CROATICA CHEMICA ACTA
CCACAA **77** (1–2) 345–351 (2004)

ISSN-0011-1643
CCA-2934
Original Scientific Paper

Characterization of 2-D Proteome Maps Based on the Nearest Neighborhoods of Spots*

Milan Randić,[a,] Nella Lerš,[b] Dejan Plavšić,[b,**] and Subhash C. Basak[c]**

[a]*National Institute of Chemistry, P.O. Box 3430, 1001 Ljubljana, Slovenia and Department of Mathematics and Computer Science Drake University, Des Moines, IA 50311, USA*

[b]*The Ruđer Bošković Institute, P.O. Box 180, HR-10002 Zagreb, Croatia*

[c]*Natural Resources Research Institute, University of Minnesota at Duluth, 5013 Miller Trunk Highway, Duluth, MN 55811, USA*

RECEIVED APRIL 24, 2003; REVISED OCTOBER 7, 2003; ACCEPTED OCTOBER 20, 2003

Key words
proteome
2-D proteome map
map invariant
neighborhood graph

A novel approach to the construction of invariants for characterization of 2-D maps, such as 2-D proteome maps, 2-D NMR spectral maps, *etc*., is put forward. The approach is based on consideration of the neighborhood of points (spots) of the map and it is sufficiently flexible to allow one to vary not only the number of nearest neighbor spots used in characterization of a map but also the density of information on the relative distance of the selected map points. The method is illustrated with a Coomassie brilliant blue stained 2-D gel electrophoresis pattern of the Fisher F344 rat liver proteome.

INTRODUCTION

In order to arrive at a numerical characterization of graphical and visual 2-D maps, we have recently outlined the steps that allow construction of 2-D map invariants.[1–10] By the word »map« we mean a region of *X*-*Y* plane with *N* discrete points or spots given by their Cartesian coordinates. Here characterization means constructing of a set of *M* numerical invariants, which are quantities that are independent of the orientation of *X*, *Y* coordinate axes and independent of the labeling of the *N* points. Thus, if two laboratories consider the same map and have selected the same set of protein spots for characterization of a proteome map, then they should arrive at the same set of values of numerical descriptors of the map. The availability of such schemes makes it possible to catalogue visual data in a digital format, which enables one to numerically characterize 2-D maps and quantify their degree of similarity. Moreover, this will make it possible to explore the relationship between the structure of foreign agents (*e.g.*, toxins) and the effects they have on the proteome.

ON MATHEMATICAL OBJECTS ASSOCIATED WITH A MAP

The underlying structure of hitherto reported characterizations of 2-D maps, all of which have been illustrated with data from proteome laboratories, is summarized in Figure 1. The first step consists in associating with a map a suita-

* Dedicated to Academician Nenad Trinajstić, one of the most distinguished Croatian scientists.

** Authors to whom correspondence should be addressed. (M.R. Permanent address: 3225 Kingman Rd. Ames, IA 50014, USA; E-mail: mrandic@msn.com and dplavsic@rudjer.irb.hr)

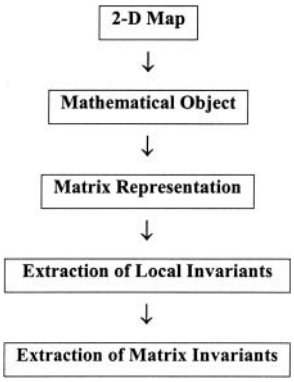

Figure 1. The underlying structure of quantitative methods for characterization of 2-D maps by map invariants.

ble mathematical object, which is then represented by a numerical matrix, which allows construction of additional matrices, all of which are used as the source of various map descriptors. In the past, the following mathematical objects were associated with 2-D maps: (i) embedded zigzag curve over a selection of N points of a map;[1–3,6] (ii) embedded graph of partial ordering on a selection of N points of a map;[4,5,7,9] and (iii) embedded cluster graph on a selection of N points of a map.[8] One can always associate with a set of N points in 2-D Voronoi polygons (in 3-D Voronoi polyhedra) and its dual, the Delaney triangulation of the plane,[11] but construction of these objects requires prior computer processing of the data. The question poses itself: Can we arrive at alternative mathematical objects that can be the basis for map characterization not being computer-intensive? As we will outline in this article, not only is the answer positive but the here proposed scheme has some advantages over the existing schemes for characterization of 2-D maps.

TABLE I. The list of coordinates (x, y) and abundance (z) of the 30 most intensive spots in the proteome map of liver cells of male Fisher F344 rats

Spot no.	x	y	z	Spot no.	x	y	z
1	2117.7	2278.6	1443.57	16	2032.7	902.8	800.15
2	2804.3	903.6	1436.30	17	2752.7	765.6	798.70
3	1183.9	959.6	1366.53	18	2334.2	980.2	727.91
4	2182.2	928.8	1272.95	19	1053.6	864.3	721.73
5	2685.6	1196.1	1185.81	20	2519.5	1365.9	694.52
6	1527.9	825.5	1149.29	21	2552.5	2409.4	677.72
7	1546.0	1352.5	1122.51	22	1214.3	620.0	648.84
8	2868.5	778.0	1088.93	23	2651.1	1149.6	610.74
9	1406.3	1118.1	982.24	24	2327.9	677.3	592.94
10	2450.2	409.2	936.01	25	2094.5	680.5	589.77
11	1474.0	665.1	900.04	26	1021.7	390.2	580.01
12	2974.9	772.8	867.30	27	1702.7	2138.3	574.00
13	2068.4	823.1	848.42	28	2070.4	929.6	554.02
14	642.2	669.8	824.92	29	2771.7	1451.0	538.96
15	2860.7	1649.9	819.65	30	2772.8	1326.9	513.47

For N points in a plane, an embedded zigzag curve introduces N–1 line segments (edges), while the embedded graph of partial ordering will have approximately N^2 edges. In contrast, the number of edges of an embedded cluster graph is variable and it depends on the selection of the critical distance. Hence, in this respect the approach based on the cluster graph has the flexibility that the other two schemes do not possess. On the other hand, it is still an open question what is the optimal »density« of a mathematical object that is sufficient to capture the most essential features of a map. Here we speak of »density« in the sense of »dense graphs« and »sparse graphs«, which is analogous to »dense matrices« and »sparse matrices«.[12] Hence, how many lines would suffice to adequately characterize a 2-D map? For example, does the use of N–1 line segments of a zigzag curve or N^2 edges of a graph of partial ordering suffice for adequate characterization of a 2-D map? As we will see, using the nearest neighborhood of each spot for characterization of maps offers an answer to the above question.

OUTLINE OF THE APPROACH

The novel approach is based on the concept of the neighborhood of a point (spot) in a 2-D map. In this respect there is some similarity between this approach and the cluster approach, but with a distinction that the present approach generates approximately uniform »density« for the embedded neighborhood graph. This is not the case of the cluster approach, which may lead to dense cluster graphs in some parts of the map and sparse cluster graphs in other parts. We will illustrate the novel approach with the data considered in our earlier work[3,9] shown in Table I, where we have listed the (x, y) coordinates of locations of protein spots, separated by the 2-D PAGE (PolyAcrylamide Gel Electrophoresis) technique. The data represent a proteome pattern from liver cells of male Fisher F344 rats. Figure 2

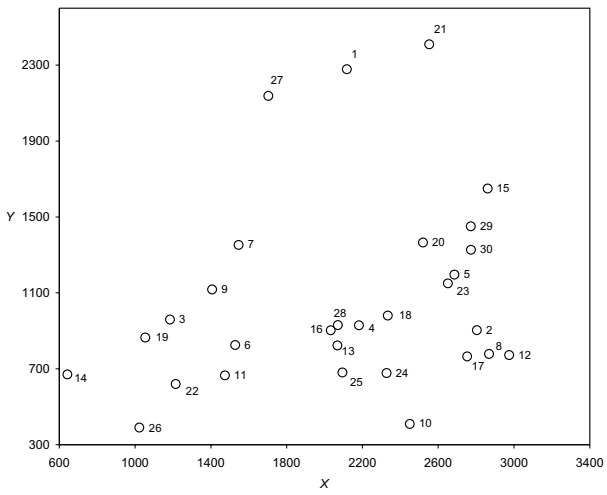

Figure 2. The simplified proteome map of liver cells of male Fisher F344 rats showing the locations of the spots of 30 most abundant proteins listed in Table I.

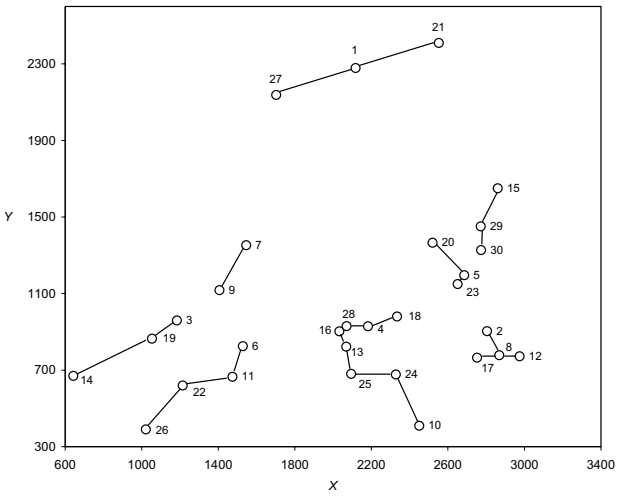

Figure 3. The neighborhood graph for the 30 protein spots of Table I for *NN* = 1.

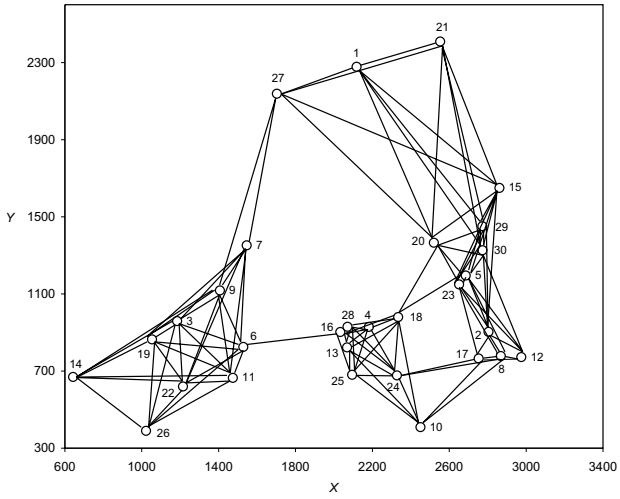

Figure 4. The neighborhood graph for the 30 protein spots of Table I for *NN* = 6.

shows the simplified proteome map, in which only the positions of the spots of 30 most abundant proteins are given.

We first calculate the Euclidean distances between protein spots and make a short list of its nearest neighbors for each spot in order to obtain the desired mathematical object, the neighborhood graph. We then select the number of nearest neighbors, *NN*, to be *NN* = 1, 2,..., 6. For a given *NN*, we find *NN* nearest neighbors for each spot separately and connect the spot considered by lines with these *NN* spots. Figures 3 and 4 show the neighborhood graphs obtained for the 30 protein spots of Table I when *NN* = 1 and *NN* = 6, respectively. Observe that in the case *NN* = 1 some vertices (spots) have degree one, while most of the other spots have degree two and one has degree three. Neighborhood graphs not only span all the spots of a map but they also have a fairly uniform density. For *NN* = 4, we have a connected graph for the first time, while for *NN* = 6 we obtain the graph depicted in Figure 4, having no cutvertex, which means that the embedded graph can be fully reconstructed from the ***AD*** matrix, in which only Euclidean distances between adjacent points are given. A vertex in a graph is called a cutvertex if its removal increases the number of components.[13]

NEIGHBORHOOD GRAPHS FOR *NN* = 1–6

Table II lists the six nearest neighbors for each of the 30 protein spots of Table I and the corresponding distances. The first entry in the second column (indicated by 1st), which is 27, means that the nearest neighbor of spot # 1 is spot # 27. Next entries in the first row: 21, 15, 20, 29, and 30 mean that the second, third, fourth, fifth and the sixth neighbors of spot # 1 are spots 21, 15, 20, 29, and 30, respectively. One should note that, *e.g.*, the nearest spot to spot 2 is spot 8 but the nearest spot to spot 8 is not spot 2 but spot 12. The number of edges in the graphs of the nearest neighborhood increases at an approximately constant rate, at least for the initial steps. For *NN* = 1 to *NN* = 6 the number of edges is: 22, 41, 57, 74, 92, and 111, respectively.

We will associate the relative distance matrix ***R***, derived from the ***AD*** matrix, with a neighborhood graph embedded in a proteome map. The ***R*** matrix is symmetric, $\boldsymbol{R} = \boldsymbol{R}^{\mathrm{T}}$, and its non-zero element $[R]_{ij}$ is defined as the quotient of the Euclidean distance between the corresponding pair of adjacent vertices (spots) (i, j) and the maximal Euclidean distance $D_{\max}$ between two spots in the map. In the case of the map in Figure 2, $D_{\max}$ is the distance between spots # 21 and # 26. For small values of *NN*, the ***R*** matrix is a sparse matrix. One should note that the present approach does not require searching for the shortest paths, as is the case with the approach using clustering of spots. Moreover, we preserve the flexibility of the cluster approach by having the opportunity to vary the number of edges in a graph by changing the number of nearest neighbors, the parameter *NN*.

In the second column of Table III, indicated by *NN* = 1, we list the row sums of the ***R*** matrix associated with the neighborhood graph for *NN* = 1. In the last row of Table III, we show the average row sum of the ***R*** matrix (0.12009), which is a map invariant. The individual row sums allow one to construct local map invariants. Suppose that we are interested in a smaller region of the proteome map, for example, a cluster of spots around the protein spot # 30, which involves spots 5, 15, 20, 23, 29, and 30. We can then consider only the row sums corresponding to these spots and view the average as a descriptor of this region. Thus, when using just the nearest neighbor, *NN* = 1, we obtain for the region: (0.11659 + 0.08599 + 0.09374 + 0.02285 + 0.13497 + 0.04898) / 6 = 0.08385; for *NN* = 2: 0.26414; for *NN* = 3: 0.45559, *etc*.

TABLE II. The six nearest neighbors for each of the 30 protein spots of Table I and the corresponding Euclidean distances in parentheses

Spot no.	Nearest neighbor protein spot (distance)					
	1st	2nd	3rd	4th	5th	6th
1	27 (432.39)	21 (459.80)	15 (973.30)	20 (999.66)	29 (1058.55)	30 (1158.79)
2	8 (141.06)	17 (147.33)	12 (214.97)	23 (289.80)	5 (315.67)	30 (424.47)
3	19 (161.43)	9 (273.10)	22 (340.96)	6 (369.21)	11 (413.39)	7 (425.03)
4	28 (111.80)	16 (151.74)	13 (155.32)	18 (161.11)	25 (263.33)	24 (290.66)
5	23 (57.90)	30 (157.20)	20 (237.53)	29 (269.05)	2 (315.67)	18 (411.38)
6	11 (169.21)	9 (316.86)	3 (362.21)	22 (374.93)	19 (475.88)	16 (510.68)
7	9 (242.03)	3 (425.03)	6 (557.51)	19 (569.07)	11 (699.22)	22 (744.25)
8	12 (106.53)	17 (116.46)	2 (141.06)	23 (430.52)	5 (456.36)	24 (549.90)
9	7 (242.03)	3 (273.10)	6 (316.86)	19 (434.52)	11 (458.03)	22 (533.82)
10	24 (294.68)	25 (447.35)	17 (467.47)	8 (557.66)	13 (563.10)	18 (584.62)
11	6 (169.21)	22 (263.59)	3 (413.39)	9 (458.03)	19 (465.21)	26 (529.29)
12	8 (106.53)	2 (214.97)	17 (222.32)	23 (496.81)	5 (512.72)	30 (589.81)
13	16 (87.33)	28 (106.52)	25 (144.97)	4 (155.32)	24 (297.65)	18 (309.65)
14	19 (455.06)	26 (471.38)	22 (574.26)	3 (614.35)	11 (831.81)	9 (885.90)
15	29 (217.90)	30 (334.75)	20 (443.93)	5 (486.41)	23 (542.43)	2 (748.43)
16	28 (46.26)	13 (87.33)	4 (151.74)	25 (230.73)	18 (311.78)	24 (571.47)
17	8 (116.46)	2 (147.33)	12 (222.32)	23 (397.21)	24 (433.88)	10 (467.47)
18	4 (161.11)	28 (268.99)	24 (304.97)	16 (311.78)	23 (358.40)	25 (385.33)
19	3 (161.43)	22 (292.42)	9 (434.52)	14 (455.06)	11 (465.21)	26 (475.17)
20	5 (237.53)	23 (253.19)	30 (256.28)	29 (266.17)	18 (426.00)	15 (443.93)
21	1 (459.80)	15 (819.65)	27 (892.00)	29 (983.15)	20 (1044.02)	30 (1104.69)
22	11 (263.59)	19 (292.42)	26 (299.84)	3 (340.96)	6 (374.93)	9 (533.82)
23	5 (57.90)	30 (215.05)	20 (253.19)	2 (289.80)	29 (324.63)	18 (358.40)
24	25 (233.42)	4 (290.66)	10 (294.68)	13 (297.65)	18 (304.97)	28 (360.50)
25	13 (144.97)	16 (230.73)	24 (232.42)	28 (250.26)	4 (263.33)	10 (447.35)
26	22 (299.84)	14 (471.38)	19 (475.17)	11 (529.29)	3 (592.05)	6 (667.63)
27	1 (432.39)	7 (862.97)	21 (892.00)	9 (1062.38)	20 (1124.17)	15 (1256.78)
28	16 (46.26)	13 (106.52)	4 (111.80)	25 (250.26)	18 (268.99)	24 (360.50)
29	30 (124.10)	15 (217.90)	20 (266.17)	5 (269.05)	23 (324.63)	2 (548.37)
30	29 (124.10)	5 (157.20)	23 (215.05)	20 (256.28)	15 (334.75)	2 (424.47)

The question arises: Is there an optimal number *NN* that suffices for characterization of a map, or do we get more and more information about the map if *NN* increases until all distances are incorporated and the neighborhood graph becomes the complete graph? As we will see, in the present approach it is sufficient to analyze any of the three initial neighborhood graphs in order to obtain map characterization, because it is possible to obtain the corresponding invariants of larger neighborhood graphs from smaller ones. Hence, if we select neighborhood graphs for *NN* = 1–3, it appears that approximately 2*N* edges may offer useful map characterization. This finding enables us to increase the efficiency of analyzing the data on proteome maps.

INCORPORATION OF ABUNDANCES INTO ANALYSIS

Thus far, we have not considered information on protein abundances in the proteome map studied. As outlined in our previous papers on graphical representation and numerical characterization of proteome maps,[1,2] inclusion of these data does not cause any difficulties. There are two alternative ways of including abundance as a third coordinate: Either (i) we construct neighborhood graphs using (x, y, z) coordinates; or (ii) we use 2-D neighborhood graphs but calculate map vectors (to be outlined later) using (x, y, z) coordinates. We prefer the latter approach for the following reasons: (i) There is no need to construct

TABLE III. Row sums of the relative distance matrices ***R*** associated with the neighborhood graphs representing the proteome map of Figure 1 for *NN* = 1, 2,..., 6

Row	Row sum of ***R***					
	NN = 1	*NN* = 2	*NN* = 3	*NN* = 4	*NN* = 5	*NN* = 6
1	0.35210	0.35210	0.73803	1.13255	1.55031	2.00763
2	0.05567	0.19865	0.19865	0.31302	0.43760	1.11691
3	0.06371	0.33923	0.78265	1.02511	1.25876	1.25876
4	0.10770	0.28224	0.34359	0. 34359	0.44751	0.44751
5	0.11659	0.17899	0.17899	0.47713	0.98416	1.14651
6	0.06678	0.19183	0.55756	0.70553	0.89334	1.35836
7	0.09552	0.60383	0.82385	1.04844	1.32439	1.61811
8	0.14365	0.14365	0.14365	0.53364	0.71374	0.93076
9	0.09552	0.32833	0.44981	1.09984	1.09984	1.66013
10	0.11630	0.29285	0.47734	0.69742	0.91961	1.15033
11	0.17081	0.17081	0.33396	0.72361	1.51144	1.51144
12	0.04204	0.12688	0.21462	0.41069	0.61304	0.84581
13	0.09168	0.13372	0.19502	0.43469	0.65688	0.65688
14	0.17959	0.36562	0.59225	0.83471	1.16299	1.51261
15	0.08599	0.54158	1.10271	1.29467	1.50874	2.30010
16	0.05273	0.20367	0.20367	0.20367	0.32671	0.67485
17	0.04596	0.10410	0.37633	0.53309	0.70432	0.70432
18	0.06358	0.16974	0.29010	0.41230	0.70350	1.23801
19	0.24330	0.35870	0.71771	0.94230	1.31371	1.31371
20	0.09374	0.19366	0.57504	0.96956	1.99341	1.99341
21	0.18146	0.50494	0.85697	1.24497	1.65700	2.09297
22	0.22236	0.33776	0.69895	0.84692	0.84692	1.35131
23	0.02285	0.20764	0.20764	0.84475	1.18694	1.32838
24	0.20841	0.32312	0.44348	0.56095	0.73218	1.23807
25	0.14894	0.41655	0.41655	0.51532	0.61924	0.61924
26	0.11833	0.30436	0.49189	0.70078	0.93443	1.19791
27	0.17064	0.51121	0.86324	1.28251	1.72617	2.22216
28	0.06238	0.21058	0.21058	0.30935	0.30935	0.45162
29	0.13497	0.13497	0.24001	0.73419	1.28007	1.49649
30	0.04898	0.32800	0.42914	0.42914	0. 42914	1.72272
Average	0.12009	0.28531	0.47347	0.72015	0.99485	1.30557

3-D neighborhood graphs to replace the already available 2-D neighborhood graphs; and (ii) Experimental errors in measuring the position of spots are separated from errors in measuring intensities of spots, and the only experimental errors affecting construction of neighborhood graphs are those of measuring the positions of spots. Proteome is not a fixed characteristic of the cell. The same cells exposed to different influences have different proteomes, but in the maps of these proteomes the same proteins are located in the same positions. Consequently, the *x*, *y* coordinates can be obtained with greater precision (after calculating the averages).

We will now construct a 2-component vector for each protein spot of the map. The first component of vector v_i associated with the protein spot *i* equals the sum of entries in the *i*-th row of the ***R*** matrix associated with the neighborhood graph embedded in the map. The second component is the protein abundance. Clearly, the magnitude of v_i, $|v_i|$, depends on *NN*, the number of the nearest neighbors considered. The components of these vectors are local descriptors. Table IV lists, for the 30 protein spots of Table I, the magnitudes of the 2-component vectors for *NN* = 1 to *NN* = 6. The entries of Table IV are obtained by combining the corresponding entries in Table

TABLE IV. Magnitudes of the 2-component vectors associated with the 30 protein spots of Table I for *NN* = 1 to *NN* = 6

Spot no.	Magnitude of 2-component vector					
	NN = 1	*NN* = 2	*NN* = 3	*NN* = 4	*NN* = 5	*NN* = 6
1	1.06018	1.06018	1.24285	1.51085	1.84485	2.24290
2	0.99652	1.01460	1.01460	1.04304	1.08694	1.49581
3	0.94877	1.03039	1.22827	1.39533	1.57499	1.57499
4	0.88767	0.92521	0.94573	0.94573	0.98824	0.98824
5	0.82972	0.84071	0.84071	0.94996	1.28193	1.41041
6	0.63830	0.81892	0.97196	1.06377	1.19662	1.57448
7	0.78343	0.98451	1.13286	1.30532	1.53579	1.79525
8	0.76782	0.76782	0.76782	0.92395	1.03843	1.19801
9	0.68709	0.75549	0.84426	1.29330	1.29330	1.79416
10	0.65875	0.71147	0.80516	0.95227	1.12521	1.32049
11	0.64645	0.64645	0.70729	0.95516	1.63499	1.63499
12	0.60227	0.61405	0.63798	0.72775	0.85836	1.03748
13	0.59483	0.60274	0.61923	0.73101	0.88142	0.88142
14	0.59900	0.67840	0.82298	1.01158	1.29580	1.61695
15	0.57426	0.78466	1.24030	1.41370	1.61204	2.36914
16	0.55697	0.59052	0.59052	0.59052	0.64341	0.87330
17	0.55503	0.56283	0.66900	0.76820	0.89555	0.89555
18	0.50823	0.53204	0.58174	0.65134	0.86555	1.33676
19	0.55601	0.61533	0.87468	1.06672	1.40563	1.40563
20	0.49016	0.51862	0.74976	1.08236	2.05065	2.05065
21	0.50332	0.68947	0.97714	1.33055	1.72222	2.14498
22	0.50022	0.56112	0.83025	0.95815	0.95815	1.42366
23	0.42370	0.47129	0.47129	0.94477	1.26009	1.39413
24	0.46060	0.52261	0.60448	0.69526	0.83953	1.30443
25	0.43485	0.58346	0.58346	0.65762	0.74187	0.74187
26	0.41885	0.50405	0.63513	0.80779	1.01715	1.26350
27	0.43270	0.64765	0.95042	1.34274	1.77138	2.25746
28	0.38882	0.43776	0.43776	0.49293	0.49293	0.59266
29	0.39700	0.39700	0.44384	0.82367	1.33341	1.54236
30	0.35905	0.48384	0.55738	0.55738	0.55738	1.75906
Average	0.60868	0.67844	0.80256	0.96642	1.19346	1.46402

III and Table I (column indicated by z whose entries have to be divided by 1443.57, the abundance value for protein # 1 in the list). For example, the first two entries in the second column of Table IV are 1.06018 = $\sqrt{0.35210^2 + 1^2}$ and 0.99652 = $\sqrt{0.05567^2 + 0.99496^2}$. As *NN* increases, the magnitudes of the vectors also increase, or remain the same if the particular protein has no additional neighbors. In the last row of Table IV, we give the average magnitudes of the 2-component vectors, which are 0.60868 for *NN* = 1; 0.67844 for *NN* = 2, *etc*. These quantities are invariants of the proteome map considered. Using the average magnitudes of the 2-component vectors for *NN* = 1 to *NN* = 6 one can construct a novel map invariant, a vector in 6-D space, which in the case of the proteome map considered reads (0.60868, 0.67844, 0.80256, 0.96642, 1.19346, 1.46402).

ANALYSIS OF MAP DESCRIPTORS

It is of considerable interest to see if there is some hidden structure in the collection of map descriptors. First, recall that each column of Table IV corresponds to one of the six neighborhood graphs such as those of Figures 3 and 4 for *NN* = 1 and *NN* = 6, respectively. While the number of edges in the neighborhood graphs increases as *NN* increases, there is no clear simple way of predicting

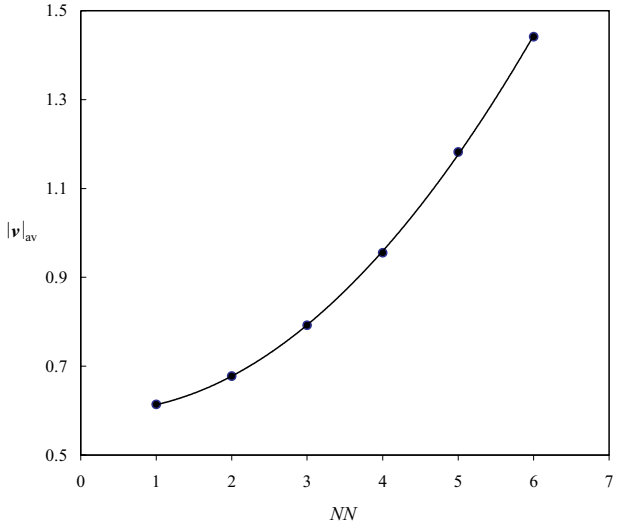

Figure 5. The plot of the average magnitude of two-component vectors $|\boldsymbol{v}|_{av}$ for different *NN* versus *NN*. The least-squares parabola is drawn in the plot.

the precise number of edges, except that they approximately increase at a constant rate (in this illustration 18.5). Thus, there is no simple relationship between the magnitudes of individual entries shown in Table IV for different *NN*, except that they steadily increase (or occasionally stay unchanged). The question arises: Are the average magnitudes of the 2-component vectors for *NN* = 1 to *NN* = 6 related? The answer is affirmative. A plot of the values in the last row of Table IV *versus NN* is shown in Figure 5. As one can see from Figure 5, the average magnitude of the 2-component vectors, $|\boldsymbol{v}|_{av}$, for *NN* = 1 to *NN* = 6 shows a quadratic dependence on *NN* of very high quality: $|\boldsymbol{v}|_{av}$ = 0.0253(±0.0007)NN^2 – 0.0059(±0.0036)*NN* + 0.5896(±0.0055), the correlation coefficient r = 0.9999, the standard error of estimate s = 0.0031, and the Fisher ratio F = 28204. This allows us to make the following two important conclusions: (i) We can reduce the number of significant descriptors from 6, the six components of 6-D vectors, to three, the parameters a, b, c defining the quadratic regressions: $y = ax^2 + bx + c$; (ii) Because the parabolic regression is of such high quality, it means that already with three neighborhood graphs, *e.g.*, *NN* = 1, 2, and 3, we have captured much of the characteristic of the parabolic regression. This then means, at least in this particular approach for characterization of proteome maps, that inclusion of a larger number of nearest neighbors does not contribute additional information on a map and hence we can work with neighborhood graphs of a relatively small number of nearest neighbors. This, of course, will be of interest when screening a large number of maps.

Acknowledgement. – This work was supported in part by the Ministry of Science and Technology of the Republic of Croatia.

REFERENCES

1. M. Randić, *J. Chem. Inf. Comput. Sci.* **41** (2001) 1330–1338.
2. M. Randić, J. Zupan, and M. Novič, *J. Chem. Inf. Comput. Sci.* **41** (2001) 1339–1344.
3. M. Randić, F. Witzmann, M. Vračko, and S. C. Basak, *Med. Chem. Res*. **10** (2001) 456–479.
4. M. Randić, *Int. J. Quantum Chem.* **90** (2002) 848–858.
5. M. Randić and S. C. Basak, *J. Chem. Inf. Comput. Sci*. **42** (2002) 983–992.
6. M. Randić, M. Vračko, and M. Novič, *J. Proteome Res*. **1** (2002) 217–226.
7. M. Randić, D. Plavšić, S. C. Basak, and B. D. Gute, unpublished paper.
8. Ž. Bajzer, M. Randić, D. Plavšić, and S. C. Basak, *J. Mol. Graphics Modell*. **22** (2003) 1–9.
9. M. Randić, J. Zupan, M. Novič, B. D. Gute, and S. C. Basak, *SAR QSAR Environ. Res.* **13** (2002) 689–703.
10. M. Randić, *Quantitative Characterization of Proteomics Maps by Matrix Invariants*, in: P. Michael Conn (Ed.), *Handbook of Proteomic Methods*, Humana Press, Totowa, NJ, 2003, pp. 429–450.
11. F. P. Preparata and M. I. Shamos, *Computational Geometry*, Springer-Verlag, Berlin, 1985.
12. M. Randić and L. M. DeAlba, *J. Chem. Inf. Comput. Sci.* **37** (1997) 1078–1081.
13. F. Harary, *Graph Theory*, Addison-Welsey, Reading, MA, 1969.

SAŽETAK

Karakteriziranje 2-D proteomskih mapa temeljeno na najbližoj okolini mrlja

Milan Randić, Nella Lerš, Dejan Plavšić i Subhash C. Basak

Predložen je novi pristup konstrukciji invarijanata za karakteriziranje 2-D mapa kao što su na primjer 2-D proteomske mape, 2-D NMR spektralne mape itd. Pristup se temelji na razmatranju okoline točaka (mrlja) mape i dovoljno je fleksibilan da ne samo da omogućava mijenjanje broja razmatranih najbližih susjeda pri karakteriziranju mape već i gustoće informacija o relativnoj udaljenosti odabranih točaka mape. Metoda je prikazana na proteomskoj mapi jetrenih stanica Fisher F344 štakora dobivenoj 2-D elektroforezom i bojanjem s Coomassie plavim.