Molecular interaction maps as information organizers and simulation guides

Kurt W. Kohn
Laboratory of Molecular Pharmacology, Division of Basic Sciences, National Cancer Institute, Bethesda, Maryland 20892

(Received 14 July 2000; accepted for publication 7 November 2000)

A graphical method for mapping bioregulatory networks is presented that is suited for the representation of multimolecular complexes, protein modifications, as well as actions at cell membranes and between protein domains. The symbol conventions defined for these molecular interaction maps are designed to accommodate multiprotein assemblies and protein modifications that can generate combinatorially large numbers of molecular species. Diagrams can either be ''heuristic,'' meaning that detailed knowledge of all possible reaction paths is not required, or ''explicit,'' meaning that the diagrams are totally unambiguous and suitable for simulation. Interaction maps are linked to annotation lists and indexes that provide ready access to pertinent data and references, and that allow any molecular species to be easily located. Illustrative interaction maps are included on the domain interactions of Src, transcription control of E2F-regulated genes, and signaling from receptor tyrosine kinase through phosphoinositides to Akt/PKB. A simple method of going from an explicit interaction diagram to an input file for a simulation program is outlined, in which the differential equations need not be written out. The role of interaction maps in selecting and defining systems for modeling is discussed. © *2001 American Institute of Physics.*
[DOI: 10.1063/1.1338126]

Our knowledge of the molecular interactions that constitute cell regulatory networks is accumulating at an astonishing pace. Organizing this information in the form of clear diagrams is essential, but made difficult by the central roles played by multiprotein complexes and protein modifications. A generally accepted method of representation is needed which can cope with these complexities. To this end, a diagram convention has been proposed,[1] and is here formalized and extended. This type of diagram, called a molecular interaction map, has some of the useful characteristics of a road map or circuit diagram. Any molecular species can readily be located by means of an index and map coordinates, and its interactions can be traced. In addition, each interaction is annotated with salient facts and references. Conventions are defined for ''heuristic'' diagrams which do not require detailed knowledge of interdependencies, and ''explicit'' diagrams which are well-defined and suitable for simulation. Heuristic diagrams help in the formulation of key questions for experiment, and in the design and interpretation of experiments. They also aid in the selection of network modules for simulation. In order to represent schemes for simulation, explicit diagrams can be prepared using a subset of the symbols defined for heuristic diagrams.[2] Explicit diagrams define a reaction input list suitable for a simulation program. Interaction maps may help to select and define systems for modeling.

I. INTRODUCTION

The unusual complexity of biological control networks arises largely from the ability of the constituent molecules to assemble in functional multimolecular complexes and to modify each other enzymatically.[3–5] The number of possible functional states of the complexes can be very large. Most of the enzymatic actions entail charge-altering modifications, such as phosphorylations and acetylations, which can switch binding ability or enzyme activity on or off. Protein kinases or phosphatases add or remove phosphate groups (which add a negative charge to the molecule). Acetylases or deacetylases add or remove acetyl groups (which mask positively charged lysines). Additional complexity arises from translocation between different compartments in the cell, such as nucleus, cytosol, cytoskeleton, and various membrane structures. Movement between compartments also is largely dependent on specific charge modifications. Moreover, the proteins often are composed of several domains, each specifying a binding or enzyme capability, or bearing a translocation signal, and each domain may be subject to modification.

In order to provide a sound foundation for further progress, we need a clear method of representation, such as a roadmap or circuit diagram. The nature of the complexities however makes the methods used for charting metabolic pathways unmanageable. A graphical method was recently proposed for mapping regulatory networks.[1] Presented here is a formalization of the mapping conventions and an extension of the symbology to accommodate actions at cell membranes and between protein domains. The symbol conventions are designed for multiprotein assemblies and protein modifications which generate combinatorially large numbers of molecular species.

An important characteristic of molecular interaction maps is that each named molecular species generally appears in no more than one place on any map. Molecular interac-

 © 2001 American Institute of Physics

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

tions and complexes are represented by means of a defined system of lines and nodes. These features make it relatively easy to find the interactions and modifications of any molecular species and to trace its possible interaction paths. This type of molecular interaction map can be thought of as a ''heuristic'' diagram, because it allows some degree of uncertainty while clarifying the known essentials. We often do not know the exact order in which multiprotein complexes are assembled, or the effects on the activity of a given molecular species caused by its possible modifications and binding states. Even if we did have all the facts, representing all of the combinatorial possibilities would be prohibitive. Heuristic diagrams allow desirable uncertainties and simplifications to be retained. On the other hand, schemes for computer simulation can be prepared using a subset of the symbols defined for ''heuristic'' maps. Such ''explicit'' diagrams include all of the interactions and contingencies of the model to be simulated.

A major issue in realistic simulations of biological regulatory systems is that much of the biological computation occurs in large multimolecular assemblies that may function akin to integrated circuits. These assemblies can be built up along many different paths whose parameters must be thermodynamically self-consistent.[3,6] Moreover, there is a stochastic aspect which confers variability in the timing of events while keeping the events in proper sequence.[7,8] In dealing with these complexities, it is often important to focus on the domains and interaction sites within individual proteins. The molecular interaction maps described here allow this intramolecular fine structure to be depicted. A formalization and extension of the previously described molecular interaction map symbols and conventions[1] is presented. Application to areas not previously included, such as events at the plasma membrane, are illustrated. Second, the linkage between explicit diagrams and simulation programs are summarized. Finally, the question is considered of how to identify biological network modules of potential interest for simulation.

II. FORMAL DESCRIPTION OF SYMBOLS AND CONVENTIONS

Schemes for computer simulation demand complete lack of ambiguity and must specify all interactions explicitly. For information mapping, however, these requirements are unnecessarily burdensome, if not totally unmanageable. Conventions were therefore devised for ''heuristic'' diagrams in which these demands are relaxed, so that it is not necessary to specify all routes to the assembly of a multimolecular complex, or the details of how different interactions or modifications of a given molecular species may affect each other. Details at this level are often unknown. For computer models, however, rules for well-defined ''explicit'' diagrams will apply.

A. Lines and nodes

Molecular interaction maps are especially convenient if each named molecule is shown only in one location on a map. Any given molecular species is then easily located and its interactions can readily be traced. In order to accomplish this objective, we need a variety of defined connecting lines, and a flexible method of representing multimolecular complexes and their modifications. We depict binary interactions by means of a variety of line types, distinguished by different types of arrowheads placed at the ends of the line. We depict multimolecular complexes or modified forms by means of ''nodes'' placed on the lines. The line-types and node definitions are summarized in Fig. 1.

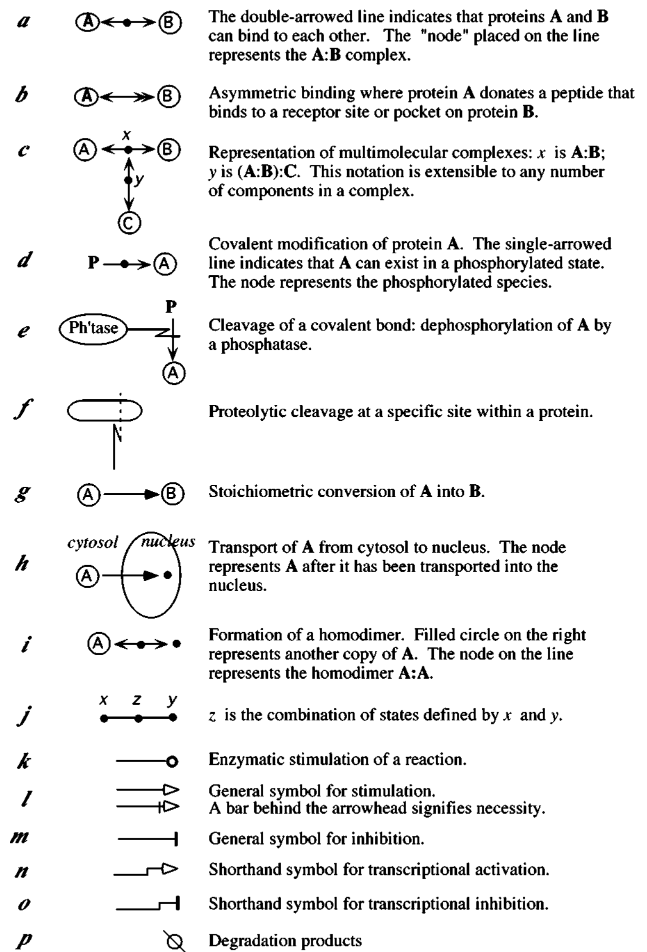

FIG. 1. Summary of symbols.

For our purposes, a ''line'' is a path from one point to another. A line may originate either at a named molecular species or at a node, and may terminate at a named species, a node, or at another line. Abbreviated forms are defined that allow lines to branch or become discontinuous, and that allow isolated nodes to function as ditto marks. When lines cross, no interaction between them is implied.

A node placed within a line (not at the ends) represents the consequence of the implied binary interaction. For convenience and clarity, multiple nodes may be placed within the same line; these nodes then refer to exactly the same thing. A node should not appear at an intersection between lines when this would create an ambiguity.

When a line indicates reversible binding [line with barbed arrowheads at both ends, Fig. 1(a)] a node within the line represents the complex formed by the combination of the species defined at the ends of the line. When the line indicates modification of a protein [line with a barbed arrow-

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

head at one end only, Fig. 1(d)], a node within the line represents the modified protein itself. When the line indicates a combination of molecular states [line without arrowheads, Fig. 1(j)], a node within the line represents the combination (set theory ''union'') of the molecular states defined at the ends of the line (e.g., a combination of modifications of a protein, or an entity consisting of two molecular species that are not necessarily bound to each other).

B. Reversible binding

As already mentioned, reversible binding between two molecular species is represented by a line having a barbed arrowhead at each end, and the resulting complex is represented by a node placed within the line [Fig. 1(a)]. This notation is extensible, since a node represents a molecular species which can be connected to other species by means of interaction lines [Fig. 1(c)]. In this way, multimolecular complexes can be built up.

When a given species can engage in multiple interactions, these interactions may or may not be interdependent. In the case of heuristic diagrams, the default assumption is that the interactions are independent, meaning that they can coexist in the same complex. Dependencies can be specified by means of qualifier symbols, to be described in the following. In the case of explicit diagrams, all molecular complexes and interactions included in the model are specified.

A multimolecular complex may be assembled by many routes, and the preferred routes may or may not be known. If we know, for example, that a particular heterodimer forms before the addition of a third component, then the manner of representation is clear. If we do not have this information, or if a complex can build up by different assembly paths, heuristic diagrams allow a compromise. Although it is possible to represent any number of assembly paths, this tends to clutter the diagram. An appropriate solution is to depict one plausible path for representation and to provide further information in the annotation list.

Binding between two proteins can be asymmetric, in the sense that one protein serves as donor of a peptide region that binds in a groove within the second protein which serves as receptor.[9] It is sometimes useful to distinguish donor from receptor, because competition of different donors may occur on the same receptor, and because it identifies opportunities for pharmacological intervention. The protein bearing the receptor can be indicated by doubling the barb on the arrowhead pointing to this protein [Fig. 1(b)].

Figure 2 illustrates a concise way to represent a pattern of interactions in which a given binding site can combine competitively with more than one species. The example shown in Fig. 2 is the pattern of competitive interactions of cyclins E, A, and B with the kinases cdk1 and cdk2. Cdk2 can bind either cyclin E or A; cdk1 can bind either cyclin A or B; and cyclin A can bind either cdk2 or cdk1. Competitive binding is represented by branched binding lines. In order to avoid ambiguity, an acute angle is employed at the branch point to distinguish the competing species from the site of competition. The individual binary complexes can be represented by appropriately placed nodes (*a, c, e*, and *g* in Fig.

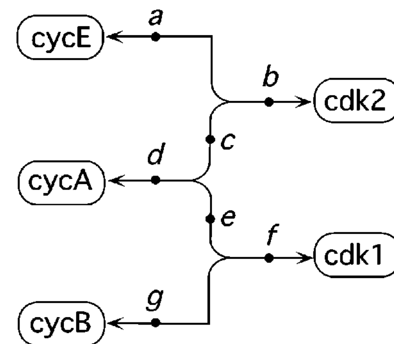

FIG. 2. Concise representation of alternative modes of binding between Cyclin E, A, or B and Cdk1 or 2. The nodes represent the following heterodimers of heterodimer combinations: (a) CycE:Cdk2; (b) Cdk2 bound to either CycE or CycA; (c) CycA:Cdk2; (d) CycA bound to either Cdk1 or 2; (e) CycA:Cdk1; (f) Cdk1 bound to either CycA or CycB; (g) CycB:Cdk1.

2). In addition, nodes can be placed to represent pairs (or groups) of complexes; e.g., node *b* represents cdk2 bound to either cyclin E or cyclin A. This simplifies the representation, for example, of reactions catalyzed by cdk2 bound to either cyclin E or cyclin A.

C. Covalent modification

Covalent modification of a protein is represented by a line having a barbed arrowhead pointing to the site of modification, while the other end of the line, at which the type of modification (e.g., phosphorylation) is indicated, has no arrowhead [Fig. 1(d)]. A node placed on this line denotes the modified protein itself.

Regulatory proteins often have multiple sites of modification, which, singly or in combination, can affect the protein's function. Multiple sites can be arrayed along the rounded box representing a named protein from N- to C-terminus, and the amino acid site can be indicated as a superscript on the modification type symbol (e.g., P^{Ser15} for phosphorylation at serine-15) [e.g., see the representation of the multiple phosphorylation sites on p53 in Fig. 6(B) of Ref. 1—or at http://discover.nci.nih.gov/kohnk/interaction_maps.html]. Multiple modification states can be denoted by using nonarrowed state-combination lines (discussed later in connection with Fig. 5).

D. Stoichiometric conversion

A reaction or conversion in which one or more species (e.g., reaction products) are generated while other species (e.g., reactants) are stoichiometrically consumed is denoted by a solid arrowhead, as shown in Fig. 1(g). The same symbology can be used to represent the transport of a molecular species from one compartment in the cell to another [Fig. 1(h)], since this is essentially a stoichiometric appearance and disappearance of molecules in different places.

E. The isolated node ditto mark

Figure 1(h) illustrates the use of a convenient abbreviation, in which an isolated node is used to represent another copy of the species represented at the other end of the line.

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

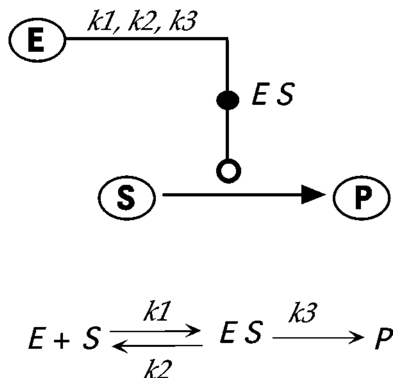

FIG. 3. Simple one-way enzymatic reaction. (If there is an energy source, such as ATP hydrolysis, it can be omitted when ATP concentration is not an important factor.) In explicit formulations, the reaction identifiers or rate constant designations can be placed on the enzyme reaction line, and the node ES can identify the enzyme-substrate species.

Another example of the isolated node convention is shown in Fig. 1(i), which depicts a homodimer. The double-arrowed line thus indicates the reversible binding of **A** to another copy of itself, and the homodimer is indicated by a node on the line. The isolated node convention often eliminates the need to duplicate the detailed interactions of a protein in different places in the same diagram.

F. Enzymatic reactions

Enzyme actions in regulatory networks are often one-way reactions, involving ATP hydrolysis, as in protein kinases, or involving bond scission, as in phosphatases or proteases. In these cases, three reactions are implied: (1) binding of enzyme (E) to substrate (S) to form ES; (2) dissociation of ES to restore E and S; (3) conversion of ES to E plus product (P). In explicit formulations, the three reactions or rate constants are marked on the enzyme action line, and a node can be placed on this line to represent the ES complex (Fig. 3). (For reversible enzymatic reactions, the line connecting S and P would have arrowheads at both ends, and appropriate rate constants could be added.)

Sometimes an enzyme catalyzes reversible binding, as in the case of a guanine nucleotide exchange factors (GEF). Thus Ras, a GEF substrate, binds either GDP or GTP tightly and almost irreversibly at a specific binding site. Interaction of Ras with GEF confers reversibility. Usually this entails dissociation of GDP and replacement by GTP (because GTP is normally present at much higher concentration than GDP). The effect of GEF on Ras can be represented concisely as shown in Fig. 4. (Figure 4 also shows how to represent the enhancement of an intrinsic enzyme activity, in this case by the Ras–GTPase-activating-protein, RasGAP.)

G. Stimulation and inhibition contingencies

Stimulation of a process is represented by an open triangle arrowhead. If the stimulatory element is *necessary* for the process, then a bar is added behind the open triangle arrowhead [Fig. 1(l) and Fig. 5]. Inhibitory effects are denoted by a bar at the end of the line [Fig. 1(m)]. (A stimulated or inhibited process could be binding, modification, enzymatic action, stoichiometric conversion, or another stimulatory or inhibitory effect.)

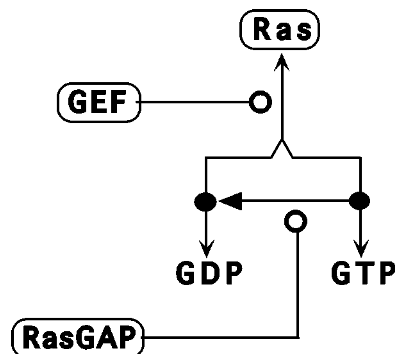

FIG. 4. Interconversions between the GTP- and GDP-bound states of Ras. (1) GDP and GTP compete with each other for binding to a site on Ras (this binding is only slowly reversible). (2) GEF (guanine nucleotide exchange factor) facilitates the binding or dissociation of GDP or GTP (the concentration of GTP normally far exceeds that of GDP). (Implicit is the reversible binding between GEF and Ras which opens the binding site for GDP/GTP exchange.) (3) Ras has an intrinsic GTPase activity that slowly converts bound GTP to bound GDP (stoichiometric conversion arrow points from the node representing Ras.GTP to the node representing Ras.GDP). (4) RasGAP (a GTPase activating protein) enhances the GTPase activity of Ras. (Implicit is the reversible enzyme–substrate binding between RasGAP and Ras.)

When more than one stimulatory (or inhibitory) effect impacts upon a process, we sometimes want to distinguish between AND versus OR logic. This can be done as suggested in Fig. 5 and explained in the legend.

H. State-combination connectors

Lines without arrowheads can be used to denote combinations of modification states or combinations of molecular species that are not necessarily bound to each other [Fig. 1(j)]. Figure 5(a) illustrates the use of a state-combination line to denote a process requiring two interacting proteins to be in phosphorylated states.

III. MULTIMOLECULAR COMPLEXES AND SCAFFOLDS: HEURISTIC AND EXPLICIT REPRESENTATIONS

Many biochemical processes are facilitated within multimolecular assemblies which may function like solid-state circuitry. Protein assemblies are favored by the multiple binding sites on multidomain proteins, by adapter and scaf-

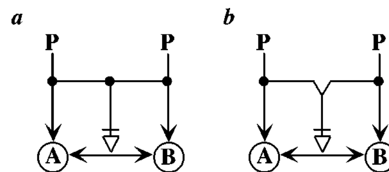

FIG. 5. AND vs OR logic for a combination of two stimulatory effects. (a) Binding between A and B requires that both are phosphorylated (AND logic). A state-combination connector (line without arrowheads) is used to indicate the combination of the two phosphorylation states. (b) Binding between A and B requires that either A or B (or both) is phosphorylated (inclusive OR).

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

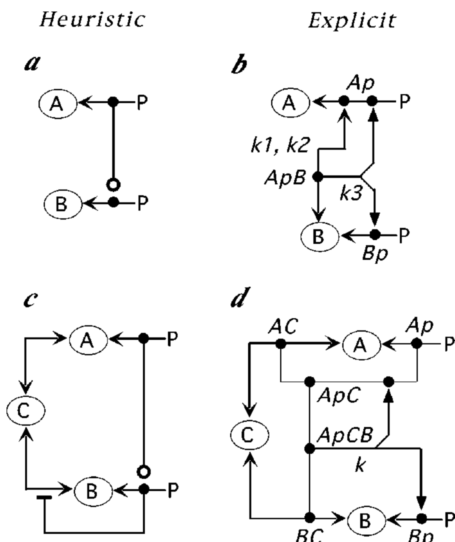

FIG. 6. One step in a phosphorylation cascade, without [(a) and (b)] or with [(c) and (d)] scaffold. Heuristic [(a) and (c)] and explicit [(b) and (d)] representations are compared. See Sec. III.

folding proteins, and by the crowded solvent conditions within the cell where protein molecules compete for water.[3] Signal transmission and computation often take place within such complexes.

In a variety of circumstances, reactions are made more efficient or specific by a scaffolding protein that brings together enzyme and substrate.[10] In the case of signaling through MAP kinases, there is a cascade of phosphorylation events, where the phosphorylated form of one protein catalyzes the phosphorylation of a second protein. These kinases generally are held together by a scaffolding protein.[11] Figure 6 shows one link of such a phosphorylation chain, with or without a scaffold protein, and comparing heuristic and explicit representations.

In the case without scaffold, the heuristic diagram [Fig. 6(a)] consists simply of a node representing phosphorylated **A** and an enzymatic action line operating to phosphorylate **B**. The explicit equivalent [Fig. 6(b)] consists of a reversible binding line between Ap (phosphorylated **A**) and **B** (rate constants $k1$ and $k2$), forming enzyme-substrate complex ApB, and a pair of stoichiometric conversion arrows indicating the conversion of ApB to product (Bp) and regenerated enzyme (Ap) (rate constant $k3$).

A scaffold protein [**C** in Figs. 6(c) and 6(d)] can simultaneously bind enzyme (Ap) and substrate (**B**), allowing the reaction to occur within the complex (by heuristic convention, **C** can bind either **A** or Ap in Fig. 6(c)]. The heuristic diagram [Fig. 6(c)] also indicates binding capabilities between **C** and **A** and between **C** and **B**. By convention, the bindings are presumed to be independent of each other and independent of the phosphorylation states of the component proteins (unless otherwise indicated by contingency symbols). A contingency is in fact needed in this example, because product must be released from the complex if the reaction is to proceed efficiently. The release of phosphorylated product Bp is shown by the inhibition operator from Bp to the **B**-to-**C** binding line.

The explicit equivalent is shown in Fig. 6(d). The conventions applying to explicit diagrams are more precise and restrictive. In heuristic diagram Fig. 6(c), node Ap meant phosphorylated **A**, regardless of other interactions in which **A** might be engaged, By contrast, node Ap in the explicit diagram means phosphorylated monomer **A** only, The complex consisting of phosphorylated **A** bound to scaffold protein **C** (node ApC) is depicted using a state-combination line connecting Ap and AC. The enzyme-substrate complex within the scaffold complex (ApCB) is depicted using a state-combination line connecting ApC and BC (C is counted only once because of the set theory union rule). The conversion of ApCB to products (rate constant k) is represented by a pair of stoichiometric conversion arrows leading to phosphorylated **B** (Bp), which is the product, and back to ApC. Note that, according to the convention applying to explicit diagrams, Bp refers to the monomeric species that is no longer bound to **C**.

IV. EXPLICIT DIAGRAMS DEFINE REACTION SCHEMES FOR SIMULATION

Each interaction in an explicit diagram can be read as an entry into a connection table suitable for input to a simulation program.[2] Figure 7 and Table I illustrate how this works. Each molecular species in the diagram [Fig. 7(a)] is first assigned an identifying number and an initial concentration; these are listed in a molecular species file (Table I). Each reaction then defines a one line entry into a reaction file, which also includes the rate constants. The set of reactions is iterated in small time steps using simple mass action laws. This is equivalent to iterating a set of ordinary differential equations: Each reaction corresponds to one term of a differential equation. The major advantage of this procedure is the direct relationship between reactions and program entries, which makes it easy to modify the reaction scheme without having to rewrite differential equations.

Another feature is that ad hoc terms, such as Michaelis–Menten steady-state approximations or Hill coefficients are avoided. All of the reactions are specified explicitly.

The example in Fig. 7 illustrates also how a simple simulation can suggest unexpected behavior amenable to experimental observations.[2] The simulation concerns the mechanism that controls the onset of DNA replication phase in the mammalian cell cycle. Many of the genes that must be activated when cells prepare for DNA replication are controlled by an E2F transcription factor. E2F proteins are held in check by binding to the retinoblastoma protein, Rb. Activation of E2F is triggered by cyclin-dependent kinases (cdk) that phosphorylate Rb with consequent release of active E2F. The conventional view is that Rb inhibits and cdk stimulates the accumulation of active E2F. The simulation shows however that, under some circumstances, Rb could enhance the production of a wave of active E2F [Fig. 7(b)]. Moreover, high concentrations of cdk could paradoxically inhibit [Fig. 7(c)]. The seemingly paradoxical behavior would arise due to Rb collecting a store of E2F, much as a condenser stores charge, and releasing active E2F quickly when all of the Rb has become phosphorylated. Several more complete and

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

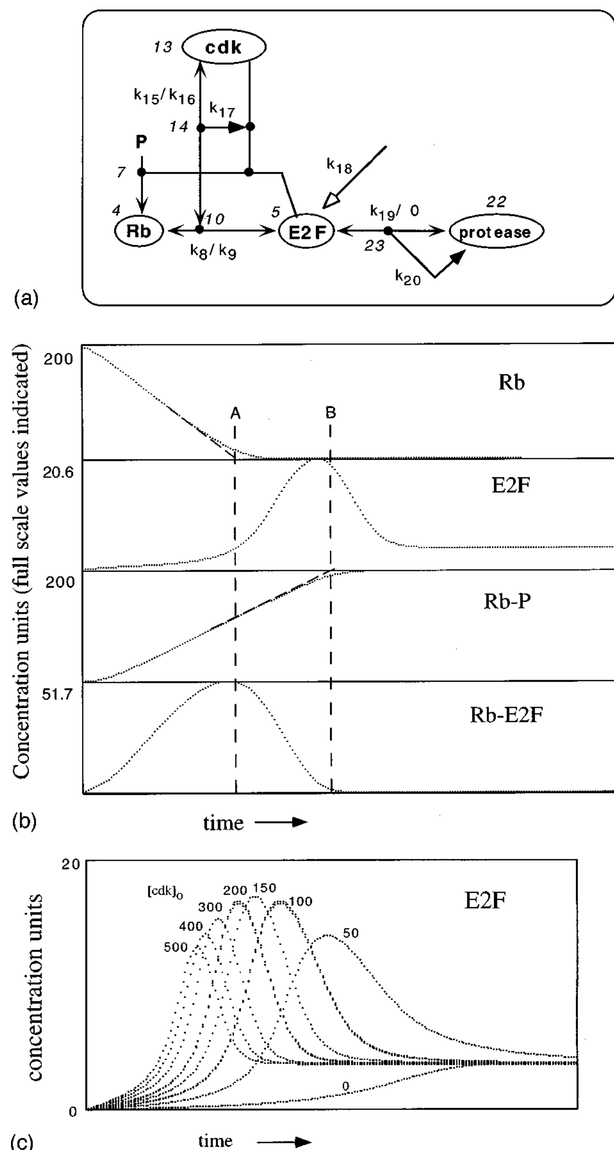

FIG. 7. Going from an explicit interaction diagram to simulation results; see also Table I. Each molecular species in the diagram (a) is assigned a number; each reaction is associated with a rate constant symbol. The translation from diagram to computer input files is shown in Table I, which also lists the rate constants and initial concentrations used for the simulation results shown. The reaction scheme models a simple configuration of components that control the onset of DNA replication. Note the large E2F activity which may arise from a paradoxical effect of Rb (b). (c) Unexpectedly shows that there could be an optimum cdk concentration for the generation of a large wave of free E2F. This example is derived from Ref. 2.

more complex variations of the reaction scheme yielded similar behavior, and some of them showed more striking inhibition by high concentrations of cdk.[2]

In this study,[2] a sequence of reaction schemes of increasing complexity was studied. The concept was to select from the known interaction network the simplest subsystem that had plausible functional capability, and then to add components stepwise. In this way a ''quasievolutionary'' sequence of simulations was obtained which seemed to give insight into what might be the fundamental or primordial part of the modern highly evolved and complex network.

TABLE I. Input files for a simulation program, corresponding to the explicit diagram in Fig. 7(a). The molecular species file assigns an identifying number and a text identifier to each molecular species occurring in the explicit interaction diagram. Initial concentrations are also assigned. The reaction file assigns an identifying number to each reaction in the diagram. For each reaction, provision is made for up to three reactant species and up to three product species. Rate constants are also assigned. Each reaction is iterated according to a simple mass action law. The minus sign prefixed to species 20 in reaction 18 flags the program not to decrement this species during the reaction, thereby allowing its concentration to be maintained constant. The missing numbers in the species and reaction identifier lists were skipped, because they identified species and reactions occurring in other simulation schemes but not in this one. Because the reactions are treated in a homogeneous manner, it is easy to modify the reaction scheme.

Molecular species file

Species number	Initial conc.	Species identifier
4	200	Rb
5	0	E2F
7	0	RbP
10	0	Rb.E2F
13	20	Cdk
14	0	Cdk.Rb.E2F
20	1	pre-E2F
22	1	protease
23	0	E2F.protease

Reaction file

Reaction number	Reactant species			Product species			Rate const (k)
8	4	5	0	10	0	0	0.1
9	10	0	0	4	5	0	0
15	10	13	0	14	0	0	0.2
16	14	0	0	10	13	0	0.2
17	14	0	0	5	7	13	0.4
18	−20	0	0	5	0	0	8
19	5	22	0	23	0	0	4
20	23	0	0	22	0	0	16

V. COMPUTATION BY INTRAMOLECULAR INTERACTIONS BETWEEN PROTEIN DOMAINS AND BY INTERMOLECULAR INTERACTIONS AT THE PLASMA MEMBRANE: SRC-FAMILY TYROSINE KINASES

We consider here two important components of molecular computation: interactions between domains within the same protein molecule and interactions between membrane-bound protein molecules. As an example, we consider the membrane-associated tyrosine kinase, Src, which is made up of several defined domains and whose structure and function have been elucidated in considerable detail.[12] Src is a member of a large family of protein tyrosine kinases that control a variety of cell functions in response to signals from outside the cell. These proteins are bound to membranes by way of hydrocarbon chains attached to their N-termini. They function in coordination with receptor tyrosine kinases, transmembrane proteins that respond to extracellular signaling molecules. For example, the EGFR receptor tyrosine kinase (also known as ErbB1) is a transmembrane protein that responds to epidermal growth factors (EGF). Figure 8 shows some of the essential interactions of the system involving Src and EGFR and reveals the circuitry that may underlie its function.

The domains of Src are shown left to right from N- to

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

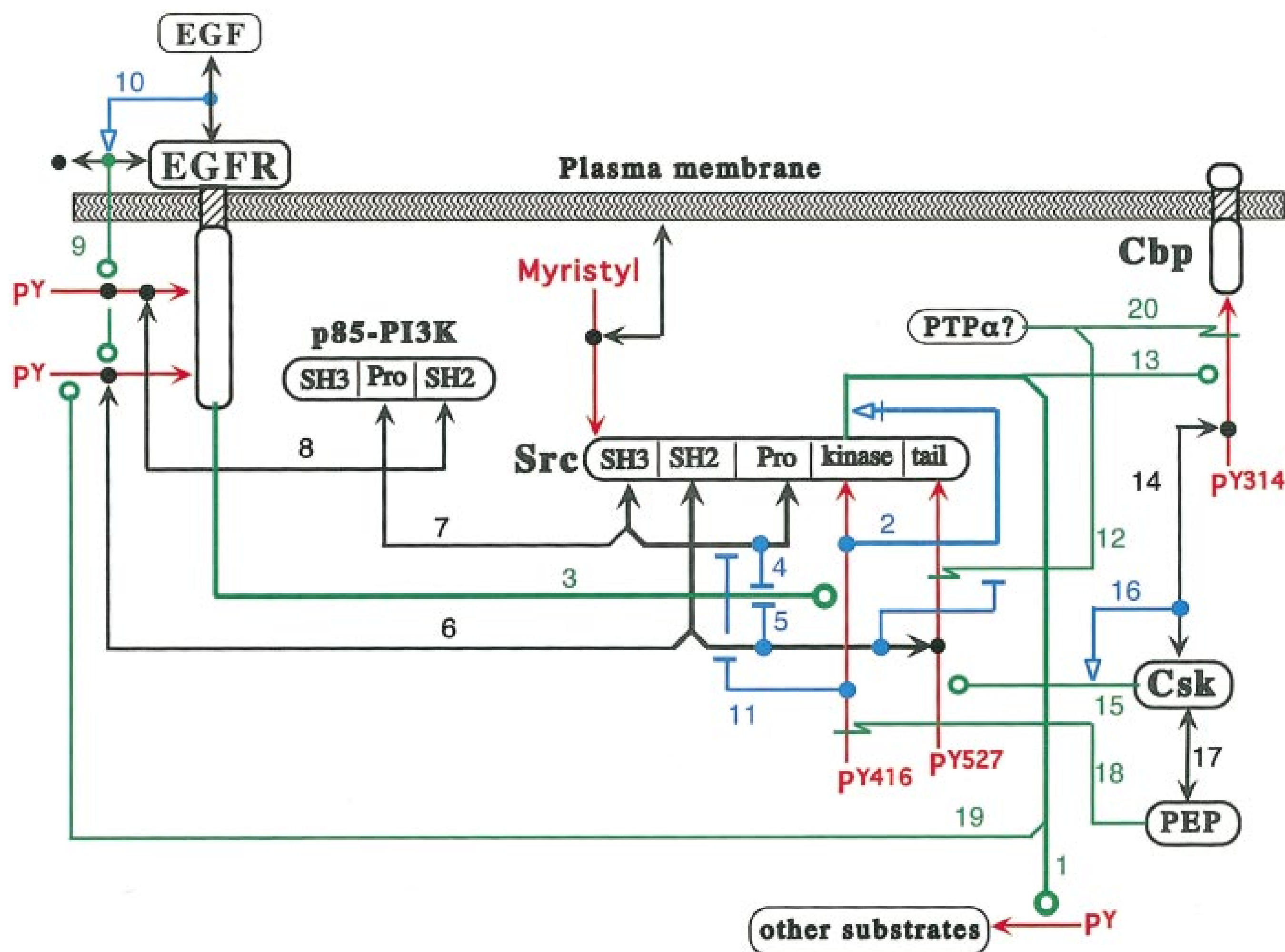

FIG. 8. (Color) A bistable regulatory system involving intra- and intermolecular interactions at the plasma membrane: interactions of Src and a receptor tyrosine kinases, EGFR. (Optional color code: black for binding interactions and stoichiometric conversions; red for covalent modifications and gene transcription; green for enzyme actions; blue for stimulation and inhibition.)

C-terminus:SH3-SH2-Pro-kinase-tail. The N-terminus bears a covalently bound hydrocarbon chain (myristyl group) which binds the Src molecule to membranes. Pro designates a proline-rich sequence that can bind the SH3 domain in the same Src molecule. The SH2 domain in Src can bind phosphorylated tyrosine-527 (Y527) in the C-terminal tail region of the same molecule. These two intramolecular linkages cooperate to maintain the Src molecule in a folded state. The kinase domain remains in an inactive conformation unless tyrosine-416 (Y416) within the kinase domain is phosphorylated. When Src is in a folded state due its two intramolecular linkages, Y416 is not accessible for phosphorylation and the kinase remains inactive. Activation requires release of the intramolecular linkages, so that the molecule can unfold and make Y416 accessible to phosphorylation. As long as Y416 remains phosphorylated, Src cannot refold.[13] These interactions and consequences are depicted in Fig. 8. The broader function of the system, also depicted in Fig. 8, will now be described.

We will trace how Src becomes activated, and then how it becomes inactivated. We will see how Src could behave in a bistable manner with hysteresis. The numbers refer to the interactions as labeled in Fig. 8. (1) The output of the system is the kinase activity of Src which can phosphorylate a variety of protein substrates at tyrosines. (2) A strict requirement for the kinase activity of Src is that Y416, which is located within the kinase domain, must be phosphorylated. (3) Y416 can be phosphorylated by a receptor tyrosine kinase such as EGFR. (4) This phosphorylation however is inhibited by the intramolecular binding between the SH3 and Pro domains, and (5) between the SH2 domain and a tyrosine-phosphorylated site (Y527) in Src's tail region. Both of these intramolecular linkages must be disrupted before the kinase domain can be activated. (6) The SH2-bound Y527 can be displaced by a phosphotyrosine on a receptor tyrosine kinase, such as EGFR.[14] (7) The SH3-bound Pro can be displaced by the regulatory p85 subunit of phosphatidylinositol-3′-kinase (PI3K) which, like Src, contains a proline-rich SH3-binding site. (These binding motifs have selectivity for amino acid context. For example, the SH3 domain of p85-PI3K does not bind the Pro of Src.) (8) p85-PI3K can be recruited to the site of action by binding via its SH2 domain to another phosphotyrosine on EGFR. (9) Several tyrosines on the intracellular domain of EGFR undergo autophosphorylation when EGFR is induced to dimerize. (10) Dimerization of EGFR is induced upon binding of EGF to the extracellular domain. The

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

binding and enzymatic capabilities of EGFR thus work together to unfold the Src molecule and phosphorylate Y416. (11) Phosphorylation of Y416 prevents the refolding of Src to its inactive conformation.[13]

(12) When Src is in the unfolded state, the phosphorylated Y527site becomes accessible to a protein tyrosine phosphatase (possibly PTPα[15]) which removes the phosphate from this tyrosine residue. This removes the SH2-binding site. While Y416 is phosphorylated and Y527 is unphosphorylated, Src cannot refold, and its kinase domain remains active even when Src dissociates from the receptor (which it can do because these bindings are reversible).

A mechanism by which active Src is returned to an inactive state has recently been clarified.[16–18] (13) Activated Src phosphorylates a recently discovered transmembrane protein, Cbp, at Y314. (14) Y314 of Cbp binds a SH2 domain on Csk. (15) Csk is a tyrosine kinase that is able to phosphorylate Y527 of Src. (16) The binding to Cbp stimulates Csk's kinase activity. (17) Csk is tightly bound to a protein tyrosine phosphatase, PEP. (18) PEP can dephosphorylate Src Y416. Thus Src, by phosphorylating Cbp, brings into play a kinase–phosphatase combination (Csk:PEP) which removes the phosphate required for Src activity (Y416) and restores the phosphate at Y527 required for refolding of Src to its inactive state.

Two additional connections may contribute to the robustness of the switch. (19) EGFR is a Src substrate, suggesting that Src may contribute to the activation of EGFR. (20) The phosphatase (PTPα) that removes the phosphate from Src Y527 may also dephosphorylate Cbp Y314; these two phosphatase activities may function coherently in tending to keep Src active. (General reference: Ref. 19).

In summary, the system input for the scheme shown in Fig. 8 is the phosphorylation of certain tyrosines on receptor tyrosine kinases (such as EGFR) at the plasma membrane. This occurs in response to binding of extracellular ligands to the receptors. PI3K is brought into play when it is recruited to the plasma membrane where all of the action takes place. PI3K can be recruited to the plasma membrane by binding (via its SH2 domain) to another phosphorylated tyrosine on activated EGFR. Thus EGFR can initiate the disruption of both of the inhibitory intramolecular interactions between Src domains. EGFR is itself a substrate for tyrosine–phosphorylation by Src,[14] which suggests a positive feedback in the activation of Src by EGFR. Src activity is turned off due to phosphorylation by Src of a membrane protein (Cbp) which recruits and activates a pair of enzymes (PEP and Csk) that remove the activating phosphate and restore the inhibitory phosphate on Src.

This example depicts part of a computational system consisting of an assembly of protein molecules. Computations involve domain interactions both within and between molecules, and are controlled in large part by phosphorylation and dephosphorylation events at specific sites. The manner of depicting interactions of membrane-bound proteins is illustrated. This example also shows how a functional part of a complex molecular network can be defined and treated as a module with inputs and outputs.

VI. TRANSCRIPTION REGULATION BY PROTEIN ASSEMBLIES AT PROMOTERS

The complexities of transcriptional regulation present special problems for interaction mapping. The nature of some of these complexities and the way they can be handled can be seen in the context of an example. Figure 9 is a heuristic map in which a group of mutually interacting proteins can bind to two different promoter elements of a gene and thereby exert either stimulatory or inhibitory effects on the initiation or elongation of the transcript RNA. Most of the actions target the transcript initiation machinery, but some alter chromatin structure and thereby affect transcript elongation.

The map (Fig. 9) encompasses part of the network that controls the E2F transcription factors involved in the control of cell cycle progression. This system has been studied extensively in mammalian cells and has yielded confusing experimental results regarding the conditions under which E2F can either stimulate or inhibit its target genes.[20–23] This scheme covers part of the previously presented network map,[1] but adds recent findings about acetylations.[24]

The representation of the promoters and transcript in Fig. 9 is similar to common molecular biology practice. In the vicinity of the transcription start, three contingencies are indicated: two stimulation symbols and one inhibition symbol. The inhibition symbol, which is the likely dominant effect, is placed last in the direction of the affected action. The first and strongest stimulatory action is due to upstream transcription factors, such as Sp1 (interaction line 9 in Fig. 9).[25–27] This action requires that Sp1 be bound to the promoter (8). (The need for Sp1 to bind to the gene in order to affect its function is implicit.) The E2F1:DP1 dimer can bind to its promoter element (2), which puts it in position to stimulate transcription (3). The main effect of E2F1:DP1 however, arises from its ability to bind pRb (4). pRb becomes tethered to the gene and thereby able to exert a strong negative effect on transcript initiation (5). In order to access the transcription initiation machinery, pRb must be tethered to the promoter region via its binding to E2F1:DP1 which, in turn, is bound to the promoter (interactions 2 and 4). A seemingly paradoxical consequence is that induced expression of E2F1 can suppress rather than enhance transcription.[23] pRb can be phosphorylated by cyclin-dependent kinases (6); this abolishes the ability of pRb to bind E2F1:DP1 (7) and serves as a major input to the network.

These actions may occur within a multimolecular assembly. Sp1 and E2F1 bind to each other and stimulate transcription cooperatively (10).[25,26] Thus all of the components of the system can bind to each other simultaneously and function as a unit or module

Two additional components of the module have recently been characterized: a protein acetylase (p300/CBP) and a deacetylase (HDAC).[24,28,29] p300/CBP can bind E2F1 (11), and HDAC can bind pRb (12). Both may function while bound to the other components of the system. Both act at three points. The first two involve acetylation/deacetylation of E2F1 (13 and 14) and of Sp1 (16 and 17). Acetylation of E2F1 or Sp1 strongly enhances their affinity for their respective promoter element (15 and 18).

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

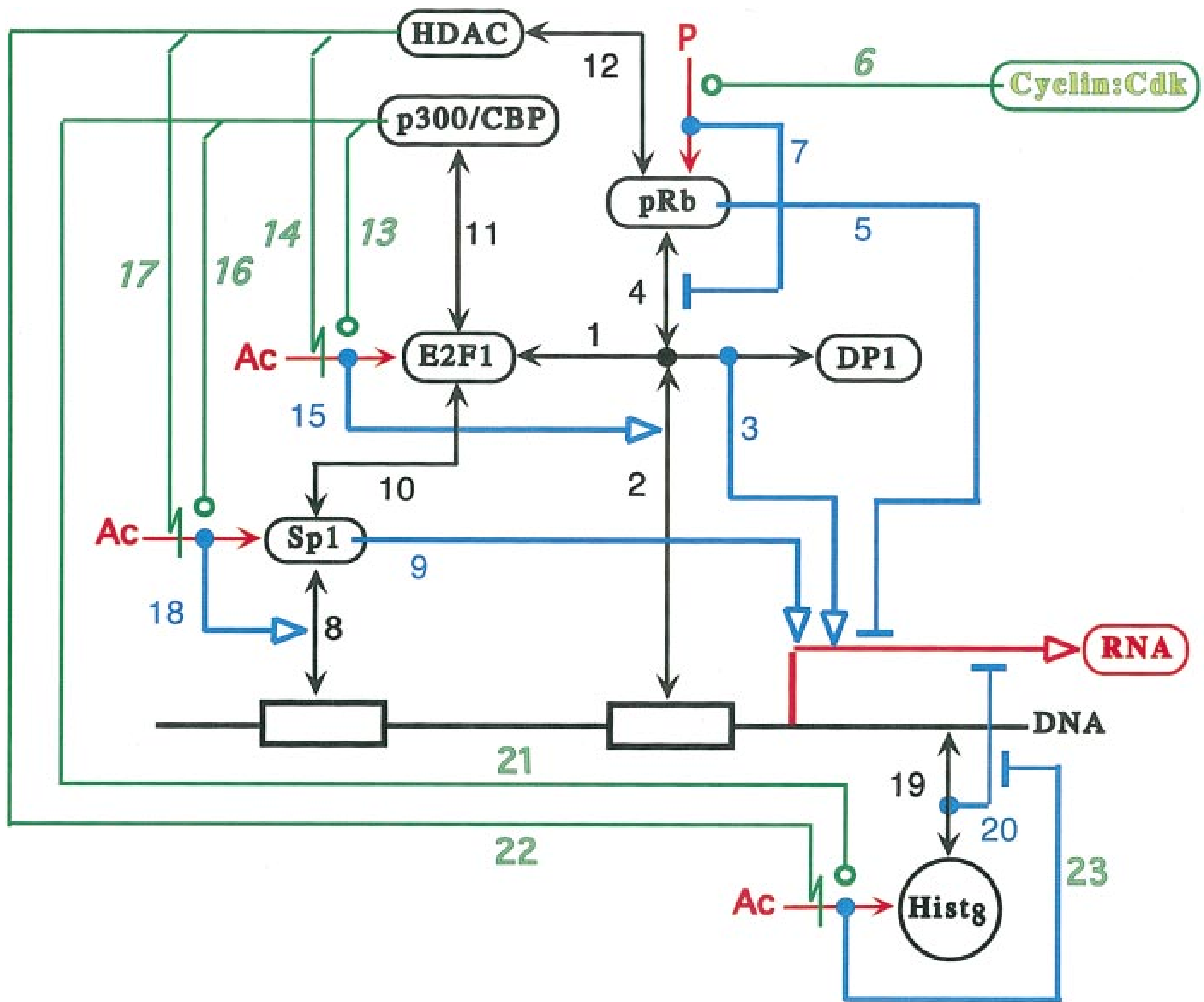

FIG. 9. (Color) Parallel modes of transcription control by a multiprotein assembly bound to promoter elements of a gene. The regulatory interactions shown here have been found in E2F-dependent genes, a class of genes that code for proteins required for cell cycle progression. E2F, together with another factor (Sp1), bind to their respective recognition sequences in the promoter regions of these genes. Both proteins have transcription activation domains. pRb binds E2F (actually a heterodimer such as E2F1:DP1), and contains a domain that blocks transcription initiation. Transcript elongation is inhibited by histone deacetylase (HDAC) through its binding to pRb and consequent deacetylation of histones. In addition, HDAC inhibits transcript initiation, due to the ability of HDAC to acetylate Sp1 (Ref. 24). (HDAC also acetylates E2F1, but the consequences are difficult to foresee.) The interactions, as numbered in the diagram, are as follows. (The numbers do not imply a temporal sequence of events.) (1) E2F1 can bind DP1 to form a stable dimer. (2) The E2F1:DP1 dimer can bind to specific promoter elements. (3) E2F1:DP1 can activate transcription. (4) E2F1:DP1 can bind pRb. (5) pRb is a powerful transcription suppressor. (6) pRb can be phosphorylated by Cyclin:Cdk complexes [details omitted here, but are included in Fig. 6(A) of Ref. 1]. (7) Phosphorylated pRb cannot bind E2F1:DP1. (8) Sp1 can bind to promoter elements in some of the same genes that have elements recognizing E2F1:DP1. (9) Sp1 is a powerful transcription initiator. (10) Sp1 can bind to E2F1. (11) E2F1 can bind the protein acetylase p300/CBP. (12) pRb can bind the deacetylase HDAC. (13) p300/CBP can acetylate E2F1. (14) HDAC can deacetylate E2F1. (15) When E2F1 is acetylated, its binding to promoter elements is enhanced (Ref. 24). (16) p300/CBP can acetylate Sp1. (17) HDAC can deacetylate Sp1. (18) When Sp1 is acetylated, its binding to promoter elements is enhanced. (19) Downstream from the transcription initiation site, the DNA is bound in nucleosomes composed of histone octomers. (20) Nucleosomes tend to block transcript elongation. (21) p300/CBP can acetylate nucleosomal histones. (22) HDAC can deacetylate nucleosomal histones. (23) Acetylation of histones relieves the transcript elongation block. (The colors used here for the different types of interaction lines are not logically required.) For a recent review, see Ref. 20.

The third point of acetylase/deacetylase action is at nucleosomal histones (21 and 22). Most of the nuclear DNA is wound around nucleosomes consisting of histone octomers (19). Nucleosomes ordinarily inhibit transcript elongation (20). Acetylation of the histones, however, relieves the elongation block (23). Histone deacetylation restores the block and is facilitated by the recruitment of HDAC through its binding to pRb.[30]

The three actions of the acetylase therefore are ''coherent'' in that they each tend to stimulate transcription, although by different paths. Similarly the three actions of the deacetylase are coherent in tending to inhibit transcription. Coherency, in this sense, is common in regulatory networks and may contribute to system robustness. Thus pRb, by tethering the deacetylase to the gene, can suppress transcript elongation as well as initiation.

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

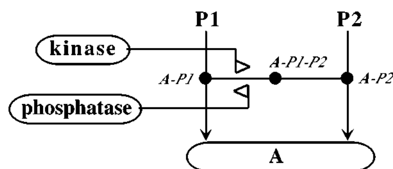

FIG. 10. Directed enzyme action symbols denote reaction from the nearest upstream node to the nearest downstream node, Molecule **A** is shown with two possible phosphorylations, **P1** and **P2**. Nodes **A**–$P1$ and **A**–$P2$ represent **A** bearing phosphates **P1** or **P2**. Node **A**–$P1$–$P2$ represents the state combination, namely phosphorylation at both sites. The kinase converts **A**–$P1$ to **A**–$P1$–$P2$, which amounts to the addition of phosphate P2. The phosphatase removes P2 from **A**–$P1$–$P2$.

This multiprotein system might be a functional building block or module within the greater network. It has a major input in the phosphorylation of pRb which abrogates the binding to E2F1:DP1 and an output of RNA transcripts of genes involved in cell cycle progression. (A sense of the greater network can be seen in Ref. 1 or at http://discover.nci.nih.gov/kohnk/interaction_maps.html.)

VII. THE DIRECTED ENZYME REACTION SYMBOL: COMPUTATION VIA PHOSPHATIDYLINOSITOL PHOSPHATES

When a molecular species is subject to multiple modifications which confer a variety of interaction capabilities, a severe challenge is presented to any diagramming method. This problem is highlighted by the regulatory functions of the various phosphorylated forms of phosphatidylinositol at the plasma membrane, a key relay molecule in the control of several basic cell functions. We will consider here a method of dealing with this problem using state-combination connectors and a new symbol to denote directed enzyme reactions. (Although all enzyme reactions technically are reversible, some reactions are essentially one-way or ''directed,'' because they involve hydrolysis of high-energy bonds. The directed enzyme symbol implies a large standard free energy change in that direction.)

A directed enzyme reaction is indicated by an open triangular arrowhead placed beside a line that connects between two defined states (Fig. 10). It is an extension of the enzyme reaction symbol, with the open circle being replaced by an open triangle. The triangle points in the direction of the reaction, which is assumed to be nearly irreversible. When used together with state-combination symbols, the directed enzyme symbol gives us a powerful and flexible method of representation.

The following new convention applies: When a directed enzyme symbol appears next to a state-combination line, it means that a reaction is being catalyzed that proceeds in the indicated direction from the nearest node on one side to the nearest node on the other side of the symbol. Thus the reaction is a stoichiometric conversion of the state defined by the nearest upstream node to the state defined by the nearest downstream node, the direction of the stream being indicated by the triangular arrowhead.

We will now apply this symbology to the signal transduction system centered on the phosphatidylinositol phosphates at the plasma membrane, a system that plays a central role in the regulation of several important cell functions. Figure 11 is a heuristic map that includes some of the major known interactions of this system. Only the interactions implicated in one of the functions of the system, namely cell proliferation control via the protein kinase Akt/PKB, are included here. An annotation list, appended to Fig. 11, provides salient facts about the individual interactions and cites recent references.

Akt/PKB phosphorylates a number of proteins with consequent positive effects on both cell proliferation and cell survival. For example, Akt/PKB phosphorylates GSK3 (reaction 36 in Fig. 11), which, by way of a regulatory module involving β-catenin, stimulates cell proliferation.[31] Akt/PKB also phosphorylates the pro-apoptotic protein, BAD, thereby blocking its inhibitory binding to the anti-apoptotic proteins, Bcl-2 and Bcl-xL.[32,33] This is one way in which Akt/PKB promotes cell survival. Full activation of Akt/PKB requires that it be phosphorylated at both Ser308 and Ser473, which requires the action of two kinases: PDK1 and PRK2. These phosphorylations are facilitated when the molecules are colocalized at the membrane through binding to PtdIns-3-phosphates, such as PtdIns[3,4]P_2 (denoted by the node on state-combination line 5) and PtdIns[3,4,5]P_3 (denoted by the node on state-combination line 7).

All three proteins have PH domains at their N-termini which bind to these PtdIns-3-phosphates. Since these phosphoinositides tend to cluster in foci within the membrane, the protein molecules, each bound to a different PtdIns-3-phosphate molecule, could have opportunities to interact with each other. If one of the proteins is in large excess, however, it could occupy most of the phosphoinositide sites and therefore inhibit reactions with the other proteins. Optimum interactions would require a balance of the concentrations of the participants. This kind of dual stimulation/inhibition relationship has been demonstrated in computer simulations.[6] Molecular interaction maps show where these kinds of phenomena may occur.

VIII. DISCUSSION AND CONCLUSIONS

There is broad consensus among cell biologists that a clear method of representing the complexities of bioregulatory networks is urgently needed.[34] The method presented here shows that it is possible to devise a workable system. It remains to be seen whether the conventions described here will be generally accepted, or whether alternatives or modifications will be proposed. Our experience suggests that the current system is very useful once the conventions have been learned. It provides a concise method of accessing a large amount of detailed information about the molecular mechanisms of bioregulation. The linkage to annotation lists allows ready access to pertinent details and references. The method provides a balance between precision and flexibility needed to accommodate the large number of combinatorial possibilities generated by multimolecular complexes and protein modification states. An index of map coordinates makes it easy to locate any molecular species and its interactions, even in large maps. The method can also be used to formal-

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

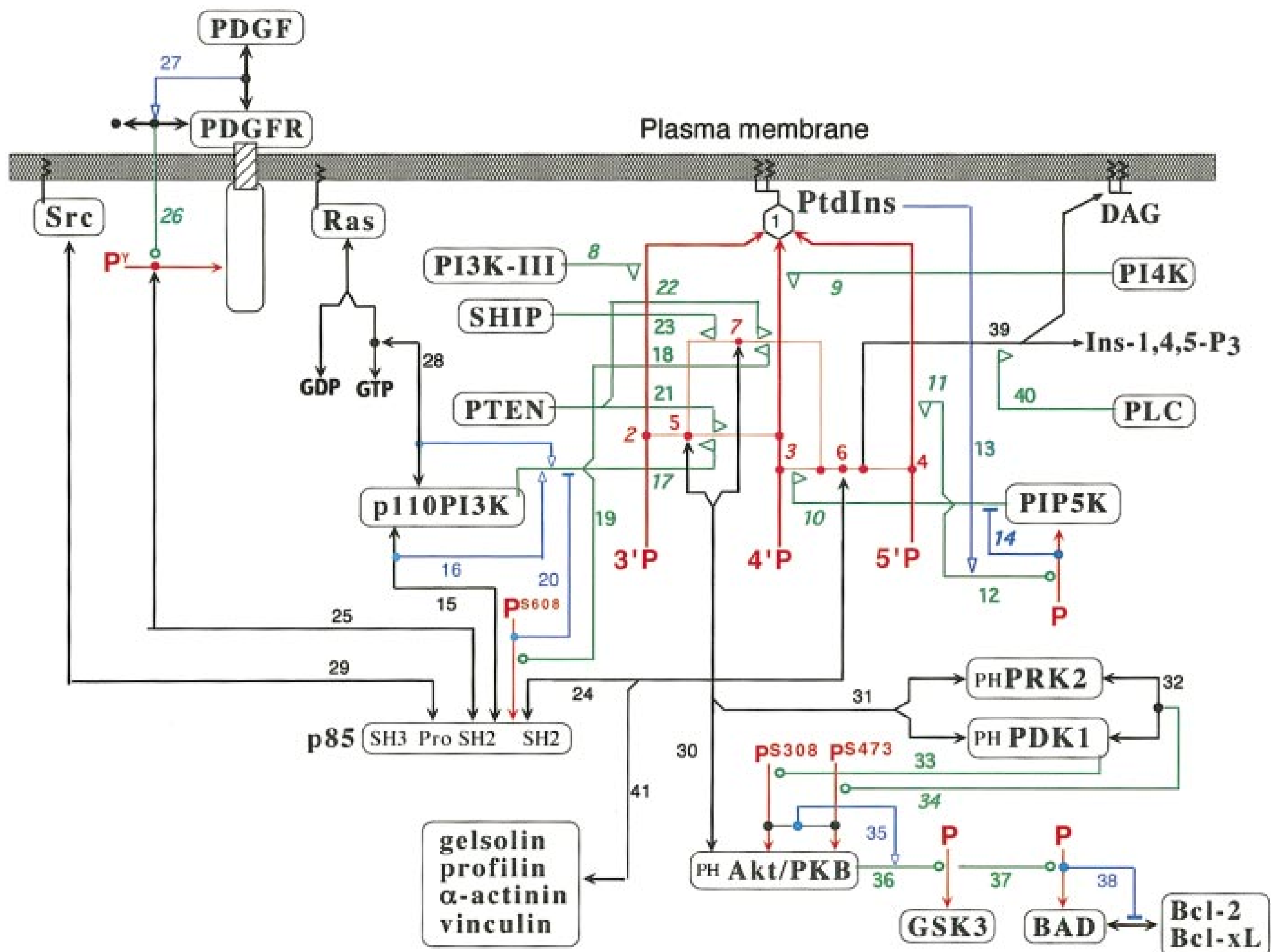

FIG. 11. (Color) Molecular interactions centered on phosphatidylinositol phosphates at the plasma membranes. Phosphatidylinositol (PtdIns) is an integral part of the inner leaflet of the membrane bilayer. It is subject to phosphorylation at three sites (designated 3′, 4′, and 5′). Various kinases and phosphatases add or remove phosphates, respectively; their specific actions are denoted by means of directed enzyme action. See Sec. VII and the Appendix.

ize unambiguous models for simulation.[2] The preparation of clear diagrams imposes a discipline for the interpretation of experiments and formulation of hypotheses. Finally, the method provides a shorthand for personal note taking.

Interaction maps do not indicate any particular sequence of events, nor do they specify biological function. They only show what the molecules, in effect, see. The interaction properties of the molecules are strictly local, and do not contain global information about what else might be going on in the network or what the intended function may be. The behavior of a biological system arises as an emergent consequence of the local interaction capabilities of the molecules of which it is composed. Since the diagrams do not impose presumed functional interpretations, they leave the field open for new interpretations. Molecular interaction maps thus can help to unravel mechanisms.

Topological interaction patterns occurring in several different interaction modules may be discerned, which could lead to module classification and functional conjectures. For example, the general pattern of control by intramolecular domain interactions seen in the case of Src (Fig. 8) can also be discerned in Akt/PKB[35] and (albeit indirectly via a 14-3-3 protein) in Raf1.[36]

Another topological generality is the common occurrence of a multiplicity of paths from a given input to the same or similar effect, as was described in the context of the control of E2F-dependent genes by pRb (Fig. 9). Although the effect may be similar from one path to another, the time dependence often differs. We call this type of parallel pathway pattern ‘‘coherency,’’ and suspect that it contributes to robustness of control.

A challenge for the future is to look for module design within the network. A module would consist of well-defined inputs and outputs, and the relationships between them. It is not clear to what extent the networks are modular, in view of their rich interconnections and rearranging multimolecular assemblies. Modularity would make it easier to understand network function. The properties of individual modules could be determined experimentally, in which case the behavior of a system of modules could be simulated by theory.

An important role of theoretical models is to demonstrate behavior that would suggest experimental possibilities

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

not otherwise obvious. Theory could generate sharply defined questions for quantitative test in appropriate systems. A first step is to define part of a regulatory network that would constitute a subsystem suitable for theoretical investigation. Molecular interaction maps can help to select and define such subsystems.

An example of this procedure was portrayed in Fig. 8, which defines a subsystem for the function of a key cell regulator component, Src. The interactions of Src have been extensively studied and the structural transitions of the Src molecule have been elucidated.[19] Figure 8 showed how one can represent interactions between domains of the same protein molecule and interactions at membranes, features that have an essential role in Src function. The interaction diagram suggests how this subsystem may function in a bistable manner with hysteresis. Although Src actually may have many inputs and outputs, the model system in Fig. 8 can be viewed in terms of a single input and a single output. With concerted effort, a physical system like this could be assembled for experimental investigation. A task of theory is to stimulate the experimentalist to make the effort and to direct attention to what might be observed.

The input in Fig. 8 is an extracellular ligand (such as EGF) which binds to receptors (such as EGFR), induces the receptor to dimerize and consequently to autophosphorylate at several tyrosine residues. This activates EGFR and allows it to interact with inactive Src, and cause the conformation of Src to change from the inactive to the active form. Three interactions are involved. Two of them serve to open the Src conformation while the third phosphorylates the Src tyrosine (Y416) that is required for activity, now made accessible by the opening of the Src structure. After its kinase activity is turned on, the conformation of Src is fixed in its active form and prevented from folding back to the inactive conformation. This is due to two factors: (1) the phosphate on Y416 not only activates the kinase domain but physically prevents the refolding;[13] (2) the phosphate on Y527, which helped to keep Src in the folded conformation, becomes accessible for removal by a phosphatase. Because it cannot refold, Src remains in its active conformation even when it dissociates from EGFR and/or associated molecules. This creates a barrier for return to the inactive state and could give rise to hysteresis. A path for the return to the inactive state involves the recently described transmembrane protein, Cbp (Csk-binding protein). Src phosphorylates Cbp at a tyrosine which then can bind a kinase-phosphatase dimer (Csk-PEP); the kinase phosphorylates Y527 of Src while the phosphatase removes the phosphate from Y416. This would efficiently accomplish the two steps required for inactivation: removal of the phosphate required for activity, and insertion of the phosphate needed for refolding. Once the molecule is folded, there is again a kinetic barrier due to the time required to reassemble Src and p85PI3K at activated receptor EGFR (Fig. 8).

Some important aspects of this system remain to be clarified, particularly regarding EGFR. There might be a positive feedback loop due to phosphorylation of EGFR by Src. Also, the control of EGFR dephosphorylation is not clear, nor is the control of the phosphatase(s) that act on Y527 of Src and on Cbp.

Another issue has to be considered in modeling interactions at cell membranes. Much of the cell's molecular computation in fact takes place in association with the plasma membrane, where receptors are located that receive input from outside the cell. These receptors however are not distributed randomly over the membrane. Rather, they tend to be localized within small specialized membrane regions, called caveolae or more generally RAFTs.[16,37] How then should the signaling processes within individual RAFTs be described and how might processes between RAFTs be coordinated? This is a new area where theory has an important role to play.

The Src subsystem described previously is a suitable candidate for modeling membrane-associated computation. This subsystem highlights the small-number-of-molecules issue, which suggests a strong stochastic component of the kinetics. Even within the limited scope of Fig. 8, we can count at least ten states in which the Src molecule can exist, depending on its phosphorylation and binding interactions. There may be of the order of 100 receptors within each RAFT. The number of Src molecules in a given state within a RAFT therefore could be small enough to be subject to stochastic fluctuation. Modeling this system may therefore require treating the molecules as individual software objects governed by interaction probabilities, a method that has been proposed by Bray.[3,6,7]

A further consideration regarding reactions in RAFTs deserves comment. A RAFT could function as a signaling unit. In the case of Src, integrated function may be dictated by the fact that the receptor tyrosine kinases are themselves substrates for phosphorylation by Src. Also, there may be coordination between RAFTs. As noted previously, activated Src could dissociate from the receptor without snapping back to an inactive state. Active Src could then diffuse from one RAFT to another, which may tend to cause the activation to spread over the membrane.

The activation and inactivation transitions of Src, which entail substantial time delays, could generate oscillation. Oscillations are known to occur in signaling systems involving calcium, but might likewise occur in other systems that do not have the advantage of a readily available detection tool such as is available for calcium. It may be that an effective signal is triggered when oscillations over a region of RAFTs becomes coherent.

These are rich areas for theoretical investigation.

APPENDIX: ANNOTATIONS FOR FIGURE 11

The following annotation list and explanations correspond to the numbered interactions and molecule descriptions in Fig. 11. For recent reviews, see Refs. 31, 38–40.

(1) Phosphatidylinositol (PtdIns) is a membrane-associated molecule made up of two hydrocarbon chains that insert in the lipid bilayer, and a phosphate-linked inositol ring that is accessible to the intracellular side of the plasma membrane.
(2–4) The inositol ring can have a phosphate group on positions 3,4, and/or 5.

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

(5–7) Multiphosphorylated forms of PtdIns are represented by nodes within state-combination lines, as follows: 5: PtdIns-3,4-P_2. 6: PtdIns-4,5-P_2. 7: PtdIns-3,4,5-P_3. [PtdIns(3,5)P_2 was omitted because little is known of its function.] Note that all of the nodes within the same line (excluding the ends if the lines) represent the same molecular species.
(8) PtdIns is the specific substrate for phosphorylation at the inositol 3 position by type III phosphoinositide 3-kinase (PI3K-III) which is the major isoform responsible for the production of PtdIns-3-P (Ref. 39). In accord with convention, the triangle arrowhead points from the substrate (unphosphorylated PtdIns) to the product (PtdIns-3-P which is represented by node 2). (A directed enzyme symbol is used here to specify that only unphosphorylated PtdIns is the substrate; an open circle enzyme symbol would imply that the indicated phosphorylation could occur regardless of phosphorylations at other sites.)
(9) The inositol 4 position is a prominent site of initial phosphorylation of PtdIns, catalyzed by phosphatidylinositol-4-kinase (PI4K) which is associated with the plasma membrane, endoplasmic reticulum, Golgi, and a variety of intracellular vesicles (reviewed by Ref. 41).
(10) PtdIns-4-P can be phosphorylated at the 5′ position by PIP5K, a type I phosphatidylinositol phosphate kinase (discussed in Ref. 42). (The directed enzyme action symbol is helpful for reactions proceeding from one node to another.)
(11) PIP5K also can phosphorylate PtdIns at the 5′ position (discussed by Ref. 42).
(12) In addition to PtdIns kinase activity, PIP5K has protein kinase activity, being capable of autophosphorylation (Ref. 42).
(13) The autophosphorylation of PIP5K is stimulated by PtdIns (Ref. 42).
(14) Autophosphorylation inhibits the activity of PIP5K (Ref. 42).
(15) The catalytic subunit of PtdIns-3-kinase type I (p110PI3K) binds to an adapter subunit (p85) (reviewed by Refs. 39 and 43). The site of binding on p85 is a region between a pair of SH2 domains.
(16) The kinase activity of p110PI3K is enhanced by binding to the p85 adapter.
(17) p110PI3K can add a phosphate to the inositol 3′ position of PtdIns-4-P to form PtdIns-3,4-P_2
(18) p110PI3K also can add a 3′ phosphate to PtdIns-4,5-P_2 to form PtdIns-3,4,5-P_3.
(19) In addition, p110PI3K has protein kinase activity, being capable of autophosphorylation at Ser608.
(20) Autophosphorylation at Ser608 inhibits the kinase activity of p110PI3K. (A similar pattern was noted above for PIP5K.)
(21, 22) The phosphatase PTEN removes the 3′ phosphates from PtdIns-3,4-P_2 and PtdIns-3,4,5-P_3 (Ref. 44).
(23) The phosphatase SHIP removes the 5′ phosphate from PtdIns-3,4-P_2 (Ref. 45, and references cited therein).
(24) The p85 PI3K-adapter subunit binds to PtdIns-4,5-P_2 and thereby recruits the p110 catalytic subunit to the site of action.
(25) The p85 adapter also binds to certain phosphotyrosine sites on activated receptor tyrosine kinases, such as platelet-derived growth factor receptor (PDGF), and can thus recruit PI3K to the plasma membrane in response to growth signals (Ref. 46).
(26) The intracellular domain of PDGF receptor is autophosphorylated (in *trans*) following homodimerization of the receptor. (The isolated node represents another copy of PDGFR, including the extracellular, transmembrane, and intracellular domains. The node within the double-arrowed binding line represents the homodimer.)
(27) Homodimerization of PDGFR is stimulated by the binding of PDGF to the extracellular domain of PDGFR.
(28, 29) Additional inputs to the PtdIns-3-P regulatory system can come from Ras via binding to p110PI3K, and from Src via binding to the p85 adapter subunit of PI3K.
(30) PtdIns-3,4-P_2 or PtdIns-3,4,5-P_3 can bind the PH domain of the kinase Akt/PKB (Ref. 39).
(31) PtdIns-3,4-P_2 or PtdIns-3,4,5-P_3 can be also bind the PH domains of kinases PDK1 and PRK2. (Phosphoinositides tend to cluster in foci within membranes, thus allowing the various bound molecules—each bound to a separate phosphoinositide molecule—to interact.)
(32) PDK1 and PRK2 can bind to each other. (This could bring together different phosphoinositide molecules in the membrane.)
(33) PDK1 phosphorylates Akt/PKB at Thr308 which is located within a loop that regulates kinase activity (Ref. 44).
(34) PRK2 phosphorylates Akt/PKB at Ser473.
(35) Full activation of Akt/PKB requires that it be phosphorylated at both Ser308 and Ser473. Activation of Akt/PKB is a major output of the PtdIns-3-P signaling network.
(36) Akt/PKB phosphorylates several regulatory proteins, including the glycogen synthase kinase, GSK3. (Further extension of the map would show that this phosphorylation inhibits the activity of the kinase and thereby relieves a restraint on cell proliferation.)
(37, 38) Akt/PKB phosphorylates the apoptosis-promoting protein BAD, and thereby inhibits its binding and inactivation of the anti-apoptotic proteins Bcl-2 and Bcl-xL.
(39, 40) A major output of signaling from PtdIns-4,5-P_2 is its convesion to two ''second messenger'' molecules, diacylglycerol (DAG) and inositol-1,4,5-P_3 which are cleavage products derived from PtdIns-4,5-P_2 through the action of phospholipase C (PLC). (Further extension of the map would show PLC binding to receptor tyrosine kinases at the plasma membrane.)
(41) Another output from PtdIns-4,5-P_2 is to the actin cytoskeleton through binding to several actin-binding proteins: gelsolin, profilin, vinculin, and α-actinin.

[1] K. W. Kohn, ''Molecular interaction map of the mammalian cell cycle control and DNA repair systems,'' Mol. Biol. Cell **10**, 2703 (1999).
[2] K. W. Kohn, ''Functional capabilities of molecular network components controlling the mammalian G1/S cell cycle phase transition,'' Oncogene **16**, 1065 (1998).
[3] D. Bray, ''Signaling complexes: Biophysical constraints on intracellular communication,'' Annu. Rev. Biophys. Biomol. Struct. **27**, 59 (1998).
[4] T. Pawson and P. Nash, ''Protein-protein interactions define specificity in signal transduction,'' Genes Dev. **14**, 1027 (2000).
[5] G. Wang, U. S. Bhalla, and R. Iyengar, ''Complexity in biological signaling systems,'' Science **284**, 92 (1999).

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp

[6] D. Bray and S. Lay, ''Computer-based analysis of the binding steps in protein complex formation,'' Proc. Natl. Acad. Sci. U.S.A. **94**, 13493 (1997).
[7] C. J. Morton-Firth and D. Bray, ''Predicting temporal fluctuations in an intracellular signalling pathway,'' J. Theor. Biol. **192**, 117 (1998).
[8] H. H. McAdams and A. Arkin, ''It's a noisy business! Genetic regulation at the nanomolar scale,'' Trends Genet. **15**, 65 (1999).
[9] P. H. Kussie, S. Gorina, V. Marechal *et al.*, ''Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain,'' Science **274**, 948 (1996).
[10] T. Pawson and J. D. Scott, ''Signaling through scaffold, anchoring, and adaptor proteins,'' Science **278**, 2075 (1997).
[11] A. J. Whitmarsh and R. J. Davis, ''Structural organization of MAP-kinase signaling modules by scaffold proteins in yeast and mammals,'' Trends Biochem. Sci. **23**, 481 (1998).
[12] W. Xu, A. Doshi, M. Lei *et al.*, ''Crystal structures of c-Src reveal features of its autoinhibitory mechanism,'' Molec. Cell **3**, 629 (1999).
[13] S. Gonfloni, A. Weijland, J. Kretzschmar *et al.*, ''Crosstalk between the catalytic and regulatory domains allows bidirectional regulation of Src,'' Nat. Struct. Biol. **7**, 281 (2000).
[14] D. A. Tice, J. S. Biscardi, A. L. Nickles *et al.*, ''Mechanism of biological synergy between cellular Src and epidermal growth factor receptor,'' Proc. Natl. Acad. Sci. U.S.A. **96**, 1415 (1999).
[15] X.-M. Zheng, R. J. Resnick, and D. Shalloway, ''A phosphotyrosine displacement mechanism for activation of Src by PTPalpha,'' EMBO J. **19**, 964 (2000).
[16] L. A. Cary and J. A. Cooper, ''Molecular switches in lipid rafts,'' Nature (London) **404**, 945 (2000).
[17] M. Kawabuchi, Y. Satomi, T. Takao *et al.*, ''Transmembrane phosphoprotein Cbp regulates the activities of Src-family tyrosine kinases,'' Nature (London) **404**, 999 (2000).
[18] S. Takeuchi, Y. Takayama, A. Ogawa *et al.*, ''Transmembrane phosphoprotein Cbp positively regulates the activity of the carboxyl-terminal Src kinase, Csk,'' J. Biol. Chem. **275**, 29183 (2000).
[19] J. S. Biscardi, D. A. Tice, and S. J. Parsons, ''c-Src, receptor tyrosine kinases, and human cancer,'' Adv. Cancer Res. **76**, 61 (1999).
[20] N. Dyson, ''The regulation of E2F by pRB-family proteins (Review),'' Genes Dev. **12**, 2245 (1998).
[21] K. Ohtani, ''Implications of transcription factor E2F in regulation of DNA replication,'' Front. Biosci. **4**, d793 (1999).
[22] H. Muller and K. Helin, ''The E2F transcription factors: Key regulators of cell proliferation,'' Biochim. Biophys. Acta **1470**, M1 (2000).
[23] S. He, B. L. Cook, B. E. Deverman *et al.*, ''E2F is required to prevent inappropriate S-phase entry of mammalian cells,'' Mol. Cell. Biol. **20**, 363 (2000).
[24] G. Marzio, C. Wagener, I. Gutierrez *et al.*, ''E2F family members are differentially regulated by reversible acetylation,'' J. Biol. Chem. **275**, 10887 (2000).
[25] J. Karlseder, H. Rotheneder, and E. Wintersberger, ''Interaction of Sp1 with the growth- and cell cycle-regulated transcription factor E2F,'' Mol. Cell. Biol. **16**, 1659 (1996).
[26] S-Y. Lin, A. R. Black, D. Kostic *et al.*, ''Cell cycle-regulated association of E2F1 and Sp1 is related to their functional interaction,'' Mol. Cell. Biol. **16**, 1668 (1996).
[27] G. Watanabe, C. Albanese, R. J. Lee *et al.*, ''Inhibition of cyclin D1 kinase activity is associated with E2F-mediated inhibition of cyclin D1 promoter activity through E2F and Sp1,'' Mol. Cell. Biol. **18**, 3212 (1998).
[28] R. X. Luo, A. A. Postigo, and D. C. Dean, ''Rb interacts with histone deacetylase to repress transcription,'' Cell **92**, 463 (1998).
[29] A. Brehm, E. A. Miska, D. J. McCanee *et al.*, ''Retinoblastoma protein recruits histone deacetylase to repress transcription,'' Nature (London) **391**, 597 (1998).
[30] Y. Takahashi, J. B. Rayman, and B. D. Dynlacht, ''Analysis of promoter binding by the E2F and pRB families in vivo: Distinct E2F proteins mediate activation and repression,'' Genes Dev. **14**, 804 (2000).
[31] S. R. Datta, A. Brunet, and M. E. Greenberg, ''Cellular survival: A play in three Akts,'' Genes Dev. **13**, 2905 (1999).
[32] B. Fadeel, B. Zhivotovsky, and S. Orrenius, ''All along the watchtower: On the regulation apoptosis regulators,'' FASEB J. **13**, 1647 (1999).
[33] I. Galetic, M. Andjelkovic, R. Meier *et al.*, ''Mechanism of protein kinase B activation by insulin/insulin-like growth factor-1 reavealed by specific inhibitors of phosphoinositide-3-kinase—Significance for diabetes and cancer,'' Pharmacol. Therap. **82**, 409 (1999).
[34] R. Brent, ''Genomic biology,'' Cell **100**, 169 (2000).
[35] M. Andjelkovic, S-M. Maira, P. Cron *et al.*, ''Domain swapping used to investigate the mechanism of protein kinase B regulation by 3-phosphoinositide-dependent protein kinase 1 and Ser473 kinase,'' Mol. Cell. Biol. **19**, 5061 (1999).
[36] G. Tzivion, Z. Luo, and J. Avruch, ''A dimeric 14-3-3 protein is an essential cofactor for Raf kinase activity,'' Nature (London) **394**, 88 (1998).
[37] G. Carpenter, ''The EGF receptor: A nexus for trafficking and signaling,'' BioEssays **22**, 697 (2000).
[38] T. O. Chan, S. E. Rittenhouse, and P. N. Tsichlis, ''Akt/PKB and other D3 phosphoinositide-regulated kinases: Kinase activation by phosphoinositide-dependent phosphorylation,'' Annu. Rev. Biochem. **68**, 965 (1999).
[39] L. E. Rameh and L. C. Cantley, ''The role of phosphoinositide 3-kinase lipid products in cell function,'' J. Biol. Chem. **274**, 8347 (1999).
[40] R. A. Anderson, I. V. Boronenkov, S. D. Doughman *et al.*, ''Phosphatidylinositol phosphate kinases, a multifaceted family of signaling enzymes,'' J. Biol. Chem. **274**, 9907 (1999).
[41] T. F. J. Martin, ''Phosphoinositide lipids as signaling molecules: Common themes for signal transduction, cytoskeletal regulation, and membrane trafficking,'' Annu. Rev. Cell Dev. Biol. **14**, 231 (1998).
[42] T. Itoh *et al.*, ''Autophosphorylation of type I phosphatidylinositol phosphate kinase regulates its lipid kinase activity,'' J. Biol. Chem. (Published electronically on 20 April 2000 as manuscript M000426200) (2000).
[43] D. A. Fruman, R. E. Meyers, and L. C. Cantley, ''Phosphoinositide kinases,'' Annu. Rev. Biochem. **67**, 481 (1998).
[44] S. J. Leevers, B. Vanhaesebroeck, and M. D. Waterfield, ''Signalling through phosphoinositide 3-kinases: The lipids take centre stage,'' Curr. Opin. Cell Biol. **11**, 219 (1999).
[45] M. F. Aman, T. D. Lamkin, H. Okada *et al.*, ''The inositol phosphatase SHIP inhibits Akt/PKB activation in B cells,'' J. Biol. Chem. **273**, 33922 (1998).
[46] S. M. Jones, R. Klinghoffer, G. D. Prestwich *et al.*, ''PDGF induces an early and a late wave of PI 3-kinase activity, and only the late wave is required for progression through G1,'' Curr. Biol. **9**, 512 (1999).

Downloaded 13 Jun 2002 to 131.215.42.77. Redistribution subject to AIP license or copyright, see http://ojps.aip.org/chaos/chocr.jsp