**A THREE-LAYER MODEL FOR DESCRIBING DEVELOPMENT OF C.ELEGANS.**

TSUGUCHIKA KAMINUMA\*, IGARASHI TAKAKO\*, NAKANO TATSUYA\*, SASAKI SHINYA\*, MIWA JOHJI\*\*

\* Division of Chem-Bio Informatics, National Institute of Health Sciences, 1-18-1, Kami-Yoga, Setagaya-ku, Tokyo 158, JAPAN

\*\* Research & Development Group and Fundamental Research Laboratoriesk, NEC Corporation, 34 Miyukigaoka, Tsukuba 305, JAPAN

Keywords: model, devolopment of c.elegans, VRML, Java, web/database

**1. Introduction**

Development of multicellular organisms has attracted researchers not only in biological fields, but also in mathematics and computer science. The eventual goal of their studies is to clarify how genes control the three-dimensional spatial organization of the cell aggregates through out the development, and to make a model for describing these processes.

However the genes themselves do not directly control the cellular organization. Cells must first divide, differentiate, and sometimes die to form proper spatial structures. It is then proteins that control these cellular events, and the cellular events are controlled by complicated molecular events in which some proteins which are called cell signaling molecules play important roles. The fate of cells could be summarized into a binary map called the cell lineage. Finally it is the genetic information which is linearly stored in the nucleic acid sequence (genes) that determines linear structure of the proteins that control the cellular events.

Thus a model that can describe developmental process should relate three different model worlds: first, the linear events of the gene expressions, second, the the cellular events such as birth and death of the cells, and third, the three-dimensional parallelly processed events of cellular arrangements. We call this model as the three-layer model for describing embryonic development.

Until recently it was difficult to implement such a model on a computer, for the system should link various information modules dynamically. However new software techniques developed particularly for the Internet enables us to implement such a system. These techniques include WWW, hyper link, VRML, Java, and Web/database interface software like CGI. With these techniques together with more conventional methods such as window, object oriented (-like) database, relational database, and 3D molecular graphics one can construct the computer system that can dynamically link linear genomic information, linear and three dimensional protein information, networks of signaling molecules, tree-like cellular information, three dimensional cell aggregates, and even their movements.

We have developed the prototype of such a system. Though this system was developed for describing early development of C.elegans, it will be relevant for describing developmental process of other organisms.

**2. C.elegans as a model organism**

Among the many popular animals used for developmental studies C. elegans is unique because each of its small number (about 1000) of somatic cells is lineally identified, the approximately 8800 connections fabricated by the total of 302 neurons are also identified, and its entire genome will be sequenced within a years or so (Information is available from A C. elegans Database (ftp://lirmm.lirmm.fr, ftp://cele.mrc-lmb.cam.ac.uk, and ftp://ncbi.nlm.nih.gov)).

The entire developmental process of C. elegans from a single-celled zygote to adulthood can be observed in the living state under Nomarski optics. The spatial positions of cells were first identified by E. Shierenberg and his colleagues and then by T. Kaminuma et al. (Kaminuma et al., 1986) up to about 100-cell stages. Later, J. White and R. Schanabel determined the cell positions at much later stages (Thomas et al., 1996).

Relatively easy isolation of mutants and the capability of performing genetic crosses provide a good genetic system from which one can gain general information on gene functions of interest . At present, several embryonically interesting genes, such as par-1 & -2, pie-1, glp-1, and emb-5 , have been cloned and characterized. Many more genes are currently under intensive study, and their properties will be identified in the near future.

Many proteins which plays important roles in development have been identified and cDNA library which already covers two third of the entire set has been provided. Though complete knowledge of cell division, cell differentiation, and cell death is not yet obtained, such knowledge is steadily increasing. Moreover this speed will be accelerated after the completion of the genome sequencing project.

Therefore, we can reasonably expect that C. elegans will be the very first animal in which the genetic implementation of the body plan is understood in molecular terms. By this reason C.elegans is the best model organism for our purpose.

**3. The Computer System**

We have designed a computer system which is conceptually based on the three-layer model.

Part of the system has been implemented on the platform machines, Silicon Graphics Indigo and O2, running on a UNIX operating system. Software has been developed modularly. It consists of a top module and submodule systems each of which corresponds to the three-layered space world. These modules are hyper-linked by internal WWW mechanisms.

**3.1 The Top Module : Main menu**

We assume all users of our system enter from their (either internal or external ) browsers. Thus it provides the home page and the main menu. From the main menu one can select either phenomena or genes. Phenomena are classified into different stages of development, i.e. gametogenesis, embryo genesis, postembryogenesis, and other topical developmental events such as neurulation. From the gene menu one can investigate certain genes that relate to the development. At present only embryo genesis module was implemented.

**3.2 The Genomic Space Module**

The first layer, the Genomic Space, is represented by two basic systems ACEDB and CSNDB (Figure3). ACEDB (Durbin and Thierry-Mieg, 1991) is an object-oriented database originally developed for the C. elegans genome project, but it has become one of the most popular database management systems for genomic data of many other organisms. It contains a wide variety of the C. elegans genomic data, including DNA base sequences, genetic and physical maps, and contigs. ( More precisely the database for the C. elegans genome is called ACeDB, while an empty data management system is called ACEDB.)

CSNDB is a data and knowledge base for cell signaling networks developed by our group (Igarashi and Kaminuma, 1997; Igarashi et al., 1996). It was constructed on ACEDB as the base (data management system ). With these basic facilities one can describe genes and their associated proteins and their interactions. Cell signaling networks for cell growth and cell cycle were described in the CSNDB.

**3.3 Cell Dictionary and Cell Lineage**

The Cellular Space for the second layer is represented by a Cell Dictionary that contains such information as cell identifications (cell names), cell positions, cell shapes, names of sister cells, and other cell characteristics (Figure 4). From these data one can generate cell lineage diagram. We used the nomenclature of the cells given by J. Sulston (Sulston et al., 1983). As for the cell positions we used the three-dimensional coordinate data determined by our group and the data which were already stored in the ACeDB. The Cell Dictionary is managed by relational databases such as ACCESS or Sybase.

**3.4 Development as a Series of Spatial Events**

About a decade ago, we developed a computerized system that could describe the spatial positions of dividing cells during early embryonic development. The system also successfully produced reconstructed images and animation of developing embryos by computer graphics.

We thus took advantage of various computer graphics technologies to represent the third layer of the Developmental Space. In our model, each cell is represented by a ball, which looks like a molecule, so that a popular molecular graphics software such as Rasmol (<ftp://colonsay.dcs.ed.ac.uk/>) or Chemscape (<http://www.mdl.com/chemscape/>) can be used. Thus, well known molecular graphics display modes, such as wire frame, ball-stick, and space-filling models, are also applicable to describe objects in the third layer. Cellular connections are defined between two sister cells that are derived from a common parental cell. This connection can be traced back to the original zygote as a single cell, and all of these connections are represented by binary trees. Conversely, one can generate an animated image of a growing binary tree from cell coordinates and cell relation tables, which corresponds to what we call a Connection Table of Molecular Graphics. One can produce such animation by Java programming.

The hyper-links between the Cell Dictionary and the three-dimensional reconstructed embryo are straightforward if we use VRML models. With VRML version 2.0, we can easily hyper-link each cell in the reconstructed embryo, which are in animation representation, to a unit object in the Cell Dictionary and the data and knowledge associated to it.

**3.5 Linking Different Spaces**

The basic linking mechanism between the first and the second layers is much more complicated than that for the second and the third layers, for we must consider gene expression at each cell separately. If cells, particularly their spatial positions, are different, their modes of gene expressions are expected to be different. The situation is very much like different members in an orchestra playing different parts of the same but an extremely complicated score. The variously resulting music (synthesized proteins) thus triggers cells to differentiate variously.

It is, however, impractical to provide each cell with a different mode of gene expression and with a different cell signaling network, because the number of cells is too large to accumulate the detailed data and knowledge per cell needed for practical and meaningful analysis. Alternatively, we have decided to assume only one universal cell signaling network database and gene expression database at the first layer and to focus on those signaling molecules or genes that characterize the cells at the second layer.

**4. Discussion**

Although the implemented system is still a prototype, it reveals an essential feature of our model system. The heart of the system is the hyper-link. It connects genomic data in ACEDB which is a object-oriented type database to the Cell Dictionary (Cellular Space) based on a relational database, and it also connects the Cell Dictionary to the reconstructed image of the embryo represented by the VRML model. This type of connection was very difficult, if not impossible, to achieve before the advent of the new Internet-related technologies, such as WWW (HTML), multimedia WWW browser, and VRML as well as database interface (CGI, Common Gateway Interface) programs between WWW and databases. These new tools now enabled us to link different sets of data and knowledge of biological systems to that of others.

With C. elegans as a model organism, we are particularly interested in the reconstruction of its embryonic development up to a stage of about 50 cells, since technically speaking it is manageable to reconstruct the embryonic development up to this stage, and scientifically speaking, before the 50-cell stage we can observe most dynamic events in morphogenesis, such as the determinations of the three body axes, the birth of all the founder cells, and the initial phase of gastrulation, which is one of the most essential events in animal development.

In the past decade, we have been gaining increasingly sophisticated knowledge about the basic molecular mechanisms that relate genes to the three-dimensional body plan of animal development, such as those for origins of the body axes of Drosophila embryos, neurulation (neural tube formation) of Xenopus and chickens, body axial formation and gastrulation of C. elegans, and neural plasticity of Aplysia, mice, and monkeys. A quite similar approach may also be applied to these processes. Thus extending our system so that it will become relevant for describing these deveopmental processes will be the next target of our studies.

**5. Conclusion**

We have developed a computerized-system that aims to represent embryonic development as the dynamic organization of a structure, as formed by a cellular aggregate. The system also allows one to link intracellular phenomena, such as gene expression, to a member of the embryonic cell aggregate at a given stage. The system is based on a three layer model: the molecular level (Genomic Space that relates to gene expression, signal transduction pathways, etc.), cellular level (Cellular Space), and organismal level (Developmental Space of three dimensional cell aggregates).

Although conceptually general in nature, the system was specifically implemented to represent the embryonic development of C. elegans. Part of the system is already available on the Internet so that it may be used by a wide range of people from students to advanced researchers.

**References**

1. Durbin, R. and Thierry-Mieg, J., 1991, A C. elegans Database. I. Users' Guide, II. Installation Guide, and III. Configuration Guide. Code and data are available from anonymous FTP servers at lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk and ncbi.nlm.nih.gov.
2. Igarashi, T. and Kaminuma,T., 1997, Development of a cell signaling networks database, in: Pacific Symposium on Biocomputing '97, R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, ed., World Scientific, Singapore.
3. Igarashi, T., Kaminuma, T., and Nadaoka, Y., 1996, A data and knowledge base for cell signaling networks, in: Computation in Cellular and Molecular Biological Systems, R. Cuthbertson. M. Holcombe, and R. Paton, ed., World Scientific,Singapore.
4. Kaminuma, T. and Matsumoto, G., eds.,1991, Biocomputers, Chapman and Hall, London.
5. Kaminuma, T., Minamikawa, R., and Suzuki, I., 1986, 3D reconstruction of spatio-temporal series of optical pictures, in: Pattern Recognition in Practice, E.S. Gelsema and L.N. Kamal, ed., Elsevier Science Publishers B. V., North-Holland.
6. Sulston, J. E., Schierenberg, E., White, J.G., and Thomson, J.N., 1983, The embryonic cell
7. Thomas, C., DeVries, P., Hardin, J., and White, J., 1996, Four-dimensional imaging: com puter visualization of 3D movements in living specimens, Science 273: 603.