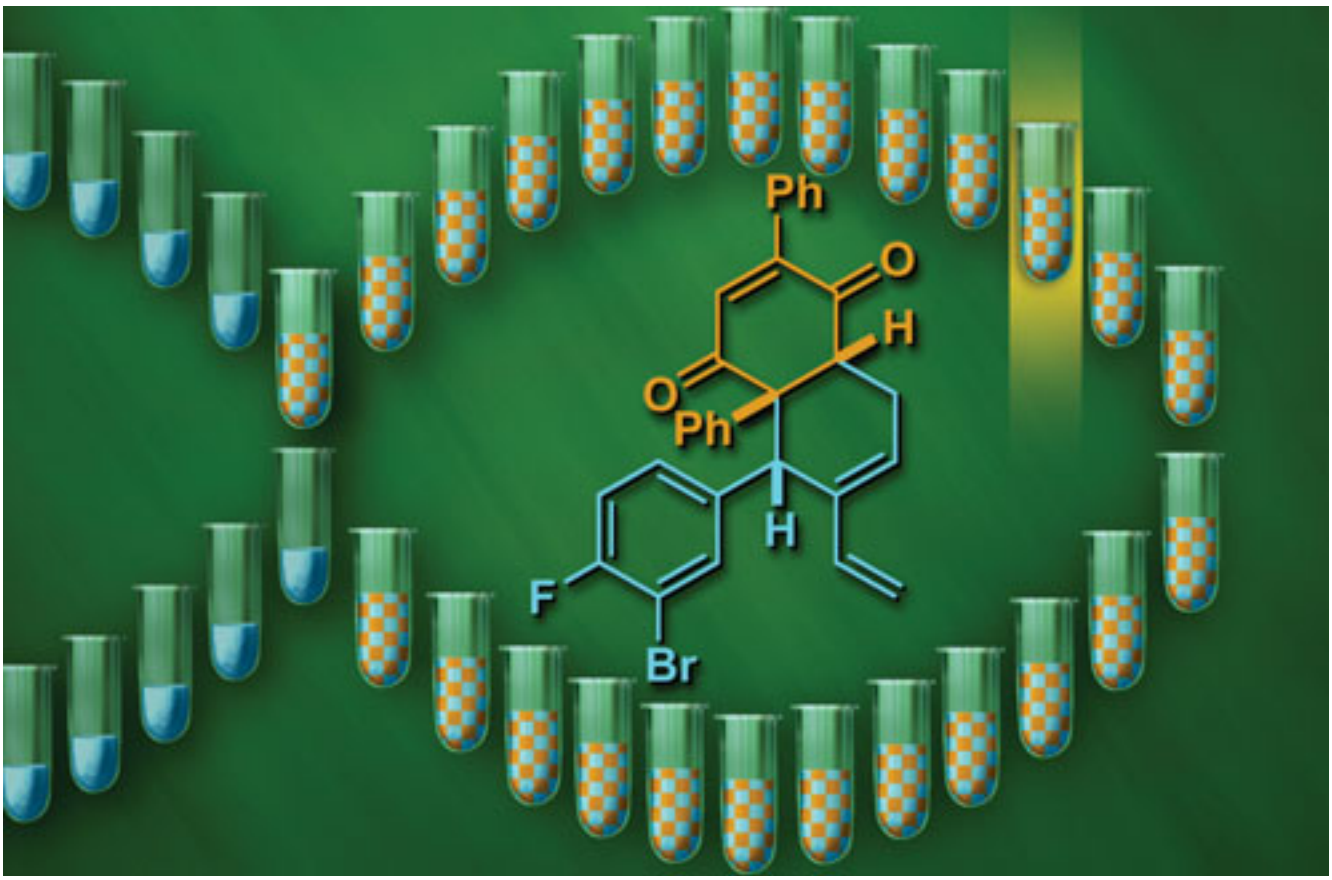


IMAGE BY CATHY NEWTON AND ERIC KELLER, HOWARD HUGHES MEDICAL INSTITUTE

GENERATING DIVERSITY **Using split-and-pool synthesis, a small-molecule intermediate is split into numerous reaction vessels prior to a subsequent step in a divergent synthesis pathway. The process yields complex and skeletally diverse small molecules, which can be screened for biological activity.**

THE SMALL-MOLECULE APPROACH TO BIOLOGY

Chemical genetics and diversity-oriented organic synthesis make possible the systematic exploration of biology

STUART L. SCHREIBER, HARVARD UNIVERSITY

Small molecules have long been associated with biological discoveries, but, in contrast to biochemical and genetic approaches, the small-molecule approach has lacked generality. Although the advances made through the use of small molecules as probes (as distinct from medicines) are impressive, they have in general come about on a case-by-case basis. Advances in diversity-oriented organic synthesis and a focus on the underlying similarities between the small-molecule and genetic approaches have increased the generality of the small-molecule approach. Although a truly systematic way to explore any and all facets of biology with small-molecule modulators has not yet been reached, these advances are beginning to influence researchers on a much broader scale. It is becoming more common for a life scientist to ask, "Should I tackle this problem with small molecules?" or even to state, "The *only* way I can tackle this problem is with small molecules."[1] The latter is becoming increasingly com-

mon as global views of biology are sought.

The discovery principles and platforms enabling this transformation constitute what Rebecca Ward at Harvard University first coined the "chemical genetic" approach on the cover of the inaugural issue of *Chemistry & Biology* nine years ago. Her term reminds us that, to understand a life process, you should perturb it and determine the consequence and that such an approach should strive to have the broad generality and power of genetics. That is, it should allow the probing of life processes in both a systematic and thorough way (analogous to the application of genetics and to the use of saturation mutagenesis, respectively). Chemical genetics is a logical outgrowth and subset of chemical biology, where chemical principles and techniques are used to dissect directly, rather than to model, biology. In this perspective, I aim to focus on the current transition from the ad hoc to systematic use of small molecules to explore the life sciences (rather than to discover new medicines) and the role of organic chemistry in mediating this transition.

The same experiments that encouraged me to explore biology with organic chemistry frustrated me as well. These frustrations had to do with, at the time, the inability of the small-molecule approach (in this context, sometimes referred to as the "pharmacological approach") to be applied with the broad generality of the reductionist-based biochemical and discovery-based genetic approaches. As a result, I was left with an uneasy feeling about the role of organic chemistry in future studies. A secure role for organic chemistry would entail its use "front and center" as a general discovery engine, rather than relying on chance opportunities provided by neighboring disciplines. Overcoming these frustrations required adapting the principles that underlie genetics (and more recently genomics) to chemistry—a process that is proving to be a fertile one for chemistry. For example, much as natural products have driven the development of both target-oriented synthesis (TOS) and synthetic methods, chemical genetics is providing a driving force for the development of diversity-oriented synthesis (DOS; see below).

From a century of genetics-based interrogations of life, we have learned that perturbing life processes and observing the consequences can provide illuminating insights. Geneticists do so through the use of gene mutations, either naturally occurring, randomly induced, or targeted. They have developed powerful analysis tools, such as "epistasis analysis" to order genes in pathways, "synthetic lethal screening" to reveal redundant elements of pathways and networks, and "modifier (suppressor and enhancer) screening" to reveal connections between pathways and networks. These principles and analysis tools are directly applicable to chemical genetics. Here, of course, the wild-type protein is used; it is the binding of a small molecule to the protein that results in a perturbation of function, either inhibition or activation. Recognizing this parallel alone, however, does not ensure a general approach to exploring biology. A research infrastructure involving new advances in chemistry must be developed and integrated into the fabric of day-to-day life science research. It must be routine and readily available to life science researchers, especially to chemists and biologists. In the sections of this perspective, I describe the use of small molecules that have illuminated life processes, the shortcomings of this type of research as a general approach, efforts to overcome these shortcomings, and an assessment of where we stand today.

AD HOC USE OF SMALL MOLECULES TO EXPLORE BIOLOGY.

The past century has yielded many examples of researchers identifying and using small molecules to probe aspects of biology. Some of these served as the subject of this article's prequel, authored 10 years earlier.[2]

Cytoskeleton. Gary G. Borisy and Edwin W. Taylor's use of colchicine, at the MRC Laboratory of Molecular Biology, in Cambridge, England, to identify the tubulin proteins is a classic illustration of using small molecules for discovery in basic biology.[3] Colchicine, along with numerous recently discovered small molecules, targets the α-tubulin/β-tubulin protein-protein interface of microtubules and thus disrupts microtubules in cells. Despite the widely held view that small molecules generally fail to disrupt protein-protein interactions, this interaction appears to be a particularly simple one. For example, in one small-molecule screen, over 300 of 16,000 small molecules were shown to have this property, while two were found to be stabilizers of the same protein-protein interaction.[4] Such molecules have illuminated the functions of microtubules as key cytoskeletal elements.

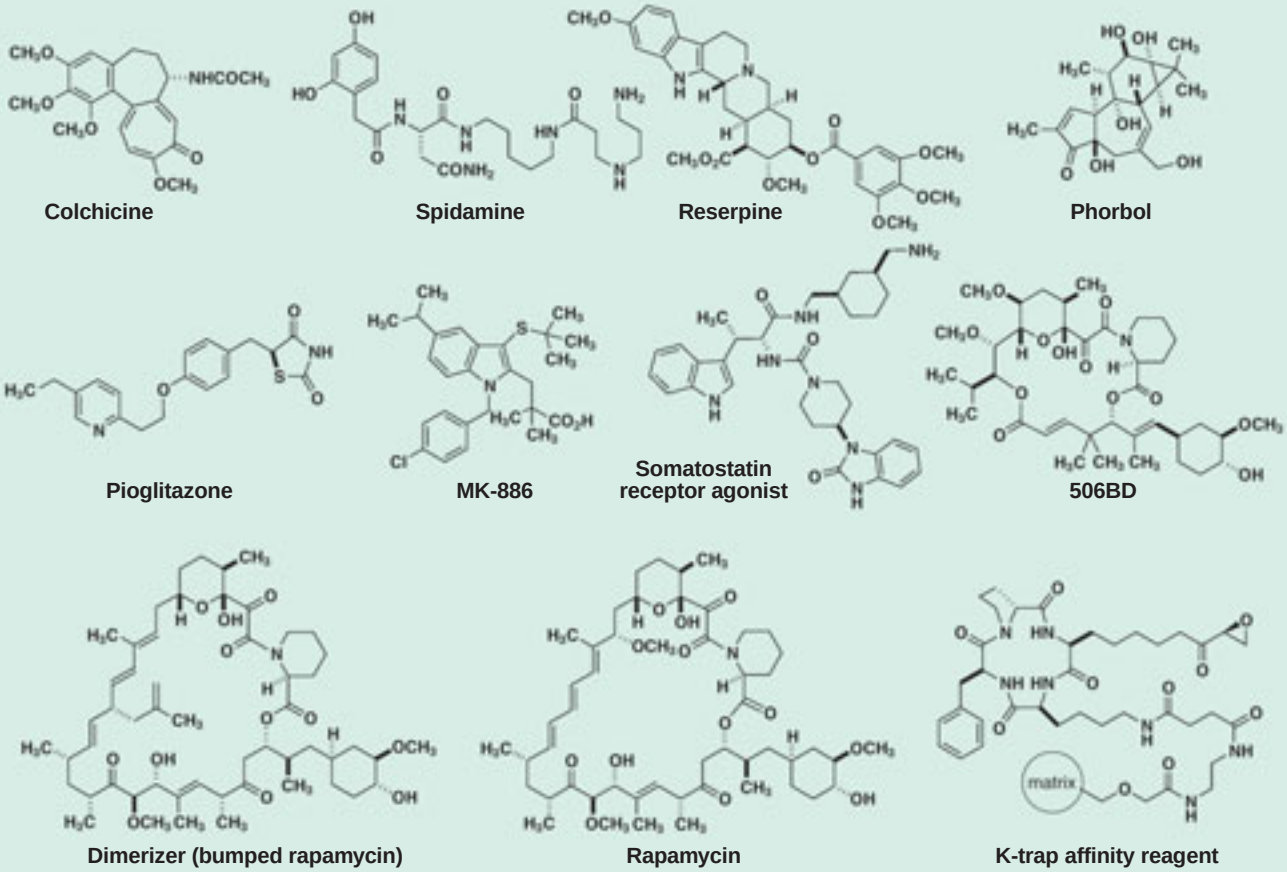
Cytoskeletal research in general has been a rich beneficiary of small-molecule probes.[3b] Microtubule-stabilizing agents such as paclitaxel (Taxol) played an important role in the identification of microtubule-associated proteins (MAPs). Actin-disrupting agents such as cytochalasin and latrunculin have played key roles in unraveling the mysteries of the actin cytoskeleton, and recently discovered small molecules that specifically target motor proteins that carry their cargo along the actin and microtubule polymers are now being used to reveal previously hidden facets of motor function.[3b] Recent experiments concerning the regulation of the cytoskeleton suggest that the use of small molecules to understand this area of research will continue in the future.

Ion channels and signaling in the neurosciences. Neurobiology has long been a beneficiary of the ability of small molecules to target the neurotransmitter receptors and ion channels that function in neurons. As a result, neurobiologists tend to be among the most eager to see advances in chemistry relevant to small molecules. Natural products, especially from snake, spider, bee, scorpion, dinoflagellate, pepper, snail, puffer fish, and soft coral, have played a particularly prominent role in these studies.[5] Scientists at Pfizer, recognizing the treasure trove of ion channel probes stemming historically from spiders, recently developed a fruitful effort in both the isolation of channel-blocking natural products from spiders and the synthesis of optimized variants. These small molecules were used to classify channel subtypes and to probe their functions in neurobiology.

At the National Institutes of Health, Arvid Carlsson's use of reserpine, L-DOPA, and chlorpromazine led to the discovery of the neurotransmitter dopamine and to its role in mediating signals within the nervous system ("chemical transmission in the brain").[6] His studies of dopamine and the dopamine receptor, and of antagonists of their interaction such as chlorpromazine, provided early hints of how extracellular factors, without entering a cell, can give rise to changes in intracellular processes—in other words, signal transduction. For these discoveries, Carlsson was awarded a share of the Nobel Prize in Physiology or Medicine in 2000.

Inner leaflet of the plasma membrane. Studies of the phorbol diesters

"There should be no problem with biology driving science unless perhaps you happen to be a chemist!"



SMALL-MOLECULE PROBES **Colchicine, a probe of tubulin[3]; spidamine, used to study glutamate receptor function[5]; reserpine, used to discover the neurotransmitter dopamine[6]; phorbol (parent alcohol of a family of diesters), used to study a family of protein kinases[7]; pioglitazone (Actos), an activator of the transcription factor PPARγ[8]; MK-886, used to discover the protein 5-lipoxygenase-activating protein (FLAP)[9]; somatostatin receptor agonist, used to study a specific receptor's physiological functions[10]; 506BD, a probe of immunophilin action[11,12]; dimerizer (methallylrapamycin), an inactive ("bumped") variant of rapamycin that, by chemical modification, gained the ability to control proximal relations of signaling proteins in cells and animals[17]; rapamycin, a probe of the nutrient-response signaling network and of the proteins FRAP and TOR[20]; K-trap affinity reagent (lysine variant of trapoxin), used to discover HDAC1.[23]**

played an important role in revealing the key functions of members of a large family of protein kinases named "PKCs" in intracellular signal transduction.[7] They also revealed a docking site used by these proteins to associate with the inner leaflet of the plasma membrane. Phorbol diesters bind to the PKCs and tether them to the leaflet, thereby creating proximal relationships with their leaflet-localized substrates. These studies demonstrate how small molecules can activate the functions of the proteins to which they bind by directing them to specific locales within cells.

Biological insights stemming from pharmaceutical research. Researchers in the pharmaceutical and biotechnology industries, while in search of new medicines, have been particularly effective at discovering new insights into life processes using small-molecule probes. As with all studies using small molecules, two approaches have been used: One emulates the underlying principles of classical genetics (sometimes called "forward genetics") and the other a modern variant of it, reverse genetics.

As an illustration of forward chemical genetics, small molecules were screened at many companies in the 1970s in search of agents capable of treating type 2 diabetes. Pioglitazone, the archetype of the "glitazones," was discovered at Takeda Chemical Industries, in Japan, during this period. Only in more recent years has the target of the glitazones been determined. In accordance with insights gained from human genetics, the glitazones bind and activate the nuclear receptor PPARγ.[8] It is now known that PPARγ plays a key role in diabetes, although that role is still mysterious, as is the molecular etiology of the disease. A second example of this approach derived from research aimed at understanding the molecular basis of inflammation. In mechanistic studies of the anti-inflammatory agent MK-886, this small molecule was used to discover 5-lipoxygenase-activating protein (FLAP) and to assign its cellular and physiological functions.[9] These studies opened the door to a new area of research involving FLAP and its role in inflammation.

As an illustration of reverse chemical genetics, small molecules have also been identified that selectively bind and activate five members (paralogs) of the somatostatin receptor family. Having predetermined the selectivity of these probes toward individual paralogs, scientists at

Merck were able to uncover the distinctive functions of the paralogs.[10] This remarkable investigation used the logic of reverse genetics, in that the researchers targeted the modulation of function (in this case, the more challenging activation of function) of individual paralogs, then searched broadly for the resulting consequences.

Signaling networks in the cytoplasm and nucleus. In my laboratory in the early and mid-1980s, a focus on target-oriented synthesis and on developing synthetic methods eventually led to studies of how the small-molecule objects of our studies perturb the functions of the proteins to which they bind. These studies using natural products allowed us to discover new principles of biological signaling networks, including the commonality of principles underlying both cytoplasmic and nuclear signaling networks. Four studies of natural products revealed basic insights into information transfer in biology.

The discovery of the FK506-binding protein FKBP12 in 1988 (independently discovered in my lab and by scientists at Merck)[11] was facilitated by target-oriented synthesis efforts aimed at, among others, the natural products FK506 and rapamycin and the nonnatural small molecules 506BD and tricyclosporin.[12] Using a number of these small-molecule probes, we determined in 1991 that FK506 and cyclosporin inhibit the activity of the phosphatase calcineurin. This occurs by an unusual mechanism: through the formation of the ternary complexes FKBP12-FK506-calcineurin and cyclophilin-cyclosporin-calcineurin.[13] This work, together with work by Gerald R. Crabtree at Stanford University concerning the NFAT proteins, led to the elucidation of the calcium-calcineurin-NFAT signaling pathway.[14] This proved to be an early example of defining an entire cellular signaling pathway from the cell surface to the nucleus, analogous to that of the Ras-Raf-MAPK pathway elucidated the following year. In subsequent years, the central roles of the calcium-calcineurin-NFAT signaling pathway in immune function, heart development, and memory acquisition were revealed by many researchers working in a variety of fields.[15]

The ability of small molecules to bind two proteins simultaneously inspired our collaboration with Crabtree and members of his laboratory in 1993 to develop "small-molecule dimerizers," which were shown to provide small-molecule regulation of transcription and of numerous signaling molecules and pathways (for example, the Fas, insulin, TGFβ, and T-cell receptors[16, 17]) through proximity effects. We demonstrated that small molecules could be used to influence signaling pathways in an animal with temporal *and* spatial control.[18] Subsequently, many researchers working on many research problems have had success with this approach, and dimerizer kits have now been distributed freely to more than 500 laboratories by Ariad Pharmaceuticals.[19] Its promise in gene therapy has been highlighted by the stable (over several years), small-molecule-induced production of erythropoeitin (EPO) in primates by treating primates with a small-molecule switch for an EPO-inducing transcription factor, and more recently, in Phase II human clinical trials for treatment of graft-versus-host disease.

Members of my group and of Solomon H. Snyder's group at Johns Hopkins University independently discovered in 1994 that the small molecule rapamycin simultaneously binds FKBP12 and the previously unknown protein we named FRAP (FKBP12-rapamycin binding protein, also known as TOR and RAFT).[20] Using chemical genetic epistasis analysis[21], diversity-oriented synthesis, and small-molecule microarrays, among other techniques, we succeeded in uncovering the nutrient-response signaling network involving TOR proteins in yeast and FRAP/TOR in mammalian cells. Small molecules such as uretupamine[22] and rapamycin were shown to be particularly effective in illuminating the ability of proteins such as FRAP, Tor1p, Tor2p, and Ure2p to receive multiple inputs and to process them appropriately toward multiple outputs ("multichannel processors"). The nutrient signaling network now appears to be key to understanding the origins of type 2 diabetes.

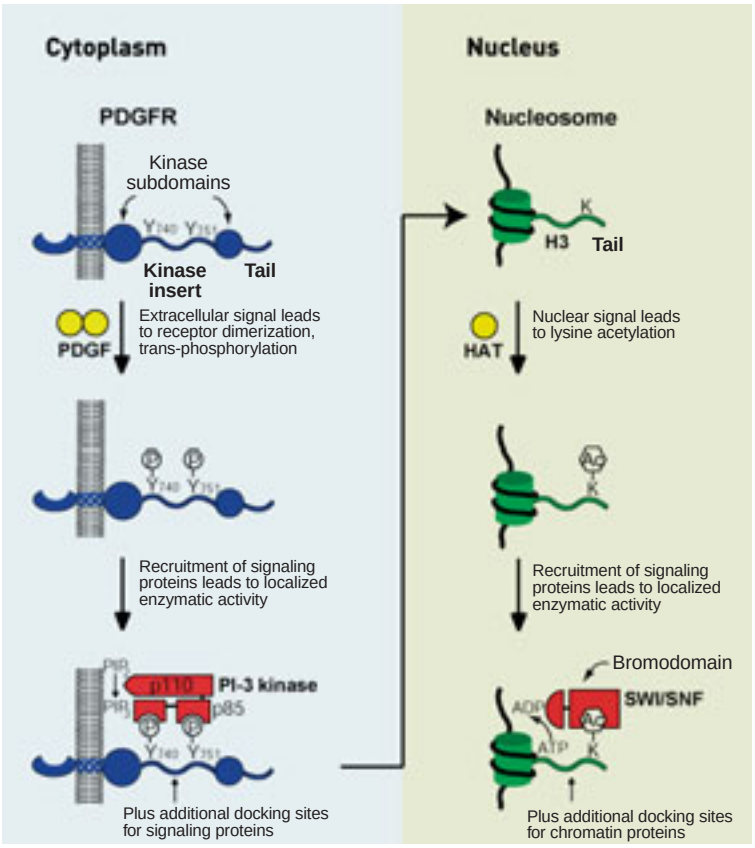
In 1996, my lab used a synthetic variant (an immobilized version of K-trap, page 53) of the natural product trapoxin to molecularly characterize the histone deacetylases (HDACs).[23] Prior to our work in this area, the HDAC proteins had not been isolated—despite many attempts by others who were inspired by Vincent G. Allfrey's detection, at Rockefeller University, of the enzymatic activity in cell extracts more than 30 years earlier. Coincident with the HDAC discovery, C. David Allis and colleagues at the University of Rochester, taking a biochemical approach, reported their discovery of the histone acetyl transferases (HATs).[24] These two contributions catalyzed much research in this area, eventually leading to the characterization of numerous histone-modifying enzymes, their resulting histone "marks," and numerous proteins that bind to these marks. By taking a global approach to understanding chromatin function, we recently proposed a "signaling network model" of chromatin and compared it with an alternative view, the "histone code hypothesis" presented by Allis.[25]

Research by many scientists in this area has shined a bright light on chromatin as a key regulatory element, rather than simply a structural element, in transcription. This research followed previous small-molecule-based studies of signal transduction that helped reveal the existence of cytoplasmic signaling networks. The signaling network model of chromatin[25] posits that both cytoplasmic signaling and chromatin signaling use key elements of networks, including feedback motifs and redundancy that ensure robustness, adaptability, and switchlike behavior. This insight was gained in part by recognizing the remarkable similarities in the principles that underlie information transfer in the cytoplasm and nucleus (page 55). Our view of information transfer in cells has been extended from the early events of signal detection, often at the plasma membrane, to chromatin, where memory of the signal is established in nondividing cells, and inheritance of the signal (epigenetics) is achieved in dividing cells. The concept of signaling pathways (in my opinion, an artifact of the reduction approach) has been evolving toward the concept of signaling networks. The fine temporal control of protein function in cells afforded by small molecules has played a key role in this transition.[25]

SYSTEMATIC USE OF SMALL MOLECULES TO EXPLORE BIOLOGY: CHEMICAL GENETICS.

From the perspective of an organic chemist participating in the study of small-molecule-based signaling networks in 1997, I could not help but be concerned with two issues. First, the skill set of an organic chemist was well suited for responding to discovery opportunities provided by biologists, but not necessarily for initiating or leading the discovery program. This is, in retrospect, not surprising, as it mirrors the role of organic chemists in the process of modern drug discovery. In contrast to drug discovery prior to the mid-1970s, with the advent of molecular biology and molecular cell biology, organic chemists are typically asked to participate in the optimization process following the decision by biologists to select a specific biological target for therapeutic intervention. This concern is admittedly a selfish one—there should be no problem with biology driving science unless perhaps you happen to be a chemist!

IMAGE COURTESY OF STUART SCHREIBER, HARVARD UNIVERSITY



SIGNALING NETWORKS **Small-molecule-based investigations of membrane-to-nucleus signaling and chromatin function suggest the existence of network motifs that ensure robustness, adaptability, and switchlike behavior, as well as other striking commonalities (the signaling network model of chromatin).[25] The figure uses the platelet-derived growth factor receptor (PDGFR) and a nucleosome to illustrate that similar principles underlie information transfer in the cytoplasm and nucleus of cells. When the extracellular growth factor PDGF binds to its receptor outside of a cell, it dimerizes the receptor. The result is to create a high effective molarity of one receptor tail in the vicinity of the other. Since the tails have tyrosine kinase activities, their proximal relationship facilitates transphosphorylation. These phosphorylations take place within flexible regions of the tails, and the phosphate groups complete binding sites for intracellular signaling proteins that have substrates that reside within the inner leaflet of the plasma membrane. The docking of the lipid kinase PI3K, for example, facilitates the phosphorylation of its substrate, the membrane component phosphatidyl inositol-4.5-bisphosphate. A series of subsequent events all proceed by this type of induced proximity, allowing the signal to eventually reach the nucleus. The new insight is that induced proximity is the key to information transfer within chromatin, and that docking sites created when the signal reaches chromatin in the nucleus mediate network behavior. For example, the cytoplasmic signal is received in the form of a histone acetyl transferase (HAT), which deposits an acetyl group on a specific lysine side chain of a nucleosome in the vicinity of a target gene. This completes a binding site for a signaling protein SWI/SNF (pronounced switch-sniff) that, after docking, remodels its now nearby nucleosome substrate. This ATP-driven motor protein mechanically loosens the nucleosome so that the transcription apparatus can access the promoter of a target gene. Small-molecule-based investigations of both networks helped illuminate their fundamental operating principles.**

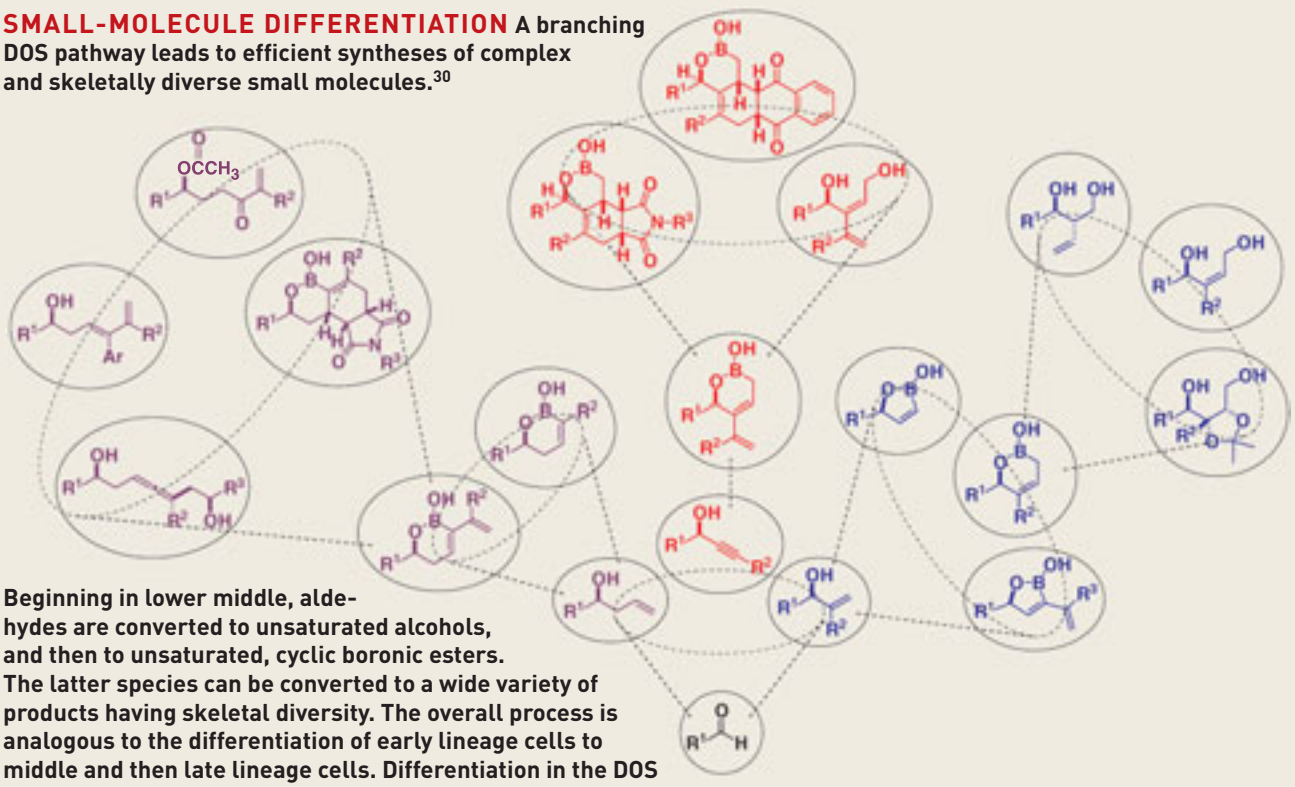
Nevertheless, there is an inevitable feeling of missing out on the front-line action.

The second issue is related, but even more personal. Exploring biology with the ad hoc use of organic chemistry draws the chemist into the seductive world of modern biology. Faced with questions about the relative role of organic chemistry and molecular biology or genetics on any given research undertaking, and wondering about what might happen with the next research undertaking, it is perhaps inevitable for chemists to increase their reliance on biological tools. But this is precisely what can lead to the personal crisis: "Will I be able to rely on my organic chemistry skill set on the next project?" Or even worse: "Am I becoming a biologist?" (Actually, "Am I becoming 'just another' biologist!")

In 1997, this angst was more than balanced by the excitement of what was emerging as a set of opportunities to transition from an ad hoc phase to a systematic one. The key was for organic chemistry to set itself on a course that would allow it to tackle any problem in biology, from the dissection of a pathway to the understanding of complex networks, and eventually even to global biology or molecular physiology (for example, what is the basis of memory and cognition?). In the resulting plan that emerged, the early challenges are proving to be largely of a purely chemical nature, although a merger with information sciences seems inevitable.

Organic chemistry as an initiator of discovery. For organic chemistry to play an initiating role in biological investigations, it will be important for organic chemists to be able to direct effectively synthetic chemistry efforts toward a set of probes—small-molecule modulators—of two sorts. In the first case, a chemist might want to prepare small molecules having overall properties never seen before. In technical terms, we say that such molecules occupy a currently poorly populated region of multidimensional, chemical descriptor space. (Chemical descriptors are computable properties of small molecules; examples include volume, charge, number of bonds with low barrier to rotation, etcetera.) By accessing these small molecules using synthesis, this virgin swath of chemical space can be interrogated. In the second case, a chemist might want to prepare small molecules targeted to a region of chemical space optimal for modulating an area of biological interest. Here, it will be important to understand the relationship of chemical space to multidimensional, biological descriptor space ("biological space").

These goals were the basis, as a first step, of an attempt to formalize a planning algorithm for diversity-oriented synthesis (DOS), analogous to retrosynthetic analysis in target-oriented synthesis (TOS).[26] DOS was able to draw upon technical developments in combinatorial synthesis, which is most often applied in TOS. In combinatorial chemistry the goal is, beginning with a small-molecule probe (or more often, a drug lead), to design a synthesis aimed to densely populate the region of chemical space occupied by the probe/lead; in collo-



SMALL-MOLECULE DIFFERENTIATION **A branching DOS pathway leads to efficient syntheses of complex and skeletally diverse small molecules.[30]**

Beginning in lower middle, aldehydes are converted to unsaturated alcohols, and then to unsaturated, cyclic boronic esters. The latter species can be converted to a wide variety of products having skeletal diversity. The overall process is analogous to the differentiation of early lineage cells to middle and then late lineage cells. Differentiation in the DOS pathway, in this particular example, is dictated by the selection of reagents ("reagent-encoded"). An alternative process ("substrate-encoded") is discussed in the text.

quial terms, "to make analogs." In contrast, in DOS, the goals are either to populate chemical space broadly, or to target broad swaths of chemical space empirically found to overlap with the biology space that characterizes an area of biology (for example, cell-cycle checkpoints) or disease (for example, cancer or diabetes).

To practice DOS in the future, several developments are required. Chemists must master their understanding of reaction transformations and hone their skills in this new type of strategic planning. The latter is particularly challenging and requires a type of strategic planning unfamiliar to chemists practicing TOS. Specifically, chemists must design branched reaction pathways that provide structurally complex and skeletally diverse small molecules in only three or four transformations. To achieve the goals of targeting these products to broad regions of chemical space, the computer increasingly will become the synthetic organic chemist's best friend. Although it has not yet been achieved, we can envision computations that will facilitate an organic chemist's selection of the reagents, building blocks and appendages, and prioritization of conceived DOS pathways that target a swath of chemical space of interest. Even more challenging, computational methods and databases for relating the chemical and biological descriptor spaces must be developed.

These considerations led in 1997 to the creation of the Harvard Institute of Chemistry & Cell Biology (ICCB), most recently sponsored by the National Cancer Institute's (NCI) Initiative for Chemical Genetics (ICG).[27] An early (but interactive) version of ChemBank[28], an NCI-sponsored suite of informatic tools and federated databases, has just been launched on the Internet (http://iccb.med.harvard.edu/chembank). The goal of ChemBank is to provide life scientists unfettered access to the above-described tools. For example, ChemBank should allow life scientists in a remote lab to tailor a DOS pathway emanating from a student's efforts in the student's lab to their own needs. This would require analysis tools at ChemBank that relate the selection of reagents and appendages to the swaths of chemical or biological spaces of interest to the remote lab. In this way, ChemBank would exploit the inherent plasticity of DOS pathways.

A primary goal of ICCB-ICG is to foster ChemBank, in part by the development of systematic ways to explore biology with small molecules, that is, the development of chemical genetics. A related goal is to be able to apply chemical genetics widely, analogous to the way that biochemistry and genetics can be applied to the dissection or interrogation of nearly any aspect of biology. Where we stand in terms of earning the name "chemical genetics" will be discussed in the final section.

Developing DOS as an effective means to populate chemical space. The efficient synthesis of complex small molecules has been accomplished repeatedly in both TOS and DOS through the use of consecutive (or at least coupled) complexity-generating reactions.[26,29] Likewise, it has been a conceptually (although not necessarily technically) simple matter to vary substituents on the resulting skeletons; the split-and-pool strategy using collections of appendages can provide an efficient solution. Far more challenging has been the conception of synthetic pathways leading to small molecules having a large variety of skeletons. Despite only limited success to date, several useful diversity-generating processes have emerged.

DOS pathways are branched or divergent, in contrast to the generally linear or convergent pathways of TOS. One effective means of achieving skeletal diversity in these branched pathways uses "reagent-encoded" processes. This strategy is reminiscent of the differentiation pathways seen in the expansion of pluripotent cells in biology (an example being a stem cell).

The early lineage cell (or stem cell) can differentiate into numerous middle lineage cells under the action of distinct differentiation factors. This process is repeated with the middle lineage cells at numerous stages, leading eventually into terminally differentiated cells (an example being a neuron). To emulate such highly branched biological pathways that lead to highly diverse products, organic chemists begin with their equivalent of a stem cell—a "pluripotent" organic functionality. Popular examples in my lab include aldehydes and terminal olefins. An example of a less useful functionality—at least given *current* capabilities in organic synthesis—closer to a "terminally differentiated state" is a methyl group lacking an adjacent activating group. Pluripotent functionality can be treated with many different reagents (analogous to the differentiation factors) to yield products having many different skeletons and, most importantly, that themselves are subject to the actions of different reagents yielding second, third, and subsequent layers of products having different skeletons. This differentiating process is illustrated on page 56, where the central aldehyde is first treated with nucleophilic reagents yielding unsaturated, secondary alcohols. This second core of functionality is treated with various unsaturated organoboronic ester reagents and a ruthenium catalyst resulting in a variety of skeletally distinct, cyclic and unsaturated organoboronic esters, which themselves are subject to the actions of a variety of electrophilic reagents, including oxidants and aldehydes.[30] Appending processes can be added to the DOS pathway leading in a more straightforward way to diversity, although generally not to skeletal diversity. By analogy, a parallel line of thought leads to branched pathways having substrate-encoded elements; that is, the middle layers of products have structural information encoded within their structures that dictate the skeletons they are capable of obtaining, with many such substrates acquiring their unique final skeleton even under one single condition. Although such diversity-generating processes have not yet been demonstrated, they are under active investigation and appear to be a fertile (and intellectually challenging) area of research (unpublished results of work done by students in my laboratory, Martin D. Burke and Eric Berger).

DOS and information science. Despite these conceptual and experimental advances, the field of DOS has not yet come close to reaching its goals. Although short and efficient DOS pathways have yielded complex and diverse small molecules with novel biological properties[22,31–37] (page 59), the structural diversity is misleading. Each small molecule depicted derives from its own DOS pathway. Even a qualitative analysis of the members emanating from a given pathway reveals that they are disappointingly similar (a notion that has been quantitated by computing molecular descriptors for each compound and comparing them with compounds derived from the same versus different pathways). Of even greater concern is that the selection of compounds has so far been guided only by the organic chemist's knowledge of candidate reactions, creativity in planning DOS pathways, and intuition about the properties likely to yield effective modulators. Retrospective analyses of these compounds show that they tend to cluster in discrete regions of multidimensional descriptor space. Although algorithms exist to identify subsets of actual or virtual compounds that best distribute in chemical space in a defined way (for example, a Gaussian distribution, or an even distribution in a region of chemical space most likely to harbor small molecules able to penetrate the blood-brain barrier), these are of little value to the planning of DOS. In order to facilitate the *planning* of DOS pathways, organic chemists require algorithms that identify the *small subset* of theoretical reactions, reagents, building blocks, and appendages that will yield *products* distributed in chemical space in a defined way.

The goals are to target broad swaths of chemical space empirically found to overlap with the biology space that characterizes an area of biology.

Further complicating matters is that we currently know little about the relationship of chemical space and biological space (other than rare exceptions such as the requirements for passing the blood-brain barrier). One hundred years of organic chemistry has not yet delivered a broadly distributed collection of small molecules that fairly represents the expansive regions of chemical space, so we cannot possibly know which regions correlate most effectively with various biological outcomes. Medicinal chemists are all too familiar with this problem as they attempt to identify predictors of "bioavailability." The scientific community has not yet developed a systematic means to assess how various small molecules perform in a set of assays reasonably well suited for measuring multidimensional biological descriptor space. At an even more primitive level, the scientific community has not yet agreed upon common standards and an ontology (consistent and logical language) that will allow us to compare structures of small molecules and their performance in biological assays.

Despite the as-yet-unsolved problems in this field, there are many reasons to be optimistic about its future. The field has matured to a point where its shortcomings can be identified, and there are encouraging signs of solutions looming on the horizon. Funding agencies have responded in a dramatic fashion, and new and exciting research and training centers are now dotting the landscape. The merging of engineering and information science with organic chemistry is already having a large effect. Information science is already being used to address the synthetic chemistry challenge noted above. ChemBank has as one of its missions the adoption of common standards (for example, a common chemical registration system) and an ontology that will allow the management and sharing of small-molecule-derived data. We hope that ChemBank will be a planning and discovery tool for chemists and biologists worldwide, the only necessities being a computer and access to the Internet. Finally, we only have to turn to our students to gain a real glimpse of the future. The fates of research fields are more in the hands of eager young scientists entering the training phase of their career than senior scientists who might have a tendency to avoid areas representing less familiarity and greater uncertainty. In this light, the challenges of DOS, although formidable, would appear to be reachable in the near future. Young students see uncertainty as opportunity.[38] My personal experience is that they also excel at applying their creative potential to this fertile area.

Making chemical genetics accessible: the development of discovery platforms. At ICCB-ICG, we have been developing an integrated set of techniques aimed at systematizing the application of small molecules, including DOS-derived small molecules, to biology. Chemistry technology platforms are key, as the products of DOS should be prepared in a way that ensures their effective integration into screening experiments and in a way that ensures high purity and accuracy of struc-

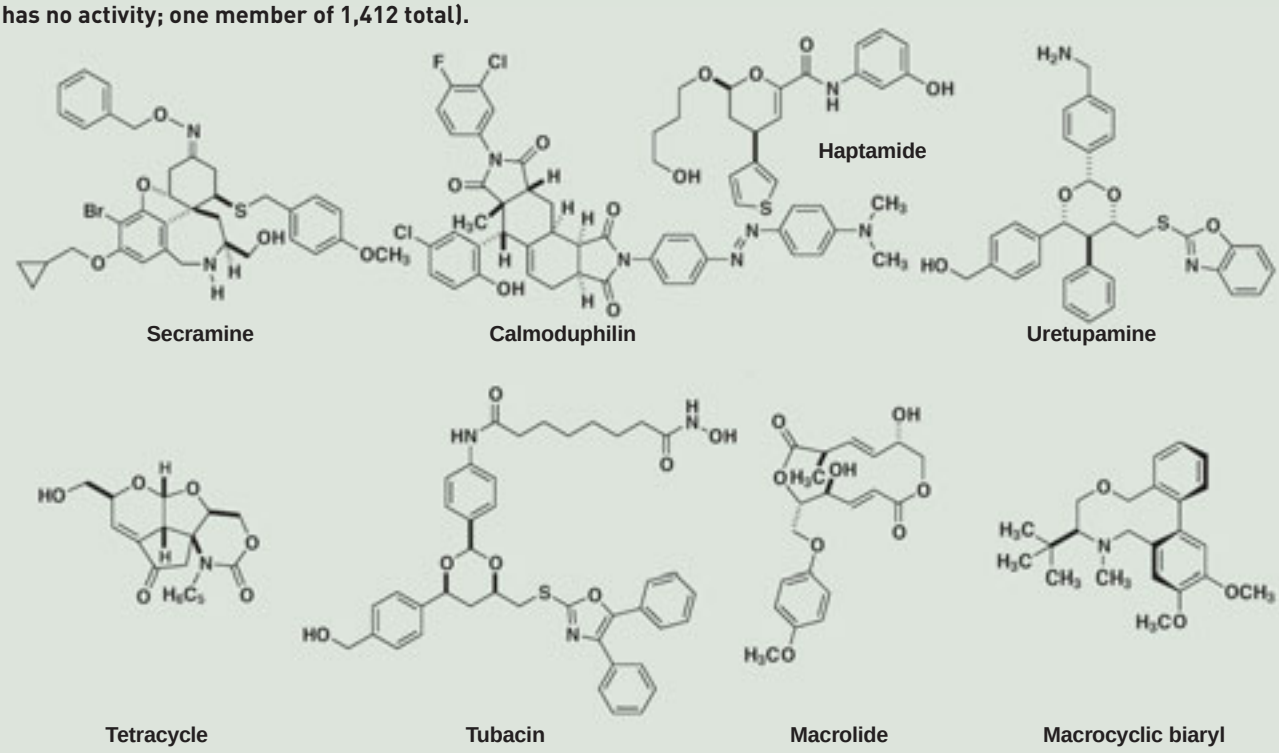
ture assignment. One chemistry technology platform, based on the "one bead/one stock solution" strategy[39a,b], has been developed at ICCB-ICG and is serving its users satisfactorily. It is being refined continuously and in ways that continue to allow its adoption in typically resource-limited academic settings. Other platforms are being developed elsewhere that might also be accessible and effective. ChemBank will benefit from these different platforms, so long as common standards and language are adopted.

In a similar way, we have developed several effective techniques for small-molecule screening, including cytoblot assays[4,40], screening-by-imaging using cells and organisms[41a,b], and small-molecule[42] and protein[43] microarrays. These techniques have already yielded new insights into biology. As an early example, in collaboration with ICCB codirector Timothy J. Mitchison and members of his lab, we used a cytoblot screen and screening-by-imaging to discover monastrol—the first small-molecule inhibitor of mitosis that does not target tubulin.[41a] Monastrol and, more recently, a more potent DOS-derived compound, were shown in the Mitchison lab to inhibit Eg5, a kinesin motor protein.[41a] This discovery provided a powerful probe of the functions of this motor protein and a new medical lead. More than 50 labs have performed more than 100 chemical genetic screens at ICCB-ICG, leading to many small-molecule probes and insights into biology.

Moving toward a chemical genomics. Most of these studies have resulted from small-molecule interrogation of specific problems in biology. In the future, small molecules will be used to probe global biology; this is an especially fertile area for organic chemistry ("chemical genomics," analogous to chemical genetics). To facilitate such studies at ICCB-ICG, we have introduced the concept of and are building a new laboratory ("ICCB-Kendall Square") for small-molecule annotation[39b] and profiling[40], additional key elements of ChemBank. Here, the outcome of all experiments tends to be more important than the outcome of any individual experiment. By analyzing the outcome of the complete matrix of a given collection of small molecules screened against a large collection of proteins and a large collection of phenotypic assays, chemists can envision many exciting outcomes. These include the development of methods to profile biological states with small molecules; to identify the protein target of a small-molecule modulator identified in a phenotypic screen; and to understand the relationship between chemical and biological descriptors and, ultimately, chemical space and biological space.

ASSESSMENT OF WHERE WE ARE AND OF PROSPECTS FOR THE FUTURE. Much like the field of DOS, the field of chemical genetics is in an early formative stage. Encouraging experiments have been recorded, challenges have been identified, and solutions are being pursued. The successes, however, have not yet matched the distinguished history of the "ad hoc stage" of small-molecule explorations of life science. However, as in the field of DOS, there are many reasons to feel optimistic about the future.

DIVERSITY-ORIENTED SYNTHESIS **Structures of representative small molecules synthesized using DOS principles and prepared as 5-mM stock solutions using the one bead/one stock solution technology platform: secramine[31] (specific modulator of protein trafficking out of the Golgi apparatus; secramine is one member of a total of 2,800 small molecules prepared using DOS); calmoduphilin[32] (K_d = 0.12 uM/calmodulin; one member of 29,400 total); haptamide[33] (inhibitor of Hap3p-mediated transcription; one member of 4,320 total); uretupamine[34] (binds to Ure2p and activates Nil1p-mediated transcription; one member of 32,000 total)[22]; tetracycle (one member of 2,500 total); tubacin[35] (tubulin deacetylase inhibitor; one member of 7,200 deacetylase-biased dioxanes); macrolide[36] (one member of 36 total); macrocyclic biaryl[37] (affects cardiovascular system during zebrafish development; enantiomer has no activity; one member of 1,412 total).**



For chemical genetics to reach its full potential, it must be embraced as a general tool for exploring biology. As an approach to dissecting biology, like the genomic approach, it is not yet viewed in the same light as biochemical or genetic approaches. Two key challenges go hand in hand: Chemical genetic methods must be accessible, and the concepts behind chemical genetics must permeate the thinking of life scientists. Efforts to address the former have been described throughout this perspective. Is the concept of interrogating biology with small molecules sinking into the mind-set of life scientists on a broader scale than in the past? Recent evidence suggests so.

For chemical genetics to reach its full potential, it must be embraced as a general tool for exploring biology.

The past several years have seen a surge both in the number of reports of biological systems being dissected with small molecules and in the number of institutional commitments to establishing the requisite infrastructure. In response to numerous queries, we have posted on the ICCB-ICG website a how-to guide for building an academic screening facility (Caroline Shamu at Harvard University, http://iccb.med.harvard.edu/screening/faq_hts_facility.htm). Last year, the Howard Hughes Medical Institute announced its plans to develop a new research campus, Janelia Farm in Leesburg, Va., with chemical genetics as one of several central elements of its unique, integrated, and collaborative structure. At ICCB-ICG alone, we have provided small molecules and performed screens for a rapidly (and somewhat alarmingly) growing number of laboratories nationwide. Basic biological insights gained from chemical genetics studies from many labs worldwide signal a noteworthy trend in the past several years. To illustrate, I have selected several representative examples from just two fields: developmental biology and cell signaling.

Developmental biology. Developmental biologists uncovered the hedgehog-signaling network, including an outline of its downstream signaling elements. This pathway is involved in the development of animals and the maintenance and repair of adult cells, and its disruption results in certain types of cancer. The mysteries of the pathway's mechanistic details are deepening. Scientists at Curis Inc.[44] and in Philip A. Beachy's lab[45] at Johns Hopkins dissected the pathway with an open embrace of genetic principles, yet using small molecules as the source of perturbation. Using chemical genetic screens, the researchers have uncovered small-molecule agonists and antagonists of hedgehog signaling. Mechanistic studies of these probes revealed that hedgehog proteins, which act through the integral membrane protein Patched, de-repress the G-protein coupled receptor Smoothened, which is targeted by the probes. These studies point to the exciting possibility of endogenous, small-molecule regulators of the Smoothened proteins. An entire issue of the *Journal of Biology* (December 2002) was recently devoted to these advances. Continuing with the theme of development and differentiation, Helene Gilgenkrantz at the Cochin Institute, Paris, and coworkers used small-molecule dimerizers in mice to ablate and regenerate, in a dose-dependent fashion, liver cells (hepatocytes).[46] In addition, Peter Schultz, at Scripps Research Institute, and coworkers in two separate studies used chemical genetic screens to identify small molecules that reverse myotube formation[47] and that promote the differentiation of mesenchymal progenitor cells into an osteoblast lineage.[48]

Cell signaling. Using the principle of chemical genetic synthetic lethal screening[40], Philip Leder, Stanley J. Korsmeyer, and colleagues at Harvard Medical School screened matched cell lines differing only in their levels of expression of the Neu oncogene.[49] A small molecule was identified that interfered with cell function only in cells overexpressing Neu. Mechanistic studies revealed that the small molecule targeted the mitochondrial proton gradient. This is an exciting experiment both in design and outcome. That an oncogene renders cells sensitive to disruptors of mitochondrial function reveals a previously hidden facet of cancer cell circuitry. That mitochondria are integral components of the apoptotic signaling network is well appreciated. However, insights into their functional role in signaling were gained in early 2003 when a new apoptotic signaling oncoprotein was discovered as the target of a small-molecule modulator, which was brilliantly uncovered using another chemical genetic screen.[50] Downstream of these proteins lies the key signal integrating protein p53. Using protein-binding screens, scientists at Pfizer identified a small molecule that stabilizes the folding of mutant forms of DNA-binding p53; these probes are illuminating the role of oncogenic mutations in cancer and the possibility of reversing their consequences through biophysical means.[51] To explore the function of another nucleic acid-protein interaction, Peter Beal and coworkers at the University of Utah identified a small-molecule probe of PKR, an RNA-dependent protein kinase thought to survey cells for evidence of viral invasion.[52] Finally, in a study that links signaling to the previously discussed cytoskeleton, fascinating insights have been gained into a signaling pathway that regulates actin nucleation and polymerization. A cyclic peptide[53a] and the small molecule wiskostatin[53b] were identified at Harvard Medical School by Marc W. Kirschner and colleagues in chemical genetic screens and found to bind to the neuronal-Wiskott Aldrich Syndrome Protein (N-WASP). Wiskostatin binding to N-WASP stabilizes an autoinhibited state and prevents N-WASP from activating a downstream actin-nucleating complex (Arp2/3), which functions to nucleate the actin polymer. The small-molecule probes revealed a previously unidentified mechanism for targeting interactions between key proteins in cells.[53]

IN SUMMARY, the systematic population of chemical space with small molecules using DOS and information science, the systematic probing of biology space with these small molecules, and the sharing of results and data using common standards and language on public databases promise a future understanding of the relationship of chemical and biological spaces. This is an exciting and important prospect, and one that will require organic chemistry to play a discovery role. Synthetic organic chemists in particular will need to meet the challenges of DOS. With such an understanding and with the availability of appropriate tools, organic chemists may in the future design synthetic pathways yielding small molecules targeted to the swath of chemical space optimal for modulating the swath of biological space of interest. Although different in its reliance on empirical observations that permit "targeting," the desired capabilities are reminiscent of those provided by natural selection, the process that yields natural products. A noteworthy difference is the anticipated time frame of the former relative to the latter!

Stuart L. Schreiber *is an investigator with the Howard Hughes Medical Institute and is chair of the department of chemistry and chemical biology at Harvard University. He is a founder of the Harvard Institute of Chemistry & Cell Biology–Initiative for Chemical Genetics.*

ENDNOTES

1. Lander, E., commenting on the challenge of understanding cell circuitry.
2. (a) Schreiber, S. L. *Chem. Eng. News,* 70, No. 43 (1992): 22–32. (b) See also Hung, D. T. et al. *Chem. Biol.,* 3 (1996): 623–640.
3. (a) Weisenberg, R. C. et al. *Biochemistry* 7 (1968): 4466–4479. (b) For an excellent account of this exciting area of research, see also Peterson, J. R., and T. J. Mitchison. *Chem. Biol.* 9 (2002): 1275–1285.
4. Haggarty, S. J. et al. *Chem. Biol.* 7 (2000): 275–286.
5. For a review of neurotoxins, see *Trends in Neuroscience,* June 1996, supplement.
6. Carlsson, A. *J. Psychopharmacol.* 4, No. 3 (1990): 120–126.
7. Blumberg, P. M. *Crit. Rev. Toxicol.* 8 (1981): 199–234.
8. Rosen, E. D. et al. *Genes Dev.* 14 (2000): 1293–1307.
9. Miller, D. K. et al. *Nature* 343 (1990): 278–281.
10. Rohrer, S. P. et al. *Science* 282 (1998): 737–740.
11. Schreiber, S. L. *Science* 251 (1991): 283–287.
12. Belshaw, P. J. et al. *Synlett* (1994): 381–392.
13. Liu, J. et al. *Cell* 66 (1991) 807–815.
14. Schreiber, S. L., and G. Crabtree. *Harvey Society Lectures* 89 (1997): 373–380.
15. Crabtree, G. R., and E. N. Olson. *Cell* 109 (2002): 867–879.
16. Yang, J. et al. *Curr. Biol.* 8 (1997): 11–18.
17. Stockwell, B. R., and S. L. Schreiber. *Curr. Biol.* 8 (1998): 761–770.
18. Spencer, D. et al. *Curr. Biol.* 6 (1996): 839–848.
19. http://www.ariad.com.
20. (a) Brown, E. J. et al. *Nature* 369 (1994): 756–758. (b) Sabatini, D. M. et al. *Cell* 78 (1994): 34–43.
21. Shamji, A. F. et al. *Curr. Biol.* 10 (2000): 1574–1581.
22. Kuruvilla, F. G. et al. *Nature* 416 (2002): 653–657.
23. Taunton, J. et al. *Science* 272 (1996): 408–411.
24. Brownell, J. E. *Cell* 84 (1996): 843–851.
25. Schreiber, S. L., and B. E. Bernstein. *Cell* 111 (2002): 771–778.
26. Schreiber, S. L. *Science* 287 (2000): 1964–1969.
27. (a) Baum, R. *Chem. Eng. News* 76 No. 46 (1998): 31–34. (b) Gura, T. *Nature* 407 (2000): 282–284, (c) For a review of ICCB-ICG, see: http://iccb.med.harvard. edu.
28. Adam, D. *Nature* 411 (2001): 873.
29. For example, see: Lee, D. et al. *Org. Lett.* 2 (2000): 709–712.
30. (a) Micalizio, G. C., and S. L. Schreiber. *Angew. Chem. Int. Ed.* 41 (2002): 152–154. (b) Micalizio, G. C., and S. L. Schreiber. *Angew. Chem. Int. Ed.* 41 (2002): 3272–3276.
31. Pelish, H. E. et al. *J. Am. Chem. Soc.* 123 (2001): 6740–6741.
32. Kwon, O. et al. *J. Am. Chem. Soc.* 124 (2002): 13402–13404.
33. (a) Stavenger, R. A., and S. L. Schreiber. *Angew. Chem. Int. Ed.* 40 (2001): 3417–3421. (b) Koehler, A., and S. L. Schreiber. Manuscript in preparation.
34. Kubota, H. et al. *Chem. Biol.* 9 (2002): 265–276.
35. (a) Sternson, S. M. et al. *Org. Lett.* 3 (2001): 4239–4242. (b) Haggarty, S. J. et al. *Proc. Natl. Acad. Sci. USA,* in press.
36. Schmidt, D. et al. *J. Comb. Chem.,* in press.
37. (a) Spring, D. R. et al. *J. Am. Chem. Soc.* 122 (2000): 5656–5657. (b) Spring, D. R. et al. *J. Am. Chem. Soc.* 124 (2002): 1354–1363.
38. Burke, M. D., and G. Lalic. *Chem. Biol.* 9 (2002): 535–541.
39. (a) Blackwell, H. E. et al. *Chem. Biol.* 8 (2001): 1167–1182. (b) Clemons, P. A. et al. *Chem. Biol.* 8 (2001): 1183–1195. (c) Blackwell, H. E. et al. *Angew. Chem. Int. Ed.* 40 (2001): 3421–3425.
40. Stockwell, B. R. et al. *Chem. Biol.* 6 (1999): 71–83.
41. (a) Mayer, T. U. et al. *Science* 286 (1999): 971–974. (b) Peterson, R. T. et al. *Proc Natl. Acad. Sci. USA* 97 (2000): 12965–12969.
42. (a) MacBeath, G. et al. *J. Am. Chem. Soc.* 121 (1999): 7967–7968. (b) Hergenrother, P. J. et al. *J. Am. Chem. Soc.* 122 (2000): 7849–7850. (c) Barnes-Seeman, D. *Angew. Chem. Int. Ed.,* in press.
43. MacBeath, G., and S. L. Schreiber. *Science* 289 (2000): 1760–1762.
44. Frank-Kamenetsky, M. et al. *J. Biol.* 1 (2002): 1:10.
45. Chen, J. K. et al. *Proc. Natl. Acad. Sci. USA* 99 (2002): 14071–14076.
46. Mallet, V. O. et al. *Nat. Biotechnol.* 20 (2002): 1234–1239.
47. (a) Rosania, G. R. et al. *Nat. Biotechnol.* 18 (2000): 304. (b) For commentary, see: Frederickson, R. *Nat. Biotechnol.* 18 (2000): 250.
48. Wu, X. et al. *J. Am. Chem. Soc.* 124 (2002): 14520–14521.
49. Fantin, V. R. et al. *Cancer Cell* 2 (2002): 29–42.
50. (a) Jiang, X. et al. *Science* 299 (2003): 223–226. (b) Nicholson, D. W., and N. A. Thornberry. *Science* 299 (2003): 214–215.
51. Foster, B. A. et al. *Science* 286 (1999): 2507–2510.
52. Carlson, C. B. et al. *ChemBioChem* 3 (2002): 859–865.
53. (a) Peterson, J. R. et al. *Proc. Natl. Acad. Sci. USA* 98 (2001): 10624–10629. (b) Peterson, J. R. et al. Unpublished results and ref. 3(b).