

# Chemical genetics: exploring and controlling cellular processes with chemical probes

Craig M. Crews and Ute Splittgerber

The new field of chemical biology brings together chemists and biologists who are seeking to understand and mimic the natural world. One research strategy in this new field is the development of biologically active small molecules as molecular probes. This approach, which has been called 'chemical' genetics, has allowed elucidation of several pathways that have been difficult to study using traditional genetic approaches.

**TWENTY YEARS AGO**, the discovery of retrovirally encoded oncoproteins revolutionized the field of cell-cycle regulation. Their identification as important signal transduction components accelerated research in this area by providing the prerequisite starting points for further biochemical investigation. In recent years, biologically active small molecules have played a similar role in cell biology. This approach, which has been called chemical genetics<sup>1,2</sup> because of its similarity to traditional genetic approaches, uses cell-permeable compounds as molecular probes to perturb intracellular processes with exquisite precision. Moreover, advances in chemistry are now facilitating the development of novel molecules to supplement the array of natural products that have already proven to be useful. Here, we examine recent studies that have used cell-permeable small molecules to explore and control cellular events. We also highlight the potential of chemical genetics for studies of protein function now that entire genomes have been sequenced.

Mechanism-based approaches to the discovery of novel, selective, small molecules that disrupt specific cellular mechanisms have dominated pharmaceutical research for decades. Such approaches have produced compounds

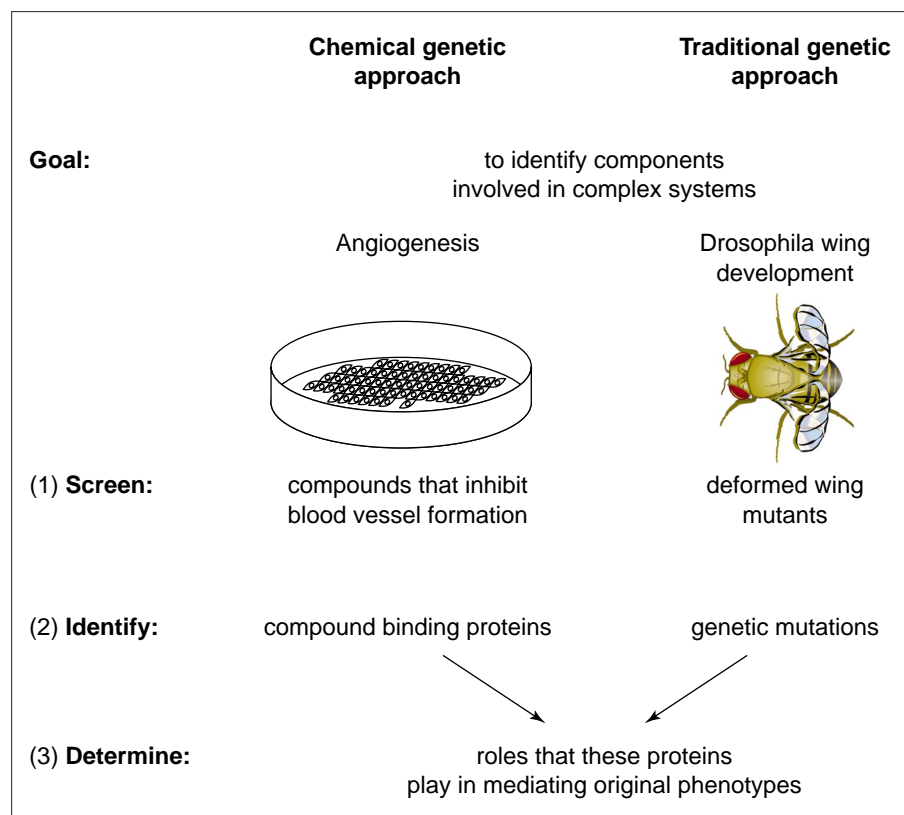
**C. M. Crews** is at the Dept of Molecular, Cellular and Developmental Biology and the Dept of Pharmacology, Yale University, New Haven, CT 06520-8103, USA; and **U. Splittgerber** is at the Dept of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520-8103, USA. Email: craig.crews@yale.edu

that affect the functions of known proteins (e.g. HMG-CoA reductase and angiotensin-converting enzyme). By contrast, a chemical genetic approach screens for compounds on the basis of their ability to perturb a cellular process or system; hence, their molecular targets are often unknown. By biochemical elucidation of the modes of action of these molecular

probes, chemical genetics aims to yield new information about complex cellular processes.

In several ways, chemical genetics resembles a traditional genetic approach. Just as the isolation of a key mutant in a genetic screen leads to the identification of the mutated gene(s), the elucidation of a small molecule's mode of action can result in the identification of interesting cellular targets. The first step for both traditional and chemical genetic approaches is to screen for a change in protein function induced by genetic mutation and pharmacological intervention, respectively (Fig. 1). Subsequent identification of the underlying genetic change or, in the case of the chemical genetic approach, identification of compound-binding protein(s) poses a new question: how does modification of critical gene products in a complex system cause the observed phenotype?

Just as traditional genetic screens can yield mutated proteins that have a variety of modified activities, small molecules can also modify the functions of target proteins in several ways. Natural products have been used to mimic null mutations (e.g. fumagillin), activating



**Figure 1**

Comparison of traditional and chemical genetic approaches. Both approaches begin with a desired phenotype that arises from either a genetic mutation or pharmacological intervention. Identification of the mutated gene product or natural product target protein leads to further investigation of the mechanisms by which perturbation of these proteins results in the initial phenotype.

mutations (e.g. phorbol esters) and gain-of-function mutations (e.g. FK506 and cyclosporin). Moreover, the addition of a biologically active small molecule to a system results in a temporary perturbation of the normal, wild-type state. Thus, these active compounds can be considered to be 'conditional alleles', insofar as they elicit a phenotype in a wild-type cell only as long as they are present. Identification of suitable biologically active compounds is one of the greatest potential obstacles in chemical genetics. Fortunately, through the efforts of natural-product chemists, many biologically active compounds with unknown molecular mechanisms have been reported. In addition, new screening methodologies and generation of chemical diversity are changing the way we identify molecular probes.

Chemical genetics complements traditional genetics insofar as it allows one to address complex systems that are not amenable to traditional genetic manipulations. For example, no good model system exists for the genetic analysis of tumor-induced neovascularization. However, the use of anti-angiogenic natural products, such as fumagillin, is beginning to provide new insights into this complex process<sup>3,4</sup>. In the past decade, this approach has also been immensely powerful in the elucidation of key immune-cell signal transduction mechanisms, using the immunosuppressive drugs FK506, rapamycin and cyclosporin<sup>5</sup>. Here, we describe three current examples of specific

protein modulators that are providing new insights into cell biology.

#### Leptomycin and nuclear transport

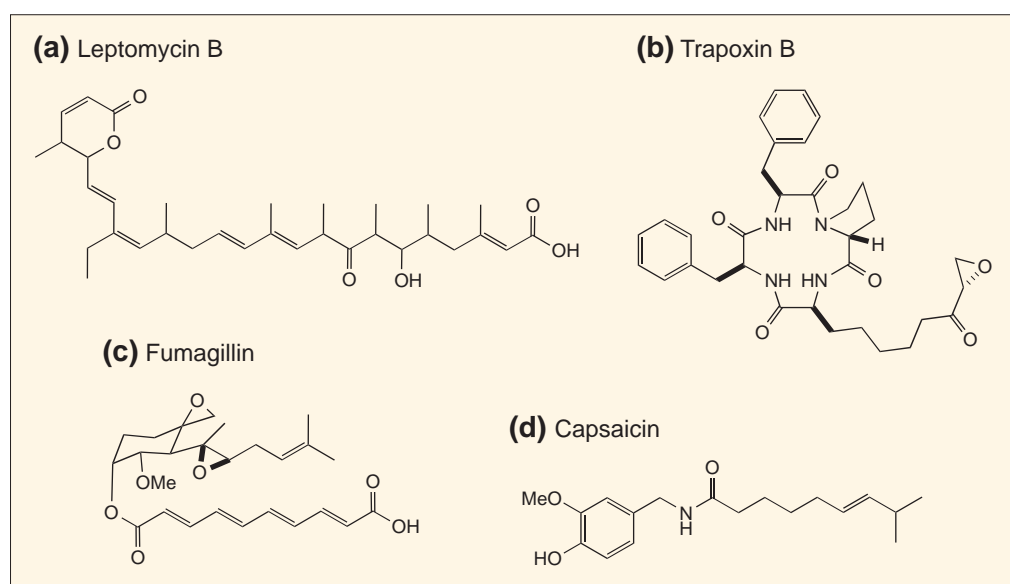
The investigation of nuclear transport of proteins has been widely studied in *Xenopus laevis* oocytes and mammalian cells, two systems in which genetic manipulation is not easy. Several researchers have used the potent cytostatic natural product, leptomycin B (LMB; Fig. 2), to overcome the inherent difficulties in the dissection of this complex process. Important insights into export mechanisms have come from studies of the HIV Rev protein, which possesses a leucine-rich nuclear-export signal (NES), and guides unspliced viral RNA out of the nucleus. Although the exact mechanism for this nuclear export was unclear, researchers expected that NES receptors existed, with the same function as the nuclear localization signal (NLS) receptor, importin  $\beta$ . A clearer picture has emerged since Wang and colleagues recently demonstrated that leptomycin B blocks nuclear export of Rev and Rev-mediated RNA export<sup>6</sup>. Building on this work, three groups<sup>7-9</sup> used LMB to conclude that the LMB-binding protein, CRM1, is a general receptor involved in the nuclear export of proteins that possess leucine-rich NES. Results from similar studies using a genetically mutated CRM1 *Schizosaccharomyces pombe* strain also led to the same conclusions<sup>10</sup>, thus validating a chemical genetic approach in genetically intractable systems.

#### Histone-deacetylase inhibitors and transcriptional regulation

Manipulation of intracellular protein function by small molecules has played a key role in recent studies linking chromatin structure and transcriptional activity. Although researchers have long believed that chromatin organization plays an important role in transcriptional regulation, traditional genetic proof of this connection has been elusive. Two natural products that inhibit histone deacetylase (HDAC1) – trichostatin A (TSA) and trapoxin B – have recently been instrumental in establishing a connection between histone deacetylation and retinoblastoma protein (RB)-mediated transcriptional repression. RB was known to inhibit cell proliferation by repressing a subset of genes controlled by the E2F family of transcription factors through binding to the *trans*-activation domain of E2F. However, because RB also interacts with proteins that are important for regulation of chromatin structure, Koonin *et al.* postulated that RB's ability to repress transcription could also be explained by interaction with chromatin-remodeling proteins<sup>11</sup>. This hypothesis was strengthened by the fact that both RbAp48 (an RB binding protein) and HDAC1 were retained on a trapoxin B affinity matrix in the initial HDAC1 purification and cloning experiment<sup>12</sup>. Thus, these results provided a physical link between histone deacetylation and transcriptional regulation. Recently, three groups<sup>13-15</sup> confirmed that RB forms a complex with E2F and HDAC1<sup>16</sup>. Through the generation of an HDAC1 conditional allele, TSA played an important role in demonstrating that HDAC activity is required for RB-mediated repression of transcription. Moreover, the groups used TSA to show that RB-mediated recruitment of HDAC represses only a subset of promoters and transcription factors, confirming that RB can repress transcription via different mechanisms.

#### Capsaicin and the nociceptive receptor

Although it was long known that sensory neurons, termed nociceptors, mediate pain signals to the brain, the specific nociceptive receptor and the mechanism underlying this signaling remained poorly



**Figure 2**

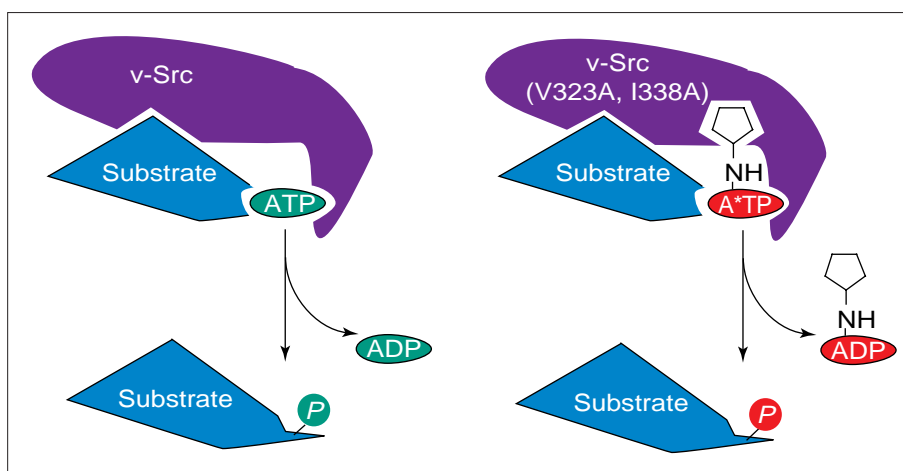
Natural products as molecular probes. (a) Leptomycin B. (b) Trapoxin B. (c) Fumagillin. (d) Capsaicin. The identification of the binding proteins of these natural products has provided an entry point into the complexities of nuclear export (leptomycin B)<sup>7-9</sup>, regulation of chromatin organization (trapoxin B)<sup>12</sup>, angiogenesis (fumagillin)<sup>3</sup> and nociceptive signal transduction (capsaicin)<sup>17</sup>.

understood. Recently, Julius and colleagues made a breakthrough using capsaicin, the active compound in chili peppers that is responsible for the painful heat sensation. Employing capsaicin as a molecular probe for pain/heat signal propagation, the authors identified the first nociceptive receptor<sup>17</sup>, vanilloid receptor 1 (VR1). VR1 is an integral membrane protein that shares homology with a family of putative calcium ion channels. It is strongly activated by capsaicin and when neurons are subjected to temperatures above a 'pain-threshold'. Thus, VR1 probably plays an important physiological role in the detection of pain stimuli. The discovery of this new noxious-temperature-sensitive cation channel represents a substantial step towards the understanding of the mechanisms involved in the heat-stimulated pain pathway. The use of capsaicin as a probe of this pathway should continue to yield new information about the mechanisms of VR1 activation.

#### Non-natural small-molecule probe development

Although natural products play an important role in chemical genetics, recent research has also focused on the development of chemically synthesized small-molecule probes. Given the enormous size of the protein kinase superfamily and the many densely interconnected regulatory kinase cascades, researchers have devoted much effort to the development of small molecules capable of modulating the activity of specific kinases. Schultz and colleagues<sup>18</sup> recently identified several potent inhibitors of the cyclin-dependent kinase CDK2 from a 2,6,9-trisubstituted purine library. These compounds strongly inhibit the activity of the CDK2-cyclin-A complex and the yeast counterpart CDK28p-cyclin-A by competing for ATP. Additionally, these inhibitors have proved to be useful in the investigation of downstream consequences of kinase activation. For example, use of high-density oligonucleotide-hybridization arrays has shown that these inhibitors affect the transcript levels of many yeast genes, including several involved in cell-cycle control.

Shokat and colleagues<sup>19</sup> recently reported a more targeted approach towards elucidation of the immediate downstream effects of a single kinase-family member within the context of an intracellular kinase network (Fig. 3). Their strategy employs both small-molecule probe development and protein



**Figure 3**

An approach for the identification of protein kinase substrates. Shokat and colleagues<sup>19</sup> mutated the tyrosine kinase v-Src (V323A, I338A) (right) so as to allow binding of a sterically bulky ATP analog (A\*TP; red) that cannot be used by the wild-type Src kinase (left) or other kinases. Thus, addition of radiolabeled modified ATP to cell lysates containing modified Src (V323A, I338A) allows for the selective <sup>32</sup>P-labeling of Src substrates. This 'bump:hole' strategy of using ATP analogs and modified kinases could become a general approach to study downstream targets of kinases.

engineering. The authors engineered the ATP-binding domain of the tyrosine kinase v-Src to accept a modified ATP analog (A\*TP) that could not be used by wild-type Src or other kinases. Despite the reconfiguration of the kinase active site, the substrate specificity of the modified kinase was not detectably altered from that of wild-type v-Src. Furthermore, the engineered v-src preferentially uses the ATP analog even in a cellular context of high (1–5 mM) ATP concentration. This approach could allow phosphorylation of specific substrates and thus facilitate identification of downstream targets of Src. Given the highly conserved nature of the kinase active site, this strategy is also likely to be versatile in identifying specific substrates of related kinases.

#### The future of chemical genetics: genomic considerations

Although current chemical genetic strategies have elucidated several novel cellular mechanisms, the discovery of natural products that are specific for any given protein is serendipitous. Thus, natural products are unlikely to meet future demand for the palette of cellular probes needed to address protein function. Fortunately, advances in chemical synthesis strategies have created many new opportunities for the generation of new reagents. The development of combinatorial chemistry in the past decade allows the easy generation of millions of unique compounds<sup>20</sup>. As in the case of natural products, such

libraries can be screened for a variety of new biological activities by the use of several assays. For example, an 'on-bead' screening strategy can identify novel peptide ligands for fluorescently labeled SH3 domains from >10<sup>6</sup> beads, each of which bears a unique short peptide sequence<sup>21</sup>. Cell-based screening of libraries are becoming routine<sup>22</sup>. Lerner and colleagues have identified new antagonists of the G-protein-coupled  $\alpha$  melanocyte-stimulating hormone ( $\alpha$ -MSH) receptor by screening a tripeptide combinatorial library, using *Xenopus* melanocytes<sup>23</sup>. Rational drug design requires detailed structural information about a protein target; by contrast, the power of a combinatorial approach lies in the presentation of a very wide variety of possible ligands to the protein target and subsequent detection of the optimal ligand that the target has 'selected'. This approach is analogous to a genetic screen, in which selection pressure is applied to a pool of randomly mutagenized proteins to identify the few mutations that have the desired effect on the function of a protein.

Although promising, these screening strategies have inherent limitations. For example, in the case of on-bead assays, the protein-binding abilities of ligands in solution and support-bound ligands can differ<sup>24,25</sup>. In addition, current cell-based screening methodologies are designed to develop novel pharmaceuticals. Therefore, ligands to a single targeted protein are identified. For combinatorial chemistry to have a greater impact on

the exploration of novel pathways in cell biology, a less focused approach will be necessary, in which compounds are identified on the basis of their ability to perturb cellular systems. Fortunately, Schreiber and colleagues<sup>26</sup> have recently taken a step in this direction: they used miniaturized mammalian cell-based assays to screen for compounds that disrupt post-translational modifications.

Combinatorial chemistry holds great potential for the development of unique chemical genetic probes; however, potential bottlenecks and stumbling blocks remain. Despite the ever-increasing possibilities of generating synthetic chemical diversity, man-made libraries are unlikely to reproduce nature's chemical diversity in the short term. For example, syntheses of intricate natural products (e.g. Taxol<sup>®</sup>, FK506 and bryostatin 1) at present require concerted efforts over several years. This complexity makes many of the more complex structures found in nature beyond the realm of current combinatorial chemical technology. Thus, at least in the foreseeable future, natural product screening will continue to play an important role in the identification of unique chemical architecture with biological activity (Fig. 4).

Issues of cell-membrane permeability, specificity and potency remain challenges in the development of new

probes. Currently, the limiting factors are not the synthesis and selection of compounds to be screened, but how to screen the large numbers of compounds generated by combinatorial chemistry. The biggest challenge in this context will be the development of efficient high-throughput screening strategies that will allow simultaneous screening of many compounds against multiple targets. Innovative collaborative efforts, such as those of Harvard's Institute of Chemistry and Cell Biology and Novartis' Institute for Functional Genomics, suggest that both industry and academia will play important roles in this next phase of chemical genetics.

### Conclusions

Our rapidly growing knowledge of the molecular details of cellular mechanisms encourages the use of more 'invasive' strategies for the manipulation of cellular processes. A manifestation of this direction in cell biology has been the advent of chemical genetics, the goals of which are the development of cell-permeable ligands that affect protein function and their use in understanding biological systems. In this regard, biologically active natural products have proven to be invaluable in the exploration of systems that cannot be studied by traditional genetic manipulations. Furthermore, in recent years, identification of target

proteins has also provided new opportunities for the development of creative approaches for controlling intracellular processes (see, for example, Ref. 2).

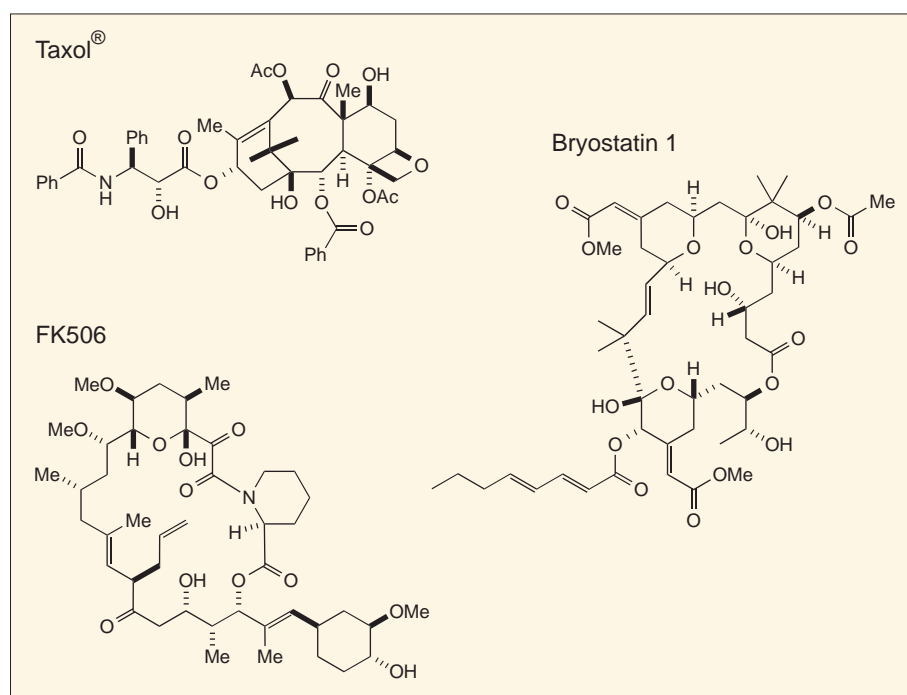
The future direction of chemical genetics will undoubtedly be greatly influenced by the results of current genome-sequencing projects. With the fruits of these projects in hand, new genome-wide analytical approaches for the assignment and investigation of protein function will be needed. Fortunately, the advent of new approaches to chemical-diversity generation, coupled with novel bioassays, has the potential to provide sets of reagents for such an endeavor.

### Acknowledgements

We thank the NIH, the Burroughs Wellcome Fund, the Donaghy Biomedical Foundation and the Fonds der Chemischen Industrie for support.

### References

- 1 Mitchison, T. J. (1994) *Chem. Biol.* 1, 3–6
- 2 Schreiber, S. L. (1998) *Bioorg. Med. Chem.* 6, 1127–1152
- 3 Sin, N. *et al.* (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 6099–6103
- 4 Liu, S. *et al.* (1998) *Science* 282, 1324–1327
- 5 Schreiber, S. L. and Crabtree, G. R. (1995) *Harvey Lect.* 91, 99–114
- 6 Wolff, B., Sanglier, J. J. and Wang, Y. (1997) *Chem. Biol.* 4, 139–147
- 7 Fornerod, M., Ohno, M., Yoshida, M. and Mattaj, I. W. (1997) *Cell* 90, 1051–1060
- 8 Fukuda, M. *et al.* (1997) *Nature* 390, 308–311
- 9 Ossareh-Nazari, B., Bachelier, F. and Dargemont, C. (1997) *Science* 278, 141–144
- 10 Stade, K., Ford, C. S., Guthrie, C. and Weis, K. (1997) *Cell* 90, 1041–1050
- 11 Koonin, E. V., Zhou, S. and Lucchesi, J. C. (1995) *Nucleic Acids Res.* 23, 4229–4233
- 12 Taunton, J., Hassig, C. A. and Schreiber, S. L. (1996) *Science* 272, 408–411
- 13 Brehm, A. *et al.* (1998) *Nature* 391, 597–601
- 14 Magnaghi-Jaulin, L. *et al.* (1998) *Nature* 391, 601–605
- 15 Luo, R. X., Postigo, A. A. and Dean, D. C. (1998) *Cell* 92, 463–473
- 16 Brehm, A. and Kouzarides, T. (1999) *Trends Biochem. Sci.* 24, 142–145
- 17 Caterina, M. J. *et al.* (1997) *Nature* 389, 816–824
- 18 Gray, N. S. *et al.* (1998) *Science* 281, 533–538
- 19 Shah, K., Liu, Y., Deirmengian, C. and Shokat, K. M. (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 3565–3570
- 20 Czarnik, A. W. and Keene, J. D. (1998) *Curr. Biol.* 8, R705–R707
- 21 Chen, J. K. *et al.* (1993) *J. Am. Chem. Soc.* 115, 12591–12592
- 22 Lam, K. S., Lebl, M. and Krchnák, V. (1997) *Chem. Rev.* 97, 411–448
- 23 Quillan, J. M., Jayawickreme, C. K. and Lerner, M. R. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 2894–2898
- 24 Morken, J. P. *et al.* (1998) *J. Am. Chem. Soc.* 120, 30–36
- 25 Liang, R. *et al.* (1996) *Science* 274, 1520–1522
- 26 Stockwell, B. R., Haggarty, S. J. and Schreiber, S. L. (1999) *Chem. Biol.* 6, 71–83



**Figure 4**

The complexity of natural chemical diversity. The challenge of synthetic chemical diversity is to match the chemical complexity found in nature. Questions remain as to whether structures as complex as these three potent biologically active natural products could be synthesized by current combinatorial chemical technology.