The Standard Graphical Notation for Biological Networks

Hiroaki Kitano

ERATO Kitano Symbiotic Systems Project, JST and The Systems Biology Institute,
Suite 6A, M31, 6-31-15 Jingumae, Shibuya, Tokyo 150-0001 Japan
Sony Computer Science Laboratories, Inc., 3-14-15 Higashi-Gotanda, Shinagawa,
Tokyo 141-0022, Japan

E-mail: kitano@symbio.jst.go.jp

Solidly defined and comprehensive graphical representation of biological networks is essential for efficient and accurate representation and dissemination of biological models. Nevertheless, there has been no standard graphical representation accepted in the community. In this paper, we intent to propose a system of graphical representation of biological networks that can be a basis of de fact standard of the community. Careful analysis of biological models revealed that a set of diagrams are necessary to fully describe interactions of proteins and genes, as well as to describe processes of biological events without compromising accuracy and readability.

1. Introduction

With abundant information on gene regulations, metabolic pathways, and signal transduction, there is pressing needs for standardized graphical notations to describe biological networks and processes. Most authors of biological papers use arrow-headed lines and bar-headed lines to indicate activation and inhibition, respectively. However, traditional diagrams are informal, often confusing, and much information is lost. There are several reasons for this problem.

First, in traditional schematics, symbols, such as arrows, are used in various semantics from transcriptional activation/inhibition, signal transduction via protein-protein interaction, transport, etc. This causes substantial confusion in the diagrams used in the papers, thus readers must read substantial part of the text to understand what does the diagram really describes.

Second, often both state transitions and relationships are described within one diagram

that made meaning of nodes and arcs confusing. One arrow may mean activation, but the other arrow in the same diagram may mean transition of the state or translocation of materials. Without consistent and unambiguous rules for representation, not only information is lost, but also misinformation could be disseminated.

The situation is further aggravated by the lack of standard for graphical notations, so that each schematic follows different rules preventing efficient and accurate dissemination of knowledge.

Recently, some researchers propose notation systems to mitigate such problems [1-5], and many research groups working on large-scale biology invent their own, but rather informal, system of describing networks and processes. Nevertheless, no standard has been formed yet, partly because proposed notations have numbers of issues to overcome, as well as lack of software tools to assist the use of such diagrams.

The goal of this paper is to propose a comprehensive system of notations for visually describing biological networks and processes. A system of graphical representation shall be powerful enough to express sufficient information in clearly visible and unambiguous way, and should be supported by software tools. It is expected that the system proposed in this paper to be the basis of standard graphical notation for Systems Biology Mark-up Language (SBML) Level-III.

2. Requirements for Graphical Notation

There are few criteria that the notation system should satisfy, such as;

(1) **Expressiveness**: The notation system should be able to describe every possible relationship among genes and proteins, as well as biological processes.

(2) **Semantically unambiguous**: Notation should be unambiguous.

(3) **Visually unambiguous**: Each symbol should be clearly identified and cannot be mistaken with other symbols. This feature should be maintained with low-resolution displays, as well as black/white printings.

(4) **Extension Capability**: The notation system shall be flexible enough to add new symbols and relationship in a consistent manner. This may include the use of color-coding to enhance expressiveness and readability, but information shall not be lost even with black and white displays.

3. What need to be represented?

Two major purposes of creating such standard are; (1) to facilitate dissemination of information on the biochemical networks, and (2) to be able to efficiently create models for simulation and analysis. Thus, the diagram should be able to visually display information necessary for these purposes, and a set of software need to be developed to support efficient creation of models based on the notation. Since biological systems are highly complex, it is not feasible or desirable to graphically represent all available information, because it will simply overload the diagram so that it will be less informative. Such information may be embedded in hyperlinks created within the diagram and show up only when it is requested.

Now, let us discuss what diagram shall represent. In traditional diagram, both informal drawings, as well as schematics proposed to be a standard notation there is only one representation modality. Traditional schemes generally describe processes of cascading activations or conversion of chemical species. In this type of diagram, only relationship between components that are involved in the process in focus are described, leaving numbers of interactions not being displayed (Figure 1A). On the contrary, Kohn's diagram describes relationships among components. However, it does not provide step-wise view of the specific biological processes. Kohn's notation constraints one species to appear only once in the diagram, and all combinations are described as arrowhead lines with dot in the middle that denotes compound (Figure 1B). This notation provides a rigid framework on constraint relationship among species. Major drawbacks of the Kohn's notation are (1) difficulties in understanding process of certain reactions, and (2) needs for experience to correctly draw and read compound formation diagram, that are described using arrowhead lines and dots.

Rad3 ——► Chk1 ——| Cdc25 ——► Cdc2 ——► G2 Progress

(A) Conventional Notation
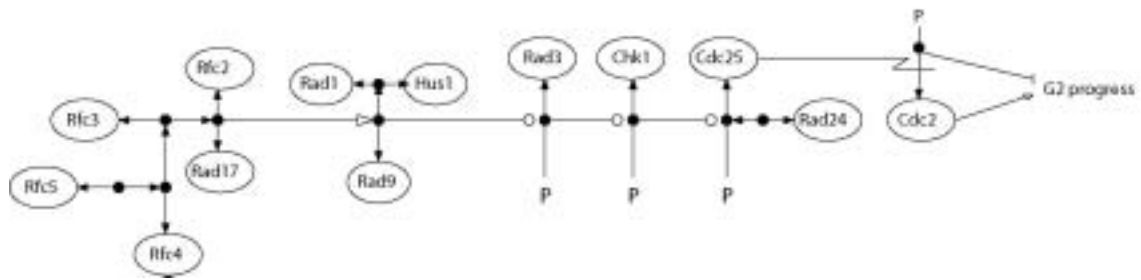
(B) Kohn's Notation

Figure 1: Conventional pathway description and Kohn's description

Trying to create a single diagram that satisfy both characteristics is neither feasible nor desirable, because they use different semantics for symbols and driven by different needs.

Our proposal in this article is that four diagrams shall compose a basic set of diagrams for describing biological networks and their basic dynamics. The four diagrams are state-transition diagram, block diagram, flow chart, and timing chart. These four diagrams are used in VLSI design process as a minimum set of graphical diagrams, a possible addition of data-flow chart, for algorithm logic design. Given that the timing chart represents results of circuit simulation, former three diagrams are considered as a basic set for representing processes in biological networks.

Three diagrams are compatible in their internal model, but visualized from the different aspects and abstraction. A state-transition diagram represents process of biochemical interactions, and visualizes flow of the state transition of the molecules in focus. Block diagrams, on the other hand, represents how each components may interacts. Two diagrams have distinct semantics. In the state-transition diagram, node represents state of the matter, and arc represents transition. In the block diagram, node represents substance and arc represents relationship between two nodes in the both end of the arc. In addition, the flow chart shall represent abstract flow of the biological processes.

4. The State-Transition Diagram

The state-transition diagram represents changes of states of molecules, such as proteins, during the biological processes. It consists of nodes that represent state of the proteins

and other biological entities, and arcs that represent transition of states. Figure 2 is an example of the state-transition diagram for a part of Yeast cell cycle process, involving MPF. Each node represents a state of MPF and other components.
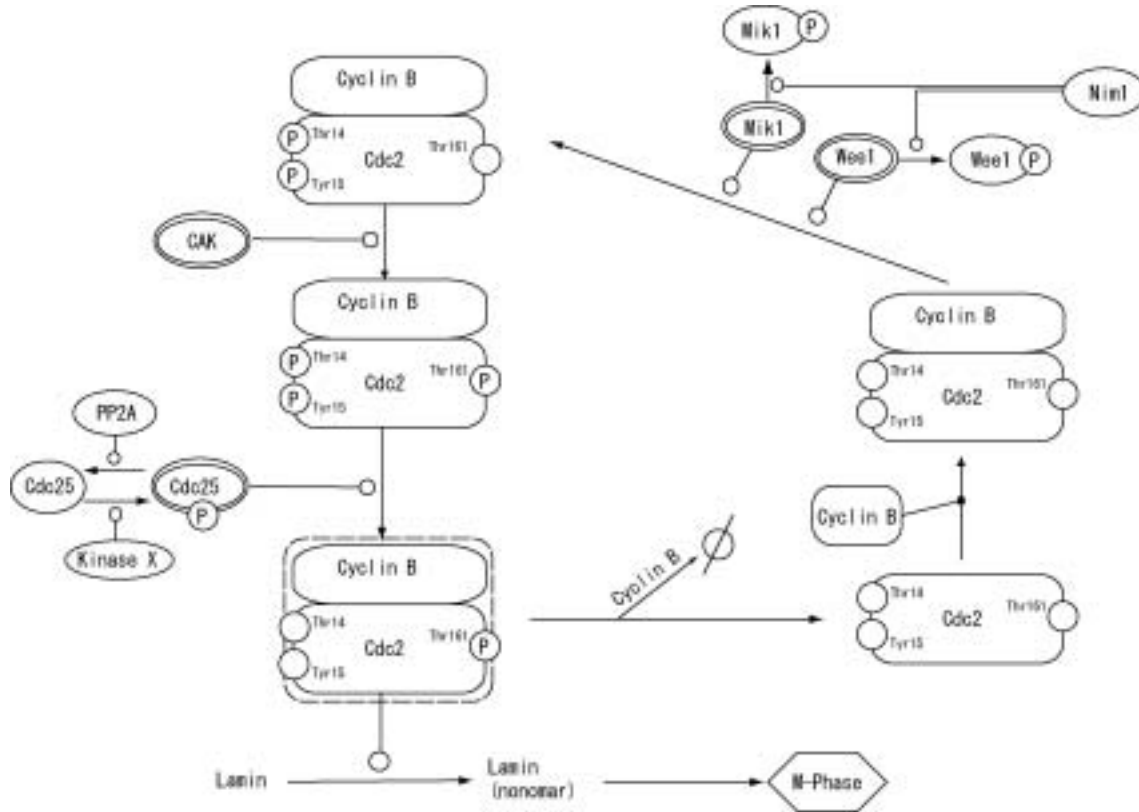
Figure 2: A State transition diagram describing a part of Yeast Cell Cycle process.

Symbols that are used in the state transition diagram are shown in Figure 3. It should be noted that there is no symbols for activation or inhibition. Active state of enzyme is denoted by dashed oval placed surrounding the oval for the enzyme.

It is useful to discuss why activation and inhibition is not on the list of symbols.

The term "activation" and "inhibition" is commonly used in biological literatures, but it must be clearly defined to create a solid notation system. To "activate protein X" actually means "to modify protein X, so that X will have kinase activity." Modifications are, in general, combination of phosphorylation, dephosophorylation, acetylation, deacetylation, etc. For example, p53 protein forms a stable complex that binds to specific DNA sequence. The formation of the complex is promoted by phosphorylation of several serine residues. The other example is M-phase Promotion Factor (MPF) that is a complex of Cdc2 and cyclin B. MPF is generally considered to be "active" when both Thr14 and Tyr15 residues are not phosphorylated and thr161 is phosphorylated. In

figure 2, wee1 changes state of MPF which is inactive, but MPF does not turn into active MPF. Wee1 only modifies two phosphorylation sites. An arrow in this transition denotes changes in the modification state of MPF, not activation.

Symbol	Legend	Symbol	Legend
gene		State transition	
RNA		Catalysis	
Anti-sense RNA		Transcriptional activation	
protein		Transcriptional inhibition	
protein active form		Translational activation	
Ion & small molecules		Translational inhibition	
Receptor		Transport translocation	
		known transitions omitted	
Ion channel		unknown transition	
		Heterodimer formation	
Plasma membrane		Homodimer formation	
Nuclear matrix			
Truncated protein		Transformation	(X)
			(X)
phenotype		Degradation	
hypothetical gene		Truncation	
hypothetical RNA			
hypothetical protein			

Figure 3: A symbol system for the state-transition diagram

Elimination of activation and inhibition symbols from the notation, ironically, makes clear semantics of the symbol system. However, we recognize that this representation is sometime inconvenient when only changes in modification state results in activation or inactivation of the protein, whereas use of activation and inhibition symbols can be conveniently used without losing accuracy of the semantics.

Thus, we can define a set of additional symbols that provides shorthand way of describing "activation" and "inhibition" (Figure 4). Notice that "activation" is shown in double arrowhead, instead of a single arrowhead for state-transition. This extension can be used only when one wish to describe activation and inhibition of protein collateral to the process in focus and when semantics are unambiguous. It shall be used with great care so that semantics of such shorthand is sufficiently clear, and shall avoid the use of such symbols within the main process in focus.
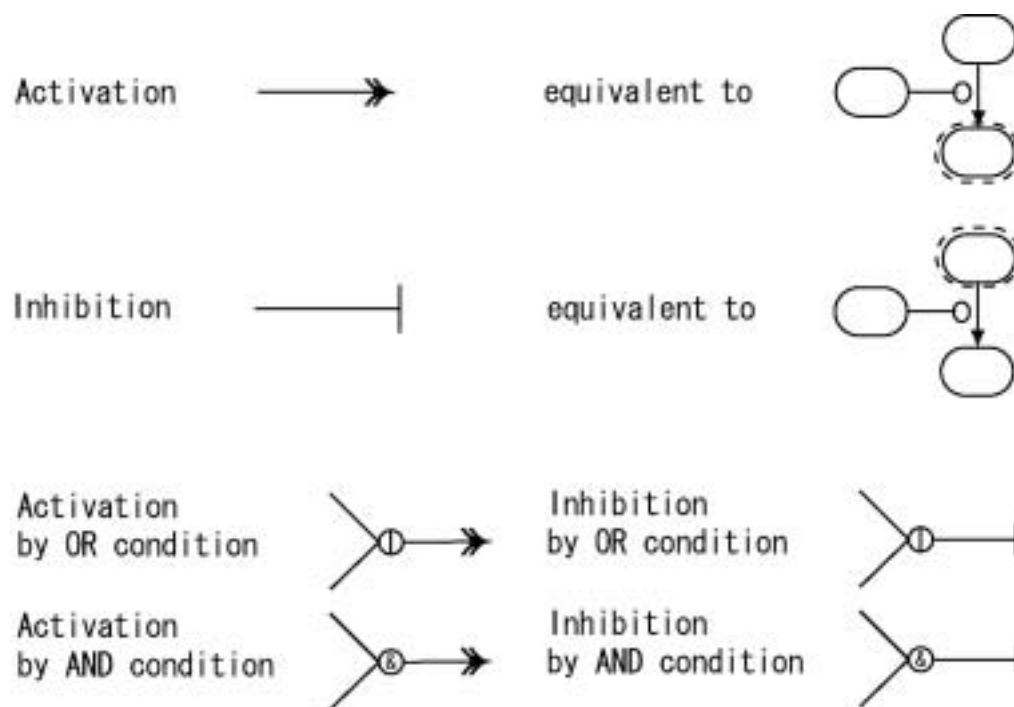
Figure 4: Additional symbols for the state transition diagram
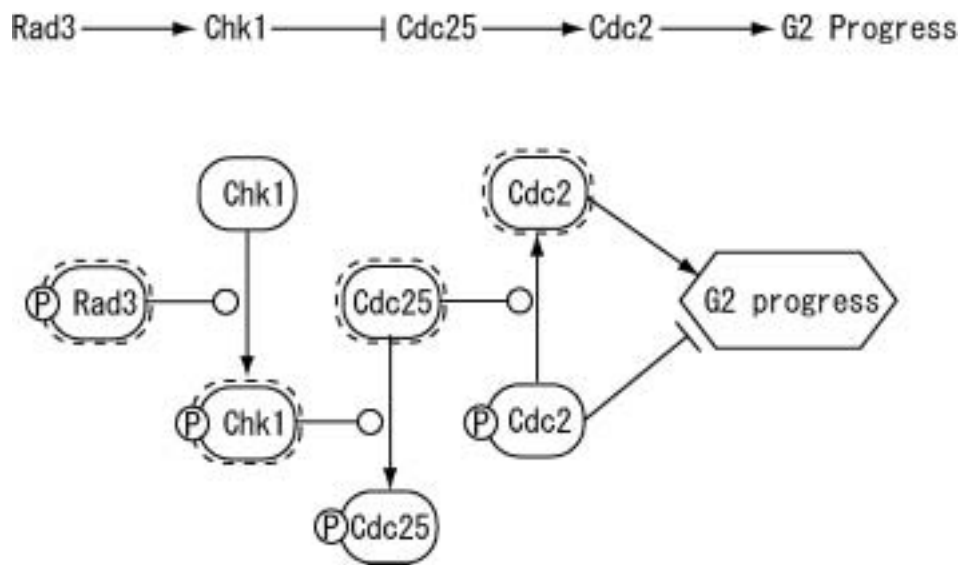
Figure 5: A shorthand representation and a regular state transition diagram

In shorthand notation, as well as in conventional diagrams, a cascade of activation from Rad3 to Cdc2 will be written as Figure 5 (A). In the state transition diagram, it will be written as Figure 5 (B). Both diagrams represent same process, but shorthand diagram do not display information on relationship between phosphorylation and activation. The shorthand diagram is simple and appears to be convenient, but it will be problematic due to loss of information. For example, it cannot indicate whether phosphorylation activates or inhibits protein. In addition, the vital problem is that such diagrams cannot properly represent transition of internal state of proteins. For example, MPF is phosphorylated by CAK without affecting activation/inactivation state. There is no proper way to represent such changes within the conventional diagram. Thus, interaction concerning MPF is usually represented by more explicit state-transition diagram. A shorthand notation using additional symbols is only allowed for limited use.

Complex can be represented with some simplification. When two or more nodes share the boundary, it means that these nodes are in binding state to form a complex. In figure 2, cyclin B and Cdc2 share the boundary, representing binding state, and form a heterodimer complex. Alternatively, a complex can be represented as one round box as shown in figure 6. Components of a complex are represented with partitioned by plus sign, such as "Cdc2+cyclin B". Residues are noted as "position/protein", such as "Thr14/Cdc2", so that which component of the complex residue is located can be unambiguously identified.
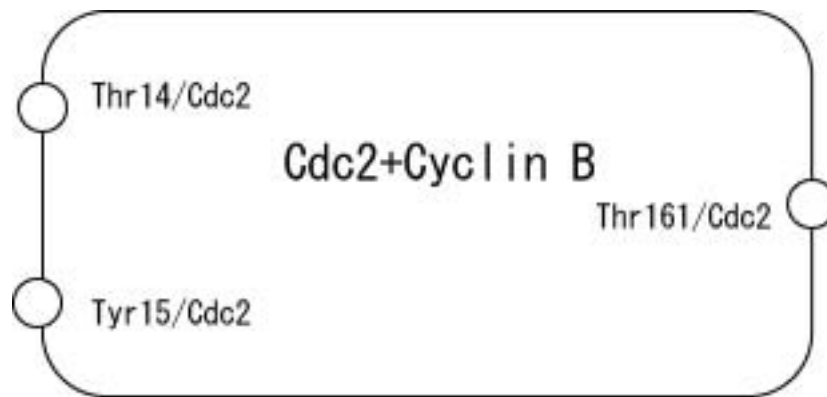
Figure 6: A simple node representation of a heterodimar complex

Round symbols on the boundary of the box indicates state of residues of interest. When each state and the activation state of the protein are known, all modifications on residues can be identified. However, it is often the case whether the states of residues affect activation state or not is unclear, the symbol "Unknown" is used. When the state of the residue does not affect activation state, "Don't Care" symbol is used to indicate this fact (Figure 7).
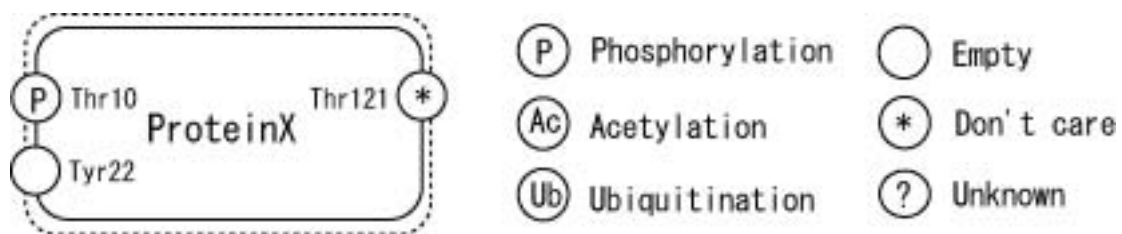
Figure 7: Residue states representation

In the Alliance for Cellular Signaling consortium (AfCS: http://www.afcs.org/), symbols proteins are shown in several different shapes, dependent on their functionalities. When such subdivisions are needed, our notation allows modifications of symbols as follows (Figure 8).

Figure 8: AfCS symbols and extended symbol set for the proposed notation

Most metabolic pathway diagrams are straightforward to fit the state-transition diagram, because current diagram conventions are basically state-transition based. A part of the reason is that chemical species involved in metabolic pathways do not have complex allosteric control, such as phosphorylation and acetylation triggered kinase activity and dimmer formation. A set of transformation symbols will cover substantial part of the metabolic pathways (Figure 9).
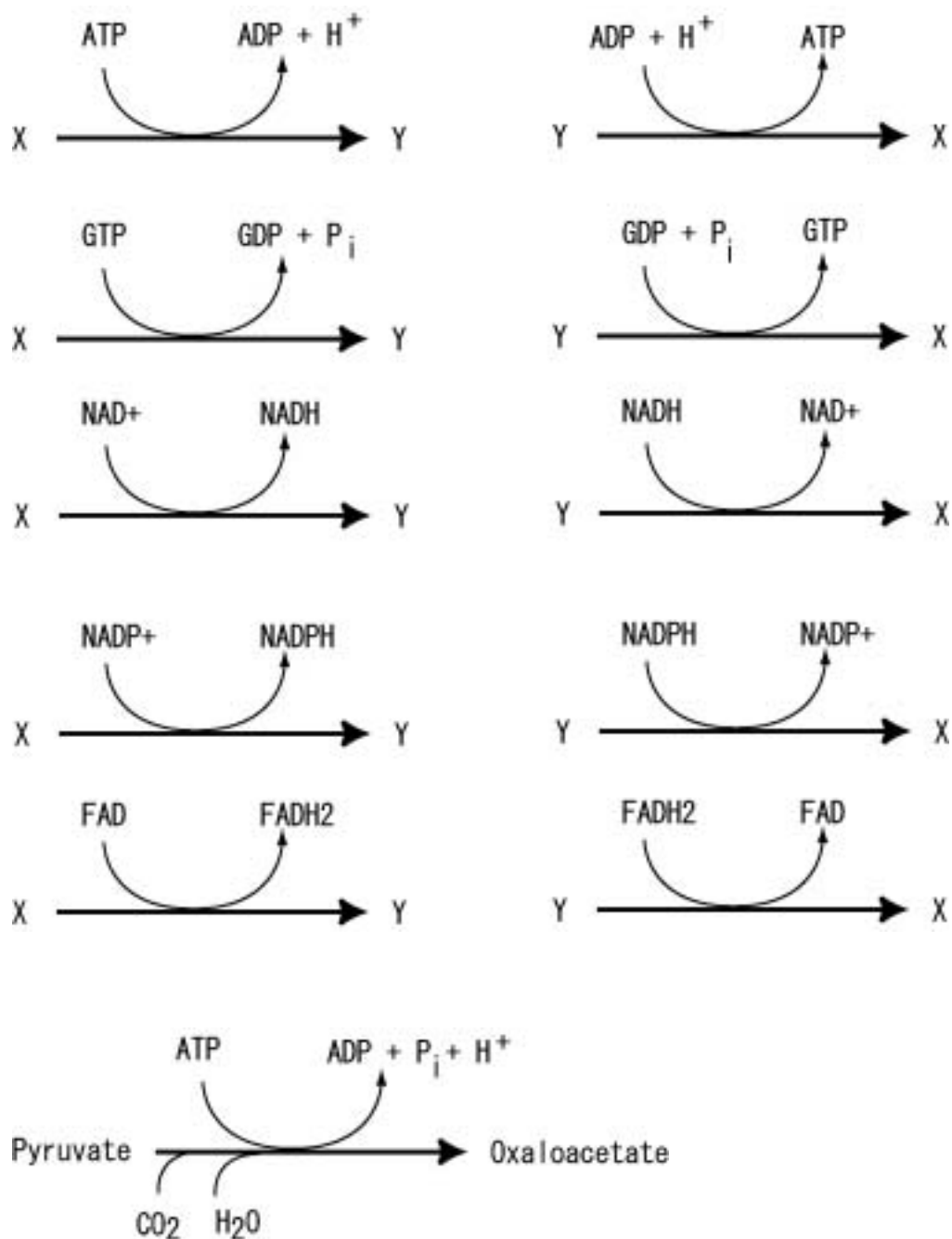
Figure 9: Transformations

Any symbols in the diagram shall have place in ontological hierarchies of biological entities and processes. Ontologically, biochemical networks can be subdivided into biochemical processes and biochemical objects. Figure 10 is a random list, not exhaustive, of that need to be described as a starting point.

Biochemical Processes

        Transcriptional activation / inhibition

        Translational activation / inhibition

        Protein activation / inhibition / degradation

        Phosphorylation / dephosphorylation

        Acetylation / deacetylation

        Metylation / demetylation

        ATP/ADP exchange

        GTP/GDP exchange

        NAD+/NADH exchange

        NADP+/NADPH exchange

        FAD/FADH2 exchange

        Ubiquitionation

        Translocation

        Diffusion

        Anchoring

Biochemical Species

        DNA

        RNA

        Protein

        Small Molecules/Ion

        Receptors

        Ion channels

Figure 10: A List of Processes and Objects to be represented

Fig 11 is an excerpt of ontology hierarchy from Signal Transduction Ontology (http://ontology.ims.u-tokyo.ac.jp/singnalontology/), which is more detailed ontology. It should be noted that there is no "activation" and "inhibition" in the ontology.

```
signal-module
    module
        energy transduction
            enzymatic reaction
                chemical modification on protein
                    chemical-group attachment
                        phosphorylation
                            Tyr phosphorylation
                            Ser-Thr phosphorylation
                        acetylation
                    chemical-group detachment
                        dephosphorylation
                        deacetylation
                guanine nucleotide exchange
                    GDP-GTP exchange
                    GTP-GDP hydrolysis
                second messenger production
                proteolytic cleavage
                    domain cleavage
                    precursor clevage
                    degradation
                        ubiquitin degradation
                        ubiquitination
                degradation
                    ubiquitin degradation
                    ubiquitination
                chemical cytokine production
                chemical hormone production
                modification of enzyme activity
        gene expression
        nuclear transport of protein
        channel transport
            channel open
            channel close
        message passing
            diffusion of second messenger
            enzynm
            transcription factor
```

Figure 11: An excerpt from Signal Transduction Ontology

5. The Block Diagram

The block diagram provides molecular species-centered view of the interaction. This diagram is heavily based on diagram proposed by Cook[4] and Kohn[2]. Rather than emphasizing the step-wise development of the process, it tries to describe all relationships among molecular species. Figure 12 shows a simple example of the relationship diagram for Cdc2. Cdc2 is represented as a round-cornered box. All modification residues are aligned upper side of the box, and all binding sites are aligned lower side of the box. CAK phosphorylates Thr161 residue, and Wee1 phosphorylates thr14 and Tyr15 residues. Cdc25 dephospohrylates Thr14 and Tyr15 residues. There is regulatory relationship described in the round-cornered box. In this example, it indicates that; (1) phosporylation of Thr161 and binding with cyclin B triggers kinase activity of Cdc2, and (2) phosphorylation of Thr14 and Tyr15 residue inhibits effects of phosophrylation of Thr161 residue.
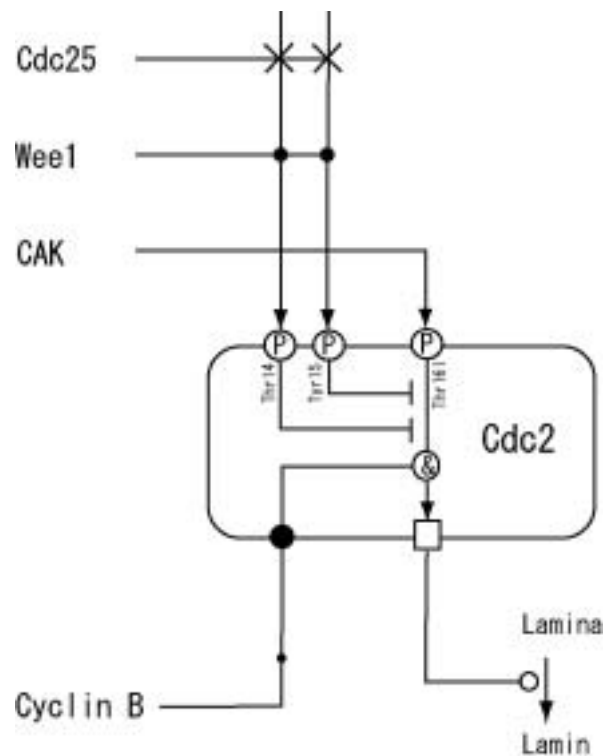
Figure 12: A relationship diagram for Cdc2
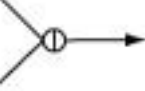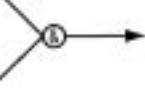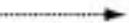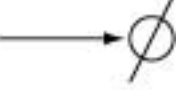
gene		Promotion	
RNA		Inhibition	
Anti-sense RNA		Absolute inhibition	
protein		Transcriptional activation	
Ion & small molecules		Transcriptional inhibition	
Receptor		Translational activation	
Ion channel		Translational inhibition	
		Internal Logic Examples: OR condition	
		AND Condition	
Plasma membrane		Transport translocation	
Nuclear matrix		known interaction omitted	
Truncated protein		unknown interaction	
phenotype		Heterodimer formation	
		Homodimer formation	
hypothetical gene		Degradation	
hypothetical RNA		Truncation	
hypothetical protein			

Fig 13: A symbol set for the block diagram

Figure 13 shows a set of symbols for the relationship diagram. Most symbols are compatible with the state-transition diagram. In this symbol set, the arrow means "promotion", instead of "activation". It is used to indicate "promote phosphorylation", "promote binding with protein X", etc. Actual effect of "activation", kinase activity, for example, is shown using a square box on the boundary of the round-cornered box.
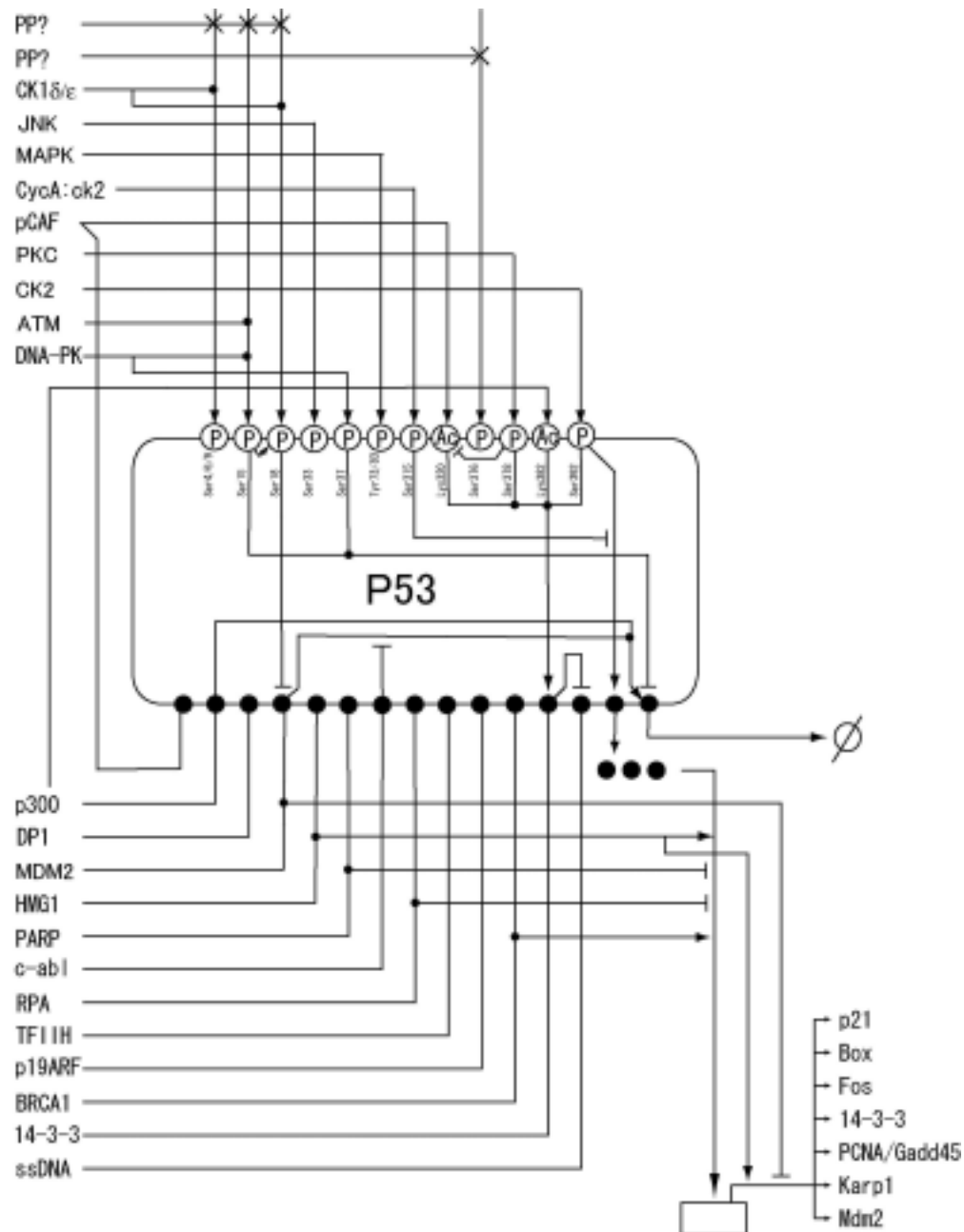

Figure 14: The block diagram for p53

Figure 14 shows more complex diagram for p53 protein. Extensive relationships are

described, but the same principle is maintained. Block diagram can be used to describe logics behind cis-regulatory region[6, 7]. Figure 15 is an example of cis-regulatory region for Endo-16 gene.
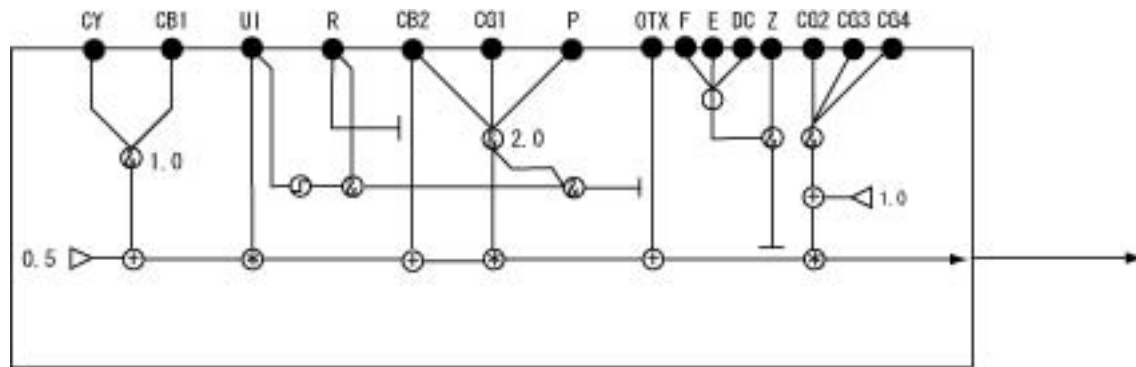

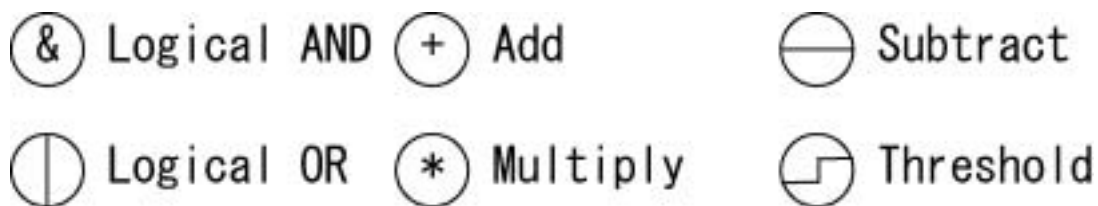Figure 15: A block diagram for endo16 cis-regulation

Figure 16: Internal logic representation

As it can be seen in the figure 15, there are substantial internal logics in the cis-regulation. This is also true for protein modification and bindings. Figure 16 lists nodes for representing internal logics. This logic is a hybrid of boolian logic and numeric calculation, so that both logical and quantitative changes of transcriptional regulations shall be represented. In some nodes, numbers are associated to indicate output value of the node when logical value of the node is true. For example, output of the node, which is logical AND with inputs from CY and CB1, is 1.0 when both CY and CB1 site is bounded by transcription factors.

6. Flow Chart and Timing chart

The flow chart represents a series of biological events that we use to intuitively describe

the process under consideration (Figure 17). This is similar to the event model in Cook's notation. Each node represents intuitively labeled biological state or process at arbitrary level of abstraction. Each node, however, need to be mapped onto the node, or nodes, in the state-transition diagram in order to ground it to the molecular level. While nodes in flow charts only describes intuitive landmarks, mere breakdown of each node and arc would not be able to reconstruct detailed molecular models. Rather this shall be used to glue multiple processes each of which are based on detailed molecular models, or to visually represent state of the biological processes at the intuitive level.

Timing chart is a graph that represents time-course changes of values involved in the model. Typically, this represents concentration levels and activation levels of protein involved.
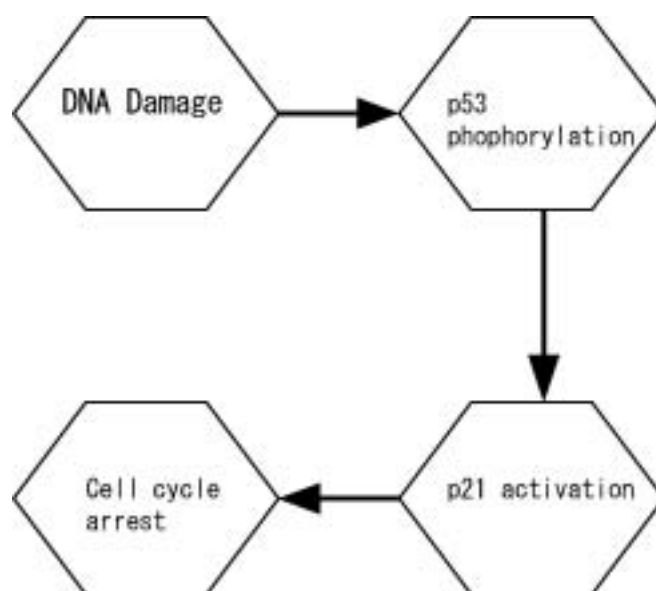

Figure 17: An example of flow chart

7. Availability and Supports

Our intention is to form a standard for graphical representation of biological networks and processes associated. Thus, we intent to submit this proposal to be the basis of further discussions to improve the system of graphical notations that can be widely acceptable to the community. The official standard documentations and review board to for further up-date and revisions shall be created. Given the proper reference is made to this original proposal and anticipated official documentations, we welcome anyone to

use this symbolic representation.

The standard may be a part of SBML or an independent standard linked with SBML and other standardization efforts.

We are also developing software tools to support this graphical representation system. SBEdit, a graphical editor, is now being developed to support efficient creation of models using diagrams in this paper.

8. Conclusion

In this paper, we proposed a graphical representation scheme for biological networks, such as gene regulatory networks and metabolic pathways. The state transition diagrams and the block diagrams are complimentary to each other, and the flow chart provides description of an abstract flow of biological processes. These diagrams are heavily based on proposals made in the past, but have solved some problems such as limitations in expression, difficulties to understand what has been described, etc. There may be addition and modification to this proposal based on the feedback from the actual use. The proposal scheme is expected to be a part of SBML level 3 definitions (or independent standard), and will be implemented as SBEdit network editor software.

References

1.   Kohn, K., *Molecular Interaction Maps as information organizers and simulation guides.* Chaos, 2001. **11**(1): p. 84-97.

2.   Kohn, K.W., *Molecular interaction map of the mammalian cell cycle control and DNA repair systems.* Mol Biol Cell, 1999. **10**(8): p. 2703-34.

3.   Pirson, I., et al., *The visual display of regulatory information and networks.* Trends Cell Biol, 2000. **10**(10): p. 404-8.

4.   Cook, D.L., J.F. Farley, and S.J. Tapscott, *A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems.* Genome Biol, 2001. **2**(4): p. RESEARCH0012.

5.   Maimon, R. and S. Browning. *Diagramatic Notation and Computational Structure of Gene Networks.* in *The Second International Conference on Systems Biology.* 2001. Pasadena.

6.   Yuh, C.H., H. Bolouri, and E.H. Davidson, *Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene.* Science, 1998. **279**(5358): p. 1896-902.

7.   Davidson, E.H., et al., *A genomic regulatory network for development.* Science, 2002. **295**(5560): p. 1669-78.