J. theor. Biol. (1992) **154**, 359–369

Methods for the Study of Cellular Sociology: Voronoi Diagrams and Parametrization of the Spatial Relationships

R. MARCELPOIL AND Y. USSON

Equipe de reconnaissance des formes et Microscopie Quantitative, Laboratoire TIM3-IMAG, USR CNRS 690B, Université J. Fourier, CERMO, BP 53X, 38041 *Grenoble Cedex, France*

(*Received on* 11 *March* 1991, *Accepted in revised form on* 12 *September* 1991)

In order to study cellular sociology, a model of parametrization and quantitation of cellular population topographies is developed here. This approach is based on space partition constructed from the set of points locating the position of cells. This spatial partition from Voronoi paving is considered to be a set of individual forms which permits calculation of three parameters which are characteristic of the population topography, (i) RFav, the average roundness factor of those forms, (ii) RFH, a measure of the roundness factor homogeneity and (iii) AD, a measure of their area heterogeneity (also called area disorder). A characterization of the space defined by the three parameters, is obtained by simulation of spatial perturbations of various theoretical populations. These theoretical populations have been subjected to factors such as aggregation and randomization of positions by an increase of spatial degrees of freedom. The use of a diagram involving RFav, RFH and AD turns out to be a powerful tool for the determination of the intrinsic disorder of a cellular population. Furthermore, it makes it possible to determine for a given set of cells, a model including its nearest homogeneous set, and the intrinsic disorder to which it refers. Finally, this model appears to be a useful way to quantify topographies and study order and disorder in many point sets by a simple reading in the parametric space defined by RFav, RFH and AD.

1. Introduction

Numerous studies of considerable scope have been either devoted to cellular interactions or to cellular pattern analysis in order to elucidate the way the cell differentiation acts on cells scattering and clustering or vice versa, to organize the living matter (Welicky & Oster, 1990). Thus, it appears that there is an increasing importance in both general comprehension of mechanisms governing cellular interactions (Viallet *et al.*, in press), and development of therapeutic applications (Tosi *et al.*, 1990), to determine the organization and the kind of relations that cells are able to set up between themselves (Sengel, 1990). A cell is not born in an information free environment, but in a universe of signals sent out by its fellows. Thus, the cell environment has to be considered as an environment organized by elementary exchanges between the population members where communication plays a structural part. Study of the way cells structure their own environment (Honda, 1978, 1983; Zahn, 1971) should

0022-5193/92/030359+11 $03.00/0

therefore increase knowledge of homeostatic tissular controls, and allow us to draw information from spatial alterations seen in pathologies.

The purpose of cellular sociology (Chandebois, 1976) is to explore the relationships existing between cellular functions, and spatial position. Before the widespread use of computers such studies were not feasible because of the vast amount of data involved. Data in the form of point sets, scattered within a region of space, are found in many fields such as biology and crystallography. It is possible to reduce many objects (i.e. cells, proteins) to points (cell gravity center), thus it is possible to treat any such data sets as spatial point patterns. Studies based on statistical analysis or graph constructions have been used to search for parameters describing order and disorder in biological structures (Hopkins & Skellam, 1954; Dussert *et al.*, 1986, 1987; Tezuka *et al.*, 1990). Here we develop a new approach which generates richer information than the above methods . This method integrates the form dimension, and is based on spatial partition from Voronoi paving (i.e. Voronoi diagram). See Appendix for a list of abbreviations used.

2. The Voronoi Diagram (Restriction to 2-D Studies)

This method requires three main steps in its development, (i) a step of construction of the Voronoi diagram which associates a polygonal form to each point of the population, (ii) a step of elimination of points whose associated form has been altered due to the construction of the partition, and (iii) a final step of parametrization and quantitation of topographical informations.

2.1. CONSTRUCTION OF THE VORONOI DIAGRAM

The Voronoi diagram is the space partition containing the most information (Toussaint, 1980). It contains all the information contained in the minimal spanning tree (Zahn, 1971), the relative neighborhood graph, and the Gabriel graph (Toussaint, 1980), that can be constructed on a given poulation. Let us briefly recall the basic definitions of this space partition.

Let S, be a set of N points in the plane, i.e. the nucleus barycenters of cells. For each point p_i in S what is the locus of points (x, y) in the plane that are closer to p_i than to any other point of S? The solution to the above problem is to partition the plane into regions [each region being the locus of points (x, y) closer to a point of S than to any other point of S]. Given two points p_i and p_j, the locus of points closer to p_i than to p_j is the half-plane containing p_i that is defined by the perpendicular bisector of $\overline{p_i p_j}$. Let us denote this half-plane by $H(p_i, p_j)$. The locus of points closer to p_i than to any other point, which we denote by $V(i)$, is the intersection of $N-1$ half-planes, and is a convex polygonal region having no more than $N-1$ sides, that is,

$$V(i) = \bigcap_{i \neq j} H(p_i, p_j).$$

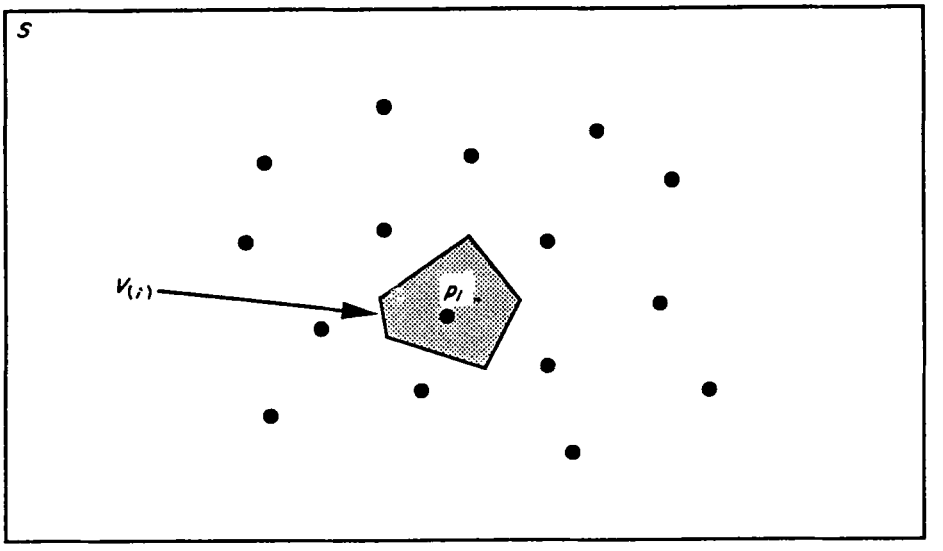

FIG 1. Given a set (S) of a few points, $V(i)$ is the associated Voronoi polygon of p_i (a point of S). The polygon associated to p_i is the locus of points (x, y) closer to p_i than to any other point of S.

Where $\bigcap$ denotes the intersection and $V(i)$ is the Voronoi polygon (Fig. 1) associated with p_i (Preparata & Shamos, 1985). The construction of the Voronoi partition associated to S, denoted Vor (S), follows an incremental method by local modification of the diagram after each insertion of a point of S (Bowyer, 1981).

2.2. ELIMINATION OF MARGINAL POLYGONS

Due to the properties of the Voronoi partition, some polygons of the paving are not statistically representative of the set of polygons. The points of S whose associated polygon is not absolutely representative of the population are considered as marginal. Those polygons are associated to points located on the border of the population, and have one or more summits which do not contain total information on their "surround". Such summits are created by points of S which belong to a half-plane that does not contain this particular summit. Therefore, every point of S, whose associated polygon satisfies one of the two following conditions, is not taken into account in the further calculations.

—The polygon is open (the point belongs to the convex hull),

—at least one of the summits of the polygon is outside the convex hull of S.

As an example (Fig. 2) let us consider a bidimensional Gaussian distribution of points and its associated Voronoi diagram. Fully representative polygons (in white) are conserved for further calculations and marginal polygons (shaded) are eliminated according to the previous rule. For convenience, only $N=500$ points are shown, although actual computation involved $N=2000$ points.

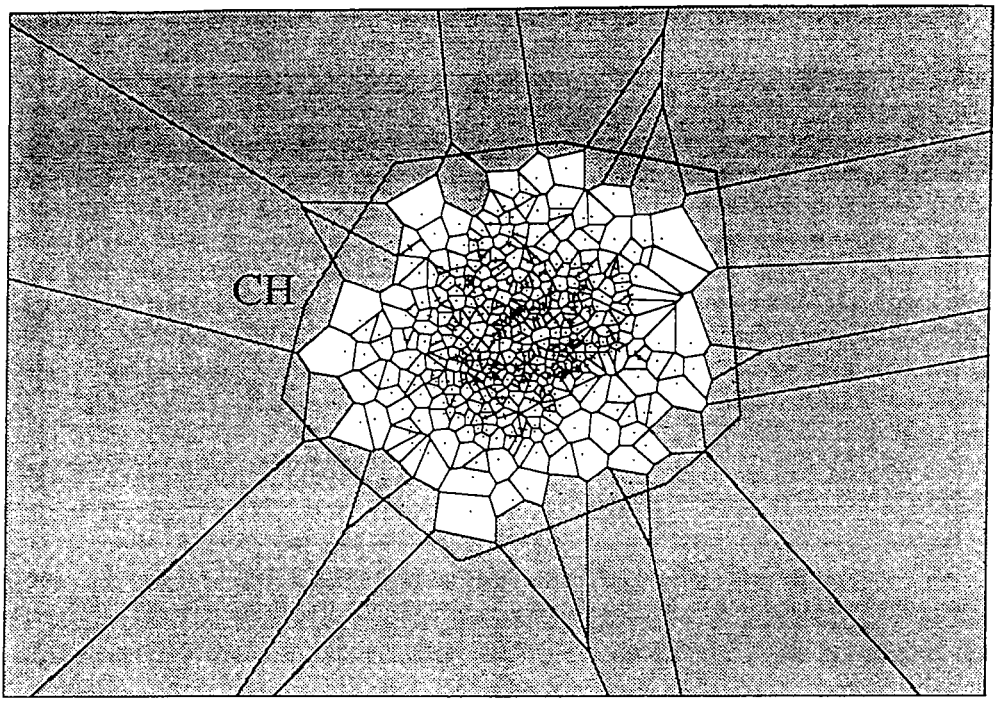

FIG. 2. Set (S) of 500 points of a bi-dimensional Gaussian distribution and its associated Voronoi diagram given that each polygon is a region of the locus of points (x, y) closer to the point of S than to any other point of S. The white polygons are representative polygons, shaded ones are marginal, they intersect the convex hull (CH) of S.

2.3. TOPOGRAPHICAL PARAMETRIZATION AND QUANTITATION

This attempt to quantitate cellular topography is based on the strong relationship which obviously links form to disorder. Therefore, we defined parameters that are descriptors of both form and disorder. These parameters are calculated on the polygonal form which has been associated to each cell of the population during the construction of the Voronoi partition.

Determination of the average type of the spatial occupation: for a convex set, X, where $A(X)$ is the area and $L(X)$ is the perimeter, it is possible to demonstrate the following isoperimetric inequality:

$$L(X)^2 - 4\pi A(X) \geq 0$$

Since Voronoi polygons are convex, the average type of S population spatial occupation is well-characterized by the average roundness factor (RFav)

$$\mathrm{RFav} = \frac{1}{N} \sum_{i=1}^{N} \frac{4\pi A(X_i)}{L(X_i)^2} \qquad (0 < \mathrm{RFav} \leq 1).$$

Example: the roundness factor [$RF = 4\pi A(X)/L(X)^2$] of a n-sided Reuleaux polygon (Regular polygon with all angles and sides equals) is $RF(n) = \pi/[n \tan(\pi/n)]$. Thus this formula makes it possible to predict the RF value of simulated polygons. The RF of a circle is 1, the RF of a line is 0.

The intrinsic disorder of the population has been expressed as two prime factors, the disorder concerning the area heterogeneity of the population and the disorder of the geometrical properties.

Determination of area heterogeneity and geometrical homogeneity of the spatial occupation inside S can be quantified by the following two parameters, area disorder AD, and roundness factor homogeneity RFH.

$$AD = 1 - \left(1 + \frac{\sigma_A}{A_{av}}\right)^{-1} \qquad RFH = \left(1 + \frac{\sigma_{RF}}{RF_{av}}\right)^{-1}$$

where σA denotes the area standard deviation, σRF the roundness factor standard deviation, Aav and RFav the mean area and the mean roundness factor.

Using two types of invariant, geometrical one and area one, it can be demonstrated that RFav, AD and RFH are uncorrelated (Marcelpoil *et al.*, 1991). These three parameters are defined on the [0, 1] interval, which makes it easier to give a graphic representation of the values for a given population. Thus RFav, RFH and AD define a three-axis graph which represents topographical informations. We tested this model with different theoretical populations.

3. The RFav vs. RFH vs. AD Diagram

The RFav vs. RFH vs. AD diagram is a three-axis graph which represents topographical informations. The three axis are defined on the [0,1] interval. Because of the intrinsic normalization process of RFH and AD, all distributions can be plotted in the RFav, RFH and AD space and easily compared with topographically well characterized populations.

For this purpose, the following simulations were made to characterize the RFav vs. RFH vs. AD diagram. Let us consider a perfect lattice with points ordered at the nodes. We consider disorder as the ability for each point to move for a maximum distance (l) in any direction of the plane from its original node. The disorder value is the diameter of the circle centered on the original node, in which the point is uniformly randomly located. This disorder is the percentage of the original length of the lattice mesh. When l equals zero [Fig. 3(a)], the set of points is perfectly ordered. When l increases, the population has an increasingly intrinsic disorder [Fig. 3(b) and (c)] and becomes randomly distributed for very large values of l [Fig. 3(d)].

Several populations corresponding to different lattices have been randomized from the hexagonal lattice, through a square to a triangular lattice (i.e. cells of a corneal epithelium), by small discrete changes. For each of these populations, AD varies when the value of the intrinsic disorder increases (Fig. 4). Note that a given value of AD corresponds to a given value of intrinsic disorder for any original population.

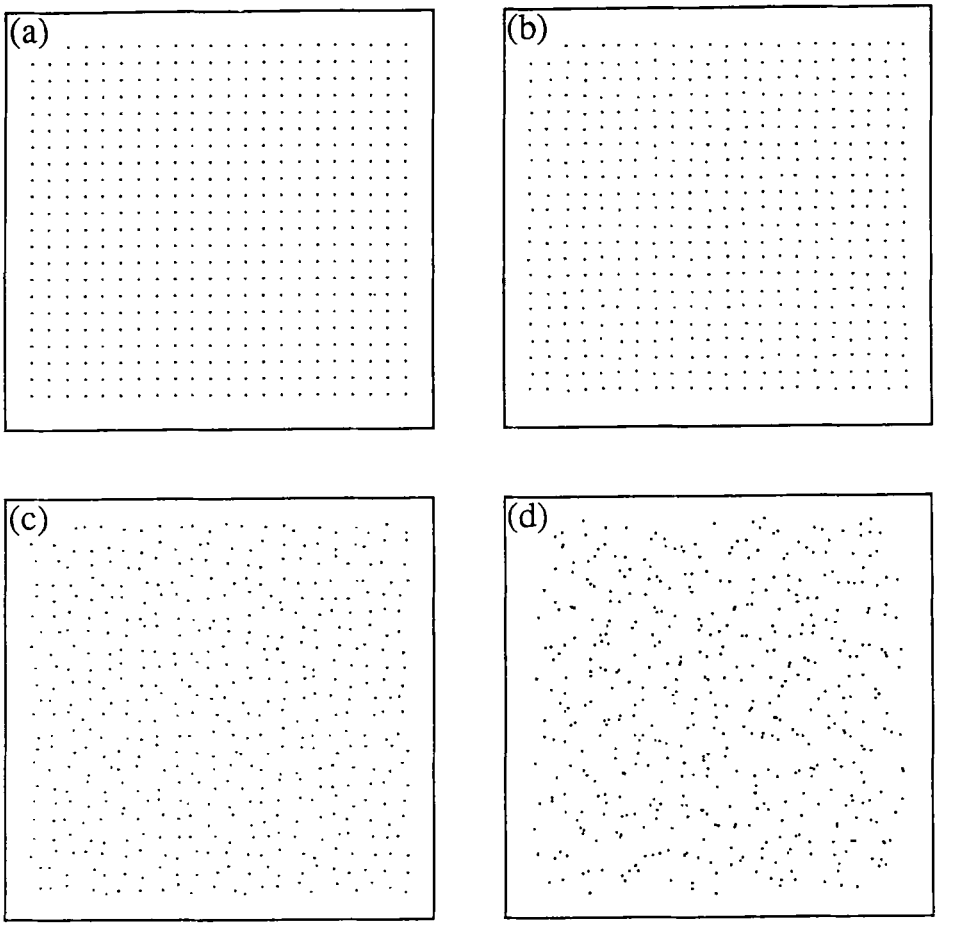

FIG. 3. (a) Set of 484 points ordered at the nodes of a perfectly square lattice. (b), (c) and (d) randomization of the arrangement of (a) giving each point a new position by its ability to move for a maximum distance ($l/2$) in any direction of the plane from its original mode. Values of l are respectively (a) 0% of the original mesh length, (b) 25%, (c) 75% and (d) 200% which tends to a uniform random distribution.

For example, let us consider a stratified epithelium. The stratum of cells next to the basal lamina are well-ordered whereas the location of these cells is randomized when they reach the apical face. With our method it becomes possible to measure and express the type of topographical variations and amount of disorder that were introduced between two successive strata during the migration process of the epithelial cells.

Whatever the population and the intrinsic disorder we consider, the AD vs. RFav diagram permits to find the initial homogeneous set and the value of the intrinsic disorder again. When the population does not have more intrinsic geometrical disorder than area heterogeneities, a RFav vs. AD or RFav vs. RFH diagram permits

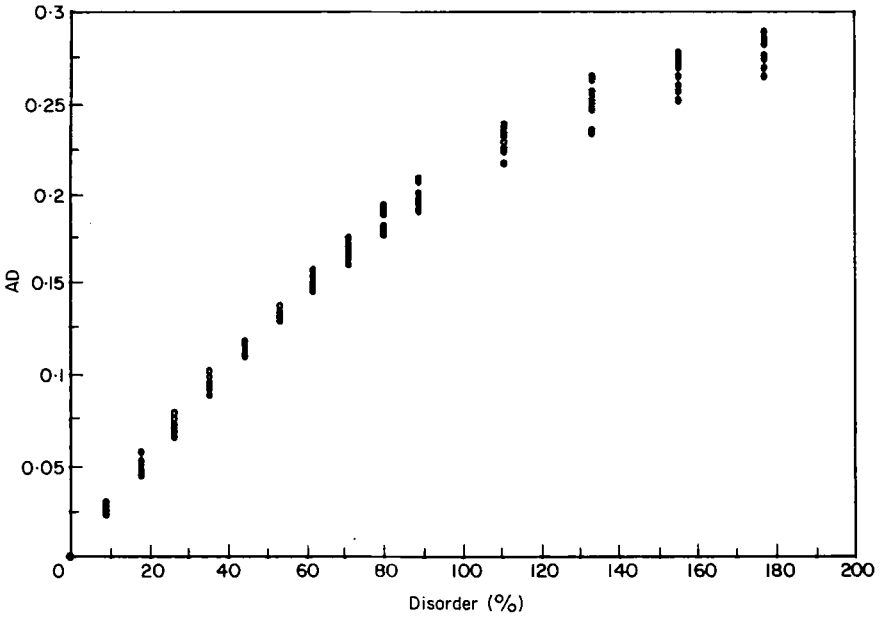

FIG. 4. Values of AD for several population corresponding to different lattices when the intrinsic disorder increases for each population. The disorder value is the diameter of the circle centered on the original node in which the point is uniformly random located. The disorder values are given in percentages of the original mesh length.

the characterization of the population topography. In fact the RFav vs. RFH vs. AD diagram is useful to quantify topography when the population has either a geometrical gradient or an area one.

Figure 5 gives the values of RFav and AD for the same populations as in Fig. 4, from AD = 0 (perfect arrangement) to AD $\approx$ 0·28 (random distribution). Note also that the path leading to disorder varies with the original homogeneous population, but the AD value remains relevant to the intrinsic disorder and so allows the topography to be well-characterized. Indeed, to characterize a population topography it is sufficient to quantitate its intrinsic disorder (simple reading of the area and geometrical disorder) and to define the analogous homogeneous population [following back to order (AD = 0 RFH = 1) the path leading to disorder].

Figure 6 shows the RFav vs. RFH vs.AD diagram and the corresponding values, RFav, RFH and AD of the previous populations corresponding to an hexagonal lattice, a square lattice and a triangular one. In particular, three specific planes and two vectors can be defined in this diagram to help a visual reading of topographies. The planes marked I, II and III correspond respectively to (i) population whose area and geometric disorder are equal, (ii) a population whose points have the same associated area but have geometrical differences, (iii) a population whose points have the same geometrical definition of the associated polygon but have different areas. The two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ correspond to an aggregation and a repulsion vector. They

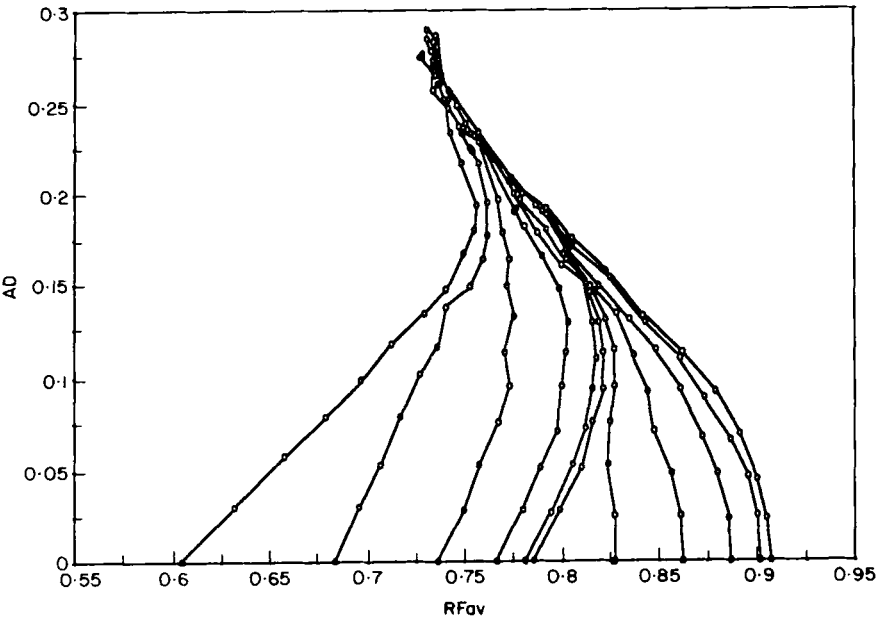

FIG. 5. The points (O) correspond to the values of RF and AD for several theoretical populations. (——) Values of RF and AD for a given population when the intrinsic disorder increases. A given value of intrinsic disorder corresponds to a given value of AD. For a given value of AD, the leftmost population is derived from an hexagonal lattice, the rightmost from a triangular lattice. All intermediary populations are obtained by discrete distortions of the hexagonal mesh, through a square, to a triangle.

lead respectively (i) to populations containing aggregates in particular locations of the space, (ii) to populations containing barren islets. As an example, let us consider (Fig. 6) the uniformly randomly distributed population (UPop), and the Gaussian population (GPop). It appears that Upop (e.g. a swarm of myxobacteria in a nutritious medium) is next to the plane marked I (its geometrical disorder is rather the same as its area disorder) whereas GPop (e.g. a swarm of myxobacteria aggregating and forming a fruiting body in response to starvation) appears *u* translated, due to the aggregation in a particular point of the space.

4. Conclusion and Perspective

The method presented here will be tested with respect to its effectiveness to analyse and quantify cellular topographies. The Voronoi diagram of a set of points, S, contains all the information contained in the minimal spanning tree, the relative neighborhood graph, and the Gabriel graph, that can be constructed on it (Toussaint, 1980). In addition to other methods based on a subgraph of this diagram (Dussert, 1986), the amount of information it contains makes it possible to decompose the intrinsic disorder of the population into two prime factors: the disorder of the area heterogeneity and the disorder of the geometrical properties. Generally, disorder is

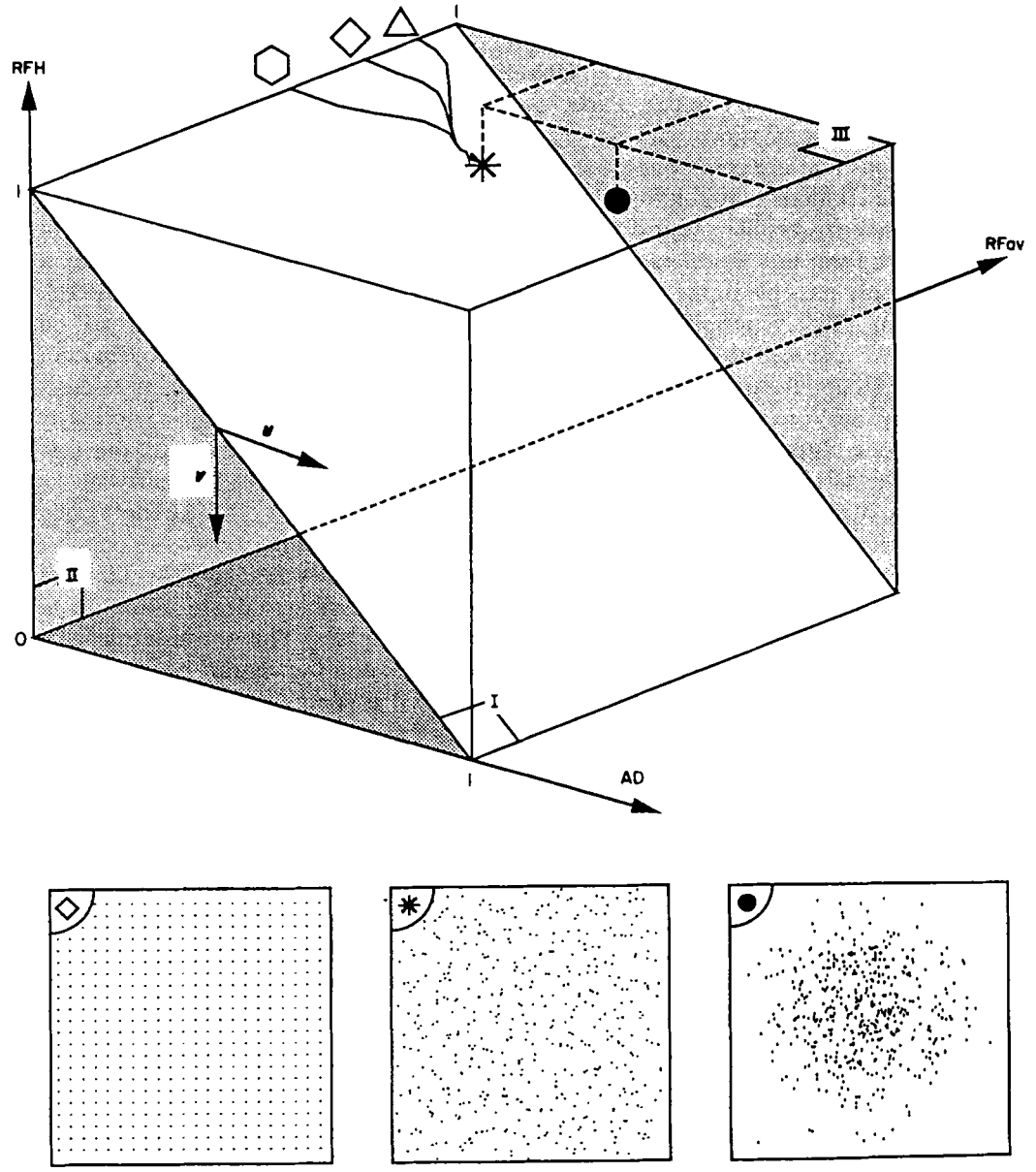

FIG. 6. RFav vs. RFH vs. AD diagram. Paths of RFav and AD of the previous populations (corresponding to an hexagonal lattice, a square lattice and a triangular one) when the intrinsic disorder increases are represented. Each location of this parametric space makes it possible to define the corresponding population. Three characteristic planes can be defined: plane I, in which the geometrical disorder equals the area disorder, plane II of geometrical invariants, and plane III of area invariants. The two vectors u and v correspond to an aggregation and a repulsion vector respectively. The location (*) correspond to the values of RFav, RFH and AD of a uniformly randomly distributed population, the location (●) corresponds to a Gaussian population and the location (◇) corresponds to an ordered population derived from a square lattice. Examples of topographies corresponding to different locations of this diagram are drawn in the three frames.

represented in one-dimension (1-D). Thus, the use of a 2-D representation of disorder (HRF vs. AD) makes now possible to represent population topography in three dimensions (RFav, HRF and AD). Quantitation of topography thus reaches accuracy and discriminatory power. Particularly, it makes it possible to distinguish very well between populations with aggregates and populations with barren islets in it, populations that are common in cell biology. Recent theories about chaos underline the subtle relations that link form to disorder (Herrmann, 1986). This new approach to characterize cellular topography uses this relation. The human approach to describe a form usually refers to a geometrical form. Therefore, the description of a form is only possible in the absolute, if this form is perfectly geometrical (Coster & Chermant, 1985). Since all forms in a Voronoi diagram are perfectly geometrical and convex, we assumed that the Voronoi diagram of a given set of cells should permit the analysis and the quantitation of its topography. The different parameters calculated from the Voronoi diagram satisfy the Hadwiger conditions of their continuity and scale, and translation and rotation invariance. Those conditions guarantee that the mathematical significance of the different parameters is not non-sensical. In addition, a physical meaning can be given to these parameters, thus strengthening the robustness of the method. For example, for given sets of cells, it becomes possible to determine a model including the nearest homogeneous set, and the inherent intrinsic disorder. Hence, this method should make it possible to describe and quantify pathological diseases that affect cellular topographies. This intrinsic disorder describes the difference between the given set of cells and the nearest homogeneous set. In physical terms, this difference can be expressed as energy. It means, the nearest homogeneous set is such that the theoretical work necessary to transform the original set in the final homogeneous set is the smaller one. For this purpose, accurate simulations to well-characterize topographies that correspond to given co-ordinates of the RF vs. RFH vs. AD diagram are currently under study. This should make it possible to determine a model of any cellular population by a simple comparison of the RFav, RFH and AD values with well-characterized populations. Thus, this method is very promising to determine an objective grading of tumors by measuring the amount of perturbations compared with normal population. This will permit the realization of planar cellular sociology. Since living systems are 3-D, we are currently extending our 2-D model to 3-D (Bertin, 1990).

REFERENCES

BERTIN, E. (1990). *Diagramme de Voronoi 3D*. Grenoble: DEA de Mathématiques Appliquées. Université Joseph Fourier.

BOWYER, A. (1981). Computing Dirichlet tesselation. *Comput. J.* **24,** 162-166.

CHANDEBOIS, R. (1976). Cell sociology: a way of reconsidering the current concepts of morphogenesis. *Acta Bioth. (Nederlands)* **25,** 71-102.

COSTER, M. & CHERMANT, J. L. (1985). Précis d'analyse d'images. *Editions CNRS*.

DUSSERT, C., RASIGNI, G., RASIGNI, M. & PALMARI, J. (1986). Minimal spanning tree: a new approach for studying order and disorder. *Phys. Rev. B.* **34,** 3528-3531.

DUSSERT, C., RASIGNI, M., PALMARI, J., RASIGNI, G., LLEBARIA, A. & MARTY, A. (1987). Minimal spanning tree analysis of biological structures. *J. theor. Biol.* **125,** 317-323.

HERRMANN, H. J. (1986). Growth: an introduction. In: *On Growth and Form, Fractal and Non-Fractal Patterns in Physics* (Stanley & Ostrowsky, eds) pp. 3-20. NIJHOFF.

HONDA, H. (1978). Description of cellular patterns by Dirichlet domains: the two dimensional case. *J. theor. Biol.* **72,** 523-543.

HONDA, H. (1983). Geometrical models for cells in tissues. *Int. Rev. Cytol.* **81,** 191-248.

HOPKINS, B. & SKELLAM, J. G. (1954). A new method for determining the type of distribution of plant individuals. *Ann. Bot.* **18,** 213-227.

MARCELPOIL, R., USSON, Y. & CHASSERY, J. M. (1991). Segmentation morphologique includant des parametres d'ordre et désordre: quantification par diagramme de voronoi et application a la sociologie cellulaire. In: *Proceedings 8th Congres RFIA*, pp. 967-972. Lyon: AFCET.

PREPARATA, F. P. & SHAMOS, M. I. (1985). *Computational Geometry.* Berlin: Springer-Verlag.

SENGEL, P. (1990). Pattern formation in skin development. *Int. J. Dev. Biol.* **34,** 33-50.

TEZUKA, F., SATO, I., HIGASHIIWAI, H., ENDO, N., ITO, K. & KASAI, M. (1990). Method for the quantitative evaluation of the distribution pattern of nuclei in normal and malignant endometrial epithelia. *Analyt. quant. Cytol. Histol.* **12,** 237-241.

TOSI, P., FILIPE, M. I., BAAK, J. P. A., LUZI, P., SANTOPIETRO, R., MIRACCO, C., SFORZA, V. & MEGHA, T. (1990). Morphometric definition and grading of gastric intestinal metaplasia. *J. Pathol.* **161,** 201-208.

TOUSSAINT, G. T. (1980). Pattern recognition and geometrical complexity. In: *Proceedings 5th International Conference on Pattern Recognition*, pp. 1324-1347. Miami Beach: IEEE Catalog No. 80CH1499-3.

VIALLET, J. P., RUBERTE, E., DU MANOIR, S., KRUST, A., ZELENT, A. & DHOUAILLY, D. (in press). Retinoic acid-induced glandular metaplasia in mouse skin is linked to the dermal expression of retinoic acid receptor βmRNA. *Dev. Biol.*

WELIKY, M. & OSTER, G. (1990). The mechanical basis of cell rearrangement. *Development* **109,** 373-388.

ZAHN, C. T. (1971). Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Trans. Comput.* **20,** 68-86.

APPENDIX

List of Abbreviations

A	area
AD	area disorder
CH	convex hull
H	half-plane
l	distance
L	perimeter
RF	roundness factor
RFav	average roundness factor
RFH	roundness factor homogeneity
S	set of points
$\boldsymbol{u}$	aggregation vector
$\boldsymbol{v}$	repulsion vector