Sensors and Materials, Vol. 14, No. 7 (2002) 395–406
MYU Tokyo

S & M 0499

Sensing of Chemical Substances Using Gene Expression Patterns in *Caenorhabditis Elegans*

Tetsuya Matsuno*, Kazuhiro Ura[1], Risa Sonoda[2], Yuji Kohara[3], Hiroko Uesugi[3], Koji Arizono[1], Taisen Iguchi[4,5] and Nobuaki Tominaga[2]

Department of Electronics and Information Engineering, Ariake National College of Technology
150 Higashihagio-machi, Omuta, Fukuoka 836-8585, Japan
[1]Division of Food and Health Environment, Faculty of Environmental and Symbiotic Sciences,
Prefectural University of Kumamoto, 3-1-100 Tsukide, Kumamoto City, Kumamoto 862-8502, Japan
[2]Department of Chemical and Biological Engineering, Ariake National College of Technology,
150 Higashihagio-machi, Omuta, Fukuoka 836-8585, Japan
[3]Structural Biology Center, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411, Japan
[4]Department of Bio-enviromental Research Center for Integrative Bioscience,
National Institute for Basic Biology, 38 Nishigonaka, Myodaiji, Okazaki, Aichi 444-8585, Japan
[5]Core Research for Evolutional Science and Technology (CREST)

(Received 28 March, 2002; accepted 31 May, 2002)

Key words: gene expression pattern, cluster analysis, endocrine disruptor, cDNA microarray, *C. elegans*

We attempted to establish a system for sensing chemical substances using gene expression patterns in *Caenorhabditis elegans* (*C. elegans*). The target chemical substances used in this study were steroid hormones: estradiol and testosterone. These hormones are known to act as endocrine disruptors when the substances which come from outside are received by biological systems as an external (xenobiotic) factor. The effects of these substances on expression patterns were measured using the cDNA microarray. The hierarchical clustering method was applied to analyse the response patterns. They were classified according to the similarity of their responses. We found that each cluster, which was a group of similar response patterns, corresponds to one kind of chemical substance. This means that the *C. elegans* cDNA microarray can be utilized as a chemical sensing system.

1. Introduction

Environmental pollution has become a serious problem because of its dangerous toxic effects on biological systems. People have produced, accumulated and distributed various

*Corresponding author, e-mail address: tetsuya@ariake-nct.ac.jp

kinds of chemicals, many of which have serious toxic effects on biological organs at extremely low concentrations. Therefore, sensing technology has been sought in order to detect these kinds of chemicals at low concentrations.

A system such as a liquid or gas chromatography system might be a representative system for the above purpose. The concept of this kind of sensor is illustrated in Fig. 1(a), in which a filter corresponds to an affinity column in this case. The known chemicals denoted A and B can be identified by that system but unknown chemical X will not be

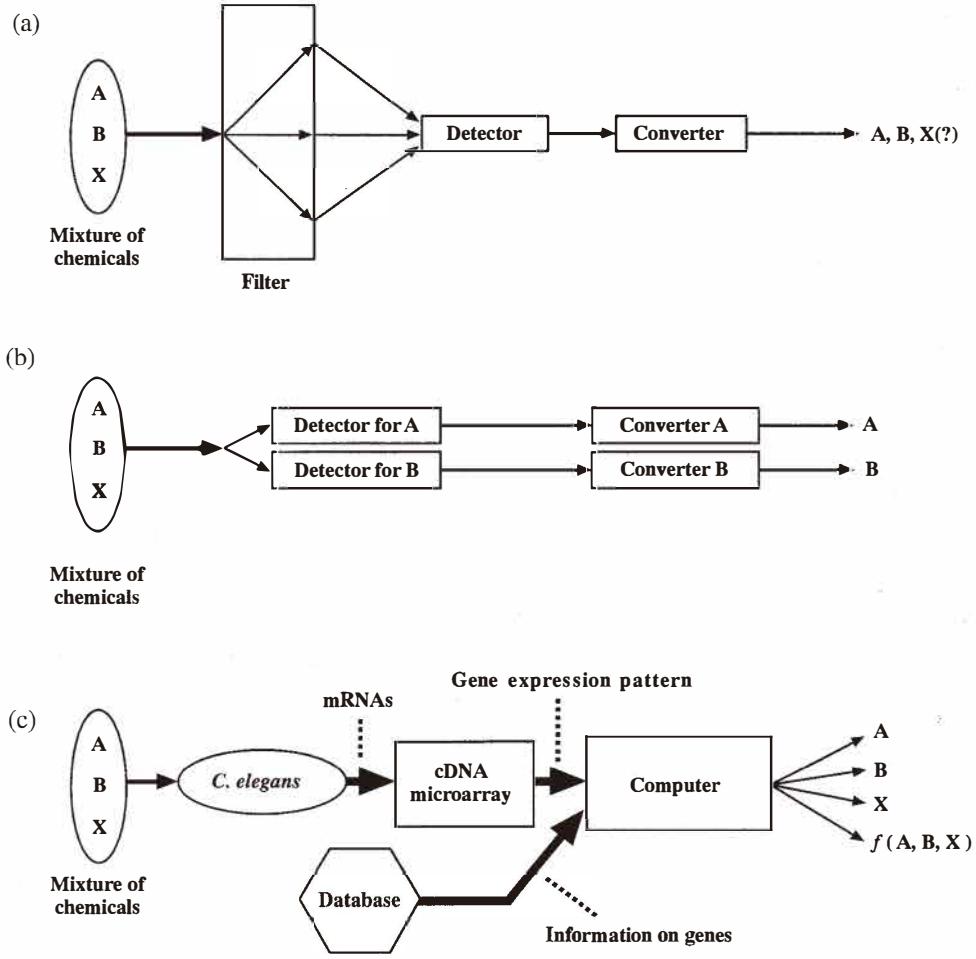

Fig. 1. Diagrams of sensor systems. A sample to be measured is assumed to be composed of known chemicals A, B and unknown chemical X. (a) Sensor system using a low selective detector together with a filter. (b) Sensor system using detectors with high selectivity. (c) Sensor system proposed in this study.

identified although the existence of something may be detected. As another type, biosensors utilizing enzymes or antibodies might be good representatives. This kind of sensor is illustrated in Fig. 1(b). The known chemicals such as A and B can be detected but even the existence of the unknown chemical X cannot be detected.

Hence, the above two kinds of sensor systems are not adequate when one wants to know whether unknown chemicals exist or not (or whether the chemicals concerned are toxic or not).

A bioassay system may be an answer to this problem. Conventional bioassay systems, in which the main observable outputs to be measured are the behaviours of animals, plants or cultured cells, need relatively high concentrations of the chemicals. In this case we observe the result or nearly the last phase of certain toxic signal transduction processes. Gene expressions would also be measured in some of the above systems, but the number of genes is limited, which means the target substances must be already specified, i.e., experimenters have to select relevant genes before the measurement. That is, conventional bioassay systems provide limited and degenerated information about the effects of the chemical substances to be measured. This means conventional bioassay systems are not adequate when one wants to detect and determine the detailed (toxic) effects of unknown chemicals on biological systems. Therefore, we need a new type of bioassay or sensor system.

If we can monitor relatively early the phase of the toxic processes, the detection at low concentrations will be possible. The effective method for the monitoring of the early phase is to observe the expression pattern (profile) of all or almost all genes of the biological system to be exposed to chemical substances.

Furthermore, it can be expected that we can obtain a large amount of information on the toxic effect of chemicals on the biological system because the number of observables is very large and because we can utilize biological databases dealing with functions and cellular roles of the genes concerned.

At present, one can observe quantitatively how much is transcribed from genes to corresponding mRNAs by using a new type of hybridization technique called a cDNA (complementary DNA) microarray, in which a deviation pattern composed of up to about 10,000 genes can be measured. There exist several studies on profiling disease (*e.g.,* cancer[1]) and on characterizing the states of yeast cells during the cell cycle[2] by using cDNA microarrays, while no study has utilized a cDNA microarray as a transducer of a sensor system for detecting unknown toxic chemical substances at low concentrations and for determining whether the effect of the concerned chemicals is toxic or not.

In this study, we tried to establish a system for sensing chemical substances using gene expression patterns in *Caenorhabditis elegans* (*C. elegans*). Figure 1(c) shows the proposed sensor system, with which it is expected that each biological effect of chemical A, B or X will be quantified at low concentrations and the composite effect of A, B and X (dnoted f(A, B, X) in Fig. 1(c)) will also be quantified.

The target chemical substances used in this study are steroid hormones: estradiol and testosterone. This study can be regarded as the first step towards the sensing of unknown general toxic chemical substances. These two kinds of hormones are known to have the effect of endocrine disruptors when biological systems receive substances containing them as foreign substances (xenobiotics). The effects of these substances on expression patterns

were measured using the cDNA microarray. The hierarchical clustering method was applied with respect to the exposure conditions. The response patterns were classified according to their similarity. We found that each cluster, which was a group of similar response patterns, corresponds to one chemical substance. This means that the *C. elegans* cDNA microarray can be utilized as a chemical sensing system.

2. Materials and Methods

2.1 C. elegans

The nematode *C. elegans* is a simple multicellular organism widely used in the field of molecular biology. The transparent body of an adult *C. elegans* is about 1mm long and is composed of about 1,000 highly differentiated somatic cells, which constitute muscle, nervous, digestive, reproductive and other kinds of systems. It has a short life cycle of 3 days at 20°C. The size of its genome is about 10^8 base pairs, which has been sequenced already, and the number of genes contained is about 19,000.[3] It easily grows on agar plates or in liquid medium containing *E. coli* as food in a laboratory.

Its ability to grow in liquid culture is advantageous to researchers when they use a group of the worms as a transducer for sensing chemicals in solutions. This is because the target liquid samples can be easily applied at various concentrations.

2.2 *Preparation of* C. elegans *and exposure to chemicals*

Worms used in this study were wild-type hermaphrodite *C. elegans*. General procedures for working with *C. elegans* were performed according to those reported by Brenner.[4] For microarray analysis, *C. elegans* were transferred from agar plate culture to liquid culture. After 3 and 4 day cultures, worms were harvested and synchronized by the alkali-bleaching method. Eggs were suspended in S-basal medium and incubated for 16 h with shaking at 100 rpm. Worms were synchronized at L1 larvae. Then *E. coli* as food was added and cultured for another 24 h. Worms grew at L2/3 larvae. One hour before chemical exposure, larvae were starved to remove and digest *E. coli*. About 200,000 worms were used for one dose of chemical exposure. Worms were resuspended in S-basal medium containing chemicals and incubated for 5 h. After chemical exposure, worms were harvested and frozen quickly by using liquid nitrogen.

2.3 *Measurements of gene expression patterns*

Total RNAs, including mRNAs reflecting the expression of corresponding genes, were extracted from the larvae with or without chemical exposure by using TRIZOL reagent (Gibco BRL) following the manufacturer's manual. Then poly(A)+RNAs were purified from total RNAs using oligo(dT) resin (Amersham Bioscience). Cy-dye labeled cDNA probes were generated by reverse transcription reaction at the presence of Cy-3 or Cy-5 conjugated dUTP in the reaction mixture for control or chemical exposure, respectively. The labeled cDNA probes were mixed together prior to hybridization to a cDNA microarray including about 8000 kinds of individual *C. elegans* cDNA spots, a procedure which was established in Dr. Kohara's NIGG lab. The mixed labeled cDNA probe hybridized to the microarray. Then the microarray was washed thoroughly and scanned using a laser

confocal scanner ScanAray (GSI Luminonics). The coefficient from the Cy-3 and Cy-5 intensities was calculated to determine the relative amounts of a particular gene transcript in the two samples. We observe an expression pattern of deviation from the normal (no exposure) condition. The experimental procedure described above is schematically illustrated in Fig. 2.

2.4 *Cluster analysis of the expression patterns*

The measurements were performed under 23 different conditions, which were 9 exposures to estradiol with various concentrations, 13 exposures to testosterone with various concentrations and no exposure (control). We selected 2847 genes out of the 8000 genes with a criterion so that the measurements were successful in greater than or equal to 15 conditions (out of the 23 measurements) for each gene. The success criterion is that the intensity of the signal from Cy-dye is sufficiently larger than that of the background signal (noise) for each spot in the microarray. The logarithm of the ratio of intensity of an exposure condition (Cy-5) to that of a normal condition (Cy-3) was used for the analysis.

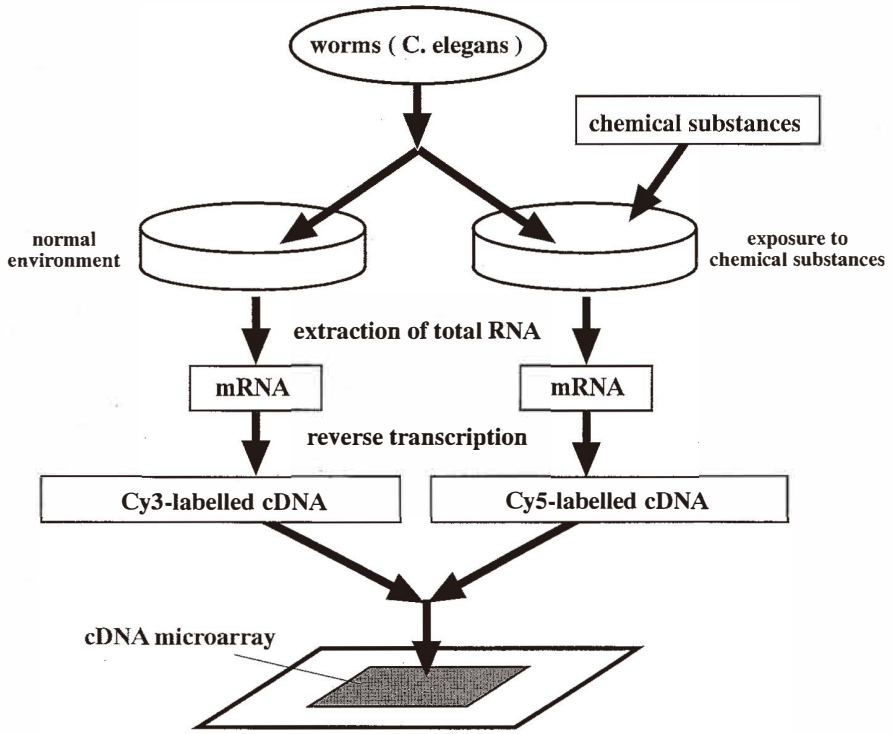

Fig. 2. Schematic illustration of the experimental procedure for preparation of *C. elegans* and exposure to chemicals. This corresponds to the transducer part of the sensor system shown in Fig. 1(c).

The obtained data of the gene expression patterns under various exposure conditions were analysed using the hierarchical clustering method, where software for the analysis written by Eisen[2] was used. We used the method known as average linkage clustering.

In general, in the result of the hierarchical cluster analysis, a lowest level cluster is a pair of observation vectors (gene expression patterns or condition patterns in this study). A higher level cluster is composed of lower level clusters and/or observation vectors. This clustering situation is usually depicted by a tree diagram called the dendrogram, in which the similarity structure of the observation vectors is illustrated quantitatively.

The cluster analysis was performed with respect to both exposure conditions and genes. In the clustering with respect to exposure conditions, similar expression patterns (composed of signals of the 2847 selected genes) were arranged in neighboring positions (rows) in the dendrogram. In the clustering with respect to genes, the similar responses, each of which consists of signals in the 23 conditions, were arranged in neighboring positions (columns).

2.5 *Annotation to genes*

There exist several databases concerning genes of *C. elegans*. We used the database WormPD,[5–7] which is one of the sublibraries of the BioKnowledge Library created and run by Proteome, Inc. In WormPD, the genes of *C. elegans* are categorized according to cellular roles and functions of proteins corresponding to them. In this study, the selected genes for measurement by the cDNA microarray were annotated by their cellular roles and functions using the categorization of WormPD in order to characterize the clusters of genes.

3. Results and Discussion

3.1 *Interpretation of response patterns*

The dendrogram representing the similarity structure of the expression patterns for 23 conditions is shown in Fig. 3. It was obtained by hierarchical cluster analysis with respect to exposure conditions. It can be found that expression patterns corresponding to the same chemicals are arranged in neighboring positions. The cluster of estradiol exposure and the cluster of testosterone exposure can be clearly seen in the dendrogram. The cluster formation according to chemical substances indicates that sensing chemicals by observing gene expression patterns is possible in the tested range of concentrations of the chemicals: estradiol and testosterone. We can conclude that the threshold value of the sensor system for estradiol or testosterone is less than 1nM because the measurements at the concentration are the constituents of the corresponding cluster of the chemical substance. That is, in Fig. 3, we can see that the conditions 13115, 13112 and 1398, corresponding to the estradiol (E2) exposure, are in the same cluster containing other E2 exposure conditions (at different concentrations), and that the conditions 16114 and 16118, corresponding to the testosterone (TS) exposure, are in the same cluster containing other TS exposure conditions. Hence, it can be expected that we can detect these chemicals at lower concentrations. The set of gene expression patterns, from which the dendrogram in Fig. 3 is obtained, of the whole selected genes for various conditions is depicted in Fig. 4.

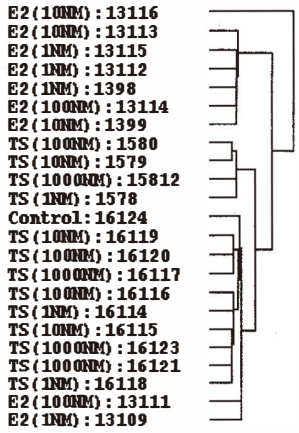

Fig. 3. The dendrogram obtained by the hierarchical cluster analysis with respect to conditions. The observation vectors corresponding to estradiol and testosterone are denoted 'E2' and 'TS,' respectively, with concentration and the measurement ID. The notation 'Control' is for the no-exposure condition (normal condition).

We can identify chemicals according to response patterns of gene expression. In the behaviors of gene expressions, however, we could not find explicit dependences on concentrations of chemical substances. At present, the reproducibility of the measurements by cDNA microarray is not very good. The error, i.e., the ratio of standard deviation to average of an absolute value of the light intensity of each spot could be around 20–30% or more (data not shown). The data of the logarithm of the ratio of the light intensity from Cy-5 to that from Cy-3 have error of the same order. This may be one of the reasons for not observing dependences on concentrations of chemicals.

It can be found in Figs. 4(b) and 4(c) that the response pattern at the non-exposure condition (denoted as 'Control' in Figs. 3, 4(b) and 4(c)) does not exhibit the "zero pattern." This appears to be strange. The zero pattern, i.e., depicted as a vertical line (column) should be black at the 'Control' in Fig. 4 because the condition of no exposure means that the two gene expression patterns corresponding to the two divided groups of worms should be identical. If there do not exist any errors in the procedures, it seems reasonable to conjecture that affinities of Cy-3 and Cy-5 for cDNAs are different and that the degree of difference is dependent on the genes. Hence, we might have to change the method for labelling mRNAs. Confirmation of the above conjecture is left as a future task.

There may be a serious problem in the result of the cluster analysis with respect to the conditions. In Fig. 3, the measurement ID is indicated for each condition. The sets of IDs 13x, 15x and 16x, where x denotes a two- or three-digit number, refer to different series of procedures for measurements. These three ID sets almost correspond to the three clusters as can be seen in Fig. 3. Then, actually, the obtained result is not desirable because we detected the effect of certain uncontrollable parameters rather than the kind of target

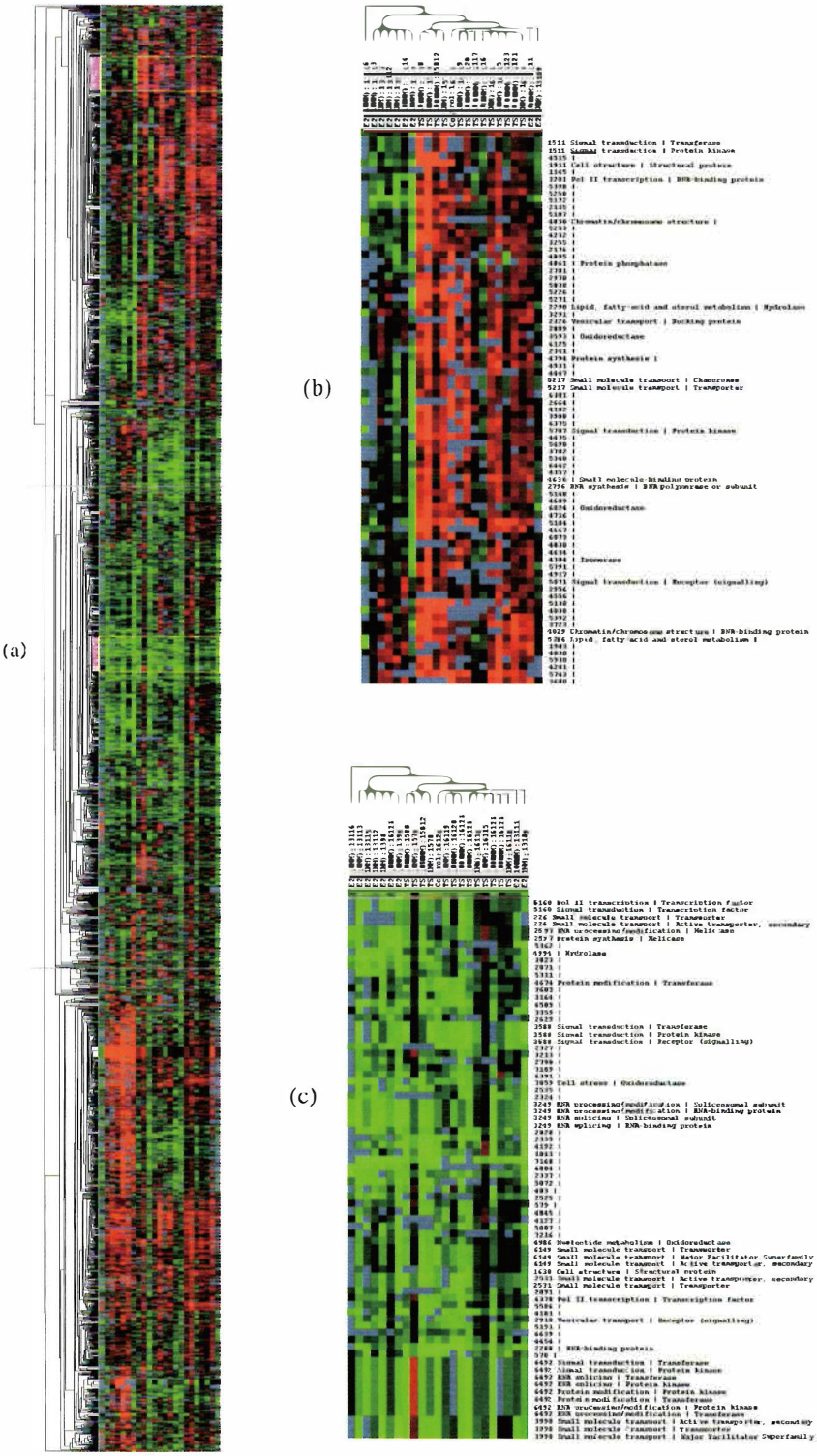

Fig. 4. (See the caption on p.403)

Fig. 4. Cluster display of data. The result of the hierarchical cluster analysis using the selected 2847 genes at 23 measurement conditions is shown. All measurements were relative to normal (no exposure) conditions for corresponding measurements. The color scale ranges from saturated green for log (base is 2) ratios –0.87 and below to saturated red for log ratios 0.87 and above. The color black represents no deviation from the normal condition. The color gray is for failed measurements. Each gene is represented by a single row of colored boxes; each condition is represented by a single column. (a) The behaviors of the whole selected genes are displayed. The dendrogram attached to the left was obtained by the cluster analysis with respect to genes. (b) Example of a magnified partial image of the whole display. The magnified region is depicted by a yellow frame depicted near the top of the whole display (shown in (a)). (c) Example of a magnified partial image of the whole display. The magnified region is depicted by a yellow frame near the middle of the whole display. In (b) and (c), the dendrogram attached to the top is the same as that shown in Fig. 3, and biological annotations are attached to the right: the number at the leftmost is an ID for a gene, an annotation to the left of the symbol ' |' is for its cellular role and that to the right is for its function. If an annotation is empty then its cellular role or function is unknown (classified as unknown in WormPD) at present.

chemical substance. However, there exists the fact that the conditions of exposure to testosterone (denoted as TS) formed two clusters according to the preparation lots but the two clusters were grouped together into one (higher level) cluster. This implies the possibility for our system to discriminate testosterone from other kinds of chemical substances, or to discriminate estradiol (E2) from other kinds of chemicals.

3.2 *Biological meaning of a cluster of genes*

The cluster analysis with respect to genes was also performed. Similar genes with respect to responses to various exposure conditions were grouped together as a cluster.The behaviors of the whole selected genes are shown in Fig. 4(a). Similar genes exhibiting similar responses are placed in nearby rows. Two examples of clusters are shown in Figs. 4(b) and 4(c), in which biological annotations (cellular role and function) are attached for each gene using WormPD.[5] We examined various clusters in order to find biological meanings of clusters of genes. However, no common characteristics of a certain cluster could be found, while in the study of gene expression patterns at various moments in the cell cycle in the budding yeast, it was found that each cluster with respect to genes consists of genes with similar cellular roles and functions.[2] This means that genes tend to be in the same cluster if the genes have similar temporal behaviors.

In the case of this study using *C. elegans*, we observed the structure of the signal transductions induced by the chemical substances. Since the gene expression patterns partially reflected the structure of signal transduction, each cluster consists of genes with various kinds of cellular roles and functions. This situation is schematically illustrated in Fig. 5. It can be assumed that each cluster corresponds to a certain kind of functional module or sub-network of signal transduction. It should be noted that the biological system concerned here is multicellular. Strictly speaking, we must consider cell specific gene expression. The observed expression patterns would mainly reflect the common responses of genes which occur in almost all the constituent cells. Genes, which express in a certain group of cells, might play important roles in the response to the chemical substances. This

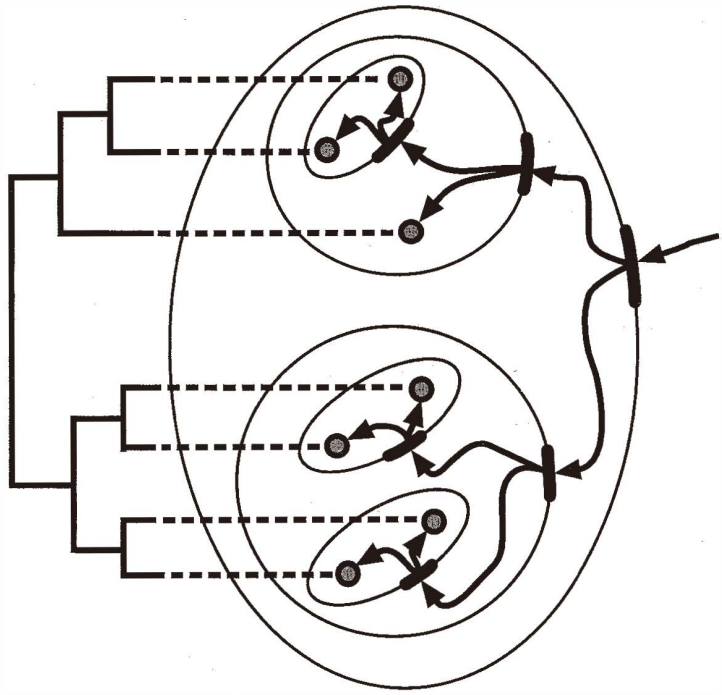

Fig. 5. A conjecture for obtained clusters with respect to genes. The inferred origin of the obtained clusters is schematically illustrated. A small gray circle represents a gene to be observed by the cDNA microarray. The small circles are enclosed by a circle, which corresponds to a cluster. The circle (corresponding to a cluster) is generally enclosed by higher level circles (corresponding to higher level clusters) hierarchically. The dendrogram corresponding to this situation is depicted on the left side. An arrow in this illustration represents a path of signal transduction. The arrow (path) may be a certain kind of chemical or electrochemical messenger. A thick part of the cluster circle represents a certain kind of transmitting molecular device which may convert or amplify the chemical or electrochemical signals.

issue may be what we have to consider in the next step of the study. We also note that one of the merits of using multicellular systems is that it will be possible to investigate the relationships among gene expression patterns, their behaviours and morphogenesis.

3.3 *Summary, further measurements and analyses*

We have proposed a new type of chemical sensing system using a cDNA microarray of *C. elegans*, together with a database and a multivariate statistical method (cluster analysis). The diagram of the proposed system is shown in Fig. 1(c). This system has a distinguishing characteristic compared to other types of sensor systems. By combining gene expression patterns and information about genes, the computer can possibly provide not only the quantities of known chemicals A and B, but also the quantity of unknown X utilizing the cluster analysis, which can be considered as a primitive method for pattern

recognition for gene expressions which reflect composite effects of chemicals A, B, and X (denoted by f (A,B,X) in Fig. 1(c)) on a biological system. It is possible that this system does not measure the substance itself but the toxic effect (this concept is similar to that in the taste sensor system measuring taste itself[8]), although our study has not yet reached the stage of demonstrating the feature.

We have shown that chemical sensing is possible by using the system, although there are some issues to be resolved such as errors in measurement of light intensity and labelling method for mRNAs. The threshold value for detecting estradiol or testosterone is shown to be less than lnM. Therefore, as one of our next steps, we have to determine the threshold for each substance by trying to perform measurements at lower concentrations. In order to clarify the common property of endocrine disruptors, we also have to try to measure other kinds of endocrine disruptors, other kinds of chemical substances or other kinds of physical or chemical stresses induced by, *e.g.*, osmotic pressure and temperature.

It is especially important to clarify the biological meanings of clusters with respect to genes when one wants to deal with unknown chemicals. In this study, we annotated each gene by categories of cellular roles and functions. However, we could not find explicit biological meaning of the obtained clusters. Further accumulation of measurement data will change the structure of the similarities (the dendrogram) and will provide various meanings for the clusters. We need information about networks of signal transduction or about interacting networks of genes or other kinds of biomolecules in order to find the explicit meanings for the clusters. Therefore, the technology of data mining concerning information about the networks[9] must be introduced in a future study. Statistical analysis using the graphical modelling may be one representative method of inferring the interacting network of molecules.[10] Observing dynamics (temporal behaviors) of genes will also provide effective information for causalities among genes and biomolecules.

Acknowledgements

This work was partially supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan, Grant-in-Aid for Scientific Research on Priority Areas (A) 411, 13027288 and Grant-in-Aid for Scientific Research (B)(1), 12555116. This work was also supported by Core Research for Evolutional Science and Technology (CREST) of the Japan Science and Technology Corporation. We thank Yuji Take and Noriyuki Watanabe for supporting the construction of the database for the analysis.

References

1 M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, N. Sampas, E. Dougherty, E. Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, J. Carpten, E. Gillanders, D. Leja, K. Dietrich, C. Beardry, M. Berens, D. Alberts, V. Sondak, N. Hayward and J. Trent: Nature **406** (2000) 536.

2 M. B. Eisen, P. T. Spellman, P. O. Brown and D. Botstein: Proc. Natl. Acad. Sci. USA (Genetics) **95** (1998) p. 14863.

3 The *C. elegans* Sequencing Consortium: Science **282** (1998) 2012.

4 S. J. Brenner: Genetics (1974) 71.

5 http://www.proteome.com
6 M. C. Costanzo, M. E. Crawford, J. E. Hirschman, J. E. Kranz, P. Olsen, L. S. Robertson, M. S. Skrzypek, B.R. Braun, K. L. Hopkins, P. Kondu, C. Lengieza, J. E. Lew-Smith, M. Tillberg and J. I. Garrels: Nucleic Acids Res. **29** (2001) 75.
7 M. C. Costanzo, J. D. Hogan, M. E. Cusick, B. P. Davis, A. M. Fancher, P. E. Hodges, P. Kondu, C. Lengieza, J. E. Lew-Smith, C. Lingner, K. J. Roberg-Perez, M. Tillberg, J. E. Brooks and J. I. Garrels: Nucleic Acids Research **28** (2000) 73.
8 K. Toko: Biomimetic Sensor Technology (Cambridge University Press, Cambridge, 2000) Chap. 6.
9 M. Kanehisa: Post-Genome Informatics (Oxford University Press, Oxford, 2000) Chap. 4.
10 T. Matsuno and N. Tominaga: Tech. Rep. IEICE (2000) 55 [in Japanese].