TM-1306

Toward Simulation Based Representation of Biological Knowledge Using Object-oriented Database Language (1)

by
M. Hirosawa (Kazusa DNA),
R. Tanaka (IMS) & H. Tanaka

July, 1994

ICOT
Mita Kokusai Bldg. 21F
4-28 Mita 1-Chome
Minato-ku Tokyo 108 Japan
(03)3456-3191~5

Institute for New Generation Computer Technology

Toward simulation based representation of biological knowledge using object-oriented database language (1)

Makoto Hirosawa

(Kazusa DNA Research Institute)

Reiko Tanaka

(IMS)

Hidetoshi Tanaka

(ICOT)

1 Introduction

Amazing advances in bio-technology now allow us to describe various phenomena occurring in the body with the language of genes or DNA. Genes encode the proteins that constitute the body. Proteins are not only the building blocks of the body, but also regulate it. These proteins are also coded in the genes. Consequently, knowledge of genes and proteins is indispensable is understanding various bodily phenomena such as the immune system.

Up until now, information describing biological phenomena has been accumulating to the extent that it explains biological phenomena to some extent. To make the explanations possible, however, we must collect information from several sources, biological text books [Albert 1994], papers and databases[Bairoch 1992]. These sources differs in their authors' interests and in their degree of meshes of description. Recently, several researchers [Goto 1993] have attempted to integrate biologial database. Their works are necessary step toward description of biological phenomena. However, they are mainly interested in integration of data and paid a little attention on the representation of biological phenomena.

As a result, non-experts in biology find it difficult to integrate these information sources to grasp biological phenomena adequately. Sometimes, even experienced biologists cannot get an integrated view of biological phenomena, because not a few biologists are only able to keep up with progress in their specialty.

It is important to express in a knowledge base the phenomena played by genes and proteins and to visualize these phenomena properly. Demonstrations of these phenomena through visualization help students of biology to understand biological phenomena in our body. Also, the visualization of the phenomena with reference to related databases facilitates research on genes and gives biologists further inspiration for new research.

As the first step in realizing a knowledge base like the one stated above, we have studied representations of biological knowledge needed to describe biological phenomena and have developed prototype knowledge base. From experience in object oriented knowledge bases[Hirosawa 1993; Tanaka 1993], we thought object oriented knowledge was suitable to describe biologial concepts. We decided to employ Quixote, an object-oriented database language [Yasukawa 1992], in this project. We expected, with the object-oriented language, that concepts like cell could be described naturally.

In the next section, the suitability of the object-oriented scheme to express biological phenomena is shown. Then, our prototype knowledge base, which represents these process inside a cell, is presented. Finally, the example of execution of our system is shown.

2 Biological phenomena and Object-oriented representation

This section shows how biological phenomena can be described naturally using an object-oriented scheme.

Cell as object

Concepts like the cell can be described naturally with an object-oriented database language. In our body, when cells receive chemical substances from outside, their behavior depends on the chemical substance. When we are only interested in the behavior of cells, we don't have to be bothered with processes inside the cells because of capsulation. But, we can describe inside processes with object-oriented language and this information can be referable.

Biological phenomena as interaction of cells

Because our body is composed of cells, if cells and interaction between cells are described properly, the biological phenomena in our body can be described. The interactions between cells are also describable as interactions between cells represented as objects. Then, it is essentially possible to describe phenomena in our body in an object-oriented scheme.

In Fig 1, in cell C1, Response R1 occurs when Stimulus S1 occurs. In cell C2, Response R2 occurs when Stimulus S2 occurs. Suppose "R1 implies S1" is true in the knowledge base. In this case , if Stimulus S1 occurs to C1, consequently, R2 occurs to C2.

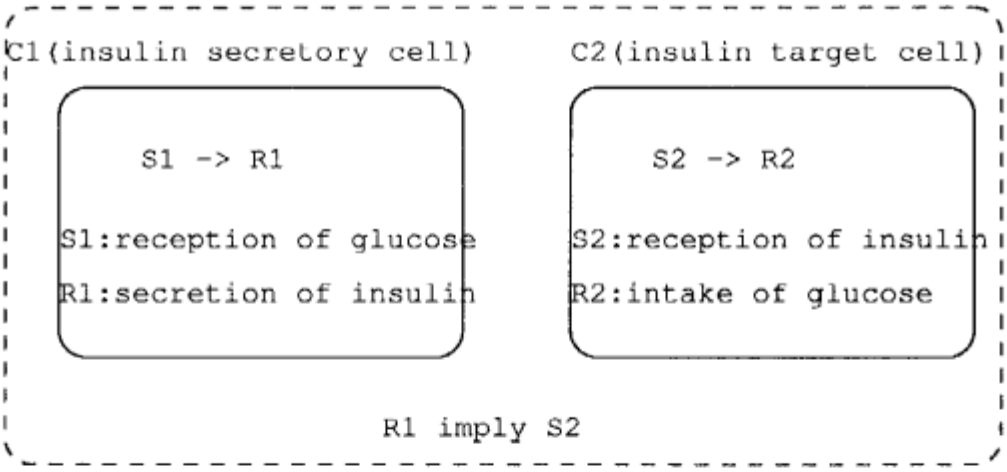

Fig 1

Species of cells using inheritance

There are many species of cells. The response of cells to a stimulus depends on the species of the cells. Similar species of cells have similar stimulus-response relationships. This situation can be represented by the use of 'inheritance' of object-oriented scheme.

Similar kinds of cells share the same stimulus-response relationship. In an object-oriented scheme, the two kinds of cells that share the same stimulus-response relationship belong to the same class that has the stimulus-response relationship. Because of inheritance of the relationship, the two kinds of cells have the same relationship.

In Fig 2, C2a, C2b and C2c are subclasses of C2. C2a has C2a1, C2a2 and C2a3 as its subclass. Because C2 has S2 > R2 as stimulus-response relationship, the lower classes

have the relationship as the result of inheritance. C2a has S2a − > R2a as its own stimulus-response relationship, which is inherited to its lower class.

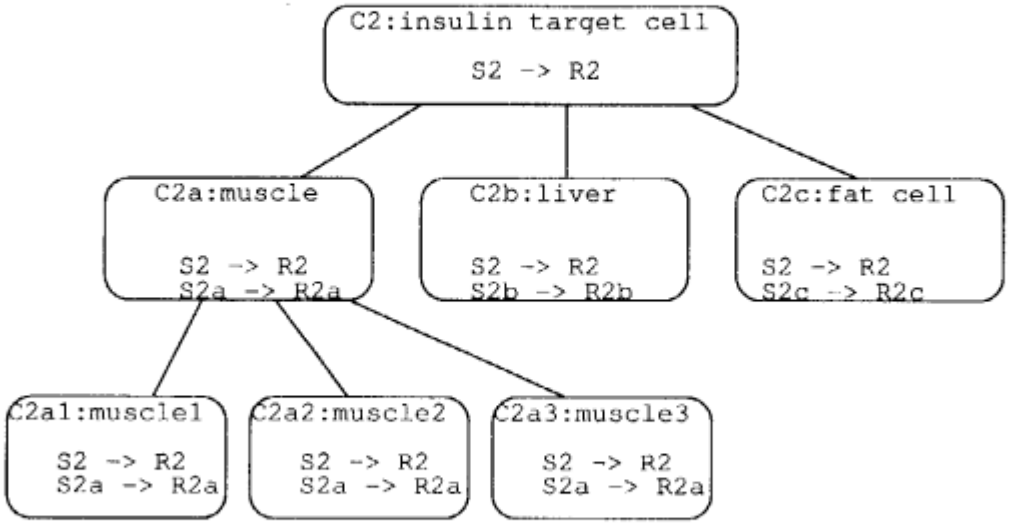

Fig 2

3 Prototype knowledge base

As indicated in the previous section, biological phenomena can be suitably described with an object-oriented language. In our strategy of research, we decided to create a knowledge base describing intracellular processes first. Secondly, we will create knowledge including intercellular processes based on the former knowledge base.

We created a prototype knowledge base using object-oriented database language. As object-oriented database language, we selected Quixote [Yasukawa 1992]. The deductive feature of Quixote is also utilized in our system.

The following is a portion of it. For the sake of simplicity, only simple knowledge is shown.

```
receive[name="EGF"]/[result=increase[name="DG"]];;   (1)
increase[name="DG"]/[result=active[name="PKC"]];;    (2)
active[name="PKC"]/[result=active[name="SRF"]];;     (3)
```

"EGF", "DG", "PKC" and "SRF" are proteins or a sort of protein. The knowledge can be read as follows. If "EGF" is received, "DG" is increased (1). If "DG" is increased, "PKC" becomes active (2). If "PKC" is active, "SRF" becomes active (3). The description is done in the form "A/[result = B];;". It signifies that if A is satisfied, then B becomes true. In this way, individual processes in a cell are described.

In the three knowledge , 'receive', 'increase' and 'active' are objects in Quixote. However,we could choose "EGF","DG","PKC" and "SRF", which correpond to entity in entity-relationship model. Researchers of database might regard the latter choice as a suitable choice. But, when we describe intracellular processes, possible events are the most essential thing that must be described. We selected 'receive', 'increase' and 'active' as object because possible events are described using these words.

To understand the collective result of each process, let us assert knowledge " receive[name="EGF"]" to the knowledge base. If we ask whether "active[name="SRF"]" is true or not, the knowledge base answers "yes". To prove "active[name="SRF"]" the above three knowledges are used. An important thing that must be noted is that only individual processes are described in the knowledge base and that the proof of "active[name="SRF"]" is done as the result of a series of inferences. Here, the deductive feature of Quixote is used.

Concepts like "EGF","DG","PKC" and "SRF" can be described in detail as follows. The knowledge (4) is a description of "EGF". It means a growth factor named "EGF" has a length of 110 and its target is epidermis.

```
growth_factor[name = "EGF", target = epidermis, length = 110] ;;   (4)
```

4 Example of execution

We will show an example of execution of our knowledge base. In this case, "what happens if insulin arrives at a muscle" is asked (a0). The knowledge base answer that two series of events will happen ((a1) - (a8), (a11) - (a21)).

Both series of events consequently decrease glucose outside muscle cells(a8, a21). In the first answer series, after reception of insulin(a1), insulin receptor tyrosine kinase, PI3 kinase and small G protein are activated(a2-a4). Then, GLUT4 is translocated to the cell membrance(a5). After intake of glucose(a6) and increase of glucose inside cell(a7), the final result happens(a8). In the second answer series, after reception of insulin(a11), series of activation(a12 - a19) happen. Then, like the first answer series, after an increase of glucose inside the cell(a20 corresponding to a7), the same final result happens(a21 corresponding to a8).

```
?- arrive[name="insulin", place = muscle ]/[result=X].                (a0)
yes.
X =    state[now= receive[name=insulin],                              (a1)
  next=state[now= active[name="insulin receptor tyrosin kinase"],     (a2)
  next=state[now= active[name="PI3 kinase"],                          (a3)
  next=state[now= active[name="small G protein"],                     (a4)
  next=state[now= translocate[name="GLUT4",location="cell membrane"],(a5)
  next=state[now= intake[name=glucose,location="inside cell"],        (a6)
  next=state[now= increase[name=glucose,location="inside cell"],      (a7)
  next=    decrease[name=glucose,location="outside cell"]]]]]]]]      (a8)

X =    state[now= receive[name=insulin],                             (a11)
  next=state[now= active[name="insulin receptor tyrosin kinase"],    (a12)
  next=state[now= active[name="Ras"],                                 (a13)
  next=state[now= active[name="Raf-1 kinase"],                        (a14)
```

```
next=state[now= active[name="MAP kinase kinase"],                    (a15)
next=state[now= active[name="MAP kinase"],                           (a16)
next=state[now= active[name="S6kinase2"],                            (a17)
next=state[now= active[name="PP1"],                                  (a18)
next=state[now= active[name="glicogen synthetase"],                  (a19)
next=state[now= increase[name=glicogen,location="inside cell"],      (a20)
next=   decrease[name=glucose,location="outside cell"]]]]]]]]]]      (a21)
```

5 Discussion

So far, a little attention has been paid to representation and visualization of biological phenomena. In our system, possible events in a cell(eg. If "EGF" is received, "DG" is increased) are stored in the knowledge base. If we want to know what happens if some stimulus comes from outside the cell, and if we ask so, inducable events are successively calculated in the system. The series of events occurred after the stimulus are then shown to user. This can be regarded as a simulation of phenomena in the body. Through the simulation, users can experience the biological processes happening in the body.

In the present research, the simulation result is shown only in the form of language. In our next project, we are planning to visualize the simulation result. We think the visualization of the simulation result enhances the understanding of biological phenomena in the body and it will contribute to the progress of biology and medicine.

Acknowledgement

This research has been done using facility of ICOT when M.Hirosawa stays at ICOT. We thank Dr. S.Uchida of ICOT and Dr. Takanami of Kazusa DNA Research Institute for realizing research collaboration between ICOT and Kazusa DNA Research Institute.

Reference

[**Albert 1994**] Albert,B.*et al.* Molecular biology of cell. Garland publishing,INC.

[**Bairoch 1991**] Bairoch,A. Prosite : A dictionary of Protein site and pattern : User manual Release 7.00, May 1991.

[**Goto** *et al.* **1993**] Goto,S. *et al.* A deductive language in object-oriented database for genome analysis. *Proceedings of International Symposium on Next Genaration Database Systems and Their Application* , pp123-129.

[**Hirosawa** *et al.* **1993**] Hirosawa,M. *et al.* Application of deductive object-oriented knowledge base to genetic unformation processing. *Proceedings of International Symposium on Next Genaration Database Systems and Their Application* , pp116-122.

[**HTakaka 1993**] Tanaka,H. A Private Knowledge Base for Molecular Biological Research. *Proceedings of the Twenty-Sixth Annual Hawaii International Conference on System Sciences*,**Vol.I** pp803-812,1993.

[**Yasukawa** *et al.* **1992**] H.Yasukawa *et al.* Objects, Properties, and Modules in Quixote. *Proc. of FGCS92, pp89-112, 1992*