Field approach to three-dimensional gene expression pattern characterization

L. da F. Costa, a) B. A. N. Travençolo, b) and A. Azredo
Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São Carlense 400, Caixa Postal 369, CEP 13560-970, São Carlos, SP, Brazil

M. E. Beletti
Instituto de Ciências Biomédicas, Universidade Federal de Uberlândia, Av. Pará 1720, CEP 38400-902, Uberlândia, MG, Brazil

G. B. Müller
Department of Theoretical Biology, University of Vienna, University of Vienna, Althausstrasse 14, A-1090 Vienna Austria

D. Rasskin-Gutman, G. Sternik, M. Ibañes, and J. C. Izpisúa-Belmonte
Gene Expression Laboratory, The Salk Institute, 10010N Torrey Pines Road, La Jolla, California 92037

(Received 1 November 2004; accepted 21 February 2005; published online 31 March 2005)

We present a vector field method for obtaining the spatial organization of three-dimensional patterns of gene expression based on gradients and lines of force obtained by numerical integration. The convergence of these lines of force in local maxima are centers of gene expression, providing a natural and powerful framework to characterize the organization and dynamics of biological structures. We apply this methodology to analyze the expression pattern of the enhanced green fluorescent protein (EGFP) driven by the promoter of light chain myosin II during zebrafish heart formation. © 2005 American Institute of Physics. [DOI: 10.1063/1.1898424]

During the development of animal organs, several processes take place such as gene activation and repression, different patterns of cell behavior, tissue inductions, and other morphogenetic events, modulated by the environment.1,2 Far from being uniform, very specific patterns of gene expression generate structured temporal and spatial patterns of varying protein concentrations. Recent advances in biochemical and imaging methods have paved the way to obtaining three-dimensional (3D) reconstructions of spatial gene activation3 which can be analyzed in order to better understand the intricate mechanisms governing tissue and organ formation.2 Among the several currently available methodologies allowing characterization of 3D gene expression, special attention has been given to enhanced green fluorescence protein (EGFP). EGFP is used as a fluorescent marker, whose expression is controlled by the promoter of the protein of interest. Thus, the same cells that express the protein of interest also express, synchronously, EGFP. This methodology can be used to demonstrate gene activity in intact cells and organisms, while taking into account the fact that the host protein is continuously synthesized, degraded, and suffering alterations within cells.3,5 As such a type of gene expression data become available, it is important to identify and develop mathematical methodologies for measuring and modeling spatial gene activation. In addition to traditional approaches (e.g., density or dispersion estimation), it is important to consider more sophisticated methods capable of addressing more directly aspects related to the dynamics of the involved biological processes, related to specific patterns of cell behavior,6,7 which play an important role during both embryonic development and pathological processes.

In this letter we characterize the spatial organization of gene expression patterns in order to assess the geometrical basis of some dynamical processes during morphogenesis. To this end, we compute a “gene expression landscape” as a scalar field ω=g(x,y,z), where ω is interpreted as the amount of expression of the protein in the spatial position (x,y,z). The same approach can be used to model and predict the dissemination of cell signaling or other factors emanating from the cell under analysis which, combined with the possibility of adopting varying values of the parameters affecting the field (e.g., the dielectric constant), defines a truly general framework for expressing field influences. In analogy with the potential dynamics of dissipative systems, we obtain the spatial trajectories (lines of force) corresponding to maximizing the gradient of gene expression. Such trajectories tend to converge to local peaks of activity, defining gene expression centers. It is proposed in this letter that the distribution of such centers provide a natural framework for characterizing and analyzing the spatial interactions between the involved developmental rudiments. The potential of such a methodology is illustrated with respect to the analysis of zebrafish heart formation from 3D gene expression data.

Zebrafish embryos have been widely used in order to study heart formation, due to their transparency and the partial independence of the cardiovascular system. Several other advantages of the zebrafish embryo for genetic studies include the ability to carry out pre- and postmeiotic mutagenesis experiments, the possibility to screen for mutants in the haploid and diploid progeny and, most importantly for this study, their amenability for live confocal microscopy.8 In vertebrates, the heart is the first organ that forms and starts operating.8 Constrictions and bending (folding) are key ele-

a) Electronic mail: luciano@if.sc.usp.br
b) Electronic mail: bant@if.sc.usp.br
ments in the early morphogenetic shaping of the heart tube. The spatial gene expression data considered in this work were acquired through the observation of 42-h post-fertilization transgenic zebrafish embryos expressing EGFP specific for heart mesoderm myosin light chain (*mlc2a; cmlc2, mylec2a - Zebrafish Information Network).*9 The zebrafish embryos were anesthetized and immobilized in agarose, and live images of the heart were taken at ambient room temperature. The image recordings were made using a Nikon Eclipse TE300 inverted microscope using 20×/0.75 NA magnification. The microscope is coupled to a Bio-Rad Radiation MP 2100 scanning multiphoton confocal system (Cambridge, MA) with a two-photon Tsunami laser (Spectra Physics, CA). The GFP was excited with the two-photon laser, at 900 nm. Here we analyze one specimen for which the total dataset is composed of 110 confocal sections every 1.15 μm. Each section is represented by a grayscale image with gray values ranging from 0 to 255.

All the 110 confocal slices were combined so as to obtain the three-dimensional volume of the heart, from which the gene expression landscape was computed as described above. It is interesting to note that this scalar field could be visualized with direct volume rendering algorithms (DVR).10 In order to minimize the spatial quantization noise implied by digital image representation, Gaussian smoothing was applied over the gene concentration data. This is done through the discrete convolution of a three-dimensional Gaussian kernel $k(x, y, z)$ with the scalar field $w$, as expressed in Eq. (1), where $1 \leq i, j, k \leq L$ for a cubic workspace of side $L$.

$$w(x, y, z) * k(x, y, z) = \sum_{i,j,k} w(i, j, k) k[(x - i), (y - j), (z - k)].$$  \hspace{1cm} (1)

The smoothed reconstruction of the 3D gene activity pattern is shown in Fig. 1(a). The gradient of this scalar field was estimated by using the enhanced finite differences scheme described in Ref. 11, by convolving the gene expression concentration with three-dimensional masks. Next, we compute the lines of force by calculating the trajectories that maximize the gradient starting from arbitrary spatial positions sampled as points uniformly distributed through spheres centered at the three-dimensional volume.

The considered lines of force would correspond, for instance, to the putative path (set of 3D coordinates) followed by an object at position $r=(x, y, z)$ with gradient dissipative dynamics:

$$\frac{\partial r}{\partial t} = \nabla [\omega(x, y, z) * k(x, y, z)],$$  \hspace{1cm} (2)

standard numerical integration was used in order to estimate such lines of force, which are illustrated in Fig. 1(b). The sampling criteria removed the lines whose scalar value of its end point were less than 10 (from a range of 0–255), eliminating those that do not reach the regions where *mlc2a* was being expressed. Small and too long trajectories were also removed, because they were influenced by noises. As expected, these lines converge to local maxima of the scalar gene expression field, which could be considered as gene expression centers. In analogy to graph theory, the total number of sampled lines of force converging to a specific center is referred to as the center degree. A total of 734 lines and 89 centers were obtained for the considered 3D gene expression data.

Figure 1(b) shows the sampled lines of force obtained by using the above described methodology, drawn in black or white according to thresholding criteria: the lines corresponding to gene expression activity centers with degrees smaller than 14 have been marked in white. Such threshold value was defined based on the relative frequency histogram of the distribution of centers degree, shown in Fig. 2. It can be seen from Fig. 1(b) that the centers of EGFP activation exhibiting higher numbers of converging lines of force (marked black) tend to concentrate along the regions subjected to the constriction and folding implied by the heart formation dynamics [marked by arrows 1 and 2 in Fig. 1(a)] as well as the sinus venosus [marked by arrow 3 in Fig. 1(a)].

The expression of GFP driven by the *mlc2a* promoter in the heart follows a uniform pattern that spans the whole heart tube during the earlier stages and persists through the adult fish in both chambers. The expression in the atrium and sinus venosus is somewhat weaker than that of the ventriculum, which might be due to differences in cell density [Fig. 1(a)].

By building the three-dimensional characterization of GFP expression as a proxy, our approach indirectly establishes local fields of activity levels for the *mlc2a* promoter throughout the heart. These local fields can have a direct effect on the molecular activity related to the *mlc2a* activation pathway. It has been experimentally proven that abnormalities in the pattern of blood flow during the developing
heart in zebrafish cause malformations such as the presence of an abnormal third chamber, diminished looping and impaired valve formation. It has also been proven that cultured myocardial cells respond to epigenetic forces such as shear stresses by changing their cytoskeletal structure and gene expression profile. In our characterization, the distribution of the node degrees is such that those with high degrees are more clustered in the chamber constriction areas [Fig. 1(a)]. Because blood flow can only originate with chamber contraction, the role of mlc2a as a major contractile component of the cardiomyocytes suggests a direct relationship between the pattern of contractility and morphogenetic processes.

The lines of force patterns indicate that the local fields of mlc2a promoter activity correlates with key morphogenetic events of heart formation, in particular with the future sites of valve formation at the atrioventricular and ventriculobulbar borders to form the characteristic looping of the early heart [arrows in Fig. 1(a)]. Taken together, our results indicate the intricate relationship between form and function during heart morphogenesis, providing a molecular underpinning to the epigenetic forces necessary for the correct growth of the embryonic heart.

While such hypotheses can only be verified through further experimental investigations, a methodology for 3D gene activity characterization has been shown to provide a natural and effective means for quantifying the spatial interactions between the biological structures involved in gene expression. Unlike differential measurements such as gradients or divergent magnitudes, the estimation of the lines of force and activity centers are integral features, indicating spatial interactions over substantial distances. It is expected that the proposed framework will prove to be useful in a number of other gene expression investigations, paving the way to a more objective understanding of the dynamics governing animal development and its pathologies.

The authors thank HFSP RGP39/2002 for funding this project. A. A. is grateful to FAPESP (02/09149-2), and B. A. N. Travençolo is grateful to CAPES and FAPESP (03/13072-8) for financial support. M. Ibañes is partially supported by the Fulbright Program and Generalitat of Catalunya. L. da F. Costa is grateful to FAPESP (Proc. 99/12765-2) and CNPq (Proc. 301422/92-3) for financial support.

1S. B. Carroll, J. K. Grenier, and S. Weatherbee, From DNA to Diversity (Blackwell Science, Malden, 2001).