



HAL
open science

Study of proteome maps using partial ordering

Milan Randić, Marjana Novič, Marjan Vračko, Dejan Plavšić

► To cite this version:

Milan Randić, Marjana Novič, Marjan Vračko, Dejan Plavšić. Study of proteome maps using partial ordering. Journal of Theoretical Biology, 2010, 266 (1), pp.21. <10.1016/j.jtbi.2010.06.008>. <hal-00613135>

HAL Id: hal-00613135

<https://hal.science/hal-00613135v1>

Submitted on 3 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Author's Accepted Manuscript

Study of proteome maps using partial ordering

Milan Randić, Marjana Novič, Marjan Vračko, Dejan Plavšić

PII: S0022-5193(10)00292-4
DOI: doi:10.1016/j.jtbi.2010.06.008
Reference: YJTBI6028

To appear in: *Journal of Theoretical Biology*

Received date: 7 October 2009
Revised date: 3 June 2010
Accepted date: 3 June 2010

Cite this article as: Milan Randić, Marjana Novič, Marjan Vračko and Dejan Plavšić, Study of proteome maps using partial ordering, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2010.06.008](https://doi.org/10.1016/j.jtbi.2010.06.008)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/jtbi

Study of Proteome Maps Using Partial Ordering

Milan Randić,^{a,*} Marjana Novič,^a Marjan Vračko,^a and Dejan Plavšić^b

^a National Institute of Chemistry, Ljubljana, Slovenia; and

^b Rudjer Bošković Institute, NMR Center, P. O. Box 180, HR-10002 Zagreb, Croatia

Key words:

Proteomics maps
Partial ordering
Peroxisome
proliferators

List of abbreviations:

PFOA	perfluorooctanoic acid
PFDA	perfluorodecanoic acid
DEHP	di(2-ethylhexyl)phthalate
QSPR	Quantitative Structure Property Relationship
QSAR	Quantitative Structure Activity Relationship

*Visiting scientist

*Corresponding author at: 3225 Kingman Rd, Ames, IA 50014. Tel: +1 515 292 7411;

fax: +1 515 292 8629.

E-mail: mrandic@msn.com (M. Randić)

Abstract

This paper describes numerical characterization of proteome maps based on partial ordering of protein spots with respect to the mass and the charge. The partial ordering diagram is embedded directly over the 2D map and the corresponding adjacency matrix is constructed. The adjacency matrix is augmented by including the information on the abundance of proteins in a gel as suitably scaled diagonal entries of the matrix. The approach is illustrated on proteome maps of Anderson *et al.* (1996) based on experimental results from liver cells of rats exposed to four peroxisome proliferators. We used the leading eigenvectors of the adjacency matrices as maps descriptors in order to determine the degree of similarity between proteome maps.

INTRODUCTION

When an experimental animal is exposed to drug or toxin, the protein content of cells will change and may change dramatically and in an unpredicted way. It is of interest to characterize variations of the relative abundance of proteins from that which characterizes a normal cell, and if possible try to related the perturbations in altered proteome map to variations in agents causing the change. Such work may eventually show the extent to which Emil Fishers "similar drugs have similar effects" extends also to the collection of proteins, cellular proteome. To facilitate characterization of changes of protein composition in cells one needs to characterize individual proteome maps by a set of map descriptors, which are mathematical invariants derived from proteome map. In this communication we will outline novel numerical map descriptors for characterization of proteome maps, which are conceptually and computationally not very involved.

Typical proteome map may display in 2D gel 2000 protein spots, proteins separated by charge and mass. The experimental information of proteome maps are usually represented as a list, the x-coordinate giving the charge and the y-coordinated giving the mass, followed by the relative abundance of individual spots. At left in Table 1 we have illustrated such results for twenty most intensive spots of normal rat liver cells, while the remaining part of Table 1 lists the abundances of the same 20 proteins for proteome maps obtained from liver cells obtained from animals exposed to four chemical agents. The experimental data listed in Table 1 are from the work of Anderson and coworkers (1996) who were studying the effect of several peroxisome proliferators on rats liver cells. Fig. 1 shows the locations of the twenty most abundant protein spots labeled 1 – 20 in 2D gel, the coordinates x, y of which are listed in Table 1.

-----Table 1 around here -----

-----Fig. 1 around here -----

Our interest is in developing a "tool" for comparative study of proteome maps. So for such work neither the number of selected points nor necessarily the criteria of selection is essential. However, because of unknown variability of experimental data we focused on the most abundant proteins, the relative experimental errors for which are expected to be the least. The question on the number of protein spots that may suffice to represent a map or cellular proteome as a whole has been considered (Randić et al., 2006; Randić, 2006) and appears to be an order of magnitude larger, but as mentioned above for the purpose of developing novel approaches for quantitative characterization of proteome maps the number of proteins considered is not critical.

Proteome maps invariant are quantities that are independent of the labeling of the proteins in proteome maps or 2D gel. Such quantities characterize a map as a *whole* or more correctly they characterize simplified map consisting of proteins selected for numerical characterization of a map. Obviously if one is to capture salient feature of information-rich proteome maps one needs several map descriptors. This paper is a continuation of our efforts to develop suitable mathematical characterization of proteome maps (Randić, 2001; Randić et al., 2001; Randić, 2002; Randić et al., 2002; Randić and Basak, 2004; Randić et al., 2005). In particular in this article we will introduce novel invariants constructed by using *partial ordering* of proteins with respect to their charge and mass and the partial ordering diagram *embedded* over proteome maps.

PARTIAL ORDERING BASED ON CHARGE AND MASS

While ordering of objects with respect to a single property is trivial, ordering of objects with respect to two or more properties is far from trivial. Scottish mathematician Muirhead around the year 1900 initiated such studies that have led to rules for construction of partial ordering (Muirhead 1903; Hardy et al., 1988). Partial ordering has not been much used in chemistry and is not widely known methodology in chemistry despite that it offers a useful tool for comparison of objects characterized by sequences. Among few applications of partial ordering in chemistry we may mention study of variations in properties of isomers (Randić and Wilkins, 1979), and the search for a characterization of pharmacophore in QSAR (Randić et al., 1987; Randić 2000). For recent developments of partial orderings in chemistry see the special issue of MATCH (Klein and Brickmann, 2000).

In this contribution we will use partial order to construct a *graph* to be associated with a proteome map. Having a graph one can construct its adjacency matrix, which is a binary matrix with zero and one as elements, the ones indicating adjacent vertices in the graph. Selected invariants of adjacency matrix can be used as map descriptors. An advantage of using partial ordering graph for construction of map descriptors is that minor variations in experimentally measured mass and charge of proteins spots will have no influence on our analysis. For earlier applications of partial ordering for characterization of proteomics maps see Randić (2002) and Randić and Basak (2004).

To obtain the diagram of partial ordering one *order* the twenty spots first of Table 1 with respect to their *charge* and then with respect to their *mass*. The units of charge and mass are those as reported by Anderson *et al.* (1996). By ordering protein spots with respect to mass we obtain the sequence: 1, 15, 20, 7, 5, 9 . . . This is the order as spots appear in the map when looking from the top of the map (proteins having the greatest mass) to the bottom (proteins having the smallest mass). When one orders the same protein spots with respect to the charge

one obtains the sequence: 12, 8, 15, 2, 17, 5 . . . which reflects the order in which protein spots appear if we go from right (the largest charge) to left (the smallest charge).

Partial ordering is defined as a set of *all* orderings involving subsets of elements of the two sequences that are true for both sequences. To find all such partial orderings it is convenient to list, as shown in Fig. 2, proteins in order of descending mass and descending charge. Once the two sequences are obtained, one connects the same protein in both sequences. Any set of non-crossing lines that one can extract from such diagram represents one partial ordering, that is, the ordering is satisfied for so selected proteins for both sequences. For example, the lines 15 - 15, 5 - 5, 18 - 18, 4 - 4, 16 - 16, 6 - 6, and 14 - 14 do not cross giving the sub-sequence $15 > 5 > 18 > 4 > 16 > 6 > 14$, which means that protein 15 has bigger mass and bigger charge than protein 5, and protein 5 has bigger mass and bigger charge than protein 18, and so on. In Table 2 we have listed only all such fragmentary orderings that start with protein spot 1. The totality of all fragmentary orderings makes the Partial Ordering for the selected twenty proteins of map of Fig. 1.

-----Table 2 around here -----

-----Fig. 2 around here -----

It is customary to graphically represent partial ordering by a directed graph or a hierarchical diagram, such as one illustrated in Fig. 3 in which points at right *dominate* points to the left. *The dominance* here means that an element at the *left* in Fig. 3 is also *below* in both sequences of Fig. 2. Fig. 3 facilitates visual inspection of the partial ordering and immediately shows which protein dominate (by the mass and by the charge) other proteins, or is dominated by others. In general when depicting partial ordering one can place vertices in arbitrary positions as

long as the top-bottom and the left-right dominance relations are preserved, but for better visibility one tries to construct diagram that will have as few as possible crossings of lines.

-----Fig. 3 around here -----

In the approach considered here we depart from the usual practice outlined above and we insist on construction of the partial ordering diagram that is directly *superimposed* on the proteome map. This is in fact the diagram already illustrated in Fig. 3. This is a critical step, because we want to model underlying proteome maps, and partial ordering is used only to arrive at suitable geometrical object for our considerations. This approach has been for the first time introduced for representation of partial ordering diagrams of proteome maps in ref. 4. Such *embedding* of graphs in 2-D space (plane) in Graph Theory implies a *fixed geometry* for the graph considered, which means that in addition to *adjacency* relationship between vertices in such graphs are also determined the *distances* between all vertices, whether adjacent or not.

Observe an important feature of the embedded graph of Fig. 3 in that all the lines in the graph have positive slope. This is a consequence of the dominance relation, and is a property that can be used for a direct construction of partial ordering diagram for a given map *without* a need for construction of Fig. 2 and the accompanying search for components of partial ordering. To obtain Fig. 3 directly from Fig. 1 one can start with the top vertex (spot 1 in Fig. 3) and connect it to the next lowest spot to left of 1 which is spot 7. One continues the same with vertex 7 and connects it to the next lowest vertex below it and to left, which is spot 3. One continues to connect 3 to 19 and finally 19 to 14. By exhausting this particular trail one return to vertex A and repeat the process: one connect 1 to 9 and then 9 to 3. In next step we connect 1 to 6 and finally 6 to 14. By backtracking one connects 6 to 11. Finally one connects 1 to 16 and 1 to 13, which are connected to 6 and 11, respectively. This has exhausted all the fragmentary orders starting

with protein spot 1, which have been listed in Table 2. The process continues with spot 15, then 8 and 12, which completes construction of the embedded graph of partial order for the map considered. In passing we may mention that the partial ordering diagram depicted in ref. [4], which considers the same twenty proteins, has a missing link that should connect protein 2 (2804.3, 903.6) and protein 16 (2032.7, 902.8), which was apparently overlooked visually because of such a small difference in the two y-coordinates,

THE ADJACENCY MATRIX OF PROTEOME MAPS

Adjacency matrix is well known binary matrix of Graph Theory [19], the matrix elements of which are defined as:

$A_{ij} = 1$ if vertices i and j are adjacent;

$A_{ij} = 0$ otherwise.

For the graph of partial ordering of the proteome map of Fig. 1, illustrated in Fig. 3, the 20x20 adjacency matrix is shown in Table 3.

-----Table 3 around here -----

Our numerical characterization of the five proteome maps of Table 1 will be based on *augmented* adjacency matrix, in which we will replace the diagonal zeros by the corresponding relative abundances for individual proteome maps. The x, y coordinate do not enter directly into analysis but determine the adjacency of the spots, while we will incorporate the abundance of proteins by augmenting the adjacency matrix by introduction of non-zero diagonal terms. Whenever experimental quantities are measured in different units, as is the case with proteome maps, which involve mass, charge and abundance, if one is to use experimentally measure

numerical values in a subsequent analysis one has to consider suitable normalization, in order that one of the properties does not dominate the other, because of use of different scales. Kowalski and Bender (1972) recommended that in such situations one should re-scale the units used for measuring these quantities and scale them both to a same interval, such as (-1, +1). In our case we have three quantities, the protein charge (coordinate x), the protein mass (coordinate y) and the protein abundance (coordinate z) that we would like to combine into our analysis.

In view that the coordinates x and y of proteome maps are of the same order of magnitude, even if they would directly enter into our analysis we could skip their rescaling. However, we have to consider the suggestion of Kowalski and Bender about scaling of the third coordinate (abundance). Observe that the approach of Kowalski and Bender does not consider a possibility that the number of points considered may also have an effect on the analysis, which is the case with use of augmented matrices (to be introduced later). Their approach is adequate only in cases when analyses are using augmented matrices of the same size. This, however, is not the case in our analyses of proteome maps, because on different occasions, and even for the same proteome maps, one may change the number of proteins considered and thus change the size of matrices considered.

For augmented matrices there is a problem that the same (0, 1) scaling will give different weight to diagonal and off-diagonal matrix elements. If diagonal entries are normalized on (0, 1) scale (assuming the we consider adjacency matrix, whose elements are 0 or 1) as the size of matrices increases the number of off-diagonal entries increases faster than the size of diagonal entries and off-diagonal entries will play more and more dominant role with respect to the role of diagonal entries. In general the number of off-diagonal entries will increase with n^2 while the number of diagonal entries will increase with n , where n is the size of adjacency matrix.

As outlined by Randić and Orel (2010) there is a way to arrive at a matrix in which both the off-diagonal entries and the diagonal entries will have balance roles. This is achieved by having the average matrix element (excluding diagonal elements) and the average trace element equal, or which is the same, having the sum of all matrix elements (excluding diagonal) equal to the sum of all diagonal elements. The sum of off-diagonal entries for the adjacency matrix of Table 3 is 72, so the abundance of Table 1 have to be multiplied by $72/(\text{Abundance sum})$, which is for the proteome maps of the control group $72/ 27758.54 = 0.002594$. The so normalized abundances for the five proteome maps of Table 1 are listed in Table 4.

-----Table 4 around here -----

In the last row in Table 4 we have listed the so scaled abundance sums for the five maps. As expected for the control group the sum equals 72, because the graph of Fig. 3 has 36 edges, each being counted twice in the adjacency matrix. The last row in Table 4 immediately shows the overall effect of peroxisome proliferators on the liver cells of experimental animals. We see an overall decrease of most abundant twenty proteins for three peroxisome proliferators (PFOA [perfluorooctanoic acid], PFDA [perfluorodecanoic acid], and clofibrate) and an increase for the peroxisome proliferator DEHP [di(2-ethylhexyl)phthalate]. Observe that proteome maps having the same proteins (the same list of the (x, y) coordinates) are scaled using the same normalization factor (based on the map of the control group).

AUGMENTED MATRICES

What makes the five proteome maps different is the information on abundance of proteins as measured by the intensity of gel spots. This information differentiates maps and will reflect

the role of drugs and other xenobiotics agents that induce changes in the proteome maps. We will incorporate the information on abundance as diagonal entries of the matrix in analogy with differentiation of heteroatoms in construction of variable connectivity indices (Randić, 1991; Randić and Dobrowolski, 1998; Randić et al., 2000). This approach has been outlined in the introductory papers on the mathematical characterization of proteome maps based on 2D zigzag pattern of proteome map and its generalization to 3-D zigzag pattern (Randić, 2001, Randić et al., 2001; Randić and Basak, 2004), however normalization used there were not chosen to parallel matrix size-dependence.

The selected twenty most abundant spots of the control group determine selection of proteins for the remaining proteome maps of PFOA, PFDA, clofibrate and DEHP. While the relative abundance to the control group steadily decrease the relative abundances of the remaining four maps oscillate at an irregular pattern illustrated at the top of Fig. 4. The most abundant appear to be protein having label 6 for liver cells exposed to PFDA (2.217), and the least abundant among proteins considered is protein 3 for liver cells exposed to DEHP (0.081).

-----Fig. 4 around here -----

From Table 4 it is interesting to observe individual variations in proteins. Thus protein 15 in the case of PFOA and PFDA has *decreased* its abundance considerably but in the case of DEHP it has *increased* its abundance. In many cases there are considerable changes in abundances in comparison with that of the control group even if the changes are within the bounds of $\pm 50\%$. In some cases abundances of proteins increased slightly after exposure to the four agents (like proteins 2 and 11) assuming that the changes are statistically significant.

Similarly some proteins diminished their abundance, although often not evenly (like proteins 7, 9, and 17). Protein 14 appears to be among the least affected by any of the four agents considered.

We are going to try to construct numerical characterization of the five proteome maps. We have selected to consider the eigenvalues and eigenvectors of the augmented adjacency matrices. In Table 5 we have listed the coefficients of the eigenvector accompanying the leading eigenvalue, the largest eigenvalue, obtained from the five augmented adjacency matrices. We will refer to the columns of Table 5 as the leading eigenvectors. Observe how the coefficients of the leading eigenvectors for the five proteome maps, which are illustrated in the bottom part of Fig. 4, parallel the relative abundances of the proteins illustrated at the top of Fig. 4. This parallelism suggests that the leading eigenvectors may offer a fair numerical characterization of the proteome maps. While the leading eigenvalues have found use as molecular descriptors for study of structure-property relationship, this is the first time that the leading eigenvectors have found useful application in chemical graph theory and bioinformatics.

-----Table 5 around here -----

In Table 6 we show the degree of similarity/dissimilarity for pair wise comparisons of the five proteome maps based on the degree of similarity/dissimilarity for the corresponding leading eigenvectors. The values in the table were computed by viewing columns in Table 4 as a 20-component vector. The Euclidean distance (in 20-dimensional vector space) gives the distance between the corresponding end-points of vectors. The smaller is the distance the more similar are vectors (or alternatively the more similar are the corresponding proteome maps). The first row gives the similarity of the four perturbed maps with the map of the control group. As one see between the four peroxisome proliferators clofibrate and DEHP cause the least perturbation of liver cell proteome, while PFDA is causing the largest perturbation. Between all five proteome

maps the most similar proteome maps are those of clofibrate and DEHP. However, above comparisons may obscure details on how each agent affects individual protein types, and the overall similarity does not imply that neither the two chemicals have necessarily similar affect on all proteins, nor that the magnitudes of the overall perturbation will manifest in similar biological behavior. That the former is the case is illustrated, for instance PFOA and PFDA, which chemically differ only somewhat in their size, in that PFOA makes little change on the abundance of protein 5, while PFDA has drastically reduced the abundance of protein 5 in liver cells.

-----Table 6 around here -----

Clearly a single descriptor, such as the 20-component vectors representing the leading eigenvector cannot capture complexity of variations occurring even for a relatively small maps such as map considered here that is based on information on only twenty proteins. It is therefore desirable to have additional descriptors, one such additional descriptor will be outlined in the next section.

THE EIGENVALUE OF MAP MATRICES

We continue to consider the five proteome maps for control groups, PFOA, PFDA, clofibrate and DEHP of Table 1, and will examine use of eigenvalues of the augmented adjacency matrices as map descriptors. In Table 7 are listed (in the reversed order) eigenvalues of the augmented adjacency matrices for the normal (control) rat liver cells and four peroxisome proliferators. One can view the columns in Table 7 as vectors in 20-dimensional vector space and one can construct the similarity/dissimilarity matrix by calculating the Euclidean distances between these vectors. Before reporting the results let us mention that there are no many

illustrations that the set of eigenvalues of structural matrices have been offering suitable descriptors for molecular structure-property studies. Even less this would be the case for characterization of complex chemical and biological systems (Randić, 2001, Randić et al., 2001; Randić and Basak, 2004). An exception are the leading eigenvalue of few matrices, which could be interpreted in term of structural concepts. Thus Lovász and Pelikan (1973) interpreted the leading eigenvalue of the adjacency matrix as an index of branching in acyclic graphs, and the leading eigenvalue of the D/D matrix (the elements of which are given as the quotient of the through space and through bonds distances) as an index of bending of chain structures (Randić et al., 1994), and more generally an index of compactness for molecular graphs in general (Randić and Plavšić, 2010). Also the leading eigenvalue of path matrices (Randić, 1998; Randić et al., 2000), which are special case of graphical matrices (Randić et al., 1997, Randić et al., 2004) is very good index of molecular branching.

-----Table 7 around here -----

When the eigenvalues of Table 7 are used as components of vectors in 209D space one obtains the entries in Table 8 as their Euclidean distances. This table can be used as an alternative measure of the degree of similarity or dissimilarity between the five proteome maps. Some parallelism but also some differences can be observed when Table 8 is compared to Table 6 in which the similarity/dissimilarity between the five proteome maps was based on the leading eigenvectors, rather than the set of eigenvalues. This time one finds as the most similar proteome maps those of PFOA and clofibrate. Other small entries in Table 8 are between the control and DEHP, PFOA and PFDA and control and clofibrate. The parallelism between the similarity/dissimilarity of the five proteome maps based on leading eigenvector and all eigenvalues is illustrated in Fig. 5.

-----Table 8 around here -----

-----Fig. 5 around here -----

DISCUSSION

That there is parallelism between the characterization of proteome maps by the leading eigenvector and the set of all eigenvalues is interesting but need not be fully surprising. The eigenvalues if placed on the diagonal of a 20x20 null-matrix represent a transformation of the initial augmented adjacency matrix into diagonal form. On the other hand the experimental abundances, which we have included as the diagonal entries, apparently have played the dominant role in the matrix eigenvalue problem. One can view the augmented adjacency matrix as a diagonal matrix that is perturbed, the perturbation being present in the form of the off-diagonal elements. This suggests that the agreement between the two characterizations of the proteomics maps will agree even more if we are further to emphasize the role of diagonal elements. The limiting case of completely negligible role of the off-diagonal matrix elements correspond to transforming the problem of characterization of proteomics maps to that of characterization of cellular proteome, because the role of the x, y coordinates has been eliminated. That the analysis of proteome map and proteome, though related, is distinct is well illustrated by a paper of Randić and Estrada (2005), who examined the reported variations of protein abundances as concentration of particular peroxisome proliferator was changes over two orders of magnitude. Thus, by ignoring the (x, y) coordinates they were able to detect a non-linear dose response at the cellular level, the phenomenon that has been known previously to holds for an organism (animal, plant) as a whole (Calabrese and Baldwin, 2003).

The similarity/dissimilarity tables such as Table 6 and Tables 8 do show some differences which illustrate the fact, that even though two approaches are based on similar methodologies, they capture somewhat different map information. As already mentioned the smallest entry in Table 6 belongs to the pair (clofibrate, DEHP), which is not the case with Table 8 where the smallest entry is associated with the pair (PFOA, clofibrate). Similar results have been found with use of “higher order” matrices and 10-D and 15-D vectors, respectively in Randić and Basak (2002). A relatively small value of the Euclidean distance entry in a similarity/dissimilarity table does not necessarily imply that the effects of the corresponding chemicals are the very similar, but it signifies that the effect of those chemicals on proteome on *average* may be similar. Even though we used a small number of proteins we have already seen that in number of instances the same proteins have been affected differently so it would be premature to speculate that higher similarity between molecules will necessarily be reflected to high the degree of parallelism, but such conclusion have to wait on experimental data derived from use of more similar compounds.

It remains to be seen modification of this or other approaches may be more suitable when comparing different proteome maps. At this stage it is important to recognize that the presented approach has some flexibilities. One should recall that here we chose to consider a particular set of invariants, the leading eigenvector and the set of eigenvalues of augmented adjacency but other matrix and graph invariants ought to be examined. Important novelty of the current approach is that we found a scheme that allows experimental data to compact into matrices that thus become the source for further exploration of comparative proteome.

As already hinted to relate similarity /dissimilarity among proteome maps with similarities/dissimilarities of agents causing perturbations would be facilitated by having data on a set of chemical compounds showing closer structural features. If there is some parallelism between the biological descriptors of proteome maps and physico-chemical descriptors inducing

changes in proteome one could test the degree to which the paradigm that similar chemical structures induce similar properties, that holds in QSPR (Quantitative Structure Property Relationship) and QSAR (Quantitative Structure Activity Relationship), also applies to proteome, even if proteome data may point to drastic changes of the cell proteome under minor perturbations of biological systems. Another testing ground for outlined numerical characterizations of proteome maps is examination of perturbations of proteome induced by gradual increase of concentrations of chemicals that produce visible perturbations of proteome, which is currently underway in our laboratory. We should add that all the calculations were made using Excel and the standard matrix manipulations available in MATLAB software (MathWorks, 2010).

CONCLUDING REMARKS

Here we have tested two measures of “proteome similarity” of the chemicals based on augmented adjacency matrices that include the information on the abundances of protein spots. One approach considers the “leading” eigenvectors (eigenvectors belonging to the leading eigenvalue which have all coefficient positive) and the other consider eigenvalues as descriptors. These two measures give similar but different pictures of the similarity of the effects of four peroxisome proliferators on rat liver cell proteome. This brings us close to the well-known problem of similarity methods for molecules: Similarity methods are not uniquely defined, they are rather user-defined and context specific (Randić, 1990, Randić, 1996). In the area of molecular similarity for relatively small molecules, Basak et al., (1988); Basak and Grundwald (1994); and Basak et al., (1999) calculated and used various molecular similarity methods based on graph invariants to select analogs, cluster large libraries of chemicals (Basak et al., 2002), and

estimate properties of chemicals from their selected neighbors (Basak et al., 1998; Basak et al., 2001) in the selected spaces. It was concluded that no universal similarity method can handle all situations. It would be interesting to see how far map invariants are able to characterize various aspects of proteome maps and their perturbation derived from the exposure of organisms to drugs and toxins.

Acknowledgments

MR thanks the laboratory for Chemometrics, National Institute of Chemistry, Ljubljana, Slovenia, for warm hospitality. This work has been supported in part by the Ministry of Science and Higher Education of Republic of Slovenia, under research grant P1-017 and the Ministry of Science, Education and Sports of the Republic of Croatia under Project “NMR Spectroscopy and Modeling of Bioactive Molecules.” We also thank Professor A. T. Balaban (Texas A&M University at Galveston) who was kind to examine the manuscript and suggested improvements that resulted in better presentations of our work.

Tables

Table 1 Gel coordinates and abundance for the twenty most intensive protein spots of rat liver cells of normal cells and the corresponding abundance for cells exposed to four proliferators

#	x	y	control	PFOA	PFDA	Clofibrate	DEHP
1	2112	2279	1444	1087	950	1471	1659
2	2804	904	1436	1556	1886	1599	1551
3	1418	960	1367	1139	1503	1636	81
4	2182	929	1272	992	731	766	1121
5	2686	1196	1186	1128	498	1099	1388
6	1528	826	1149	1924	2216	1661	1806
7	1346	1353	1123	587	389	732	771
8	2869	778	1089	26	507	459	1169
9	1406	1118	982	911	830	842	929
10	2450	409	936	832	629	799	1094
11	1474	665	900	1293	1124	1127	1194
12	2975	773	867	707	787	1058	1163
13	2068	823	848	738	455	719	974
14	642	670	825	740	745	847	885
15	2861	1650	820	161	165	601	1490
16	2033	903	800	773	801	760	1008
17	2753	766	798	208	131	388	538
18	2334	982	728	764	527	556	774
19	1054	864	722	780	604	468	781
20	2520	1366	695	378	161	572	713

Table 2 Fragmentary orderings of proteins spots holding with respect to charge and mass respectively that start with protein 1.

1	1 ⇒	7 ⇒	3 ⇒	19 ⇒	14
2	1 ⇒	9 ⇒	3 ⇒	19 ⇒	14
3	1 ⇒	13 ⇒	11		
4	1 ⇒	13 ⇒	14		
5	1 ⇒	16 ⇒	6 ⇒	11	
6	1 ⇒	16 ⇒	6 ⇒	14	
7	1 ⇒	16 ⇒	6 ⇒	14	
8	1 ⇒	16 ⇒	19 ⇒	14	

Table 3 The adjacency matrix for the partial ordering graph of Fig. 3

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

Table 4 Scaled Abundance values

Spot	Control	PFOA	PFDA	Clofibrate	DEHP
1	5.200	3.916	3.423	5.299	5.976
2	5.174	5.604	6.794	5.760	5.586
3	4.923	4.102	5.413	5.895	0.292
4	4.582	3.572	2.632	2.761	4.038
5	4.272	4.063	1.793	3.956	5.000
6	4.140	6.933	7.982	5.983	6.506
7	4.044	2.114	1.402	2.636	2.777
8	3.923	0.940	1.828	1.654	4.209
9	3.539	3.284	2.989	3.033	3.348
10	3.372	2.996	2.267	2.877	3.940
11	3.242	4.659	4.048	4.058	4.301
12	3.124	2.549	2.835	3.810	4.189
13	3.056	2.659	1.638	2.591	3.510
14	2.972	2.665	2.683	3.051	3.190
15	2.953	0.581	0.594	2.164	5.367
16	2.883	2.785	2.885	2.739	3.633
17	2.876	0.749	0.472	1.398	1.939
18	2.622	2.751	1.900	2.003	2.789
19	2.600	2.809	2.175	1.686	2.814
20	2.502	1.363	0.581	2.059	2.568
Sum	72	61.095	56.334	65.417	75.974

Table 5 Leading eigenvector for the Twenty Most Intensive Protein Spots of Rat
Liver Cells of the Normal Cells and Cells Exposed to Four Chemicals

Protein	Control	PFOA	PFDA	Clofibrate	DEHP
1	0.5314	0.3195	0.2374	0.5065	0.5287
2	0.3059	0.1973	0.1853	0.2823	0.2471
3	0.2157	0.0608	0.0447	0.1918	0.0417
4	0.2134	0.1052	0.0573	0.1073	0.1358
5	0.1629	0.0431	0.0138	0.0800	0.1067
6	0.3039	0.7737	0.8699	0.5646	0.5716
7	0.2404	0.0675	0.0396	0.1429	0.1119
8	0.0231	0.0041	0.0033	0.0080	0.0109
9	0.2501	0.0919	0.0524	0.1701	0.1443
10	0.0919	0.0207	0.0081	0.0404	0.0507
11	0.1284	0.2663	0.2012	0.1987	0.1904
12	0.0192	0.0053	0.0039	0.0121	0.0108
13	0.2802	0.1915	0.1162	0.2326	0.2490
14	0.1478	0.1887	0.1679	0.1796	0.1681
15	0.1287	0.0359	0.0258	0.0746	0.1344
16	0.3021	0.2711	0.2335	0.2935	0.3162
17	0.0896	0.0302	0.0237	0.0514	0.0479
18	0.1026	0.0333	0.0124	0.0438	0.0558
19	0.1282	0.0951	0.0659	0.1042	0.0907
20	0.5314	0.0360	0.0165	0.0785	0.0823

Table 6 Similarity/Dissimilarity Among Perturbations of Abundances of Proteome Rat Liver Cells for the Control and the Four Peroxisome Proliferators based on 20-component leading eigenvectors vectors of Table 4

	Control	PFOA	PFDA	Clofibrate	DEHP
Control	0	2.2793	2.7583	1.1820	1.2639
PFOA		0	0.6508	1.0973	1.1020
PFDA			0	1.5997	1.5404
Clofibrate				0	0.4925
DEHP					0

Table 7 Eigenvalues for the Twenty Augmented Adjacency Matrices Belonging to the Most Intensive Protein Spots of Normal Cells Rat Liver Cells and Cells Exposed to Four Chemicals

Eigenvalue	Control	PFOA	PFDA	Clofibrate	DEHP
20	-0.0830	-1.1439	-1.7941	-0.6431	-0.8055
19	0.6326	-0.6731	-0.7304	-0.0715	0.4254
18	0.7767	0.4575	-0.0998	0.3382	0.7267
17	1.4262	0.6936	0.3285	0.6511	1.3338
16	2.2079	0.8606	0.5743	1.5114	1.7088
15	2.3826	1.3571	0.9712	1.7600	2.3483
14	2.5278	1.6779	1.4737	2.0168	2.8501
13	3.0976	2.4354	1.8390	2.4977	3.4341
12	3.1487	2.5309	2.1415	2.5601	3.6500
11	3.4520	2.8756	2.4984	3.0014	4.0585
10	3.5987	3.0562	2.6893	3.3902	4.1709
9	3.9179	3.4333	3.2123	3.4929	4.1991
8	4.1029	3.5074	3.3452	4.0193	4.3161
7	4.6763	4.3467	3.8255	4.2165	4.7448
6	4.8083	4.6362	4.2210	4.3761	5.1266
5	5.2633	4.8190	4.2995	5.2386	5.2619
4	5.5240	5.2513	4.7327	5.6414	6.0093
3	6.3278	6.0646	6.3835	6.5875	6.4306
2	6.4204	6.6249	7.4747	6.7621	7.3740
1	7.7914	8.2840	8.9477	8.0701	8.6109

Table 8 The Similarity/Dissimilarity Table Between the Control Data and Data for the Four Chemicals Based on the eigenvalues

	Control	PFOA	PFDA	Clofibrate	DEHP
Control	0	12.2991	20.1990	8.5783	7.1881
PFOA		0	8.4263	5.8563	14.8792
PFDA			0	12.2635	20.5157
Clofibrate				0	11.1962
DEHP					0

References

- [1] Anderson, N.L., Esquer-Blasco, R., Richardson, F., Foxworthy, P., Eacho, P., 1996. The effect of peroxisome proliferators on protein abundances in mouse liver. *Toxicol. Appl. Pharmacol.* 137, 75-89.
- [2] Randić, M., Witzmann, F.A., Kodali, V., Basak, S.C., 2006. Dependence of a characterization of proteomics maps on the number of protein spots considered. *J. Chem. Inf. Model.* 46, 116-122.
- [3] Randić, M., 2006. Quantitative characterization of proteome: Dependence on the number of proteins considered. *J. Proteome Res.* 5, 1575-1579.
- [4] Randić, M., 2001. On graphical representation of proteomics and their numerical characterization. *J. Chem. Inf. Comput. Sci.* 41, 1330-1338.
- [5] Randić, M., Zupan, J., Novič, M., 2001. On 3-D graphical representation of proteomics maps. *Chem. Inf. Comput. Sci.* 41, 1339-1334.
- [6] Randić, M., Witzmann, F., Vračko, M., Basak, S. C., 2001. On characterization of proteomics maps and chemically induced changes in proteome using matrix invariants: Application to peroxisome proliferators. *Med. Chem. Rev.*, 10, 456-479.
- [7] Randić, M., 2002. A graph theoretical characterization of proteomics maps. *Int. J. Quantum Chem.* 90, 848-858.
- [8] Randić, M., Novič, M., Vračko, M., 2002. On characterization of dose variations of 2-D proteomics maps by matrix invariants. *J. Proteome Res.* 1, 217-226.
- [9] Randić, M., Lerš, N., Vukučević, D., Plavšić, D., Gute, B.D., Basak, S.C., 2005. Canonical labeling of proteome maps. *J. Proteome Res.* 4, 1347-1352.

- [10] Randić, M., Basak, S.C., 2004. On similarity of proteome maps, *Medicinal Chem. Res.* 13, 800-811.
- [11] Randić, M., Basak, S.C., 2002. A comparative study of proteomics maps using graph theoretical biodescriptors. *J. Chem. Inf. Comput. Sci.* 42, 983-993.
- [12] Muirhead, R. F., 1903. Some Methods Applicable to Identities and Inequalities of Symmetric Algebraic Functions of n Letters. *Proc. Edinburgh Math. Soc.* 21, 144-157. [R. F. Muirhead (1860-1941)]
- [13] G. H. Hardy, J.E. Littlewood, G. Pólya. *Inequalities*, second ed., Cambridge University Press, Cambridge, England, 1988, pp. 44-48.
- [14] Randić, M., Wilkins, C.L., 1979. On a graph theoretical basis for ordering of structures. *Chem. Phys. Lett.* 63, 332-336.
- [15] Randić, M., Wilkins, C.L., 1979. Graph theoretical ordering of structures as a basis for systematic searches for regularities in molecular data. *J. Phys. Chem.* 83, 1525-1540.
- [16] Randić, M., Jerman-Blažič, B., Rouvray, D.H., Seybold, P.G., Grossman, S.C., 1987. The search for active substructure in structure-activity studies. *Int. J. Quantum Chem: Quantum Biol. Symp.* 14, 245-260.
- [17] Randić, M., 2000. On characterization of pharmacophore, *Acta Chim. Slovenica* 47, 143-151.
- [18] Klein, D. J.; Brickmann, (guest editors). *Partial Orderings in Chemistry, MATCH, (Mathematical Chemistry Communication)*, 2000, No. 42,
- [19] F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA, 1969.
- [20] Kowalski, B.R., Bender, C.F., 1972. A powerful approach to interpreting chemical data. *J. Am. Chem. Soc.* 94, 5632-5639.

- [21] Randić, M., Orel R., (submitted). On numerical characterization of proteomics maps based on partitioning of 2-D maps into Voronoi regions, *J. Math. Chem.*
- [22] Randić, M., 1991. Novel graph theoretical approach to heteroatom in quantitative Structure-Activity Relationships. *Chemometrics Intel. Lab. Systems* 10, 213-227.
- [23] Randić, M., 1991. On computation of optimal parameters for multivariate analysis of structure-property relationship. *J. Comput. Chem.* 12, 970-980.
- [24] Randić, M., Dobrowolski, J. Cz., 1998. Optimal molecular connectivity descriptors for nitrogen-containing molecules. *Int. J. Quantum Chem.* 70, 1209-1215.
- [25] Randić, M., Mills, D., Basak., S.C., 2000. On use of variable connectivity index for characterization of amino acids. *Int. J. Quantum Chem.* 80,1199-.
- [26] M. Randić, M. Vračko, M. Novič, Eigenvalues as molecular descriptors, in: M.V. Diudea (Ed.), *QSPR/QSAR Studies by Molecular Descriptors*, Nova Science Publishers, Huntington, N. Y. 2001, pp.147-211.
- [27] Lovász, L., J. Pelikán, J., 1973. On the eigenvalues of trees. *Periodica Math. Hung.* 3, 175-182.
- [28] Randić, M., Kleiner, A. F., De Alba, L. M., 1994. Distance/Distance matrices. *J. Chem. Inf. Comput. Sci.* 34, 277-286.
- [29] Randić, M. and Plavšić D., work in progress.
- [30] Randić, M., 1998. On structural ordering and branching of acyclic saturated hydrocarbons. *J. Math. Chem.* 24, 345-358.
- [31] Randić, M., Guo, X., Bobst, S., 2000. Use of path matrices for characterization of molecular structure. *DIMACS Series in Disc. Math. Theor. Comput. Sci.* 51, 305-322.
- [32] Randić, M., Plavšić, D., Razinger, M., 1997. Double invariants. *MATCH – Commun. Math. Comput. Chem.* 35, 243-259.

- [33] Randić, M., Basak, S.C., Plavšić, D., 2004. Novel graphical matrix and distance-based molecular descriptors. *Croat. Chem. Acta* 77, 251-257.
- [34] Randić, M., Estrada, E., 2005. Order from chaos: Observing hormesis at the proteome level. *J. Proteome Res.* 4, 2133-2136.
- [35] Calabrese, E.J., Baldwin, L.A., 2003. Toxicology rethinks its central belief. *Nature* 412, 691-692.
- [36] MATLAB (abbreviation for Matrix Laboratory) is a product of The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760.
- [37] M. Randić, Design of molecules with desired properties. Molecular similarity approach to property optimization, in: M. A. Johnson, G. Maggiora (Eds.), *Concepts and Applications of Molecular Similarity*, Wiley, New York, 1990, pp. 77-145.
- [38] M. Randić, Similarity methods of interest in chemistry. In: S. Kuchanov, (Ed.), *Mathematical Methods in Contemporary Chemistry*, Gordon & Breach Publ. Inc., New York 1996.
- [39] Basak, S.C., Magnusaon, V.R., Niemi, G. J., Regal, R.R., 1988. Determining structural similarity of chemicals using graph-theoretic indices, *Discrete Appl. Math.* 19, 17-44.
- [40] Basak, S.C., Grunwald, G.D., 1994. Use of topological space and property space in selecting structural analogous. *Math. Modelling Sci. Computing* 4, 464-469.
- [41] S. C. Basak, B.D., Gute, G.D., Grunwald, A hierarchical approach to the development of QSAR models using topological, geometrical, and quantum chemical parameters, in: Devillers, J., Balaban, A.T. (Eds.), *Topological Indices and Related Descriptors. QSAR and QSPR*, Gordon and Breach Science Publishers, The Netherlands, 1999, pp. 675-696.
- [42] Basak, S.C., Mills, D., Gute, B.D., Grunwald, G.D., Balaban, A.T., Applications of topological indices in property/bioactivity/toxicity prediction of chemicals, in: D.H.

Rouvray, R.B. King, (Eds.), *Topology in Chemistry: Discrete Mathematics of Molecules*. Horwood, Chichester, UK, 2002, pp. 113–184.

[43] Gute B.D., Basak, S.C., 2001. Molecular similarity-based estimation of properties: A comparison of three structure spaces. *J. Mol. Graphics* 20, 95-109.

[44] Basak, S.C., Grunwald, G.D., Host, G.E., Niemi, G.J., Bradbury, S.P., 1998. A comparative study of molecular similarity, statistical, and neural network methods for modes of action of chemicals. *Environ. Toxicol. Chem.* **17**, 1056–1064.

Accepted manuscript

Figure Captions

Fig. 1 Location of twenty most abundant proteins for proteomics maps of the control group

Fig. 2 Ordering of twenty proteins relative to their mass (left) and their charge (right).

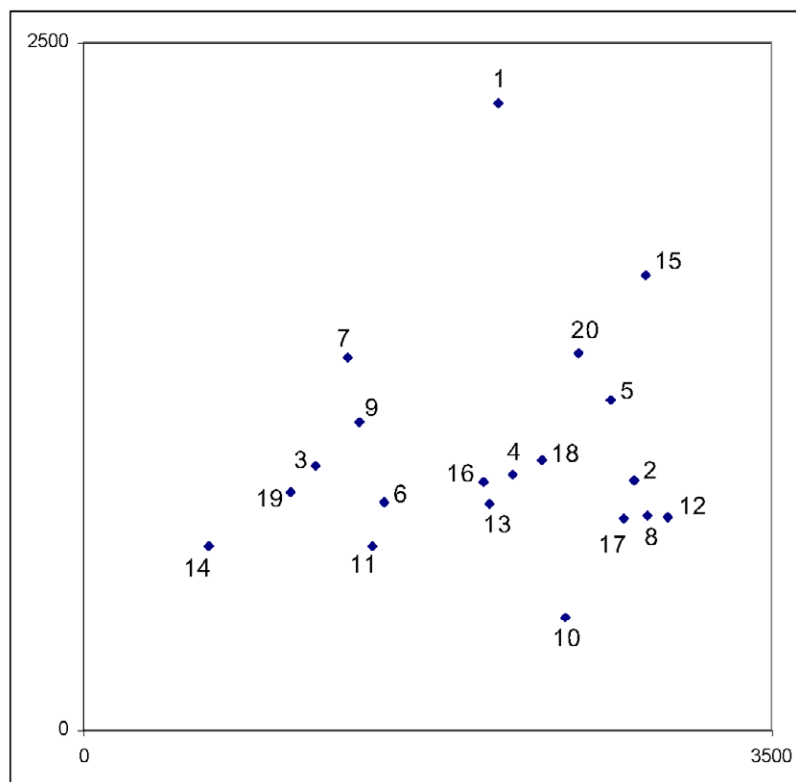
Fig. 3 Partial ordering diagram for twenty proteins of Fig. 1

Fig. 4 The experimental and calculated abundances for twenty most abundant proteins of the control group

Fig. 5 The parallelism between the corresponding similarity/dissimilarity values of Table 6 and Table 8

Accepted manuscript

Fig.1



Accept

Fig.2

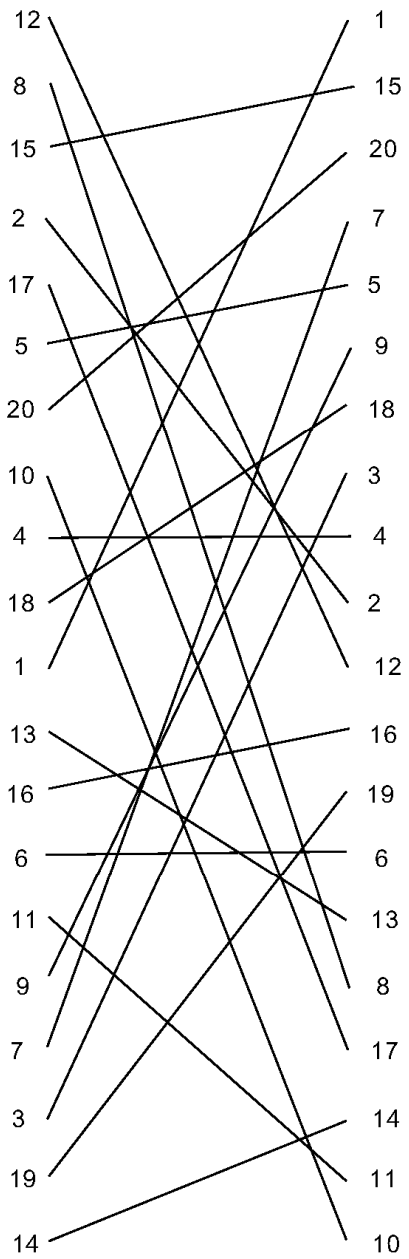
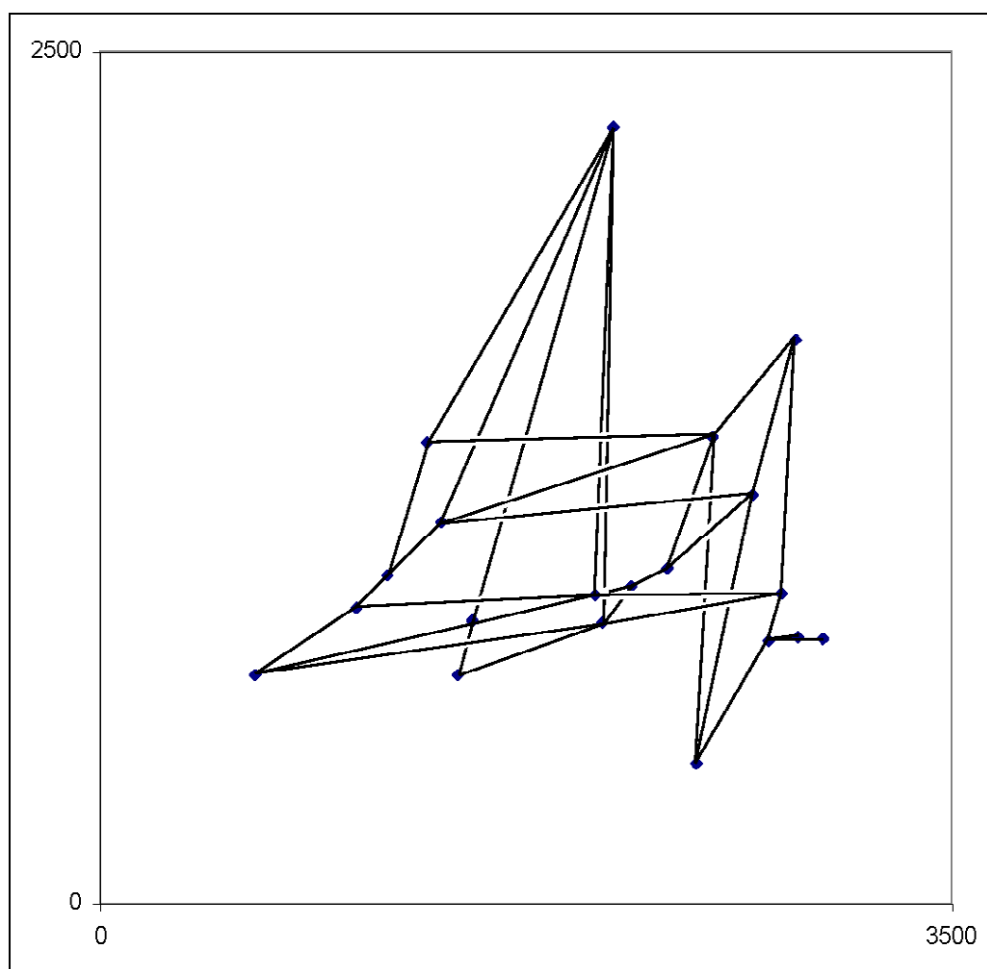


Fig.3



Acc

Fig.4

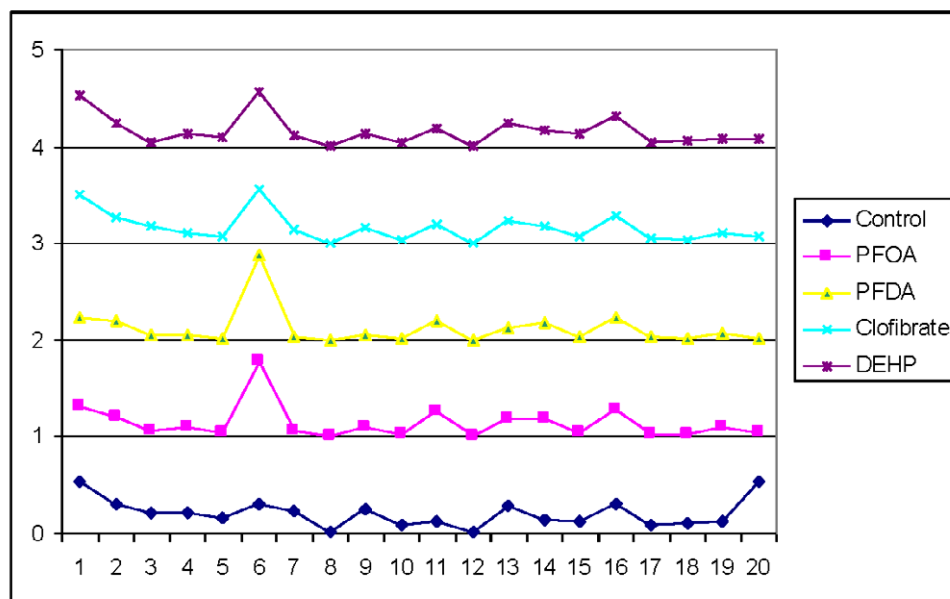
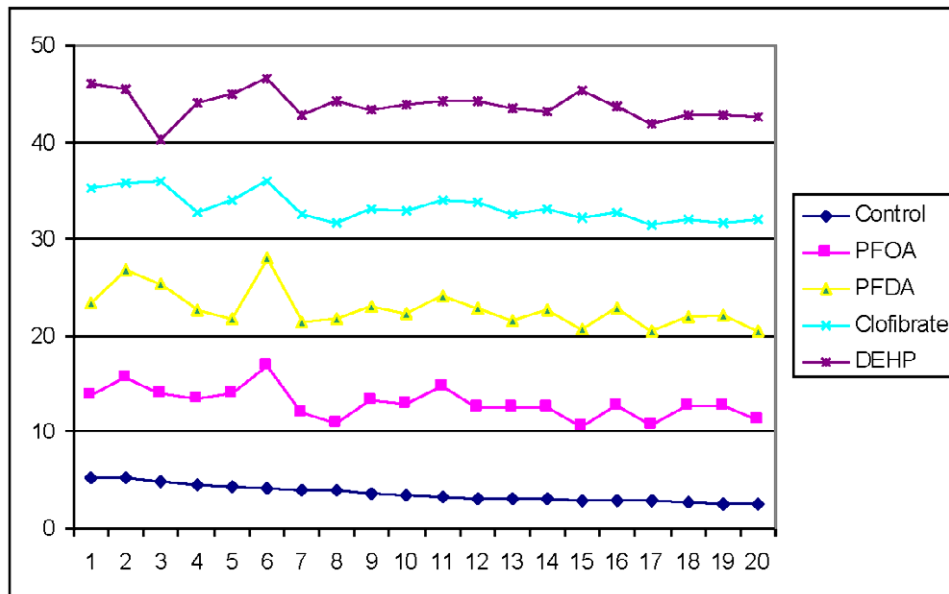
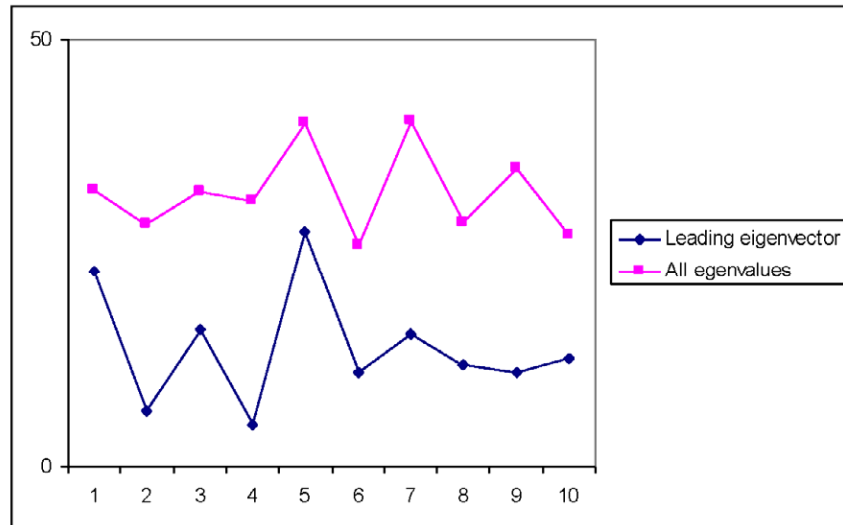


Fig.5



Accepted m.