General strategy for understanding intracellular molecular interaction cascades that elicit stimulus-invoked biological processes

By Hiroto OKAYAMA*1,†

(Communicated by Nobutaka HIROKAWA, M.J.A.)

Abstract: Recent advances in biology have been driven by chemical analyses of the substances that form living organisms. Such analyses are extremely powerful as way of learning about the static properties of molecular species, but relatively powerless for understanding their dynamic behaviors even though this dynamism is essential for organisms to perform various biological processes that perpetuate their lives. Thus, attempts to identify individual species and molecular interaction cascades that drive specific responses to external stimuli or environmental changes often fail. Here I propose a general strategy to address this problem. The strategy comprises two key elements: functional manipulation of a given protein molecule coupled with close monitoring of its biological effect, and construction of a knowledge base tailored for conjecture-driven experimentation. The original idea for this strategy co-evolved with and greatly helped a series of studies we recently performed to discover critical signal cascades and cellular components that regulate the cell cycle transition from G1 to S phase.

Keywords: new strategy, signal cascades, reverse genetics, conjecture

Introduction

Virtually all modern studies have been guided by two conceptually distinct principles that René Descartes articulated in the book “Discours de la méthode” written in the 17th century.1) One is “neither to accept as true, nor to make judgements on, anything unless you are convinced it to be absolutely true”. The other is “to separate objects into parts, examine those first and then assemble all the findings obtained on each part, while confirming their mutual consistency, toward understanding the objects”. These two principles laid the intellectual foundation for all fields of today’s natural sciences

*1 Department of Biochemistry and Molecular Biology, Graduate School and Faculty of Medicine, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, Japan.
† Correspondence should be addressed: H. Okayama, 6-22-16 Honkomagome, Bunkyo-ku, Tokyo 113-0021, Japan (e-mail: okayama-tky@umin.net).

Abbreviations: GTP: guanine nucleotide triphosphate; GDP: guanine nucleotide diphosphate; GAP: GTPase activating protein; GEP: guanine nucleotide exchange protein; TSC: Tuberous sclerosis complex; mTOR: mammalian target of rapamycin; mTORC1: mTOR complex 1; ROCK: Rho-associated kinase; Cdk4/Cdk6: cyclin-dependent kinase 4/cyclin-dependent kinase 6; CycA: cyclin A; Cdk2: cyclin-dependent kinase 2; p27: p27kip1; preRC: prereplicative complex; ORC: origin recognition complex.

doi: 10.2183/pjab.92.372 ©2016 The Japan Academy
molecular interactions among the proteins and other constituents in response to external stimuli and internal signals. It is estimated that 50–100 thousand protein species are produced in humans. These are encoded by spliced mRNAs transcribed from ~23000 genes.\textsuperscript{6} One basic question is: which protein molecules activated by stimuli dynamically interact with one another, and how do these interactions elicit particular biological consequences? Analytical approaches that have been used to tackle this problem are \textit{in vitro} or quasi \textit{in vivo} studies on protein-protein interactions such as physical binding and complex formation, and identification of protein modifications such as phosphorylation, ubiquitination and etc. under the name of proteomics.\textsuperscript{7–10}

One intrinsic obstacle to this approach, however, is that many interactions and modifications detected \textit{in vitro} and even \textit{in vivo} on a particular protein may be nearly irrelevant to its most critical biological functions. There are scientifically credible data to support this. For instance, many proteins are phosphorylated at multiple sites targeted by more than one kinase, only one or two (or none) of which are biologically important.\textsuperscript{11,12} Furthermore, kinases themselves are not intrinsically equipped with any clear co-evolved mechanism that distinguishes between physiologically relevant and irrelevant substrates as exemplified by the observation that a chicken-derived tumor-causing tyrosine kinase expressed by genetic engineering very efficiently phosphorylates never-met proteins in \textit{E. coli}.\textsuperscript{13} Consequently, it is hard to determine the biological effector molecule for a given protein kinase without carrying out laborious and time-consuming kinase-effector matching studies \textit{in vivo} to identify the correct effector, and the multi-functionality of some proteins makes the problem even more complex. One effective complement to such an analytical approach is to guess effector molecule candidates based on information about the kinase and every conceivable effector molecule before actually implementing examination of their interactions. If the conjecture is right, this approach could markedly cut labor and shorten the time required to find the right effector.

We are good at guessing, and do this many times a day. Unconsciously perceived intuition and antipodal scientifically deliberated conjecture both involve guessing. We all know that this speeds certain tasks. In science, once a conjecture is made with deep deliberation, simple and quick experiments can be planned and carried out to support or defeat an idea. If the initial conjecture is disproved, another conjecture can be made and new experiments can be performed. This process can be repeated until support for a hypothesis is obtained. Consequently, making a right conjecture at the beginning is extremely important to cut labor and time to reach a goal. In general, the probability of making a right conjecture is proportional to the richness of the available knowledge base on the properties of possible players. Fortunately, for the past four decades, the molecular properties of most of the biological materials comprising humans and other organisms have been revealed thanks to the concurrently developed recombinant DNA, genetic engineering and biotechnology.\textsuperscript{14–16}

In this article, I wish to propose a general strategy to understand intracellular interaction cascades critically involved in stimulus-invoked biological processes, along with a general thought for making right conjecture and carrying out experimentation to quickly and clearly prove or disprove the conjecture.

Proposal

General assumptions for addressing biological problems. In this proposal, two empirically well supported assumptions are laid for solving biological problems. The first assumption is that not all physical interactions among biological materials and their chemical modifications, in particular protein-protein interactions and protein modifications, are equal in relevance to the biological phenomena to be understood. Some interaction or modification is crucial for the protein to execute its recognized function, but some others may be totally dispensable. The second is evolutionary conservation of many biologically processes at molecular levels. For example, many cellular factors that are critical for cell proliferation and survival are, if not identical, well conserved from yeast up to human in their structures, biological functions and even regulatory mechanisms. These two assumptions are taken as grand rules for contemplation throughout all the processes in this proposal.

Two technical elements

1. Functional manipulation of a given protein molecule coupled with close monitoring of its biological effect. As a basis for understanding the importance of the first technical element in this proposal, it would be extremely useful to depict a few easily imaginable interaction cascades (signal pathways) that lead to the biological effects in
response to particular stimuli in organisms and their individual cells. Figure 1 schematically presents a total of 11 such cascade patterns in two groups, Cascade I and Cascade II. Each group consists of 8 or 4 distinct patterns with one to three sensors for the stimulus and one effector for Cascade I or two for Cascade II that trigger the response. In this figure, arrows stand for promotion of the flow of interactions and bars for inhibition of this flow. Cascade I-1 shows a single interaction cascade that conveys the stimulus signal to a single effector molecule to elicit the biological response. This cascade may often be controlled by a positive feedback loop or an activating or inhibitory signal. In I-2 and I-3, two independent cascades mediate the single stimulus signal through a single sensor molecule S [I-2] or two different sensors Sa and Sb [I-3]. In Cascade I-4 and I-5, unlike in I-3, activation of both sensors is required for the biological response. In I-4, signals from Sa and Sb cooperate to activate the cascade that triggers the response whereas in I-5, the signal from Sb is not conveyed to the effector molecule but rather controls the flow of the signal from Sa by turning off the molecule that blocks the cascade from Sa. I-6 is a complex variant of I-5 and requires 3 independent signals for activation of the effector. In all the cascades so far illustrated, the interaction signals originate from the sensors, but in I-7, it originates elsewhere (shown as X here) and activates the effector that invokes the response. The signal from the sensor merely controls the flow of this cascade. In I-8, unlike in the rest, the stimulus signal invokes response by inactivating the effector molecule. In the Cascade II group, activation of two effectors is required for full response to the stimulus. In II-1, the signal from the single sensor branches into two and activates E1 and E2, respectively. In II-2, two distinct signals originating from each sensor are separately conveyed so as each to activate E1 and E2, respectively, but the signal from Sa flows into the other pathway. In II-3, signals generated from both sensors crossover and flow into the others, so that activation of either one of the sensors is sufficient to produce a full response. The last II-4 is a double effector version of I-5. The only difference is that in II-4, the Sb signal is required for activation of only E2. More complex cascades activated by more than two distinct stimuli and/or targeting more than two effector molecules can be generated by simply combining two or more of these 12 cascade patterns. However, it should be emphasized that these cascade patterns are only easily imaginable ones and some of those that occur in nature could be very different from these.
Given this background, the first technical element proposed here is functional manipulation of a candidate molecule with close monitoring of its effects on the biological response to the stimuli (Fig. 2). Let assume that factor A mediates the signal from the stimulus to induce the biological response in a single intracellular interaction cascade (Fig. 2A). If factor A is inactivated or forced to disappear (A⁻), the biological response would disappear. By contrast, if a non-conditionally activated form of factor A (A*) is expressed, the cell would show the same biological response as stimulated despite no stimulation. In the reverse situation, if one finds a molecule that alters in activity or physical state upon the stimulation, such as chemical modification or association with other molecules, eliminates only the biological response when inactivated and inversely, invokes the same response as stimulated despite the absence of the stimulus when non-conditionally activated, one could conclude that with a high probability, the molecule is an active component of the interaction cascade responsible for the stimulus-invoked biological processes and that the cascade is single at least at this molecule.

In the above argument, factor A is assumed to be a positive regulator activating the immediate downstream factor. However, not all the components of signal cascades in living organisms are positive regulators. For example, small GTP-binding proteins (small G) occur in two forms, GTP-bound and GDP-bound, and activate particular protein kinases only when they are in the GTP-bound form. GTPase activating proteins (GAPs) facilitate conversion of the GTP-bound form to a GDP-bound form, thereby inactivating small G, whereas guanine nucleotide exchange proteins (GEP) activate small G by exchanging the bound GDP with GTP. Many signal cascades contain more than one such a GAP-small G-protein kinase set as critical components. In this set, GAP inactivates small G until it is itself inactivated by an upstream factor. Thus, GAP forms a double negative regulator (negatively regulated negative regulator). Consequently, in overall effects on cascade signaling, inactivation (A⁻) for such negative regulators is equivalent to non-conditional activation (A*) for positive regulators and vice versa, forming a reverse relationship with the two groups (Fig. 2B). Notably, this relationship holds for any component of a double negative cascade region created by the presence of two simple negative regulators. Double negative regulators or regions are in general active until inactivated by upstream

![Fig. 2. The basic concept for the experiments to identify unknown cascades responsible for the stimulus-invoked biological processes. In this experiment, original cells respond with emergence of specific biological processes to a particular stimulus. Cells lacking expression of factor A (A⁻ cells) or expressing its non-conditionally activated form (A* cells) are then constructed and examined for the presence or absence of a proper response or alike with or without stimulation. A. If factor A is a critical component of this stimulus-invoked response cascade and activates the immediate downstream component, cells lacking A expression are expected to show no response to the stimulus whereas cells expressing its constitutively activated form are to show the same biological process as the response in the absence of the stimulus. B. If factor A is a double negative-type single molecule regulator or a component of a double-negatively controlled pathway region in a cascade, cells having similarly manipulated A as would show totally reversed phenotypes in the same experiment, with A⁻ cells to exhibit stimulus-free response whereas A* cells no response to the stimulus. Manipulated factor A in the leftmost cascade in trios is a component of a positive signal cascade; the one in the middle a single molecule double negative regulator; the one in the rightmost a component in a double negative region made of two single negative regulatory components. C. If factor A is a single negative, cells having similarly manipulated A as would show the same responses as positive regulators with A⁻ cells to exhibit no response to the stimulus and A* cells stimulus-free response. Manipulated factor A in the left in pairs is a positive regulator; the one in the right in pairs a single negative regulator that controls the immediate downstream double negative regulator (empty mark).]
factors. Among other known double negative regulators made of a single molecule are protein inhibitors for cyclin-dependent kinases that control cell cycle progression.\(^{17}\) This type of double negative regulators acts as such ones by controlling the flow of a signal originated elsewhere.

Obviously, there are many simple negative regulators working in living organisms. They include various protein kinases that inactivate target molecules such as GAP. These protein kinases are not innately destined to function as negative regulators. They basically work as both positive and negative regulator, but their target molecules determine them to function as either one. Consequently, unlike the double negative-type, this type is generally inactive until activated by upstream factors and behaves just like positive regulators with loss of stimulus-invoked response when inactivated and conversely, emergence of stimulus-free response when forcibly activated (Fig. 2C).

Inactivation of factor A can be done by treatment with a specific inhibitor such as for protein kinases\(^{18}\) if one is available, by ablating the gene that encodes it via knockdown, knockout or genome editing,\(^{18},19\) or by overexpressing a dominant negative-form if one can be made. Alternatively, a natural or man-made null mutant cell for the factor could be used if it can be confirmed that re-expression of the factor can restore the wildtype phenotype. On the other hand, its non-conditionally activated form could be obtained by genetically modifying its molecule by introducing a point mutation or a regional deletion. Many protein kinases and small GTPases have been extensively studied and their regulatory mechanisms are well understood.\(^{20},21\) Consequently, if factor A is one of these, it may be easy to obtain its non-conditionally activated form. Alternatively, overexpression of a natural activator for factor A, such as small G for paired protein kinases, can be used as its unconditionally active form. Furthermore, if A disappears upon stimulation, simple enforced expression of original A itself may be sufficient to generate the same effects as its constitutively activated form. If factor A is a double negative-type regulator, its non-conditionally activated form can also be made by inactivation of the upstream regulator or by creating a similar situation.

To see how useful a data set of the biological effects of inactivation and non-conditional activation of a candidate molecule can be for identifying the interaction cascades under your search, take the Fig. 1 cascades as a model. First, assume that the original cell has factor A, that it responds to a particular stimulus with emergence of a particular phenotype, and that both cells lacking A expression (A-) and cells expressing a constitutively active A (A\(^*\)) are available. Both cells are then examined for the presence (R+) or absence (R−) of response or the same phenotypic change with or without the addition of stimulus. As shown in Table 1 left half, because each experiment yields either R+ or R−, a unique combination of a total 4 results is obtained for each factor. The right half of this table shows the expected results of such experiments if factor A is located at the places numbered 1–22 in Cascade I, 1–10 in Cascade II or bolded 0 outside the cascades (see Fig. 1). This table is useful for identifying whether or not your examining factor plays a role in the cascade under your study, and if yes, what the cascade would look like and where it could be located in the cascade. One simple way to use this table is to find a location number that matches your results. To make this easier, data in the table are reorganized so as for all possible combination of results to be listed first and then all the numbered A’s to be placed in the list they match, as shown in the expanded table just beneath. Evidently, all the A’s located in various cascades are clustered only in several combinations and many of the rest of the mathematically generated combinations are judged to be incompatible and therefore, non-existent. In 2+4+X+X, both A- and A\(^*\) cells show “no response” despite stimulation whereas in X+X+5+7, they show a “full response” to no stimulation. Thus, in both cases, A is clearly involved in the cascade, yet A- and A\(^*\) show the same phenotype, which starkly contradicts with the definition of factor A for which an active form can be constructed. For some combinations of possible assay results, such as those marked $ and % (Table 1 lower part), there is no match-up location in the Fig. 1 cascades. Nevertheless, cascades and locations for factor A consistent with those two combinations can be drawn although they both have stimulus-invoked translocation of factor A from one site to another with no clear biological benefits for the presence of such cascades foreseeable (Fig. 3). Consequently, these cascades may be non-existent in the real world. From the common feature of the cascades where the assayed A’s are located and the specific combinatory results owned by each of the assayed A’s, one could infer what the studied cascade looks like and whether the cascade is partly redundant, non-redundant or with an inhibitory signal flow-in or the presence of double negative regulators by finding...
the cascades and locations in Fig. 1 that have the same combinatory result as the factor under analysis. In this regard, it should be noted that the double negative regulator in Cascade I-7 and the one in Cascade II-1 yield the same combinatory results. In another word, in this assay alone, cyclin-dependent kinase inhibitor-type double negative regulators that control the flow of a signal originated elsewhere cannot be distinguished from small G-GAP complex-type regulators that automatically generate a positive

Table 1. Anticipated stimulus-invoked responses of cells lacking expression of, or inversely expressing a non-conditionally activated form of, factor A, which is located at various places in Cascade I and Cascade II

| Stimulus | Genotype | Response No | Location in Cascade I | Location in Cascade II | Comments |
|----------|----------|-------------|-----------------------|------------------------|----------|
| +        | A        | R+          | 1                     | 0.3,4,6,7,8,12,15,17,20| 1,5,7,8,10|
|          | A-       | R+          | 2                     | 1.2,3,4,5,6,7,8,9,10,11,13,14,16,18,19,21,22 | 2,3,4,6,9 |
|          | A*       | R+          | 3                     | 0.1,2,3,4,5,6,7,8,9,10,11,13,14,16,18,19,21,22 | 2,3,4,5,6,7,8,9 |
|          | A-       | R-          | 4                     | 4,12,15,17,20          | 1,10     |
|          | A        | R-          | 5                     | 7.20                   | all      |
|          | A-       | R-          | 6                     | 0.1,2,3,4,5,6,8,9,10,11,12,13,14,15,16,17,18,19,21,22 | 2,3,4,5,6,7,8,9,10 |
|          | A*       | R+          | 7                     | 1.2,3,5,6,8,10,21,22   | 2,4,7,8   |
|          | A-       | R-          | 8                     | 0.4,7,9,11,12,13,14,15,16,17,18,19,20 | 1,3,5,6,9,10 |

Combinatory response Location in I Location in II Comments

1+3+5+7 incompatible
1+3+5+8 7 redundant pathway with double negative regulator
1+3+6+7 3,6,8 7,8 redundant pathway; positive signal flow-in
1+3+6+8 0 5
1+4+5+7 incompatible elsewhere originated signal to activate effector controlled by inhibitory signal with double negative regulator; double negative regulator in non-redundant pathway
1+4+5+8 20 1 inhibitory signal flow-in; double negative regulator
1+4+6+7 %
1+4+6+8 4,12,15,17 10 inhibitory signal flow-in; double negative regulator
2+3+5+7 incompatible
2+3+5+8 $2+3+6+7 1.2,5,10,21,22 1.2,4 non-redundant pathway
2+3+6+8 9,11,13,14,16,18,19 3,6,9 non-redundant pathway that conveys part of signals required for full response
2+4+5+7 incompatible
2+4+5+8 incompatible
2+4+6+7 incompatible
2+4+6+8 incompatible

In this proposed experiment, the original cell responds (R+) to the stimulus (+). The cell is then modified so as to have no expression of A (A-) or expression of a constitutively activated A (A*) and asked for their effects on the response in the presence and absence of the stimulus. If A- or A* cell shows a full response to the stimulus, it scores R+ and otherwise R-. If A is located as numbered in Cascade I or Cascade II in Fig. 1, the responses A- or A* cells are expected to exhibit with or without the stimulus are described in the upper table. In the lower table, data are reorganized so as to view all the possible combination of each assay data and the actual combined result for each numbered A. See the text and Fig. 3 for % and $ in Comments.
signal when inactivated by upstream signals. This perhaps is a reflection of the topological equality of these two types of regulation. Consequently, it is always required to consider these two types equally whenever the presence of a double negative regulator is suggested in this assay. Likewise, whether stimulus-invoked response is brought about by activation or inactivation of the effector molecule(s) cannot be determined in this assay (see locations 1 and 22 in Cascade I in Table 1). Biochemical analysis of the status of the effector(s) in the presence or absence of the stimulus is therefore needed.

As well noted, many biochemical interactions and modifications detected in vitro and even in vivo on a particular protein may be nearly irrelevant to its most critical biological functions. Therefore, one important task of this assay is to quickly determine if a candidate protein is actually involved in the cascade under study. Any factors entirely unrelated to this cascade should produce no effect on the cellular response irrespective of the presence or absence of the stimulus is therefore needed.

In this experiment, not only the biological response but also the response of the effectors E1 and E2 are examined and scored. R+ stands for full biological response (response of both E1 and E2); R− for total lack of response (no response of both E1 and E2); Rp for partial response (response of only either E1 or E2); Re1− for lack of E1 response; Re2− for lack of E2 response. Some location numbers are underlined to draw your attention (see text).

**Table 2.** Anticipated responses of cells lacking A expression or expressing a non-conditionally activated A, as examined for the responses of each effector molecule to the stimulus

| Stimulus | Genotype | Response | Location in Cascade II |
|----------|----------|----------|------------------------|
| A        | R+       | 0.1,5,7,8,10 |
| A-       | R+       | 2         |
|          | Rp Re1-  | 3,4       |
|          | Rp Re2-  | 6,9       |
| A*       | R+       | 0.2,3,4,5,6,7,8,9 |
| A-       | R-       | 1         |
|          | Rp Re1-  |           |
|          | Rp Re2-  |           |
| A        | R-       | 1         |
| A-       | R-       | 0.2,3,4,5,6,7,8,9,10 |
|          | Rp Re1-  |           |
|          | Rp Re2-  |           |
| A*       | R+       | 2,4,7,8   |
| A-       | R-       | 0.1,9,10  |
|          | Rp Re1-  | 5,6       |
|          | Rp Re2-  | 3         |

In this experiment, not only the biological response but also the response of the effectors E1 and E2 are examined and scored. R+ stands for full biological response (response of both E1 and E2); R− for total lack of response (no response of both E1 and E2); Rp for partial response (response of only either E1 or E2); Re1− for lack of E1 response; Re2− for lack of E2 response. Some location numbers are underlined to draw your attention (see text).
In addition, examination of each effector response is also useful for finding whether they are controlled by a redundant or non-redundant cascade, or whether or not their pathway branches off to activate more than one effector like Cascade II-1 (see 2 and 3 in Table 2). Whereas A* at No. 2 in this cascade activates both E1 and E2 in the absence of stimulus, A* at No. 3 activates only E1 under the same condition. Similarly, whereas A- at No. 2 eliminates stimulus-invoked activation of both E1 and E2, A- at No. 3 eliminates activation of only E1. Although the proof of branch-off requires an additional experiment to show that A* at No. 2 cannot induce stimulus-free E1 activation in cells with A- at No. 3, the finding of a cascade branch-off cannot be made without examining each effector response. If more than two effectors are involved in the biological response under study, the same cell assays with A- and A* as in Tables 1 and 2 can be done for each effector to elucidate its controlling cascade first and then all cascades elucidated for individual effectors be assembled to draw an entire picture on the cascades that elicit the stimulus-invoked biological response. In summary, by performing the original and modified sets of cell assays, one could determine whether or not a certain factor is involved as an active component in a particular stimulus-invoked interaction cascade, what the cascade looks like and in some case, the actual location of the factor in the cascade if it is involved.

2. Building a solution-oriented personal knowledge base. The second key element of this strategy is to build a personal mini-scale knowledge base suitable for solving the biological problems that challenge you. The current knowledge base and experimental data related to your particular biological question have first to be re-evaluated on the basis of the availability of firm scientific evidence. Review articles are the most popular knowledge source for many researchers who seek to know the current status of the research fields they are engaged or interested in. However, many of these reviews are written on a flat basis that treats all the reported molecular interactions with nearly equal scientific significance even if proper controls were missing and the cell-based biological materials used for analysis were selected poorly. The first step for building the knowledge base needed is to correct one’s own assumptions. In one’s mind, information may be classified into three categories: “True, Undetermined and False” or if it concerns personal need, “Accept, Pending and Ignore” as described in Fig. 4. No matter whether it is for general matters or personal need, the current knowledge base in one’s mind is certainly far from being error-free because we tend to believe what we hear and read repeatedly even in the absence of solid supporting evidence and furthermore, we can be misguided by illusions or careless misunderstandings that often originate from our brain’s innate activities. In the field of experimental science, I believe that three items can be classified into “Accept” in the personal knowledge base after evidence-based correction: 1) Firmly established knowledge; 2) Carefully executed, mutually consistent experimental results obtained with right materials and strict controls; 3) Scientifically
Among the three, the second is far more important for building an error-less knowledge base not only because information in this category is enormous and highly valued but also because the necessity to use right materials has often been overlooked unconsciously. For many studies in life sciences, live animal cells are a central material when addressing various questions, particularly on physiological intracellular processes of higher eukaryotes. However, historically, immortalized lines of rodent cells were commonly used. These were established to be immortal for ease of handling and maintenance by constantly selecting cell populations with higher proliferative potential under the laboratory culture conditions. Worse yet, for the studies of human cells, cancer cells obtained from patients were similarly derived and used simply because unlike rodent counterparts, primary human cells are extremely hard to immortalize and culture. Differing from the original primary cells, the longtime cultured rodent cells and human cancer cells are very likely to contain genetic and epigenetic alterations that may seriously affect the experimental results and therefore their credibility if you aim to understand physiological cellular processes, particularly those related to cell proliferation and its control. We often encountered this problem with established cell lines. Consequently, it is safer to exclude any cell proliferation-related studies with established lined cells or cancer cells from this category unless there is a reason to believe that no compromise occurred on the credibility of the experimental data or such cells were indeed the object for the study.

In this proposal, the item “any conclusions meaning merely consistent with” is categorized into “Ignore” because it simply means that the experimental data obtained are necessary but may not be sufficient to conclude so. Unless other possibilities are carefully considered and excluded, such a conclusion is of little value and it may cause a misunderstanding that the data shown are sufficient to conclude so. On the other hand, one category named “special attention needed” is created for this solution-oriented knowledge base and contains two items. One is “any unexpected results, particularly in your research field, obtained with the right materials and strict controls, yet clearly reproducible and consistent with all of facts and carefully executed experimental results”. Precautions should be taken against unexpected exciting results published particularly in high profile journals because they often prove un reproducible. The other is “newly developed methods and techniques that are likely to aid your current studies”. Special attention should be paid to these two items when information corresponding to either one of them comes open because they might provide a critical clue or an incomparable aid to solving your problem.

**Conjecture-driven experimentation**

The design of experiments governs the entire course and outcome of research the enterprise. Therefore, experiments should be well thought out and designed to minimize the time and effort involved without compromising the credibility of the data obtained. Whenever this is possible, readily available tools and reagents should be used. If initial conjecture is not disproved, perform in-depth analysis to obtain solid evidence in support of the conclusion (Fig. 5). If it is, make another conjecture and perform a quick analysis. If it is not disproved, perform in-depth analysis on the second conjecture and so on until scientifically firm conclusions on your question are obtained.

In the quest for intracellular molecular interaction cascades, the first step is to search for and pick

---

**Fig. 5. Flow-chart of conjecture-driven experimentation.**
the best candidate based on your own experimental data and/or all the “Accepted” information on known molecules in your newly built personal knowledge base (Fig. 4). The molecules that change their biochemical properties or dynamic behaviors in response to the stimulus are among primary candidates. When the best candidate is found, the next step is to design and implement experiments to quickly prove or disprove it at least tentatively. If a specific chemical inhibitor is available, its use is the first choice. If the inhibitor is effective or no inhibitor is available, transient RNAi-mediated knockdown of the candidate can be performed. To make unequivocal judgements in these experiments, confirmation of the effectiveness of the treatments is always required.

If the candidate is not disproved in these quick experiments or no such quick experiments can be done, you could proceed to preparing cells lacking expression of the candidate molecule and cells expressing its non-conditionally activated form, preferably in an artificially inducible fashion. As described, there are several established methods for preparing cells completely lacking expression of any gene. Alternatively, a natural or man-made null mutant cell for the candidate could be used if it retains the original mutant phenotype and re-expression of the factor can restore the wildtype phenotype. On the other hand, preparation of cells expressing a non-conditionally activated candidate molecule may not be possible if no information is available on how to construct its constitutively active form. In such cases, molecules immediate upstream or downstream of the candidate could be manipulated instead if they are already known and cells expressing such manipulated molecules could be used for analyses. Moreover, in case where expression of the candidate molecule is turned off upon stimulation, simple enforced expression of the original molecule could effectively substitute for expression of its non-conditionally activated form. In most cases, whether or not the candidate is a component of the studied cascades can be determined by examining the effects of its elimination in the presence of the stimulus and/or its active form expression in the absence of the stimulus. If it is negative in this assay, yet you still suspect it to be involved, you could perform the elaborate assays described in Tables 1 and 2. Finally, various layers of biochemical and molecular biological confirmatory experiments are needed to draw scientifically well supported conclusions on the candidate molecule and its partners.

Discussion

As described, the central part of the proposed strategy is pairwise use of cells lacking a certain molecule and those expressing its non-conditionally activated form for analysis. By using this strategy, one could find whether or not any given molecule is a component of the interaction cascades that invoke particular biological responses to a certain stimulus and moreover, what the cascade looks like if this is so. But such pairwise use may not always be possible due mainly to lack of information on how to prepare a non-conditionally activated form for some molecules. However, in many cases, such pairwise use may not be needed to make new findings. The strategy may prove powerful even when such cell pairs are unavailable however.

Besides those of the hematopoietic origin, virtually all the cells comprising adult animals require an anchorage to the extracellular matrix mainly made of fibronectin and collagen for their cell cycle onset and survival. We have long been interested in understanding how the anchorage signal controls this fundamental cell attribute because transformed (i.e., malignant) cells proliferate and survive without anchorage and acquiring this ability is thought to underlie their tumorigenicity and metastatic potential. Figure 6 summarizes our recent findings about interaction cascades mediating an anchorage signal that controls the cell cycle start. One key finding is that Cdk6 protein, known as an ATP-dependent remodeling factor that assembles prereplicative complexes onto the origin recognition complex (ORC)-bound replication origins in chromosomes, also activates Cdk2 inactivated by binding of p27Kip1 the protein inhibitor through its ATP-dependent removal as well as obstructs apoptosis assembly by ATP-dependent association with activated Apaf1 molecules, thereby securing cell proliferation when cells have committed chromosomal replication. The other is the signal cascade that controls the G1-S transition in response to cellular anchorage to extracellular matrix proteins.

When deprived of anchorage, virtually all cells comprising solid organs of adult mammals arrest in G1 with inactivation of both the G1 cyclin-dependent kinases Cdk4/Cdk6 and Cdk2. Eventually they die of apoptosis. A decade ago, during a search for cell cycle factors affected by anchorage loss, we incidentally found that in addition to inactivation of the G1 cyclin dependent kinases, Cdk6 the AAA+ ATPase essential for the onset of chromosomal replication...
Fig. 6. Cdc6 as a trifunctional AAA+ ATPase that critically controls the G1-S cell cycle transition and interaction cascades that mediate a cellular anchorage signal to control this transition. **A.** Cdc6 as a remodeling factor that critically controls the mammalian G1-S transition. Cdc6 assembles prereplicative complexes (preRC) by loading minichromosome maintenance (MCM) complexes onto the origin recognition complex (ORC)-bound replication origins with consumption of an energy. In parallel, Cdc6 activates Cdk2-cyclin A complexes inactivated by the protein inhibitor p27Kip1 by removing the bound p27Kip1 in an ATP-dependent manner. Activated Cdk2 phosphorylates several factors that bind prereplicative complexes and finally promotes recruitment of DNA polymerase α and the onset of chromosomal replication. Meanwhile, Cdc6 protein binds induced Apaf1 proteins and obstructs their assembly into active apoptosome to ensure safe cell proliferation. This figure was a modification and combination of those published previously. **B.** For the original cascades elucidated, see ref. 30. Rho-activated ROCK under the presence of cell’s anchorage to the extracellular matrix proteins inactivates TSC to activate Rheb that activates mTORC1. Activated mTORC1 activates Cdk4/Cdk6, which in turn activate E2F-regulated genes such as those for Cdc6, cyclin A, Emi1 and Apaf1. Expressed Emi1 inactivates the APC/CCdh1 ubiquitin ligase that degrades Cdc6 and cyclin A, resulting in stabilization of these proteins (not shown in this figure, see ref. 30). The Cdk2 complexed with cyclin A remains inactivated by association with p27Kip1. Meanwhile, active ROCK phosphorylates the C-terminus of the Cdk2-bound p27Kip1. Expressed and stabilized Cdc6 protein activates replication origins and also removes the C-terminally phosphorylated p27Kip1 from the Cdk2, thereby activating this cyclin-dependent kinase required for the onset of chromosomal replication. In parallel, Cdc6 forms stable complexes with E2F-dependently induced and cytochrome C or else-activated Apaf1 molecules to prevent apoptosome assembly and subsequent cell death, securing cells from death during proliferation. All the three functions of Cdc6 are its ATPase-dependent.
disappears completely upon anchorage loss.\textsuperscript{26} We
began to search for the underlying mechanism with rodent embryonic fibroblasts forcedly expressing Cdc6 as an assay host. When this expressor was treated with rapamycin a highly specific inhibitor of mTOR complex 1 (mTORC1) (A- of mTORC1), Cdc6 protein disappeared just as it did when cells were deprived of anchorage, implying that the mTOR pathway might mediate a signal for cellular anchorage. To confirm this possibility, we chose to use tsc2\textsuperscript{−/−} rat embryonic fibroblasts and prepared the cells similarly forcibly expressing Cdc6 protein.\textsuperscript{27} TSC2 forms the GAP complex TSC with TSC1 to inactivate Rheb the small GTPase that activates mTORC1 the evolutionally highly conserved serine/threonine-targeting protein kinase (Fig. 6B). Because the functional GAP is missing, mTORC1 is constitutively active in tsc2\textsuperscript{−/−} cells. Thus, this manipulated cell is equivalent to the one expressing A\textsuperscript{*} of mTORC1. As expected, Cdc6 was stable in this cell during anchorage loss. If forcibly expressed Cdc6 disappears upon anchorage deprivation when TSC2 is re-expressed in this cell, one could conclude that the mTOR pathway mediates an anchorage signal at least for controlling Cdc6 protein stabilization. This was indeed the case.\textsuperscript{27} Then we examined what part of the anchorage loss effects was reversed by forced activation of mTORC1. Reversed was the activity of the Cdk4/Cdk6-E2F cascade, resulting in activation of E2F-regulated genes including those for Cdc6, cyclin A and E and a natural inhibitor of the APC\textsuperscript{Cdh1} ubiquitin ligase that degrades Cdc6, cyclin A and others.\textsuperscript{27} Thus, the major mechanism for anchorage-controlled stabilization of Cdc6 protein was understood.

The next question we addressed is how the anchorage signal originating from integrins in the plasma membrane controls the mTORC1 pathway. As proposed above, we initially tested chemical inhibitors for protein kinases that have been implicated to participate in integrin-linked cytoskeleton organization controls and a chemical that disrupts actin filaments and inhibits cell division. Of 6 chemicals tested, staurosporine an inhibitor for protein kinase C, myosin light chain kinase and focal adhesion kinase, Y27632 a highly specific inhibitor for Rho-associated kinase (ROCK), and cytochalasin D that disrupts actin filaments were found to cause the same effects as anchorage deprivation on the mTORC1 pathway and the G\textsubscript{1} cell cycle factors including Cdc6.\textsuperscript{28} To narrow down candidates further, we tested the tsc2\textsuperscript{−/−} cell complemented with the shortest TSC2 isoform for its sensitivity to these three chemicals because as described above, we knew that this cell is susceptible to anchorage loss. Quite luckily, unlike original rat embryonic fibroblasts predominantly expressing TSC2 isoform 5 that contains one extra exon, this cell turned out to be sensitive to the ROCK inhibitor but not to staurosporine or cytochalasin D and in addition, we noted that anchorage loss inactivates ROCK.\textsuperscript{28} These results strongly suggested that ROCK is the molecule that connects the anchorage signal to the mTOR pathway. This suggestion was confirmed by effective reversal of anchorage loss-induced mTORC1 inactivation by expression of a non-conditionally activated ROCK (A\textsuperscript{*}), which was constructed by deleting its C-terminal region.\textsuperscript{29}

The next question we addressed is where the ROCK target site is in the Tsc complex. There are 35 sites in the complex that match up to the minimal consensus sequences for ROCK phosphorylation, but only one site in TSC2 is evolutionarily conserved from human through fruit fly and present even in the shortest TSC2 isoform. By conducting construction and functional analysis of an unphosphorylatable mutant of this site, we concluded that this site is phosphorylated \textit{in vivo} by ROCK and is essential for an anchorage signal to activate the mTORC1 pathway.\textsuperscript{28} The key turning point that led us to the conclusion without time consumption was played by the use of the shortest form (lacking two exons) of TSC2 for complementation of tsc2\textsuperscript{−/−} cells and by the decision to follow the general assumptions in the proposal and pick the unique evolutionarily conserved consensus sequence present in TSC2 as the best candidate for the physiological ROCK phosphorylation site, dispensing with a highly laborious orthodox phosphorylation site analysis. In this regard, it is noteworthy that this shortest TSC2 molecule is a partially inactivated form (partial A-) of a double negative regulator, which effectively eliminates the need of staurosporine/cytochalasin D-suppressible signal flow-in for inactivating the TSC complex.

In the studies with Rock, we noted that unlike those lacking TSC, cells expressing a constitutively activated ROCK (A\textsuperscript{*}) retain active Cdk2 during anchorage deprivation and can proliferate in anchorage-free methylcellulose semisolid medium though slowly. The activated ROCK in this situation is equivalent to the factor A\textsuperscript{*} located at No. 2 in Cascade II-1, whose signal branches off and activates an additional effector E2 in the absence of stimulus.
(Fig. 1). Consequently, this finding prompted us to search for the mechanistic basis for the ROCK-involved activation of Cdk2. There were several articles that reported C-terminal phosphorylation of p27\textsuperscript{Kip1} the major inhibitor of Cdk2 and argued its possible involvement in nuclear exclusion of this inhibitor.\textsuperscript{30} But one article provided direct evidence for the importance of this modification for Cdk2 activation,\textsuperscript{31} which prompted us to pay attention to this site phosphorylation and perform a series of studies to find that phosphorylation of this site is required for activation of Cdk2, that an active ROCK phosphorylates this site\textit{in vitro} as well as\textit{in vivo} and that Cdc6 is additionally required for Cdk2 activation.\textsuperscript{32} These three pieces of information led us to finally identify an elaborate mechanism that controls Cdk2 activity during the G\textsubscript{1}-S transition. In this mechanism, Cdc6, the AAA+ ATPase previously identified as the assembly factor for the formation of preRC on ORC-bound chromosomal replication origins, removes the bound p27\textsuperscript{Kip1} from p27\textsuperscript{Kip1}-bound inactive Cdk2-Cyclin A/E complexes through its ATP-dependent remodeling function after anchorage-activated ROCK phosphorylates the C-terminal threonine of the bound p27\textsuperscript{Kip1}.\textsuperscript{32,33}

This regulatory cascade corresponds to the pattern of Cascade I-6 in Fig. 1, which requires three cascade signals originating from one stimulus for invoking one particular response.

The last development in this series of studies was discovery of an additional unexpected function for Cdc6 the AAA+ ATPase: efficient obstruction of apoptosome assembly and consequent cell death by its ATPase-dependent formation of stable complexes with activated Apaf1 molecules.\textsuperscript{34} Thus, Cdc6 has a total of three distinct biological functions that promote the G\textsubscript{1}-S transition while suppressing cell death, all relying on its ATP-dependent remodeling activity.\textsuperscript{34,35} This finding could not have been made without use of cells forcedly expressing Cdc6, which are equivalent in this situation to those expressing a non-conditionally activated form of a given molecule discussed throughout.

As already clear, this proposed strategy is not intended to be used as a tool for the proteomics-type approach, which relies on detection of physical interactions under non-physiological conditions and produces massive data in a short period of time as an extension of an orthodox analytical science. Instead, this is intended to be used as a tool for the reverse genetics-type approach, which relies on detection of functional interactions always in cells under proper physiological conditions and moves step by step with deep deliberation. In this regard, as shown in our recent work on the G\textsubscript{1}-S transition control, use of A* for positive regulators and A- or partial A- for double negative regulators appeared to better serve for solving the problems than other combinations. The reason is simple and as follows. Any molecules involved in a particular interaction cascade are of two types: dynamic components which directly serve to transfer the signal from the immediate upstream component to the immediate downstream component, and static components which support dynamic components to function properly and effectively. Any molecules for which a non-conditionally activated form is makeable can be dynamic components if they are components of a particular cascade and therefore, would greatly facilitate identification of the basic structure of the cascade without time consumption. In addition, use of A* for positive regulators or A- or partial A- for double negative regulators if present is essential to find redundant pathways. As shown by Combinatory response 1+3+6+7 in Table 1 lower part, components in redundant pathways can be detected only by analysis of A* cell in the absence of stimulus.

Acknowledgment

I thank Dr. Michael J. Brownstein for critical reading of the manuscript and useful suggestions and comments.

References

1) Descartes, R. (1994) Discours de la méthode (ed. Heffernan, G.). University of Notre Dame Press, Notre Dame, Indiana.

2) Beaney, M. (2015) Analysis, In The Stanford Encyclopedia of Philosophy (ed. Zalta, E.N.), http://plato.stanford.edu/archives/spr2015/entries/analysis/.

3) The human genome project: https://www.genome.gov/10001772/all-about-the-human-genome-project-htgs/.

4) Novelline, R. (1997) Squire’s Fundamentals of Radiology. Harvard University Press, Cambridge, MA. ISBN 0-674-83339-2.

5) Betzig, E., Patterson, G.H., Sougrat, R., Lindwasser, O.W., Olenych, S., Bonifacino, J.S., Davidson, M.W., Lippincott-Schwartz, J. and Hess, H.F. (2006) Imaging intracellular fluorescent proteins at nanometer resolution. Science 313, 1642–1645.

6) Abdellah, Z. et al. (2004) Finishing the euchromatic sequence of the human genome. Nature 431, 931–945.
interactome. Proc. Natl. Acad. Sci. U.S.A. 98, 4569–4574.
8) de Mol, N.J. (2012) Surface Plasmon Resonance for Proteomics. Methods Mol. Biol. 800, 33–53.
9) Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. Science 298, 1912–1934.
10) Xu, G. and Jaffrey, S.R. (2013) Proteomic Identification of Protein Ubiquitination Events. Biotechnol. Genet. Eng. Rev. 29, 73–109.
11) Pelizon, C., Madine, M.A., Romanowski, P. and Laskey, R.A. (2000) Unphosphorylatable mutants of Cdc6 disrupt its nuclear export but still support DNA replication once per cell cycle. Genes Dev. 14, 2526–2533.
12) Vervoorts, J. and Luscher, B. (2008) Post-translational regulation of the tumor suppressor p27Kip1. Cell. Mol. Life Sci. 65, 3255–3264.
13) Gilmer, T.M. and Erikson, R.L. (1981) Rous sarcoma virus transforming protein, p60src, expressed in E. coli, functions as a protein kinase. Nature 294, 771–773.
14) Jackson, D.A., Symons, R.H. and Berg, P. (1972) Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc. Natl. Acad. Sci. U.S.A. 69, 2904–2909.
15) Hood, L. (1988) Biotechnology and medicine of the future. JAMA 255, 1837–1844.
16) Baltimore, D., Berg, P., Botchan, M., Carroll, D., Charo, R.A., Church, G., Corn, J.E., Daley, G.Q., Doudna, J.A., Fenner, M., Greeley, H.T., Jinek, M., Martin, G.S., Penhoet, E., Puck, J., Sternberg, S.H., Weissman, J.S. and Yamamoto, K.R. (1999) An essential part for Rho-associated kinase in the transcellular invasion of tumor cells. Nat. Med. 5, 221–225.
17) Sherr, C.J. and Roberts, J.M. (1995) Inhibitors of the mammalian target of rapamycin pathway. J. Biol. Chem. 270, 6275–6283.
18) Okayama, H. (2006) C-terminal phosphorylation controls the stability and function of p27Kip1. EMBO J. 25, 5159–5170.
19) Itoh, K., Yoshioka, K., Akiyama, H., Uehata, M., Ishizaki, T. and Narumiya, S. (1999) An essential role for Cdc6 in anchorage-dependent growth in HeLa cells grown in suspension culture for the selective assay of cells transformed by polyoma virus. Virology 235, 291–294.
20) MacPherson, I. and Montagner, L. (1964) Agar suspension culture for the selective assay of cells transformed by polyoma virus. Virology 23, 291–294.
21) Milo, G.E. and DiPaolo, J.A. (1978) Neoplastic transformation of human diploid cells in vitro after chemical carcinogen treatment. Nature 275, 130–132.
22) Jinno, S., Yageta, M., Nagata, A. and Okayama, H. (2002) Cdc6 requires anchorage for its expression. Oncogene 21, 1777–1784.
23) Arakawa-Takeuchi, S., Kobayashi, K., Park, J.-H., Urunbileg, B., Yamamoto, H., Jinno, S. and Okayama, H. (2010) Mammalian target of rapamycin complex 1 signaling opposes the effects of anchorage loss, leading to activation of Cdk4 and Cdc6 stabilization. FEBS Lett. 584, 2779–2785.
24) Park, J.H., Arakawa-Takeuchi, S., Jinno, S. and Okayama, H. (2011) Rho-associated kinase connects a cell cycle-controlling anchorage signal to the mammalian target of rapamycin pathway. J. Biol. Chem. 286, 23132–23141.
25) Itoh, K., Yoshioka, K., Akiyama, H., Uehata, M., Ishizaki, T. and Narumiya, S. (1999) An essential part for Rho-associated kinase in the transcellular invasion of tumor cells. Nat. Med. 5, 221–225.
26) Okayama, H. (2012) Cell cycle control by anchorage signaling. Cell. Signal. 24, 1599–1609.
27) Kossatz, U., Vervoorts, J., Nickeleit, L., Sundberg, H.A., Arthur, J.S., Manns, M.P. and Malek, N.P. (2006) C-terminal phosphorylation controls the stability and function of p27Kip1. EMBO J. 25, 5159–5170.
28) Urunbileg, B., Yamamoto, H., Park, J.-H., Mohanty, A.R., Arakawa-Takeuchi, S., Jinno, S. and Okayama, H. (2012) Cdc6 protein activates p27Kip1-bound Cdk2 only after the bound p27 protein undergoes C-terminal phosphorylation. J. Biol. Chem. 287, 6275–6283.
29) Mohanty, A.R., Kan, Q., Srivastava, S., Urunbileg, B., Arakawa-Takeuchi, S., Fujita, N. and Okayama, H. (2012) Successive phosphorylation of p27Kip1 at Serine-10 and C-terminus crucially controls its potency to inactivate Cdk2. J. Biol. Chem. 287, 21757–21764.
30) Niimi, S., Arakawa-Takeuchi, S., Urunbileg, B., Park, J.-H., Jinno, S. and Okayama, H. (2012) Cdc6 protein obstructs apoptosis assembly and consequent cell death by forming stable complexes with activated Atp1 molecules. J. Biol. Chem. 287, 18573–18583.
31) Okayama, H. (2012) Cdc6: a trifunctional AAA+ ATPase that plays a central role in controlling the G1-S transition and cell survival. J. Biochem. 152, 297–303.

(Received May 29, 2016; accepted Aug. 9, 2016)