correlation coefficients drop, i.e., the selected spots contribute an essential part to the correlation between similarity indices and biological parameters. Furthermore, the similarity indices calculated from selected spots can be used in advanced QSAR models as biodescriptors. In this way we can include the biological parameters into QSAR models, not knowing the precise mechanism of interaction between a chemical and protein. On the other hand, the selected spots can be used in further biochemical analyses to indicate the possible mechanisms of activities. This work is still in progress.

ACKNOWLEDGMENT

M.V. gratefully thanks Ministry of Education, Science and Sport of the Republic of Slovenia that supports this work under contract P1-0017. This material is based on research sponsored by the Air Force Research Laboratory, under agreement number F49620-02-1-0138 and cleared for public release as document ASC-01-0868. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. This is a contribution 391 from the Center for Water and Environment of the Natural Resources Research Institute.

REFERENCES AND NOTES

(1) Crebelli, R.; Andreoli, C.; Carere, A.; Conti, L.; Crochi, B.; Cotta-Ramusino, M.; Benigni, R. Toxicology of halogenated aliphatic hydrocarbons: structural and molecular determinants for the disturbance of chromosome segregation and the induction of lipid peroxidation. *Chem.-Biol. Interact.* **1995**, *98*, 113−129.

(2) Geiss, K. T.; Frazier, J. M. QSAR modeling of oxidative stress in vitro following hepatocyte exposure to halogenated methanes. *Toxicol. in Vitro* **2001**, *15*, 557−563.

(3) Roszak, S.; Koski, W. S.; Kaufmann, J. J.; Balasubramanian, K. Structure and energetics of CFCl, CFBr, and CFI radical ions. *J. Chem. Phys.* **1997**, *106*, 7709−8813.

(4) Bajzer, Z.; Randic, M.; Plavsic, D.; Basak, S. C. Novel map descriptors for characterization of toxic effects in proteomics map. *J. Mol. Graphics Modell.* **2003**, *22*, 1−9.

(5) Basak, S. C.; Balasubrimanian, K.; Gute, B. D.; Mills, D.; Gorczynska, A.; Roszak, S. Prediction of cellular toxicity of halocarbons from computed chemodescriptors: A hierarchical QSAR approach. *J. Chem. Inf. Model.* **2003**, *43*, 1103−1109.

(6) Lilley, K. S.; Razzak, A.; Cupree, P. Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantification. *Curr. Opin. Chem. Biol.* **2002**, *6*, 46−50.

(7) Witzmann, F. A. Proteomics applications in toxicology. In *Comprehensive Toxicology, Vol. XIV: Cellular and Molecular Toxicology*; Vanden Heuvel, J. P., Greenlee, W. F., Perdew, G. H., Mattes, W. B., Eds.; Elsevier: New York, 2002; pp 539−558.

(8) Vracko, M.; Basak, S. C. Similarity study of proteomic maps. *Chemom. Intell. Lab. Syst.* **2004**, *70*, 33−38.

(9) Ross, A.; Senn, H. Automation of measurements and data evaluation in biomolecular NMR screening. *DDT* **2001**, *6*, 583−593.

(10) Randic, M. On the graphical representation of proteomics and their numerical characterization. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1330−1338.

(11) Randic, M.; Zupan, J.; Novic, M. On 3-D graphical representation of proteomics maps. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1339−1344.

(12) Randic, M.; Witzmann, F.; Vracko, M.; Basak, S. C. On characterization of proteomics maps and chemically induced changes in proteomes using matrix invariants: application to peroxisome proliferators. *Med. Chem. Res.* **2001**, *10*, 456−479.

(13) Hamori, E. Graphical representation of long DNA sequences by methods of H curves, current results and future aspects. *BioTechniques* **1989**, *7*, 710−720.

(14) Nandy, A. New graphical representation and analysis of DNA sequence structure. I. Methodology and application to globin genes. *Curr. Sci.* **1994**, *66*, 309.

(15) Randic, M.; Vracko, M.; Nandy, A.; Basak, S. C. On 3-D graphical representation of DNA primary sequences and their numerical characterization. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1235−1244.

(16) Randic, M.; Gou, X.; Basak, S. B. On the characterization of DNA primary sequences by triplet of nucleic acid bases. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 619−626.

(17) Randic, M.; Vracko, M.; Zupan, J.; Novic, M. Compact 2-D graphical representation of DNA. *Chem. Phys. Lett.* **2003**, *373*, 558−562.

(18) Randic, M.; Vracko, M.; Lers, N.; Plavsic, D. Analysis of similarity/dissimilarity of DNA sequences based on novel 2-D graphical representation. *Chem. Phys. Lett.* **2003**, *371*, 202−207.

(19) Trohalaki, S.; Pachter, R.; Geiss, K. T.; Frazier, J. M. Halogenated aliphatic toxicity QSARs employing metabolite descriptors. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1186−1192.

(20) Witzmann, F. A.; Clack, J. W.; Geiss, K.; Hussain, S.; Juhl, M. J.; Rice, C. M.; Wang, C. Proteomic evaluation of cell preparation methods in primary hepatocyte cell culture. *Electrophoresis* **2002**, *23*, 2223−2232.

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

(22) Basak, S. C.; Gute, B. D.; Witzmann, F. A. Information-theoretic biodescriptors for proteomics maps: Development and applications in predictive toxicology. *WSEAS Trans. Inf. Sci. Appl.* **2005**, *7*, 996−1001.

(23) Seglen, P. Preparation of isolated rat liver cells. In *Methods in Cell Biology*; Prescott, D. M., Ed.; New York: Academic Press: 1976; pp 29−83.

(24) Neuhoff, V.; Arnold, N.; Taube, D.; Ehrhardt, W. Improved staining in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **1988**, *9*, 255−262.

(25) The VITROBOX method was patented in 2003 − #US6593136-B1.

(26) Slater, A. F. G.; Stefan, C.; Nobel, I.; van den Dobblsteen, D. J.; Orrenius, S. Signaling mechanisms and oxidative stress in apoptosis. *Toxicol. Lett.* **1995**, *82*−*83*, 149−153.

(27) Basak, S. C.; Mills, D. Prediction of mutagenicity utilizing a hierarchical QSAR approach. *SAR QSAR Environ. Res.* **2001**, *12*, 481−496.

(28) Basak, S. C.; Mills, D.; Balaban, A. T.; Gute, B. D. Prediction of mutgenicity of aromatic and heteroaromatic amines from structure: a hierarchical QSAR approach. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 671−678.

(29) Mazzatorta, P.; Vracko, M.; Benfenati, E. ANVAS: artificial neural variables adaptation system for descriptor selection. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 335−346.

(30) Ros, F.; Pintore, M.; Chretien, J. R. Molecular descriptor selection combining genetic algorithms and fuzzy logic: application to database mining procedure. *Chemom. Intell. Lab. Syst.* **2002**, *63*, 15.

(31) Goldberg, D. E. *Genetic algorithms in search, optimization & machine learning*; Addison-Wesley: New York, 1989.

CI0502597