HAL open science

A new tool to study genetic expression using 2-D electrophoresis data: the functional map concept

Thierry Rabilloud, Pierre Vincens, Philippe Tarroux

To cite this version:

Thierry Rabilloud, Pierre Vincens, Philippe Tarroux. A new tool to study genetic expression using 2-D electrophoresis data: the functional map concept. FEBS Letters, 1985, 189 (2), pp.171-178. 10.1016/0014-5793(85)81018-8 . hal-03739131

HAL Id: hal-03739131

https://hal.science/hal-03739131v1

Submitted on 29 Jul 2022

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.

A new tool to study genetic expression using 2-D electrophoresis data: the functional map concept

T. Rabilloud, P. Vincens and P. Tarroux*

UA 686, Ecole Normale Supérieure, Laboratoire de Zoologie, 46 rue d'Ulm, 75005 Paris, France

Received 20 June 1985, revised version received 18 July 1985

A method derived from general computerized data analysis techniques is used here first to tabulate the characteristics of individual peptide spots observed by two-dimensional electrophoresis and to compare these characteristics among spots. Multivariate analysis of such data arrays then leads to the grouping of spots in an *n*-dimensional space, according to their expression characteristics. Such maps display zones which are characteristic of the tissue studied. These functional maps provide a powerful analytical tool for describing such functional interactions and also for predicting the function of unidentified peptides observed on the electrophoresis gels.

Data analysis *Developmental regulation* *2-D electrophoresis*

1. INTRODUCTION

Analysis of proteins using two-dimensional (2-D) electrophoresis plays a prominent role in the study of gene expression [1–4].

This technique is now the most powerful means of analyzing genetic expression in a cell or in a tissue. Up to several thousand peptides can be observed in a single analysis, and therefore about the same number of gene activities can be monitored together. Some methods have been developed to routinely collect data on posttranslational modifications (e.g. acetylation, phosphorylation) or biochmeical characteristics (associated lipids or polysaccharides). These data, together with the determination of the molecular mass and isoelectric point of each peptide, consist of a set of parameters whose values characterize each spot on a given gel.

However, the enormous quantity of information to be manipulated leads to a self-limitation on the exhaustive use of 2-D electrophoresis in the study of all the peptides that can be observed on a 2-D gel. The use of computer methods is needed to handle such amounts of data. Here we describe a method to extract the significant relationship concerning protein functions and functional interactions from the collected data. The data are thus presented as functional maps.

* To whom correspondence should be addressed

2. MATERIALS AND METHODS

2.1. *Sample preparation*

Tissues were incubated in Landureau medium in the presence of 0.05 mCi [^{35}S]methionine (800 Ci/mmol) for 3 h. After labeling, tissues were rinsed and collected by centrifugation. They were then homogenized in extraction buffer (0.12 M NaCl, 0.5% SDS, 50 mM Tris-Cl, pH 7.5). After centrifugation to remove non-homogenized tissue and cell debris, the supernatant was carefully collected and then precipitated with one volume of cold saturated ammonium sulfate. The precipitate was pelleted and solubilized in O'Farrell extraction medium.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/85/$3.30 © 1985 Federation of European Biochemical Societies

2.2. *Protein electrophoresis*

Electrophoresis analyses were performed as suggested by O'Farrell with minor variations in the geometry of the apparatus and the experimental protocol. The pH gradient in the first dimension ranged from 3.5 to 10 pH units and acrylamide concentration in the second dimension from 7.5 to 14.5%. Isoelectric focusing was performed for 9000 V·h at 400 V and second dimension electrophoresis at 18 mA/gel. After electrophoresis slabs were rinsed in methanol acetic acid, stained with G-250 Coomassie brilliant blue, photographed and then treated for fluorography using the technique described by Laskey and Mills, and that described by Bonner and Laskey. Gels were dried and autoradiographed at 200 K for one week to one month depending on the amount of radioactivity.

2.3. *Computer image treatments*

A computer based 2-D gel image treatment system is used to digitalize and analyze gel images. After a set of preliminary filtering treatments devoted to cleaning of the image and to reduction of the noise, the main shape parameters and the volume of each detected spot is calculated according to the fitting of these parameters to a Gaussian-derived spot model. A spot list is thus established including a reference number, each parameter of the model, the volume and the intensity of each spot.

A set of spot lists corresponding to a series of experiments is matched to determine the homologous spots from one gel to another and a data array is then constituted. It includes the intensity of each spot throughout the series of gels.

2.4. *Data analysis*

We used different approaches to perform data analyses. The main technique we used was the principal component analysis (Kendall et al. [11]). This analysis is a multivariate technique for examining relationships between several variables. It is useful to reduce the number of initial variables to their principal components (a linear combination to the initial variables, the coefficients of-which are the eigenvectors of the correlation or covariance matrix). In our case we used the correlation matrix calculated from the data. A constant feature of this type of data analysis is that it gives the broadest distribution of the objects through the variable space.

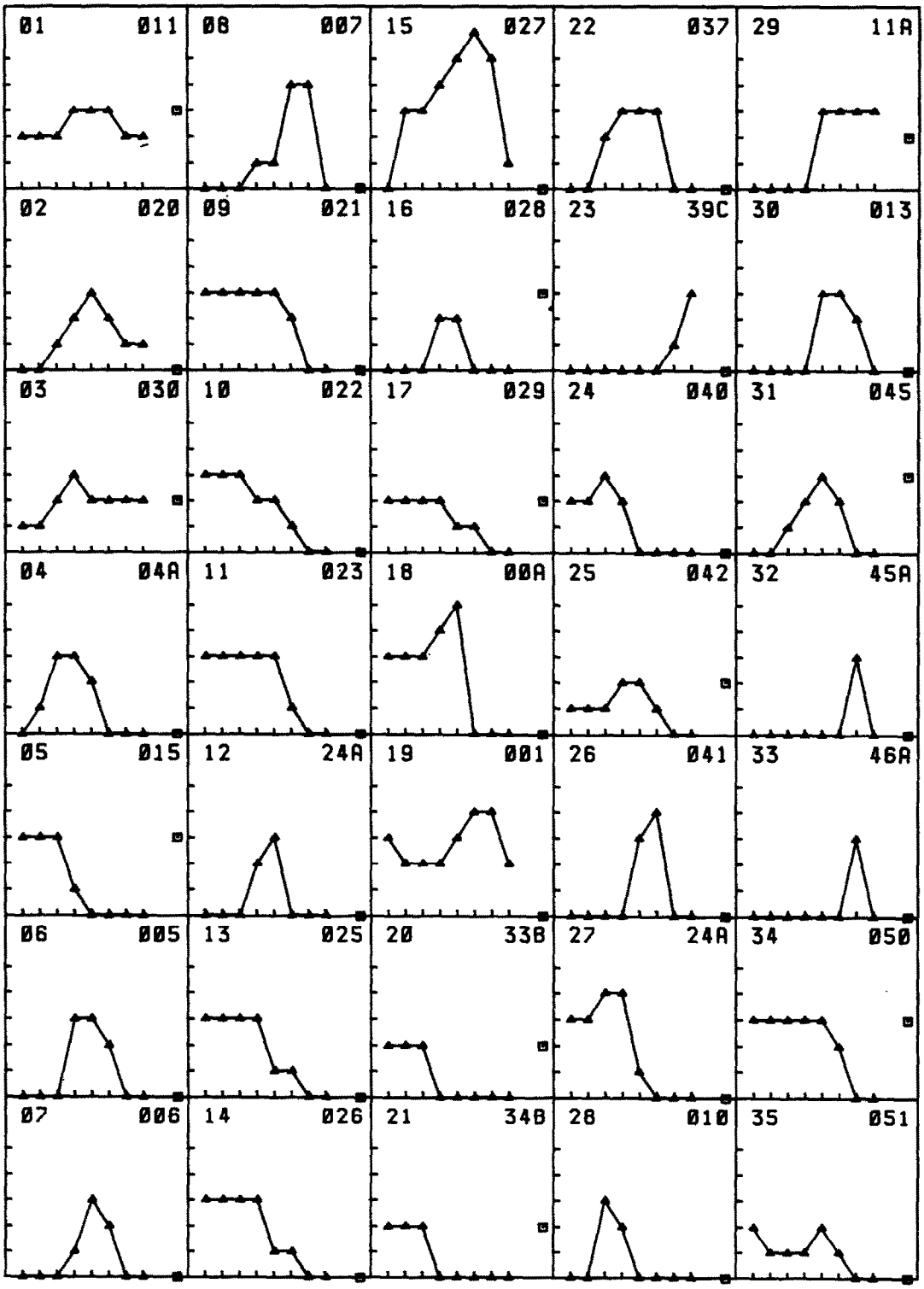
3. RESULTS

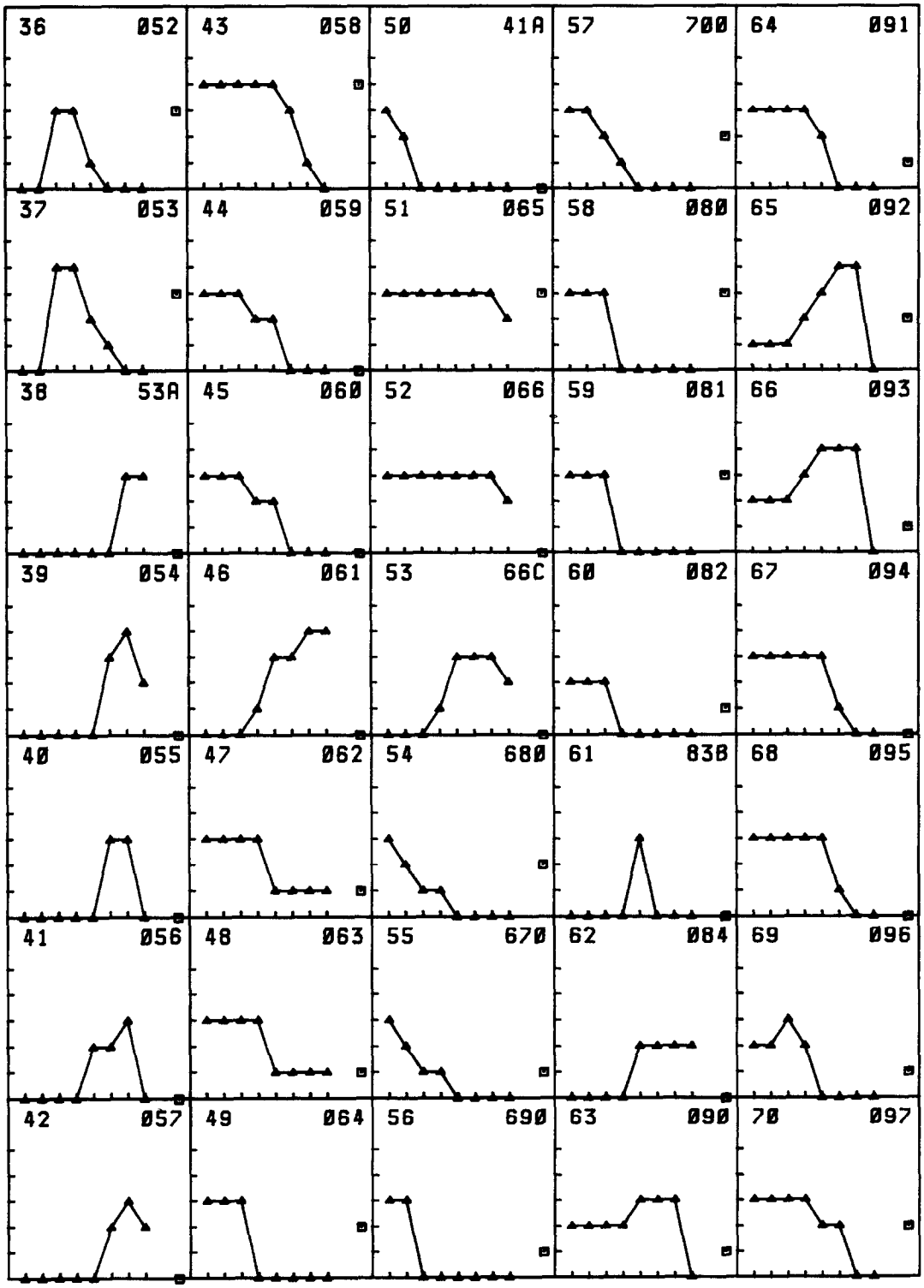
The global approach used here was made possible by the recent development of several techniques and systems concerning the automatic analysis of 2-D gels [1,5–7]. These techniques reduce the information contained in a gel to a spot list which includes spot coordinates, intensities and other physical characteristics such as shape, diffusion, tailing, etc.

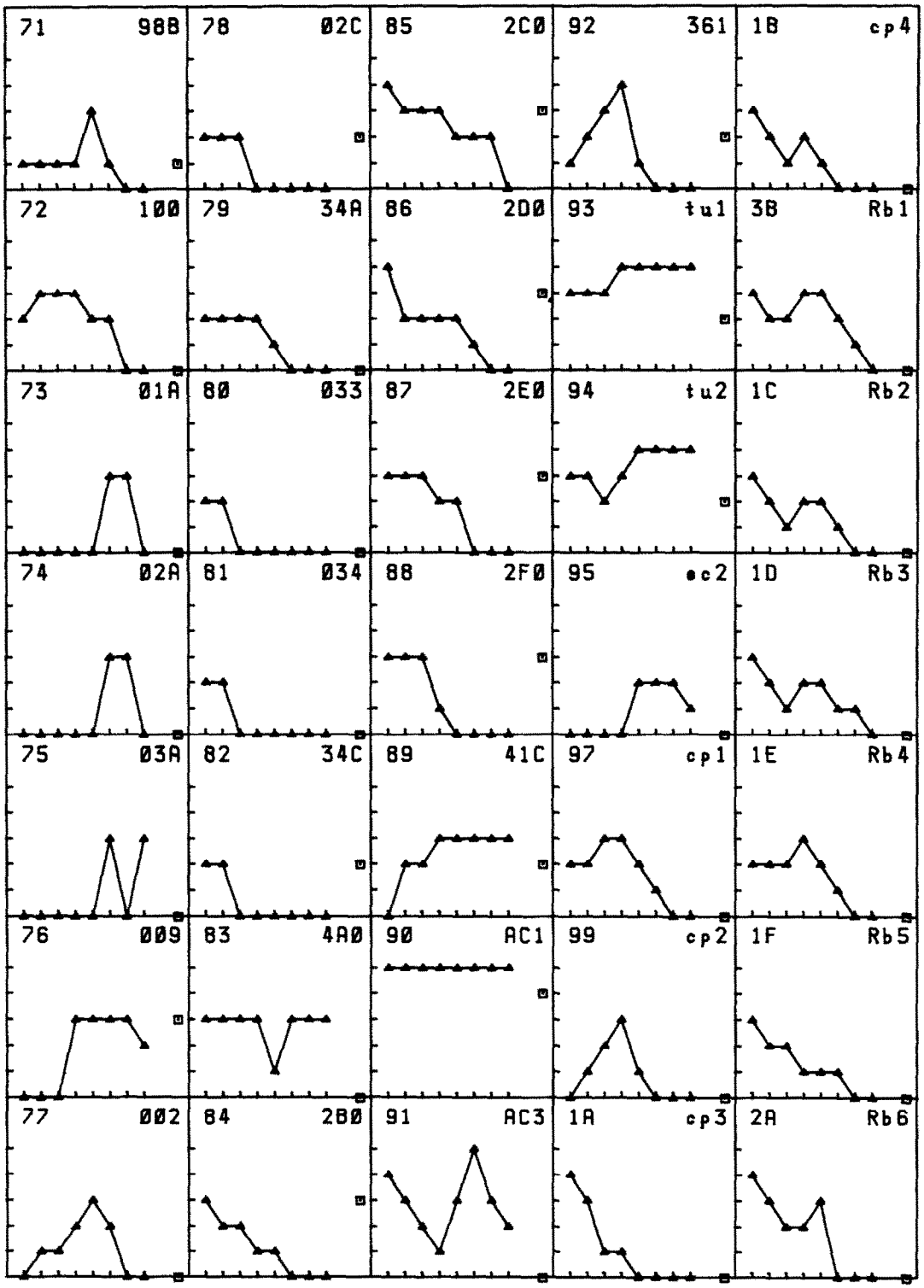
Obviously, the regulation mechanism of gene expression lies in a complicated network of interactions including the expression of numerous polypeptides. In other words, the expression of many genes is interrelated. The methods described here can detect these critical interactions between the expression of several genes in a given tissue. Here we show the method by describing the setting up of such a functional map for a particular case, the terminal differentiation of an insect tissue, the wing imaginal disc of the Lepidopteran *Pieris brassicae*. We performed a set of several electrophoretic analyses of proteins extracted every 24 h when the nymphal development proceeds [8], and we established an equal number of spot lists using our computer graphic system [8]. A data array, constituting a 2-D electrophoresis data base for this special tissue, was then established (fig.1). The corresponding data array consists of a descrip-

→

Fig.1. Data array constituted from 2-D electrophoresis results. The dissected wing imaginal discs of *Pieris brassicae* nymphs were incubated with [^{35}S]methionine (>800 Ci/mmol, Amersham) at 50 μCi/ml. The tissue was homogenized in 50 mM Tris-Cl containing 0.125 M NaCl and 0.5% SDS, and the proteins were precipitated with ammonium sulfate. After analysis by 2-D electrophoresis, the spots were submitted to a semi-quantitative analysis in each gel. The results are reported in this data array. Each protein is referenced by a global number (top left) and a local number (top right) used to reference it in a given gel. This local number is sometimes replaced by a mnemonic (e.g. tu for tubulin, AC for actin, etc.). Each curve displays the evolution of the protein labeling along the developmental stages.


01 011
08 007
15 027
22 037
29 11A
02 020
09 021
16 028
23 39C
30 013
03 030
10 022
17 029
24 040
31 045
04 04A
11 023
18 00A
25 042
32 45A
05 015
12 24A
19 001
26 041
33 46A
06 005
13 025
20 33B
27 24A
34 050
07 006
14 026
21 34B
28 010
35 051


36 052
43 058
50 41A
57 700
64 091
37 053
44 059
51 065
58 080
65 092
38 53A
45 060
52 066
59 081
66 093
39 054
46 061
53 66C
60 082
67 094
40 055
47 062
54 680
61 83B
68 095
41 056
48 063
55 670
62 084
69 096
42 057
49 064
56 690
63 090
70 097


71 98B
78 02C
85 2C0
92 361
1B cp4
72 100
79 34A
86 2D0
93 tu1
3B Rb1
73 01A
80 033
87 2E0
94 tu2
1C Rb2
74 02A
81 034
88 2F0
95 ec2
1D Rb3
75 03A
82 34C
89 41C
97 cp1
1E Rb4
76 009
83 4A0
90 AC1
99 cp2
1F Rb5
77 002
84 2B0
91 AC3
1A cp3
2A Rb6

tion of the proteins in an n-dimensional space, n being the number of different stages used in the analyses. It is then possible to submit these objects (the proteins), described by an n-dimensional set of variables (their expression), to multivariate analysis [9–11]. From these computations, we obtain a linear combination of the n starting coordinates, which leads to the best separation of the different peptides along each variable in the resulting n-dimensional space. At this stage, it is of peculiar interest to understand how this separation is obtained. Each protein differs in the shape of its expression pattern during development. Some are very different (nos 39 and 85 in fig.1), and some are closely related (nos 59 and 60) or even essentially identical (nos 5 and 88). These similarities between synthesis patterns reflect the expression of all the interactions formed during tissue differentiation. Some aspects put the peptides together in the same area of the variable space, while others separate them. In such cases, peptides will be aggregated along one dimension of variable space and separated along another. The result of these computations is shown in fig.2. Only the main planes of the variable space have been plotted. The proteins are separated on these planes as spread clouds, each corresponding to common characteristics for the expression of the enclosed spots. It is of interest to note that this separation reflects functional differences. In fact, we identified some functions of the separated spots. Group 1 of fig.2 contains cuticular peptides, group 2 cytoskeletal proteins (see legend), groups 3a and 3b ribosomal proteins and group 4 ecdysone-induced proteins (EIP). A significant observation is that the projection of each identified functional area retains its cohesion properties on each plane in the variable space. This feature means that each functional zone consists of a compact volume in the n-dimensional space. The juxtaposition of these functional volumes gives a true functional map of the tissue. This approach yields an overall description of the interactions between these functions in a given tissue. For example, EIP seem to have no features in common with cytoskeletal proteins, because the cytoskeleton is a constant element of the studied tissue. Experiments not reported here revealed that ecdysteroids were unable to induce tubulin or actin synthesis in the imaginal disc. On the other hand, cytoskeleton and cuticle areas

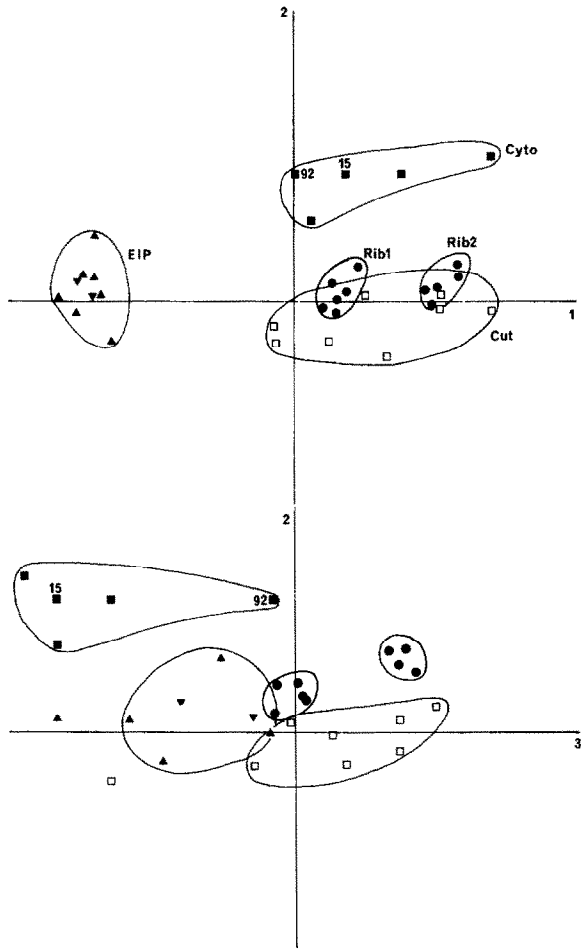


Fig.2. Multivariate analysis of the proteins contained in the data base. The planes constituted by the 1–2 (top) and 2–3 (bottom) axes are represented. Group 1 (cuticle), □; group 2 (cytoskeleton), ■; group 3a and b (ribosome), ● and group 4 (EIP), ▲.

display common features in plane 1–3 (46% of initial information) and plane 1–4 (44%), represented in fig.3.

As we observed elsewhere (data not replotted), it is a characteristic of this tissue that the cytoskeleton plays a prominent rôle in the elaboration of cuticular structures. The interaction between these two clouds is essentially realized through protein 92 (acidic form of actin). This protein, usually associated with the cytoskeleton, seems to be present in cuticular protein prepara-

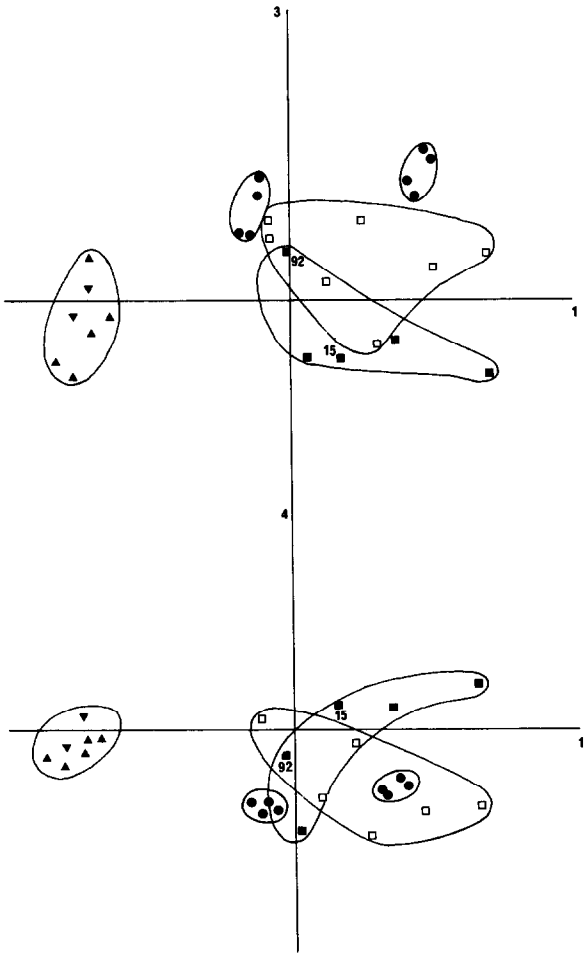


Fig.3. Same as in fig.2, but the 1–3 and 1–4 planes are represented.

tions (not shown) and could represent a cytoskeletal element embedded in cuticular structures. We also observed that protein 15 is always present within the projection of the cytoskeleton volume onto each plane. This observation led us to believe that this protein could be associated in the cell with the cytoskeleton. Further preparation and analysis of the entire cytoskeleton revealed the presence of protein 15 in this set of peptides.

Groups 3a and 3b enclose ribosomal proteins and exhibit a strong association of the aggregated peptides. We cannot explain why ribosomal proteins are split into two groups in this tissue, but this characteristic is constant over all the planes of the variable space. However, the members of each group are different from one plane to another. Ribosomal proteins then, are sorted into the same, or different, groups according to various aspects of their expression patterns. This feature probably hides some characteristics of the functional specialization of ribosomal proteins.

4. DISCUSSION

The approach presented here has a descriptive role and can allow a more global approach both to the cell functions and to the interactions between them. The visualization of functional zones can help to understand how, in a given tissue, the regulation of cell functions is integrated. Such a map achieves a true *n*-dimensional separation in a functional space, adding several facets to the 2-D space of the electrophoretic separation. However, our method also plays a predictive role. In this work, we can quote the discovery of the interaction of actin with cuticular structures and the discovery of protein 15 as a cytoskeletal component. The bifunctional characteristics of ribosomal proteins raise additional interesting questions; the close relationship between peptides found in group 4 allows us to predict that some peptides of this group should be tested for the sensitivity of their genes to ecdysteroids. A promising way consists of the establishment of such functional maps for several tissues to observe whether the relative position of the different functions in variable space varies from one tissue to another. Numerous applications can be derived from this technique, especially using data on posttranslational modifications or physicochemical characteristics. The comparison of functional maps obtained from different cell lines with closely related characteristics may also be very useful.

So, the data analysis allows sorting of some aspects of protein expression that were difficult to establish with the more conventional biochemical approach. Therefore, the usefulness of such functional maps resides in that they highlight certain elements of protein expression, further study of which could be profitably undertaken using more conventional biochemical tools. Such predictive ability derives from the large quantity of information contained in a single 2-D gel and from the maximum exploitation of this information, this

technique being only a tool to handle these data. The fact that functional sorting is achieved through the use of protein expression patterns suggests a relation between the regulation of the protein expression during development and the function of these proteins in the cell. This means that protein expression is coordinated to perform defined functions. However, such maps are only predictive. For example, replacement of one protein by another to perform the same function yields two opposing expression profiles and a separation of the two peptides in the map. But it must be remembered that, if these two proteins perform the same function, they will be subject to the same regulatory processes. If the data array used to perform the analysis covers a sufficiently wide range of conditions, the similarities induced by the common regulation should overcome the initial discrepancy. It appears therefore that the power and accuracy of this approach increase with the quantity of available data, and also with the relevance of the data collected.

ACKNOWLEDGEMENTS

The authors are gratefully indebted to J.A. Meyer for his kindly advice during the elaboration of this paper and to D. Szafarz and A. Ishida for the careful reading of the manuscript.

REFERENCES

[1] Garrels, J.I. (1979) J. Biol. Chem. 254, 7961–7977.
[2] Unterreger, G., Zang, K.D. and Issinger, O.G. (1983) Electrophoresis 4, 303–311.
[3] Lepesant, J.A., Kejzlarova-Lepesant, J. and Garen, A. (1978) Proc. Natl. Acad. Sci. USA 75, 5570–5574.
[4] Greenberg, R.M. and Adler, P.A. (1982) Dev. Biol. 89, 273–286.
[5] Taylor, J., Anderson, N.L. and Anderson, N.G. (1981) in: Electrophoresis '81 (Allen, R.C. and Arnaud, P. eds) pp.383–410, De Gruyter, Berlin.
[6] Scolnick, M.M., Sternberg, S.R. and Neel, J.V. (1982) Clin. Chem. 28, 969–978.
[7] Vo, K.P., Miller, M.J., Geiduscheck, E.P., Nielsen, C., Olson, A. and Xuong, N.H. (1981) Anal. Biochem. 112, 258–291.
[8] Tarroux, P. (1983) Electrophoresis 4, 63–70.
[9] Gnanadesikian, R. (1977) in: Methods for Statistical Data Analysis of Multivariate Observations, J. Wiley and Sons, New York.
[10] Diday, E., Lemaire, J., Pouget, J. and Testu, F. (1982) in: Elements d'Analyse des Données, Dunod, Paris.
[11] Kendall, M. and Stuart, H. (1973) in: The Advanced Theory of Statistics, vol.2, Griffin.