Journal of Molecular Graphics and Modelling 22 (2003) 1–9

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Novel map descriptors for characterization of toxic effects in proteomics maps

Željko Bajzer [a,*], Milan Randić [b], Dejan Plavšić [c], Subhash C. Basak [d]

[a] *Mayo Clinic Rochester, Rochester, MN, USA*
[b] *National Institute of Chemistry, Ljubljana, Slovenia*
[c] *Institute Rudjer Bošković, Zagreb, Croatia*
[d] *NRRI, University of Minnesota at Duluth, Duluth, MN, USA*

Received 12 August 2002; accepted 22 October 2002

Abstract

We consider a novel numerical characterization of proteomics maps based on the construction of a graph obtained by connecting all protein spots in a proteomics map that are at distance equal to, or smaller than, a critical distance D_c. We refer to the so constructed graph as a *cluster graph* and we calculate four associated characteristic matrices, previously considered in the literature: (1) the Euclidean-distance matrix *ED*; (2) the neighborhood-distance matrix *ND*; (3) the path-distance matrix based on the shortest paths between connected spots *PD*; and (4) the quotient matrix *Q*, the elements of which are given as the quotient of the corresponding elements of *ED* and *ND* matrices. Numerical descriptors for proteomics maps include in particular the leading eigenvalue of the *Q* matrix and the family of associated "higher order" matrices defined as powers of *Q*. These map descriptors show considerable sensitivity to perturbations of proteomics maps by toxicants.

Keywords: Graph-theoretical descriptors; Matrix invariants; Proteomics map; Quotient matrix; Peroxisome proliferators

1. Introduction

In this contribution, we consider development of novel numerical parameters for the characterization of proteomics map as a whole. Introduced parameters can be viewed as biodescriptors in an analogy to various chemodescriptors, which are represented by a set of molecular descriptors derived from molecular skeleton by various methods of discrete mathematics (combinatorics and graph theory in particular). Characteristic parameters are in essence map invariants, which can be also constructed for other maps as well. Map invariants allow quantitative comparisons of maps which are often compared only visually, that is qualitatively.

Typical protemics map is the result of horizontal separation of proteins by electrophoresis and vertical separation by chromatography, so that proteins on the left have greater charge and proteins at the top have greater mass. The map is then formatted into so-called "bubble diagram" in which a point (two-dimensional gell spot) representing charge and mass of a protein is the center of a circle with radius corresponding to the abundance of that protein (see Fig. 1 in [1]). Data from a bubble diagram represent an experimental *output* that will be viewed by us as *input* data to be mathematically processed. Alternatively, bubbles can be represented by a point in three-dimensional space where *x*, *y*, *z* coordinates are proportional to charge, mass, and abundance, respectively. Kowalski and Bender [2] recommended that distinct experimentally measured quantities (such as charge, mass and abundance) should be re-scaled to the interval $[-1, 1]$, so that they are given the same weight. We followed this recommendation, however, rather than scaling data to interval $[-1, 1]$, we selected a normalization that for the average charge, the average mass and the average abundance gives the value of 1.

To obtain a numerical characterization of a map, we follow the strategy outlined in our earlier work [3–9], which implies the following steps:

(1) represent a proteomics map by some mathematical object of *fixed geometry*, such as graph, or a set of lines;
(2) for the selected mathematical object, construct its numerical representation in the form of a *matrix* or set of matrices;
(3) from obtained matrices extract a set of *invariants*, which will serve as numeric map descriptors.

* Corresponding author. Tel.: +1-507-284-8584; fax: +1-507-284-9420.
E-mail address: bajzer@mayo.edu (Ž. Bajzer).

1093-3263/02/$ – see front matter
PII: S1093-3263(02)00186-9

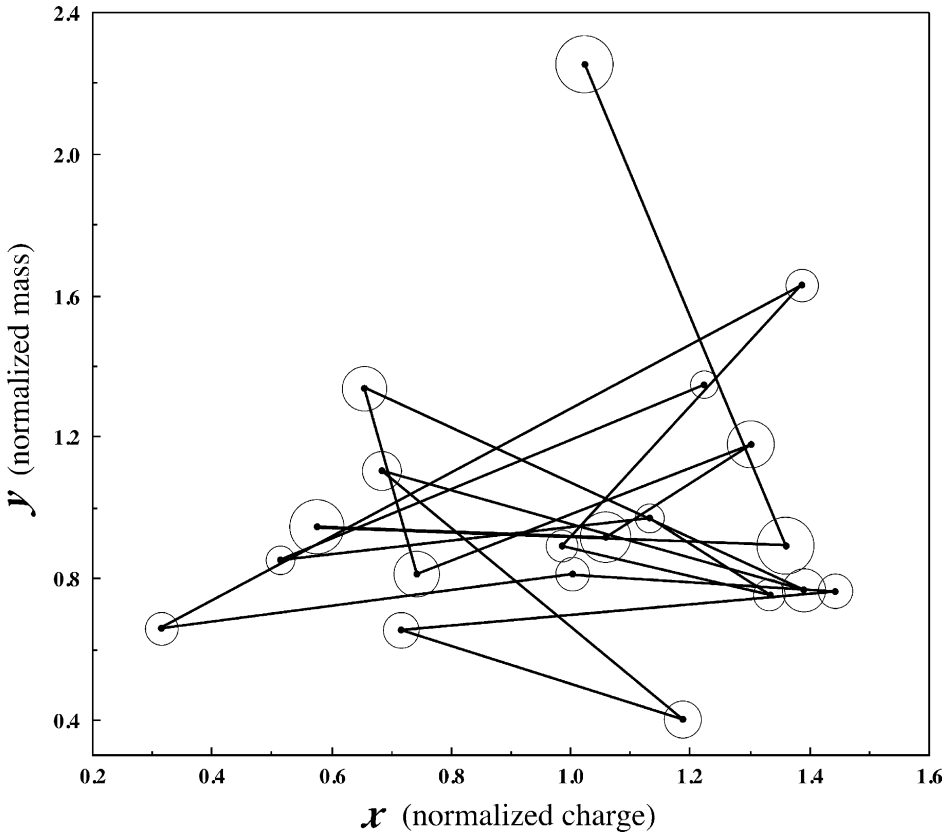

Fig. 1. The "bubble diagram" with zigzag line for simplified proteomics map given by 20 protein spots with coordinates shown in Table 1. Spot coordinates are graph vertices and lines are its edges.

Mathematical invariants of a system (a structure, a sequence, a map) are quantities that are independent of labels used to identify its components; hence mathematical invariants belong to a system as a whole. They are analogous to physico-chemical *properties* of a system, except that they are mathematical by nature, in contrast to properties, which are obtained by various physical or chemical measurements. Map invariants, being of mathematical origin, have advantages and disadvantages similar to advantages and disadvantages of the so-called topological indices [10–12]. These indices are used in QSAR (the quantitative structure–activity relationship) [13,14], and in other aspects of the structure–property–activity studies, including those using virtual combinatorial libraries, and structure-similarity/dissimilarity studies among molecules [15,16]. Furthermore, these topological indices are used for characterization of strings of symbols (such as DNA primary sequences) [17–23].

2. Map representation

In the previous work in which construction of map invariants was initiated [1,3–8], map descriptors were constructed from two mathematical objects: (a) graph consisting of a *zigzag line*, obtained by ordering protein spots of a proteomics map by abundance [1,3,4,7] and (b) *embedded graph* obtained by partial ordering of proteins spots with respect to charge and mass [5,6]. To illustrate these mathematical objects we considered 20 spots corresponding to most abundant proteins in rat liver cells [1], which can be viewed as map of a local proteomics region. In Table 1, related normalized *x*, *y*, *z* coordinates are listed, and Fig. 1 shows zigzag line which connects spots in order from the highest to the lowest protein abundance. In Fig. 2, we illustrate embedded graph depicting partial ordering of spots, i.e. in this diagram only those protein spots are connected that either dominate or are dominated in both the mass and the charge by the neighboring spots. Observe that all the lines of the embedded graph shown in Fig. 2 have positive slope, which facilitate construction of the graph.

2.1. Representation of a map by matrices

Having selected mathematical object of fixed geometry associated with a map, in the next step one proceeds to construct various matrices, which incorporate the information on the distances and the adjacency between spots. In particular, we will consider the following matrices: the Euclidean-distance matrix *ED*; the neighborhood-distance matrix *ND*, the path-distance matrix *PD*, and the quotient

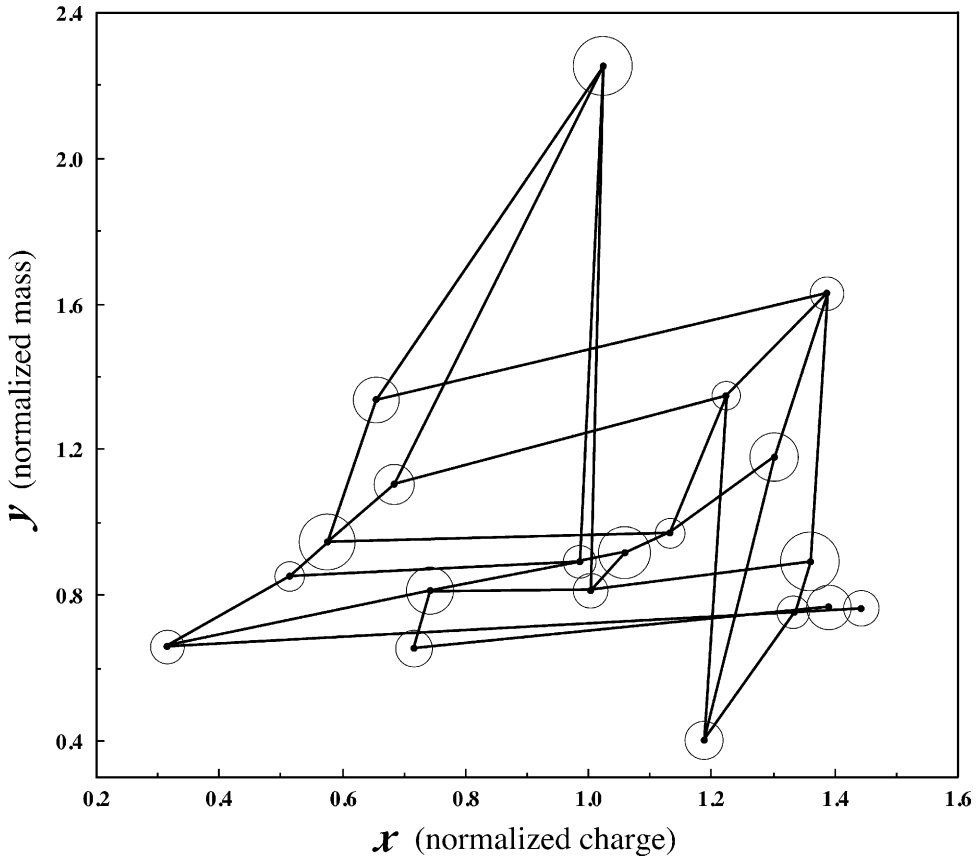

Fig. 2. The embedded graph depicting the partial ordering of the 20 protein spots of Table 1, overlaid on corresponding "bubble diagram".

Table 1
Normalized charge (x) and mass (y) two-dimensional gel coordinates for 20 protein spots ordered by their abundance for proteomic map of control rat liver cells [5]

Protein spot	x	y	z
1	1.023	2.255	1.44457
2	1.35869	0.89414	1.43730
3	0.57360	0.94956	1.36748
4	1.05728	0.91908	1.27283
5	1.30118	1.18358	1.18663
6	0.74027	0.81686	1.15009
7	0.65214	1.33835	1.12329
8	1.38979	0.76986	1.08958
9	0.68136	1.10640	0.98292
10	1.18713	0.40492	0.93666
11	0.71416	0.65814	0.90066
12	1.44134	0.76471	0.86790
13	1.00214	0.81449	0.84901
14	0.31115	0.66279	0.82549
15	1.38601	1.63263	0.82022
16	0.98485	0.89335	0.80070
17	1.33369	0.75759	0.79902
18	1.13092	0.97192	0.72841
19	0.51047	0.85526	0.72223
20	1.22070	1.35161	0.69500

matrix Q, all to be briefly outlined in the following sections.

2.1.1. The Euclidean-distance matrix ED

The matrix elements (i, j) are given by the Euclidean-distance in the three-dimensional space in which the (x, y) coordinates give the location of the protein spots on the two-dimensional gel, while the z coordinate gives the abundance of the corresponding spots

$$[ED]_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

2.1.2. Neighborhood-distance matrix ND

The matrix elements (i, j) of ND are different from 0 only if protein spots i and j are within a certain neighborhood, i.e. if their Euclidian-distance is smaller or equal to some critical distance D_c. The magnitude of non-zero elements are given by the Euclidean-distance in three-dimensional space between the spots i and j, based on their (x, y, z) coordinates (such as those of Table 1)

$$[ND]_{ij} = \begin{cases} [ED]_{ij} & \text{if } [ED]_{ij} \leq D_c \\ 0 & \text{if } [ED]_{ij} > D_c \text{ or if spots } i \text{ and } j \text{ are not connected} \end{cases}$$

2.1.3. Path-distance matrix PD

The matrix element (i, j) are given by the sum of the Euclidean-distances (in three-dimensional space) along the shortest path through connected spots between (i, j) pair

$$[PD]_{ij} = \min_{P_{ijk}}([ED]_{ik_1} + [ED]_{k_1k_2} + [ED]_{k_2k_3} + \cdots + [ED]_{k_mj}), \quad [PD]_{ii} = 0$$

where $P_{ijk} = \{i, k_1, k_2, \ldots, k_m, j\}$ are the sets (indexed by k), representing connected spots that define contiguous paths from spot i to spot j. One can find the length of the shortest path by using Dijkstra's algorithm [24].

2.1.4. Quotient matrix Q

The matrix elements (i, j) of the quotient matrix Q are given by the quotient of the corresponding elements of the Euclidean-distance matrix *ED* and the path-distance matrix *PD*. The matrix elements for all adjacent pairs of spots are necessarily equal to 1 and for spots that are not adjacent, matrix elements are necessarily <1, except if it happens that spots are strictly collinear (which is not likely to occur)

$$[Q]_{ij} = \frac{[ED]_{ij}}{[PD]_{ij}}; \quad i \neq j, \ [Q]_{ii} = 0$$

In the case of embedded molecular graphs, the quotient matrix has been known as the *D*/*D* matrix [25]. Its elements have been defined as the quotient of the Euclidean-distance between a pair of vertices (atoms) divided by the number of bonds (edges) between the atoms, which is the graph-theoretical distance between vertices [26].

It is important to observe that matrix elements of the Q matrix combine the information on the ad*jacency* as well as on the *distances* between the selected spots into a single matrix. By virtue of the fact that its elements are <1 or at most equal to 1, Q matrix does not require additional normalization when matrices derived from it by Hadammard multiplication are considered (to be discussed later).

2.2. Higher order map matrices

Although a *single* matrix, such as the quotient matrix Q, can be a source of numerous matrix and structure invariants, as has been illustrated by various topological indices [9], the complexity of proteomics maps may be better characterized by additional matrices. One way to obtain additional matrices is to consider a *family* of structurally related matrices that has been outlined in several publications related to matrices designed for molecular graphs [27–31]. Hence, as shown in the literature, for a given matrix M, one can always construct two families of structurally related matrices by either using the standard matrix multiplication of linear algebra, or by using the Hadammard matrix multiplication for raising matrix elements to ever increasing powers. The Hadammard multiplication of a matrix M leads to matrices ${}^nM$, the elements of which are defined as $(m_{ij})^n$, where m_{ij} is the matrix element of M.

Because the matrix element of the quotient matrix Q are <1, or at most equal to 1, by raising Q to ever increasing powers, in the limit as $n \to \infty$, all the matrix elements <1 will become 0, while those equal to 1 will remain constant. As a result we obtain a binary matrix, which is the ordinary adjacency matrix A of the embedded graph.

3. Cluster approach

In this contribution, we will introduce an *embedded cluster graph* as the mathematical object on which we will base the construction of map matrices and extraction of map invariants. As before, to clearly illustrate the approach, we selected the twenty most intensive spots of a proteomics map [1] shown in Table 1 and in Fig. 3A. The embedded cluster graph, presented in Fig. 4, is obtained by making connections between the protein spots that are separated by Euclidian three-dimensional distances shorter or equal to a given critical distance D_c. In comparison with Fig. 1 (the embedded zigzag line) and Fig. 2 (the embedded partial ordering graph), the cluster graph of Fig. 4 is rather *dense*, i.e. it has considerably larger number of lines. Hence, the cluster graph involves much more information about the map, than the zigzag line, or the graph associated with the partial ordering of spots. It has yet to be seen if such additional information is useful or to some extent redundant. However, the cluster approach in contrast to the approaches based on the zigzag line or partial ordering, has the flexibility in the critical distance which can be viewed as a variable parameter. Thus, one can vary the density of the graph, which can be decreased or increased at will.

In this study, we selected as the critical distance $D_c = \max_{i=1,\ldots,20}(\min_{j=1,\ldots,20}[ED]_{ij})$, i.e. the largest distance that results in a graph that is necessarily connected. This is the distance between spot nos. 1 and 15 in Fig. 4, which in fact represents the (x, y) projection of the three-dimensional cluster graph that also includes z coordinate (information on abundance). Hence, the projection of the critical distance, the lines 1–15 in Fig. 4, is not necessarily the shortest distance in (x, y) plane, but it is the shortest distance in three-dimensional space.

In Table 2, we illustrate the structure of the neighborhood-distance matrix *ND* for the graph of Fig. 4. We indicate the sites for non-zero matrix entries by a bold dot, while the 0 matrix elements are indicated by zeros. The structure of the quotient matrix Q is related to that of the neighborhood-distance matrix; wherever in *ND* was 0, now we have matrix elements <1, except for the diagonal elements that remain 0 in both matrices. On the other hand, all the non-zero entries of the *ND* matrix (that correspond to adjacent vertices of Fig. 4) are now equal to 1.

3.1. Map invariants

We will examine the leading eigenvalue of the quotient matrix Q as one of the invariants that we use as

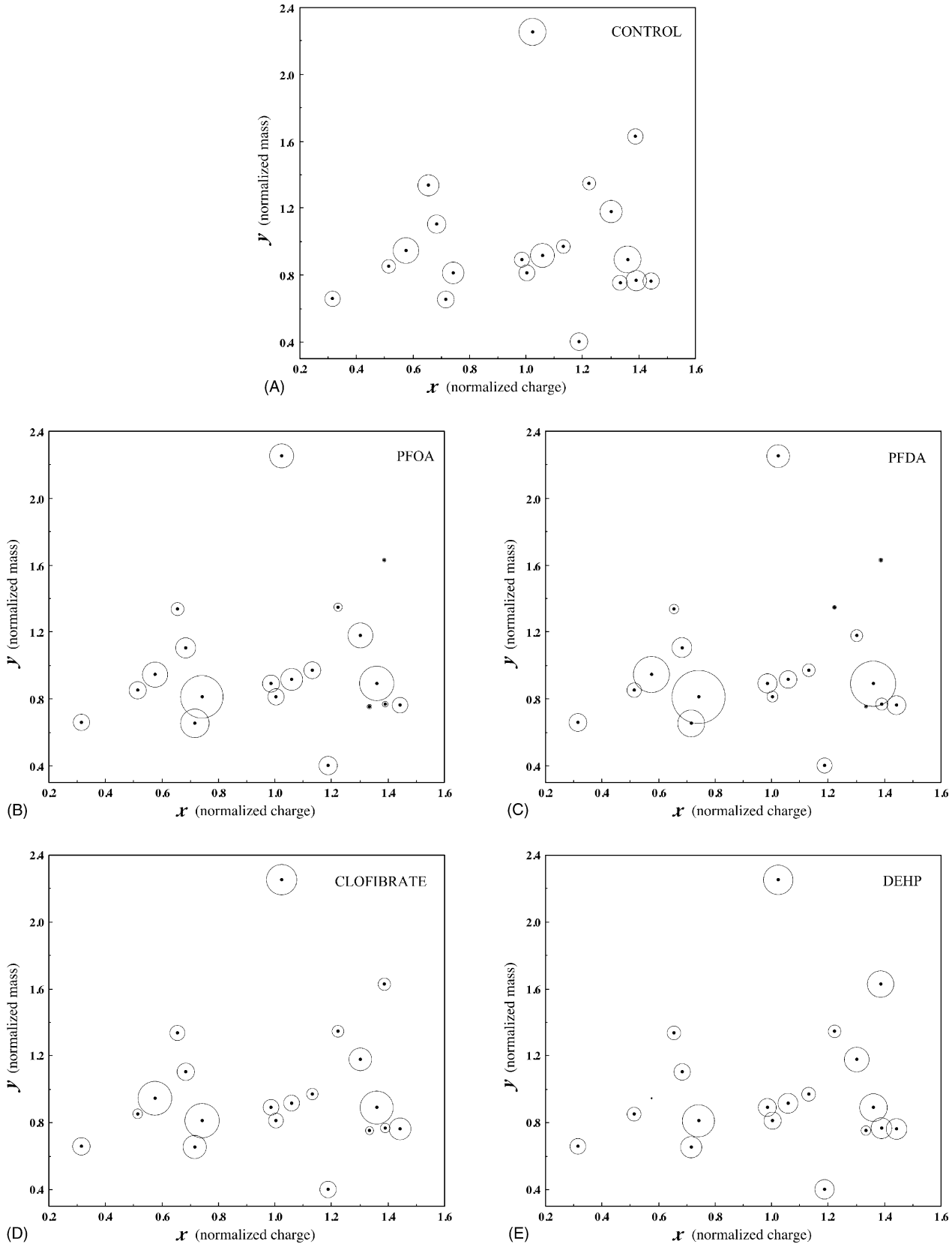

Fig. 3. The bubble maps obtained from liver cells of rats [5]; panel (A): control cells (corresponding to Table 1); panels (B–E): cells exposed to peroxisome proliferators: (B) PFOA (perflourooctanic acid); (C) PFDA (perflourodecanic acid); (D) clofibrate; and (E) diethylhexyl phthalate (DEHP).

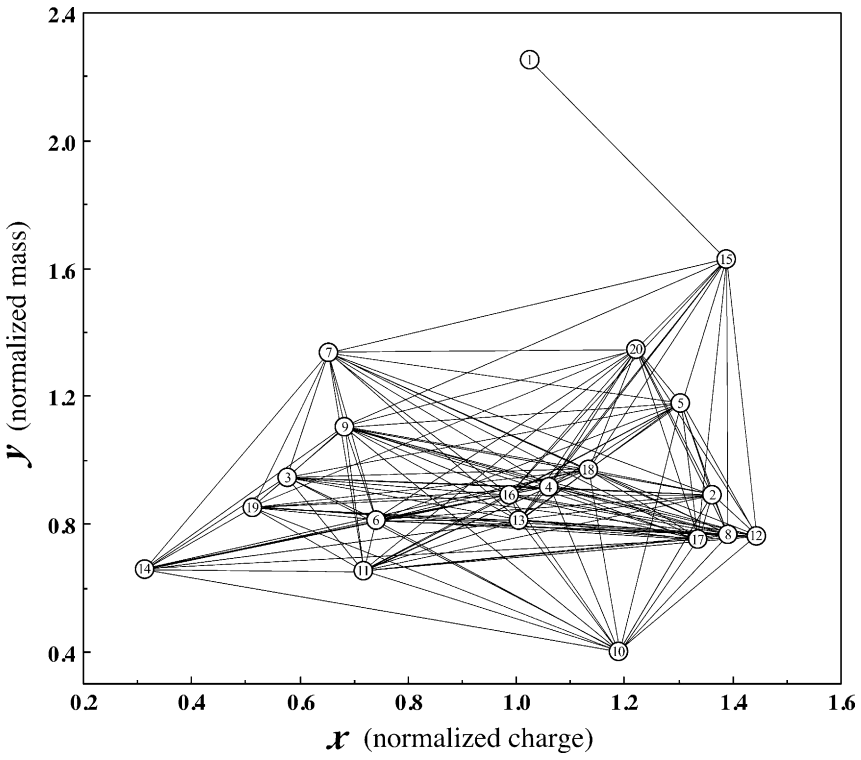

Fig. 4. The cluster graph obtained by making connections between the protein spots, which are separated by the critical distance D_c, or shorter distances.

Table 2
The structure of the neighborhood-distance matrix *ND* for the cluster graph of Fig. 4 (the non-zero matrix entries are shown as bold dots)

$[ND] =$

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	•	0	0	0	0	0	1
2		0	•	•	•	•	•	•	•	•	•	•	•	0	0	•	•	•	0	•	2
3			0	•	•	•	•	•	•	•	•	0	•	•	0	•	0	•	•	0	3
4				0	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	4
5					0	•	•	•	•	•	•	•	•	0	•	•	•	•	0	•	5
6						0	•	•	•	•	•	•	•	•	0	•	•	•	•	•	6
7							0	•	•	0	•	0	•	•	•	•	•	•	•	•	7
8								0	•	•	•	•	•	0	•	•	•	•	0	•	8
9									0	•	•	•	•	•	•	•	•	•	•	•	9
10										0	•	•	•	•	0	•	•	•	•	0	10
11											0	•	•	•	0	•	•	•	•	•	11
12												0	•	0	•	•	•	•	•	•	12
13													0	•	•	•	•	•	•	•	13
14														0	0	•	•	0	•	0	14
15															0	•	•	•	0	•	15
16																0	•	•	•	•	16
17																	0	•	•	•	17
18																		0	•	•	18
19																			0	•	19
20																				0	20

Symmetrical

•: Number <1.

Table 3
The leading eigenvalue for ${}^{n}[Q]$ matrices for the first 10 powers and for selected higher powers

Power	The leading eigenvalue	λ_n/λ_∞
1	18.5642	1.16544
2	18.2719	1.14709
3	18.0654	1.13413
4	17.9116	1.12447
5	17.7916	1.11694
6	17.6939	1.11081
7	17.6115	1.10563
8	17.5401	1.10115
9	17.4769	1.09718
10	17.4199	1.09361
50	16.5160558	1.036862822
100	16.2146863	1.017943122
500	15.9412476	1.000776891
1000	15.9300323	1.000072804
3000	15.9288765	1.000000245
4000	15.9288762	1.000000226
∞	15.9288762	1.000000000

map descriptor. For the leading eigenvalue, we obtain $\lambda_1 = 18.5642$. As additional map invariants, one could consider the leading eigenvalues of the two types of the "higher order" Q matrices: (1) ${}^{n}[Q]$ matrices are obtained by use of the Hadammard product of matrix by itself, i.e. (i, j) element of ${}^{n}[Q]$ is given by $([Q]_{ij})^n$, $n = 1, 2, 3, \ldots$; (2) $[Q]_n$ matrices are obtained by use of standard matrix multiplication of Q matrix by itself.

In the left column of Table 3, we have listed the leading eigenvalue for ${}^{n}[Q]$ matrices for the first 10 powers, followed by list of the leading eigenvalues for selected large values of n. As we see, the rate of convergence of the leading eigenvalue to its limit is rather slow. This limiting eigenvalue corresponds to the leading eigenvalue of the adjacency matrix related to graph of Fig. 4 when viewed as a *graph*, rather than an *embedded graph* of fixed geometry. The adjacency matrix is structurally equal to neighborhood-distance matrix of Table 2 with dots replaced by ones.

The magnitude of the leading value, which is according to Gerschgorn circle theorem [32] bounded by the maximal and the minimal matrix row sums, will depend on the size of a matrix. In order to make comparisons between matrices of different size easier, we include in the right column of Table 3 the quotient λ_k/λ_∞, where λ_∞ is the limiting leading eigenvalue.

Before giving an illustration, we would like to add that there are other ways of comparison of the shortest path graphical metric and the Euclidean metric [33,34]. As outlined by Bytautas et al. [34], the comparison between two metrics on a common set of vertices is conveniently expressible in terms of ratios of their value at different arguments, which leads to novel distance-ratio matrices. Following this procedure, using the leading eigenvalue and eigenvector, one can extract novel invariants closely related to the average matrix element.

3.2. Illustration

To assess the sensitivity of the proposed map descriptors, we have re-examined data considered in [1]. These data consist of protemics maps for rat liver cells of control group (Fig. 3A, Table 1) and four proteomics maps shown in Fig. 3B–E, obtained from liver cells of rats exposed to four peroxisome proliferators labeled as PFOA, PFDA (for perflourooctanic acid and perflourodecanic acid, respectively),

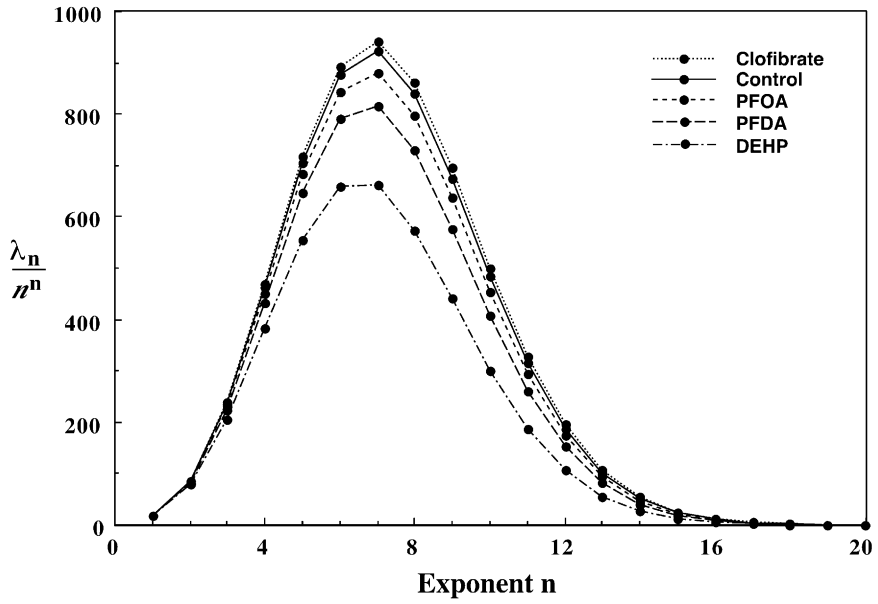

Fig. 5. The dependence of the normalized leading eigenvalues of the $[Q]^n$ matrices (obtained by standard matrix multiplication) on the exponent n, for the control group, and animals exposed to peroxisome proliferators.

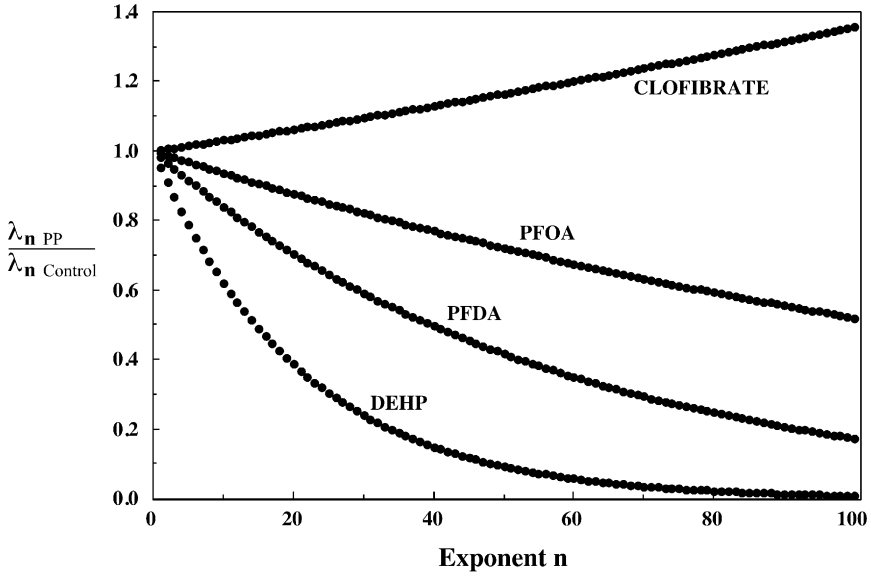

Fig. 6. The quotients $\lambda_{n\,\mathrm{PP}}/\lambda_{n\,\mathrm{CONTROL}}$ for the four peroxisome proliferators as the exponent n increases $n = 1$–100.

clofibrate and DEHP (for diethylhexyl phthalate). By visual inspection of Fig. 3, one finds it difficult to characterize these maps and to assess which of these maps are more similar or dissimilar. Better insight is given by leading eigenvalues λ_n of corresponding $[Q]^n$ matrices, normalized with factor $1/n^n$ to reduce their fast growth with n, and plotted against exponent n (Fig. 5). All of the five curves of Fig. 5 have the same general shape, but display different heights and different half-widths. By inspection of Fig. 5, we see that the eigenvalue curves for the control group and clofibrate are very close to each other, suggesting that the corresponding proteomics maps show more similarities than they both do with the other proteomics maps. Once this is seen from Fig. 5, it can be to some extent confirmed from the original maps of Fig. 3. Further, from Fig. 5 we see that the PFOA and PFDA eigenvalue curves are more similar to each other than to the DEHP curve, which departs most from others. In contrast to qualitative visual inspection of Fig. 3, eigenvalue curves can be characterized with their peak height and their width at half maximum, which can provide some objective quantitative measure of the degree of similarity.

In order to even better discriminate the differences in the overall dependence of the leading eigenvalues on the exponent n in Fig. 6, we plot the quotients $\lambda_{n\,\mathrm{PP}}/\lambda_{n\,\mathrm{CONTROL}}$ for the four peroxisome proliferators (PP) as a function of n. For the leading eigenvalue of Q matrix $\lambda_{1\mathrm{PP}}/\lambda_{1\mathrm{CONTROL}}$ is essentially one for all PP. However, Fig. 6 shows clearly that a different initial slope and different values for $n = 100$ characterize considered PP. Both Figs. 5 and 6 suggest that novel invariants based on the cluster graph and leading eigenvalues of powers of quotient matrix (via matrix multiplication) are quite sensitive to abundance variations in proteomics maps and thus appear as promising biodescriptors for these maps. When the Hadamard multiplication is used for the powers of Q, we found that the sensitivity to variations is less emphasized.

4. Concluding remarks

Just as has been the case with QSAR, the quantitative structure-activity relationship, where hundreds of molecular descriptors continue to be considered in multivariate regression analysis and PCA (principal component analysis), we anticipate design and use of numerous proteomics map descriptors. The proposed descriptors here, namely the leading eigenvalues of the quotient matrix based on the embedded cluster graph, complement the already reported similar map invariants based on the zigzag line and the partial ordering of spots. However, the present approach has an interesting advantage over the existing approaches: it is flexible in selecting the critical cluster distance D_c, which allows the construction of conceptually similar additional map descriptors corresponding to different critical distances D_c. Depending on the critical distance, map descriptors may capture more or less of the information content of a proteomic map.

Acknowledgements

The authors would like to thank Professor D.J. Klein (Galveston, TX) for numerous useful comments related to this manuscript.

References

[1] M. Randić, F. Witzmann, M. Vračko, S.C. Basak, On characterization of proteomics maps and chemically induced changes in proteomics

using matrix invariants: application to peroxisome proliferators, Med. Chem. Res. 10 (2001) 456–479.
[2] B.R. Kowalski, C.F. Bender, A powerful approach to interpreting chemical data, J. Am. Chem. Soc. 94 (1972) 5632–5639.
[3] M. Randić, On graphical and numerical characterization of proteomics maps, J. Chem. Inf. Comput. Sci. 41 (2001) 1330–1338.
[4] M. Randić, J. Zupan, M. Novič, On 3-D graphical representation of proteomics maps and their numerical characterization, J. Chem. Inf. Comput. Sci 41 (2001) 1339–1344.
[5] M. Randić, A graph-theoretical characterization of proteomics maps, Int. J. Quantum Chem. 90 (2002) 848–858.
[6] M. Randić, S.C. Basak, A comparative study of proteomics maps using graph-theoretical biodescriptors, J. Chem. Inf. Comput. Sci. 42 (2002) 983–992.
[7] M. Randić, M. Novič, M. Vračko, On characterization of dose variations of 2-D proteomics maps by matrix invariants, J. Proteome Res. 1 (2002) 217–226.
[8] M. Randić, S.C. Basak, Canonical labeling for protein spots and proteomics maps, J. Chem. Inf. Comput. Sci., in press.
[9] M. Randić, Topological indices, in: P.V.R. Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner (Eds.), Encyclopedia of Computational Chemistry, Wiley, Chichester, 1998, pp. 3018–3032.
[10] A.T. Balaban, A personal view about topological indices for QSAR/QSPR, in: M.V. Diudea (Ed.), QSPR/QSAR Studies by Molecular Descriptors, Nova Science Publication Inc., Huntington, NY 2001, Chapter 1, pp. 1–30.
[11] M. Randić, The connectivity index 25 years after, J. Mol. Graph. Modell. 20 (2001) 19–35.
[12] A.T. Balaban, Historical developments of topological indices, in: J. Devillers, A.T. Balaban, (Eds.), Topological Indices and Related Descriptors in QSAR and QSPR, Gordon and Breach, Amsterdam, The Netherlands 1999, pp. 403–453.
[13] S.C. Basak, G.D. Grunwald, G.J. Niemi, Use of graph-theoretic geometrical molecular descriptors in structure–activity relationships, in: From Chemical Topology to Three-Dimensional Geometry, Plenum Press, New York, 1977, pp. 73–116.
[14] L.B. Kier, L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.
[15] D.R. Flower, DISSIM: a program for the analysis of chemical diversity, J. Mol. Graphics Modell. 16 (1998) 239–253.
[16] G. Grassy, B. Calas, A. Yasri, R. Lahana, J. Woo, S. Iyer, M. Kaczorek, R. Floch, R. Buelow, Computer-assisted rational design of immunosuppressive compounds, Nat. Biotech. 16 (1998) 748–752.
[17] M. Randić, A. Nandy, S.C. Basak, D. Plavšić On numerical characterization of DNA primary sequences, J. Math. Chem., submitted for publication.
[18] M. Randić, Condensed representation of DNA sequences by condensed matrix, J. Chem. Inf. Comput. Sci. 40 (2000) 50–56.
[19] M. Randić, M. Vračko, On the similarity of DNA primary sequences, J. Chem. Inf. Comput. 40 (2000) 599–606.
[20] M. Randić, M. Vračko, A. Nandy, S.C. Basak, On 3-D graphical representation of DNA primary sequences and their numerical characterization, J. Chem. Inf. Comput. Sci. 40 (2000) 1235–1244.
[21] M. Randić, S.C. Basak, Characterization of DNA primary sequences based on the average distance between bases, J. Chem. Inf. Comput. Sci. 41 (2001) 561–568.
[22] X. Guo, M. Randić, S.C. Basak, A novel 2-D graphical representation of DNA sequences of lower degeneracy, Chem. Phys. Lett. 350 (2001) 106–112.
[23] M. Randić, A.T. Balaban, On 4-dimensional representation of DNA primary sequences, J. Chem. Inf. Comput. Sci., in press.
[24] L. Råde, B. Wedergoen, Beta Mathematics Handbook, CRC Press, Boca Raton, FL, 1990.
[25] M. Randić, A.F. Kleiner, L.M. DeAlba, Distance/distance matrices, J. Chem. Inf. Comput. Sci. 34 (1994) 277–286.
[26] F. Buckley, F. Harary, Distance in Graphs, Addison-Wesley, Reading, MA, 1989.
[27] M. Randić, Molecular shape profiles, J. Chem. Inf. Comput. Sci. 35 (1995) 373–382.
[28] M. Randić, Molecular profiles: novel geometry-dependent molecular descriptors, New J. Chem. 19 (1995) 781–791.
[29] M. Randić, Molecular bonding profiles, J. Math. Chem. 19 (1996) 375–392.
[30] M. Randić, G. Krilov, On characterization of 3-D sequences of proteins, Chem. Phys. Lett. 272 (1997) 115–119.
[31] M. Randić, G. Krilov, On characterization of the folding of proteins, Int. J. Quantum Chem. 75 (1999) 1017–1026.
[32] F. Gantmacher, Theory of Matrices, vol. II, Chelsea Publishers, New York, 1959 (Chapter 13).
[33] D.J. Klein, Similarity and dissimilarity in posets, J. Math. Chem. 18 (1995) 321–348.
[34] L. Bytautas, D.J. Klein, M. Randić, T. Pisanski, Foldeness in linear polymers: a difference between graphical and Euclidean-distances, DIMACS Ser. Discrete Math. Theo. Comput. Sci. 51 (2002) 39–61.