On 3-D Graphical Representation of Proteomics Maps and Their Numerical Characterization

Milan Randić,* Jure Zupan, and Marjana Novič

National Chemistry Institute of Slovenia, Ljubljana, Hajdrihova 19, Slovenia

Received January 3, 2001

We consider numerical characterization of proteomics maps by representing a map as a three-dimensional graphical object based on x, y coordinates of the spots and using their relative abundance as the z coordinate. In our representation the protein spots are first ordered based on their relative abundance and labeled accordingly. In the next step a 3-D path is constructed connecting spots having adjacent labels. Finally a matrix is constructed by assigning to each pairs of labels (i, j) matrix element, the numerical value of which is based on the quotients of the Euclidean distance and the distance along the 3-D zigzag between the two points. The approach has been illustrated on a fragment of a proteomics map and compared with 2-D graphical representation of proteomics maps.

INTRODUCTION

In the preceding paper, a novel approach to analysis of proteomics maps has been outlined in which a proteomics map is "transformed" into a geometrical pattern of line segments obtained by first ordering spots relative to their abundance and then connecting spots with adjacent labels.[1] A result is a rather complex 2-D zigzag path that crosses itself several times that has been referred to as the map "fingerprint". It appears that the fingerprint pattern is characteristic for a map in the sense that different maps are expected to yield distinct fingerprint patterns. Important advantage of such novel graphical view of proteomics maps is that the zigzag graphical representation is susceptible to rigorous mathematical analysis. Randić, Kleiner, and DeAlba[2] have developed an approach in chemical graph theory[3] which offers numerical characterization of molecular skeletons and mathematical curves embedded in a space based on a set of structural invariants derived from suitably constructed matrices associated with molecular skeletons or mathematical curves. In this paper we want to generalize the initial characterization of proteomics maps based on 2-D fingerprint patterns by considering representation of proteomics maps in 3-D, where the third coordinate indicates relative abundance. The 2-D representation of proteomics maps only indirectly considers the relative abundance of protein spots via the ordering of spot and assignment of labels. Now instead of semiqualitative representation of proteomics maps we will consider fully quantitative representation of protemics maps in which numerical values of relative abundance is taken into account.

ON 3-D REPRESENTATION OF A MAP

Currently proteomics maps that are reported as experimental gel photographs are often reproduced as "bubble" diagrams by a computer software program in which protein spots are represented by circles of different size. In the preceding paper we have illustrated one such bubble map and have also listed the (x, y) coordinates and abundance for the 20 most abundant proteins. In Figure 1 we show a 3-D zigzag path connecting these 20 most intensive spots. The zigzag path is descending from the maximal abundance value at about 144.4 70 to the minimal value of 72.2. Projection of the zigzag curve on the x, z plane gives the map the fingerprint, that was illustrated in ref 1 and was the basis for 2-D representation of the proteomics map. The problem to consider is how can one arrive at a quantitative characterization of maps given either as bubble diagrams or defined by a 3-D zigzag path that may facilitate comparison of different maps and even associate with such maps some numerical characterization. We decided to expand on the idea of fingerprint patterns recently proposed for 2-D graphical representations of proteomics maps by considering an abundance of protein spots as the third coordinate in a 3-D space.

NUMERICAL CHARACTERIZATION OF 3-D CURVE

One can arrive at a numerical characterization of a curve, chemical structure, or any object having a well-defined periphery and having a fixed geometry or being embedded in a space (or even 2-D plane), by constructing the so-called D/D matrix.[2,4−6] The D/D matrix combines information on distances between points that characterize the object considered and the information on adjacency. The element (i, j) of the D/D matrix corresponding to two points is obtained as a quotient of the Euclidean distance between the points divided by the distance measured along the path connecting the two points. In the case of molecular graphs each edge of a graph (or a chain) contributes one unit of length, thus the distance along the paths is simply given by the number of edges between the two points. Here instead of segments of unit length we have segments of variable length. In Table 1 we show the Euclidean distances as measured in 3-D space for the first 10 most abundant protein spots. From this information one can construct the D/D matrix by considering quotients of the corresponding distances through the space

* Corresponding author fax: (515)292-8629; e-mail: milan.randic@ki.si. Current address: 3225 Kingman Rd., Ames, IA 50311.

10.1021/ci0001684 CCC: $20.00
Published on Web 07/14/2001

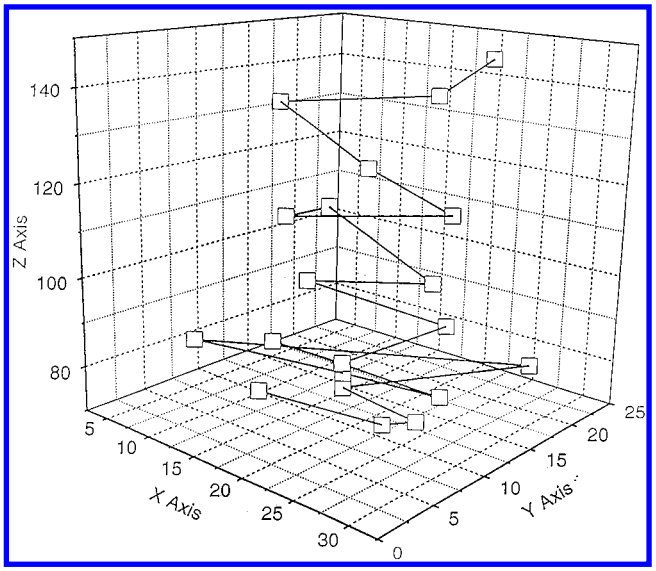

Figure 1. Zigzag line connecting the first 20 most intensive spots of Figure 1 in 3-D.

Table 1. Distances (in Arbitrary Units) between Ten Most Intense Spots for Simplified Proteomics Map of Figure 1

	1	2	3	4	5	6	7	8	9	10
1	0	15.67	17.59	22.20	28.68	33.63	34.28	39.36	48.24	51.53
2		0	17.45	17.46	25.20	31.52	35.07	34.73	47.55	49.70
3			0	13.83	23.59	22.10	24.75	32.65	38.56	44.94
4				0	10.39	14.19	18.11	19.62	30.15	33.58
5					0	13.18	15.48	10.68	24.21	24.54
6						0	6.84	15.23	17.00	24.05
7							0	17.42	14.45	22.39
8								0	18.67	15.91
9									0	12.84
10										0

and along the path. The part of the so constructed D/D matrix is shown in Table 2, again for the first 10 most abundant protein spots.

Observe that the table of distances (Table 1) does not contain information on the adjacency (defined by labels that follow the relative magnitude of abundance), while the table of cumulative path distances (which can be easily computed using diagonal entries of the distances shown in Table 1) incorporates an adjacency relationship but has no information on the relative distance of spots, except those adjacent in the 3-D zigzag path. Matrix D/D thus combines information on distances and adjacency in a single matrix. After constructing a D/D matrix we can consider various invariants of this matrix and use them to construct a vector, the components of which can be viewed as numerical descriptors of the original proteomics map. For example, by adding all the entries above the main diagonal in Table 2 we obtain an invariant analogous to the Wiener number,[7] W, which gives for the 10 × 10 portion of the proteomics map 27.11764. From the row sums of the D/D matrix one can compute an index analogous to the Balaban's J index,[8] which in this case (disregarding normalization) gives the value 1.641435. Large number of such mathematical descriptors can be constructed from a given matrix. Among these mathematical invariants the leading eigenvalue occupies perhaps a special place, in part because of its interpretation as a measure of the foldedness of a curve.[2,9] The leading eigenvalue is the largest positive eigenvalue of a matrix and is associated with an eigenvector which has no nodal surfaces (where change of signs for the coefficients of eigenvector occur). According to Frobenius-Perron theorem[10] it is bounded from above by the largest row sum and from the below by the smallest row sum. In the case of 10 × 10 matrix of Table 2 the upper bound of λ_1 is 6.14717 and the lower bound is 4.51283. The exact value for λ_1 is found to be 5.47266. All the eigenvalue calculations here reported were obtained using MATLAB,[11] a software, which is dedicated to the mathematical manipulations of matrices.

HIGHER ORDER ^{M}D/^{M}D MATRICES

In view of the complexity of proteomics maps clearly we need more than just a few invariants. Moreover, just as is the case with structure−property−activity studies[12] it appears desirable instead of an ad hoc collection of invariants that have little relations one to another, to construct invariants that are structurally related. In the case of QSAR (quantitative structure−activity relationship) such are, for instance, the paths of different length,[13] the connectivity indices,[14,15] and the valence connectivity indices,[16] the kappa shape indices,[17] the path/walk shape indices,[18] and the molecular "profiles"[19−21] derived for molecular skeletons of fixed geometry embedded in space. Molecular profiles can be derived from matrices ^{m}D/^{m}D, which are closely related to the D/D matrix, and thus appear suitable for our purpose here. The elements of ^{m}D/^{m}D matrix are obtained from the corresponding elements of the D/D matrix by raising it to the power m, m being an integer.

In Table 3 we have listed for $m = 1$ to $m = 20$ the leading eigenvalue λ_1 of ^{m}D/^{m}D matrices for the first 10, 15, and 20 most intensive spots of Figure 1. As m tends to infinity the ^{m}D/^{m}D matrix in the limit becomes a binary matrix which is an adjacency matrix for a chain of length n. The leading eigenvalues of the limiting matrices are given by $2\cos[\pi/(n + 1)]$, which for 10 × 10, 15 × 15, and 20 × 20 matrices are 1.918986, 1.961575, and 1.977662, respectively.

As we can see from the Table 3 by considering only the first 20 members of the map "profile" the convergence is rather slow. Thus if necessary the profile "length" can be extended considerably before reaching the limiting value, even when we confine characterization to 5−6 significant figures. In Table 4 we show few higher members of the map "profiles" illustrating the possibility of using vectors having 100, 200, and more components for characterization of individual maps or their sections. It appears from Table 4 that maps involving a large number of spots will allow construction of vectors with a considerably larger number of components.

DISCUSSION

In ref 1 there was a brief discussion of the sensitivity of map "profiles" to minor changes in the assignment of the relative abundance. Also in ref 1 there was an illustration of map fingerprint regions when drastic changes in the relative intensities of the protein spots occur. In both case it was shown that map "profiles" appear as a promising tool for numerical characterization of proteomics maps. A comparison between the original semiqualitative approach and here considered quantitative approach is of some interest. What have we gained by including the information on the relative

Table 2. Part of the D/D Matrix for Ten Most Intense Spots of the Proteomics Map of Figure 1

	1	2	3	4	5	6	7	8	9	10
1	0	1	.53090	.47279	.50015	.47691	.44316	.41526	.42523	.40801
2		0	1	.55824	.60469	.57469	.56844	.43898	.48631	.44931
3			0	1	.97411	.59082	.55939	.52947	.48009	.48239
4				0	1	.60195	.59561	.41017	.45339	.42336
5					0	1	.77343	.28529	.43150	.35594
6						0	1	.62782	.39593	.43135
7							0	1	.40039	.45757
8								0	1	.50490
9									0	1
10										0

Table 3. Leading Eigenvalue of the D/D Matrix and the Higher Order ^{m}D/^{m}D Matrices for Map of Figure 1 When 10, 15, and 20 Spots Are Considered

	10 × 10	15 × 15	20 × 20
1	5.47266	7.14803	8.07988
2	3.86618	4.42875	4.61066
3	3.11573	3.29808	3.34032
4	2.74716	2.80387	2.81817
5	2.55224	2.57055	2.57812
6	2.44077	2.44779	2.45492
7	2.37221	2.37577	2.39743
8	2.32706	2.32951	2.38652
9	2.29528	2.29748	2.38306
10	2.27146	2.27389	2.38108
11	2.25255	2.25585	2.37971
12	2.23679	2.24259	2.37865
13	2.22316	2.23526	2.37778
14	2.21099	2.23227	2.37701
15	2.19992	2.23076	2.37631
16	2.18966	2.22975	2.37567
17	2.18007	2.22895	2.37506
18	2.17102	2.22827	2.37548
19	2.16243	2.22765	2.37392
20	2.15424	2.22708	2.37337

Table 4. Illustration of the Convergence of the ^{m}D/^{m}D Matrices as the Exponent *m* Tends to Infinity

	10 × 10	15 × 15	20 × 20
25	2.11815	2.22448	2.37076
50	2.00931	2.21277	2.35873
100	1.94005	2.19092	2.33611
250	1.91937	2.13432	2.27597
500	1.9189864	2.06542	2.19719
1000	1.91898594723010	1.99722	2.09655
2500	1.91898594722899	1.96550	1.99814
5000		1.96192	1.97946
10 000		1.96167	1.97768354
25 000			1.977661652493
50 000			1.97766165245026
limit	1.91898594722899	1.96157056080646	1.97766165245026

abundance in construction of our 3-D graphical representation of proteomics map?

First we ought to point out that ordering of protein spots does not depend on the exact but only on the relative magnitudes of the abundance. Thus the partial order (discussed at some length in ref 1) based on 2-D and 3-D map representations are the same. The 3-D representation clearly has all the information of the 2-D representation of ref 1. Hence, everything that can be derived from the 2-D representation, which is just the projection of the 3-D representation on the (x, y) plane, can also be derived from the 3-D representation. Other advantages of the 3-D map representation that may be of interest are the following: For smaller portions of maps the 3-D representation apparently offers a larger number of vector components for the same number of significant digits. Also the limiting binary matrix derived form ^{m}D/^{m}D as *m* increases indefinitely in the case of 3-D representation will almost always lead to adjacency matrix for a chain of length n, while in the case of 2-D representation the presence of collinear fragments on adjacent sections of a zigzag path are more likely. When collinear spots occur, one will obtain the so-called line-adjacency matrix, which differs from the adjacency matrix by having additional nonzero binary entries.

Another flexibility that 3-D representation offers (absent in 2-D representation of proteomics maps) is optional adjustment of the scale used to measure spots position in the gel plane relative to the scale used for abundance. In this presentation we have used data from Table 1 of ref 1, but one could for example enhance or diminish the role of abundance of the spots by introducing scale factor k, which multiplies entries of the last column of Table 1. One advantage of such a factor is that when one compares different maps one can give more weight either to positions of the protein spots (by considering factors $k < 1$) or intensities of the spots (by considering factors $k > 1$). Alternatively, one can normalize the abundance by using unit $(I_{max} - I_{min})/n$, where I_{max} and I_{min} are the largest and the smallest abundance value in the map or in the fragment of the map considered, and n is the number of spots (points) considered.

ON THE SENSITIVITY OF THE APPROACH

To obtain insight into the sensitivity of the outlined approach we will closely examine the leading eigenvalues of the D/D matrix when gradually the abundance of one of the points is decreased. We selected a 5 × 5 matrices corresponding to the five most intensive points of Figure 1 and changed intensity for the middle protein spot having coordinates (12, 10, 136.7). We decreased the intensity of this spot in steps of 1.9 so that in five steps we reach the intensity 127.2, which equals that of the next spot at position (22, 9, 127.2). The paths connecting all these intermediate intensities are illustrated in Figure 2. In Table 5 we have listed the leading eigenvalues for computed map profiles based on the first 15 powers of ^{m}D/^{m}D matrices. As we see even these small changes in the abundance induce visible (but not dramatic) changes in the leading eigenvalues of the ^{m}D/^{m}D matrices. The limiting leading eigenvalue for all 5 × 5 matrices is $\sqrt{3}$.

From Table 5 we see that constructed map profiles are sufficiently sensitive to reflect even minor changes in intensities of the protein spots. For the particular case of 5 × 5 matrices (based on the five most intensive spots) when

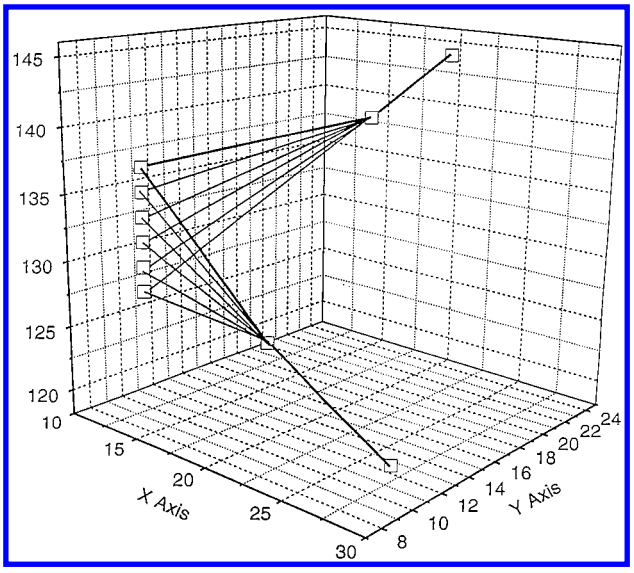

Figure 2. Zigzag line connecting the first five most intensive spots for the case of decreasing intensity for the middle point.

Table 5. Illustration of the Sensitivity of the Map Profiles on Gradual Changes of the Intensity of a Single Spot (Shown in the Top Row)

	136.7	134.8	132.9	131.0	129.1	127.2	127.2[a]
1	3.08515	3.09366	3.09367	3.08289	3.06067	3.02701	3.02205
2	2.63121	2.63488	2.62737	2.60588	2.56988	2.52098	2.54021
3	2.40120	2.39659	2.37982	2.34828	2.30220	2.24526	2.27891
4	2.27913	2.26651	2.24112	2.20091	2.14738	2.08629	2.12345
5	2.20967	2.19023	2.15751	2.11026	2.05146	1.98883	2.02358
6	2.16671	2.14162	2.10284	2.05004	1.98782	1.92547	1.95548
7	2.13771	2.10793	2.06415	2.00710	1.94293	1.88211	1.90686
8	2.11638	2.08266	2.03476	1.97453	1.90958	1.85112	1.87090
9	2.09944	2.06237	2.01107	1.94855	1.88372	1.82814	1.84358
10	2.08513	2.04517	1.99109	1.92701	1.86298	1.81058	1.82239
11	2.07246	2.02999	1.97364	1.90863	1.84592	1.79684	1.80569
12	2.06088	2.01622	1.95804	1.89260	1.83161	1.78589	1.79236
13	2.05006	2.00349	1.94386	1.87843	1.81944	1.77702	1.78163
14	2.03982	1.99158	1.93084	1.86576	1.80897	1.76977	1.77293
15	2.03005	1.98036	1.91880	1.85437	1.79991	1.76378	1.76583

[a] The last column shows the alternative map profile obtained by changing the path 1−2−3−4−5 to 1−2−4−3−5 in view that points 3 and 4 have the same intensity.

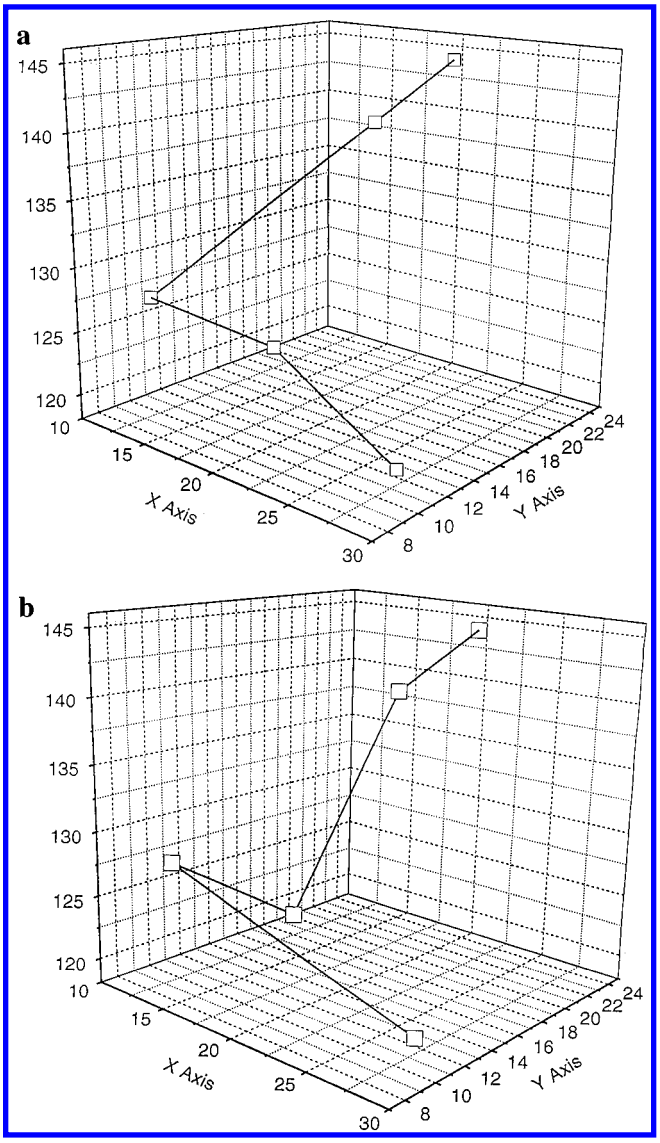

Figure 3. Two alternative zigzag lines showing connected path for the first five most intensive when two points have the same intensity.

m is about 25−30 the last two components of the leading eigenvalue vector agree to two decimal places, when *m* is about 100 the last two components of the leading eigenvalue vector agree to three decimal places, and when *m* is about 200 the last two components of the leading eigenvalue vector agree to four decimal places. Hence, if a comparison of different maps would require characterization of maps by vectors with many components the outlined approach could offer such characterization. On the other hand it is also important to recognize that this approach may characterize different maps with vectors having quite a few components.

A comparison of the last two columns in Table 5 is also instructive. Both columns correspond to a fictitious map in which two spots have the same intensity (127.2). As we see the two different paths (illustrated in Figure 3a,b) have distinctive map characterizations. Which one to choose? Such ambiguities can be resolved by adopting a convention that gives preference to one of the points based on their (x, y) coordinates. We can assign a smaller label to the spot belonging to a protein having a larger mass (large *y* value) or to a protein having a larger charge (large *x* value).

CONCEPTUAL QUESTIONS TO BE ADDRESSED

There are a number of conceptual questions that need clarifications if the proposed approach is to find wide application. Arguments presented in this work clearly hold for any 2-D matrix, but experimental content and conditions may vary considerably from case to case. For instance 2-D NMR maps are symmetrical and both x and y coordinates use units of the same dimension, hence the unit of path distance is constant, i.e., independent of a position in the x−y plane. This is not the case with proteomics maps where the nature of the analyzed experiment dictates the dependence of the unit of path distance upon the x−y position of the electrophoretic band (spot). Change of units along the x-axis (charge) or y-axis (mass) will have an impact on the numerical characterization and may influence a choice of invariants. Moreover, the unit along the z-axis (abundance) also can be changed (as already mentioned). A way to reduce the role that relative magnitudes of different units can have on the analysis is to follow Kowalski and Bender[22] who suggested rescaling of different units so that they cover the same range. In particular they suggested the intervals −1 <

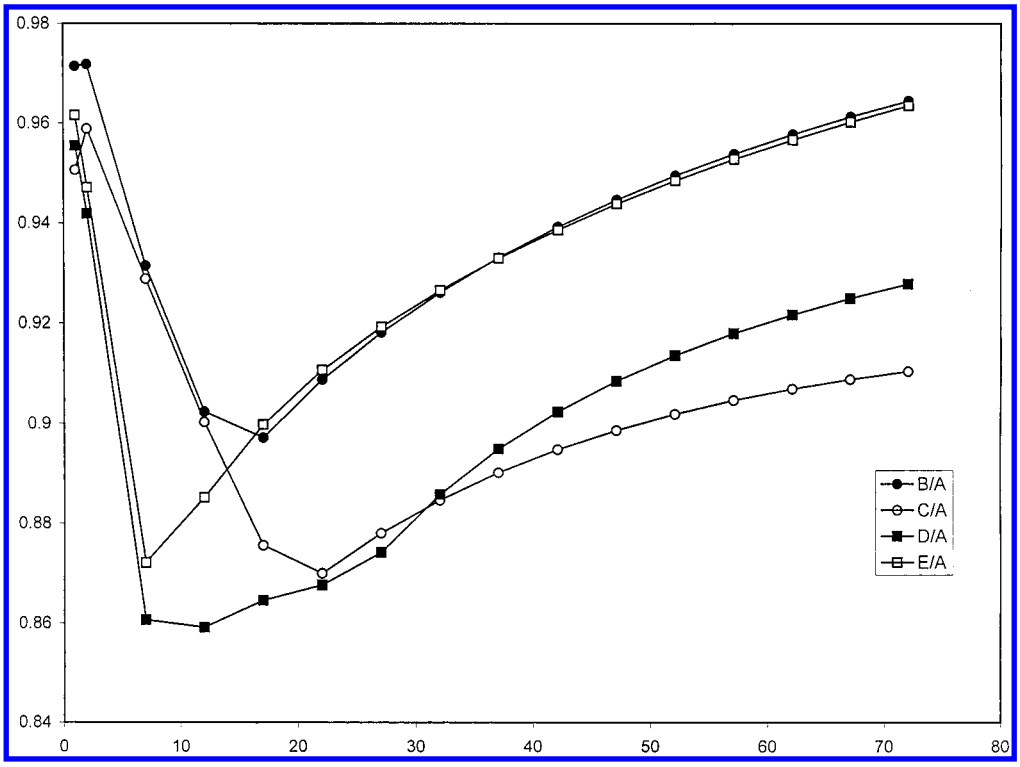

Figure 4. Dependence of the quotient of the leading eigenvalue of proteomics maps corresponding to four chemicals (peroxisome proliferators, B−E) with the control group (A), taken from ref 23.

x, y, z < +1. We have implemented this recommendation in a follow up paper on characterization of proteomics maps in which here the outlined approach is applied to characterization of proteomics maps in which changes were induced by chemical causing peroxisome proliferators.[23]

The goal of this work is to present formalism that (i) can quantitatively encode relevant information in the experimental data given as 2-D maps in a way that allows one to extend this information toward elucidation of the biological activities of molecules and (ii) enable compacting the relevant large scale information into a manageable database with possibility of fast searches and fast comparisons. From the outlined description of the method developed here it appears that the approach has that potential even if this was not demonstrated. However, besides showing that the approach is sensitive to changes in the primary data in order to have a wide applicability the approach ought to be efficient. This would be demonstrated by showing that new information can be extracted from the processed primary information. That indeed this is the case has been described in the paper on peroxisome proliferators[23] in which the plots of the quotient of the leading eigenvalues of ${}^{m}\mathrm{D}/{}^{m}\mathrm{D}$ matrices against m were considered for the control experimental data and the four chemicals inducing peroxisome proliferators. In Figure 4 we illustrate the plot for the four cases discussed in ref 23, which show different initial slopes, different minima, and different limiting behavior for the four chemicals considered. The magnitude of the slopes, the positions of minima, and the rate of convergence represent novel conceptual parameters for discussing and describing the role of different chemodescriptors.

Finally, we should address the case when in the course of analysis of a series of inputs a new spot will appear in the primary data. This is actually a very important situation in the type of experiments that we mentioned as the primary target of our method. By the very nature of the presented formalism such a new entry may have a significant impact on the results. When new spots are detected for proteomics maps of a series one should view the new spots as spots that have abundance zero in the control map, and hence such spots should be added at the end of the zigzag path.

CONCLUDING REMARKS

The outlined 3-D representation of proteomics maps, together with 2-D representations and accompanying map "profiles", graphical "fingerprints", and partial order diagrams offer tools for quantitative analysis of proteomics maps that were hitherto mainly visually inspected. We have outlined one particular set of map invariants, but if a parallel is to be taking with mathematical characterization of molecules, one can expect expansion of mathematical characterization of proteomics maps and anticipate construction of additional map invariants. Here introduced map profiles based on 3-D representation of a map do incorporate all available information that can be read from proteomics map. However, one has to be aware of the fact that any characterization of a system by invariants is accompanied with some loss of some initial information. This is the case whether one considers molecular structure, DNA primary sequence, or proteomics map. The critical step when considering a different set of invariants may be the selection of invariants at one's disposal. Just as in QSAR different invariants will to a different degree capture information on the most critical aspects of the system considered. It is not easy to know in advance which invariant is a better descriptor for different structural features of objects considered. Comparison of characterizations based on different descriptors will eventually allow one to recognize merits of individual

descriptors. We have offered here a set of invariants for characterization of proteomics maps, that have yet to be found useful, which we cannot guarantee. But if we are to draw a parallel with the demonstrated use of molecular profiles in structure−property studies we can guarantee that the proposed approach has potential and may allow novel mathematical characterization of data that has hitherto been given only as a list of 2-D position and abundance.

REFERENCES AND NOTES

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

(2) Randić, M.; Kleiner, A. F.; DeAlba, L. M. Distance/distance Matrices. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 277−2866.

(3) Trinajstic, N. *Chemical Graph Theory;* CRC Press: Boca Raton, Fl, 1992.

(4) Randić, M.; Razinger, M. On characterization of three-dimensional molecular structure. In *From Chemical Topology to Three-Dimensional Geometry;* Balaban, A. T., Ed.; Plenum Press: New York, 1996; pp 159−236.

(5) Bytautas, L.; Klein, D. J.; Randić, M.; Pisanski, T. Foldedness in linear polymers: A difference between graphical and Euclidean distances. *DIMACS Ser. Discrete Mathematics Theor. Comput. Sci.* **2000**, *51*, 39−61.

(6) Randić, M.; Vračko, M.; Novič, M.; Basak, S. C. On ordering of folded structures. *MATCH* **2000**, *42*, 181−231

(7) Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.

(8) Balaban, A. T. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **1982**, *89*, 399−404.

(9) Randić, M.; Krilov, G. On characterization of the folding of proteins. *Int. J. Quantum Chem.* **1999**, *75*, 1017−1026.

(10) Gantmacher, F. *Theory of matrices*; Chelsea Publishers: New York, 1959; Vol. II, Chapter 13.

(11) MATLAB (version edu-MATLAB 87-91); The MathWorks Inc.: Natick, MA.

(12) Randić, M. *Topological Indices.* In *Encyclopedia of Computational Chemistry;* Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, III., H. F., Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998; pp 3018−3023.

(13) Platt, J. R. Influence of neighbor bonds on additive bond properties in paraffins. *J. Chem. Phys.* **1947**, *15*, 419.

(14) Randić, M. On the characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609−6615.

(15) Kier, L. B.; Murray, W. J.; Randić, M.; Hall, L. H. Molecular connectivity V. Connectivity series concept applied to density. *J. Pharm. Sci.* **1976**, *65*, 1226−1230.

(16) Kier, L. B., Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research;* Academic press: New York, 1976.

(17) Kier, L. B. A shape index from molecular graphs. *Quant. Struct.-Act. Relat.* **1985**, *4*, 109−116.

(18) Randić, M. Novel shape descriptors for molecular graphs. *J. Chem. Inf. Comput. Sci.* (submitted).

(19) Randić, M.; Razinger, M. On characterization of molecular shapes. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 594−606.

(20) Randić, M. Molecular profiles − Novel geometry-dependent molecular descriptors. *New J. Chem.* **1995**, *19*, 781−791.

(21) Randić, M.; Krilov, G. On characterization of molecular surfaces. *Int. J. Quantum Chem.* **1997**, *65*, 1065−1076.

(22) Kowlaski, B. R.; Bender, C. F. A. A powerful approach to interpreting chemical data. *J. Am. Chem. Soc.* **1972**, *94*, 5632−5639.

(23) Randić, M.; Witzmann, F.; Vračko, M.; Basak, S. C. On characterization of proteomics maps and chemically induced changes in proteomics using matrix invariants: Application to peroxisome proliferators. *Med. Chem. Res.* (in press).

CI0001684