Thinking big with small molecules

60 years of **JCB**

Karen Mruk[1,2] and James K. Chen[1,2]

[1]Department of Chemical and Systems Biology and [2]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305

Synthetic chemistry has enabled scientists to explore the frontiers of cell biology, limited only by the laws of chemical bonding and reactivity. As we investigate biological questions of increasing complexity, new chemical technologies can provide systems-level views of cellular function. Here we discuss some of the molecular probes that illustrate this shift from a "one compound, one gene" paradigm to a more integrated approach to cell biology.

Since the alkaloid colchicine was described as a mitotic poison in 1934 (Lits, 1934), chemical probes have played an integral role in cell biological research. Laboratories are commonly stocked with the transcriptional inhibitor actinomycin D, the translational blocker cycloheximide, and other "workhorse" small-molecule modulators that target key cellular processes. More specialized aspects of cell function are also routinely studied with pharmacological agents, ranging from the adenylate cyclase activator forskolin to the mTOR inhibitor rapamycin. The rapid and often reversible actions of these chemical tools make them valuable complements to genetic technologies, and over the past several decades, chemistry-driven cell biology has evolved into what is now commonly referred to as chemical genetics.

Chemical genetics shares several parallels with its nucleic acid–based namesake. Forward chemical genetic screens survey the proteome with diverse compound libraries, just as mutagenesis screens strive to stochastically generate genomic alterations. Reverse chemical genetics interrogates protein function through specific, small-molecule modulators, akin to the targeted disruption of individual genes by CRISPR/Cas9 and RNA interference reagents. Chemical genetic screens can also identify pharmacological suppressors or enhancers of existing phenotypes. Moreover, the single-nucleotide resolution of genetic technologies has inspired efforts to match that precision with chemical probes. Vast collections of natural products and structurally diverse synthetic reagents have been assembled with the hope of identifying one or more specific modulators for each gene product. This reductionist "one compound, one gene" focus has resonated with researchers interested in chemical mechanisms. Indeed, the beauty of deconstructing a cellular process

into individual molecular interactions or reactions has attracted many chemists to the biological sciences.

However, the sequencing of entire genomes has shown that even a comprehensive list of individual parts and annotated functions yields an incomplete picture of cell biology. A more holistic view requires an understanding of how systems with shared structural elements, chemical reactivity, and/or spatiotemporal dynamics contribute to emergent properties. Several genetic and biochemical technologies have been developed to capture this "big picture." RNA sequencing can provide a global perspective of the transcriptomes associated with specific cell states, and ribosomal profiling can reveal which of these transcripts are actively translated at a particular moment. Various regimens of biochemical cross-linking, immunoprecipitation, and sequencing have been used to obtain genome-wide views of protein–nucleic acid interactions and chromatin structure. Central to these methods is the facility with which DNA and RNA can be modified, amplified, and analytically characterized—technological advances that have been largely limited to nucleic acids. As a result, we are comparatively in the dark about how proteins and other biomolecules function at the systems level. Illuminating their collective activities in cells will require techniques that can target them in a structure- or reaction-dependent manner, and innovation at the chemistry–biology interface can help us meet this challenge.

Several recent advances illustrate the power of thinking big with small molecules. One of the breakthroughs in this approach is activity-based protein profiling, in which substrate-like probes are used to covalently tag specific enzyme classes for visualization or purification (Fig. 1 A; Cravatt et al., 2008). By treating live cells or cell lysates with these reagents, one can gain global insights into enzyme families that recognize a given substrate. For example, biotinylated long-chain fluorophosphonates have been used to obtain serine hydrolyase activity signatures for specific tissue types, exploiting the general reactivity of these suicide substrates toward this enzyme class (Liu et al., 1999). Through this approach, expression of the membrane-associated serine hydrolyase KIAA1363 was found to strongly correlate with cancer cell invasiveness (Jessani et al., 2002). Other nucleophile-containing enzyme families can be targeted by electrophilic reagents, including cysteine proteases,

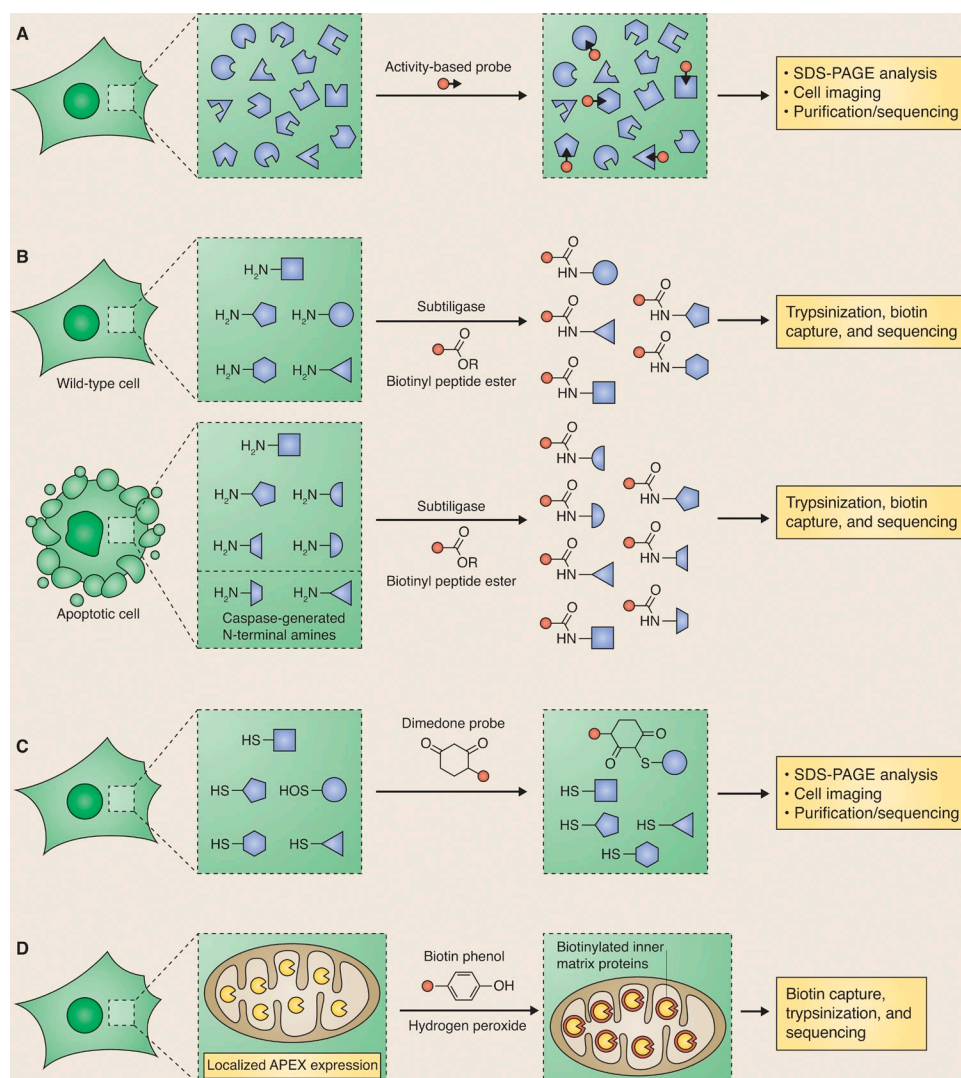Correspondence to James K. Chen: jameschen@stanford.edu.

Figure 1. **Representative chemical technologies for systems-level analyses of cell biology.** (A) Activity-based profiling of specific enzyme classes using biotinylated activity-based probes. (B) Subtiligase-based profiling of apoptosis-dependent proteolytic targets. (C) Dimedone-based profiling of protein oxidation states. (D) APEX peroxidase-mediated proteomic mapping of the mitochondrial inner matrix.

deubiquitinases, kinases, and glycosidases. Photoreactive groups can extend activity-based protein profiling to an even broader spectrum of targets. Many of these probes can be applied in live cells, taking advantage of azide/alkyne cycloaddition, azide/phosphine ligation, or other bio-orthogonal tagging chemistries that do not cross-react with endogenous molecules (Patterson et al., 2014). Chemical methods have also been devised to tackle the converse systems-level question: what is the ensemble of substrates for a given enzymatic activity? Using the engineered enzyme subtiligase to biotinylate free protein N termini, 333 caspase-like cleavage sites within 292 protein substrates were identified in apoptotic cells (Fig. 1 B; Mahrus et al., 2008). A cyclin-dependent kinase 1 (Cdk1/cyclin B) mutant and complementary ATP-γ-S analogue were also used to thiophosphate-label substrates for subsequent covalent capture, revealing >70 direct Cdk1 targets (Blethrow et al., 2008).

A variety of other posttranslational alterations can now be comprehensively surveyed in cells by deploying chemical probes with unique reactivities. Protein oxidation states can be

assessed using azide- or alkyne-based dimedone analogues, which selectively react with the redox-sensitive intermediate cysteine sulfenic acid (Fig. 1 C; Paulsen and Carroll, 2013). Protein prenylation and fatty acylation can be chemically monitored through the metabolic incorporation of azide- or alkyne-functionalized lipids into proteins (Hannoush and Sun, 2010). Metabolic labeling with peracetylated azido-*N*-acetylglucosamine (GlcNAc) also enables the profiling of O-GlcNAc–modified proteins (Prescher and Bertozzi, 2006). By coupling these methods with bio-orthogonal tagging and mass spectrometry–based sequencing, whole collections of posttranslationally modified proteins have been characterized. Importantly, these techniques have also identified several proteins previously unknown to bear these functional groups. How these populations vary between cell types or change in response to specific perturbations can then be readily assessed.

Chemical approaches can even enhance our "big picture" view of DNA and RNA regulation by targeting nucleic acid modifications that are challenging to discern through

genetic techniques alone. For instance, the bacteriophage enzyme β-glucosyltransferase has been used selectively couple glucose to 5-hydroxymethylcytosine, a common DNA oxidation product that may have functional roles in development and disease. By comparing TET protein-assisted bisulfite sequencing of the glucose-modified DNA to traditional bisulfite sequencing results, 5-hydroxymethylcytosine sites associated with specific cellular states can be determined with genome-wide, single-base resolution (Yu et al., 2012). Our ability to comprehensively characterize RNA–protein interactions through UV cross-linking, immunoprecipitation, and sequencing has similarly benefited from synthetic reagents. In this case, metabolic labeling of cellular RNAs with 4-thiouridine has been found to dramatically increase cross-linking efficiency and RNA recovery, as well as generate distinguishing thymine-to-cytosine transitions at the site of 4-thiouridine–protein coupling (Hafner et al., 2010).

Finally, it is worth noting a recent chemical strategy that targets biomolecules according to their subcellular localization rather than their inherent chemical properties. The segregation of cellular components into functional ensembles has typically been studied through biochemical fractionation or cell imaging, and the two techniques have complementary strengths and limitations. In comparison, this new proteomic mapping method combines the best of both worlds: proteome-wide detection and spatiotemporal resolution in live cells. By targeting an engineered form of ascorbate peroxidase (APEX) to the mitochondrial matrix and then pulse-treating the cells with biotin-phenol and hydrogen peroxide, proteins within that compartment have been selectively biotinylated through phenoxyl radical coupling (Fig. 1 D; Rhee et al., 2013). This approach identified 495 components of the mitochondrial matrix proteome, including 31 factors not previously associated with mitochondria. Because APEX is active in all subcellular domains, this chemical mapping procedure should be generally applicable to other organelles of interest.

The advances cited here are far from a comprehensive list, but they illustrate the exciting new capabilities and insights that can be obtained when chemical approaches are applied to questions of cell biology. Moreover, they demonstrate how chemistry can go beyond the targeted perturbation of individual gene products and capture the complexity that underlies cellular behavior. Future challenges include the development of new molecular probes to more broadly cover the chemical space within cells. These include technologies that facilitate the profiling of chromatin marks, three-dimensional genomic architectures, nonenzymatic protein families, protein–glycan interactions, or cellular metabolites, to name a few. Methods that are amenable to live-cell analyses or even in vivo applications will be particularly valuable.

Achieving these goals will require the collaborative efforts of chemists and biologists. Fortunately, this partnership has been fostered by recent developments, including: (1) new PhD programs at the chemistry–biology interface, particularly those that provide research opportunities spanning synthetic chemistry, cell biology, and animal models; (2) institutional compound screening and synthesis services that make chemistry

more accessible to biologists; (3) complementary genomic resources that make complex biological systems more accessible to chemists; and (4) scientific journals and conferences that support the chemical biology community as a whole. Building upon this logistical framework, it will be important to navigate cultural differences between the two communities. In many respects, chemists and biologists speak different "languages" that stem from distinct scientific traditions and training practices. "Bilingual" scientists who have had substantive experiences in both disciplines will play important roles in bridging the two. By combining an in-depth knowledge of chemistry and a sophisticated understanding of cell biology, we can realize a vision that neither perspective can achieve alone.

References

Blethrow, J.D., J.S. Glavy, D.O. Morgan, and K.M. Shokat. 2008. Covalent capture of kinase-specific phosphopeptides reveals Cdk1-cyclin B substrates. *Proc. Natl. Acad. Sci. USA*. 105:1442–1447. http://dx.doi.org/10.1073/pnas.0708966105

Cravatt, B.F., A.T. Wright, and J.W. Kozarich. 2008. Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. *Annu. Rev. Biochem.* 77:383–414. http://dx.doi.org/10.1146/annurev.biochem.75.101304.124125

Hafner, M., M. Landthaler, L. Burger, M. Khorshid, J. Hausser, P. Berninger, A. Rothballer, M. Ascano Jr., A.C. Jungkamp, M. Munschauer, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. 141:129–141. http://dx.doi.org/10.1016/j.cell.2010.03.009

Hannoush, R.N., and J. Sun. 2010. The chemical toolbox for monitoring protein fatty acylation and prenylation. *Nat. Chem. Biol.* 6:498–506. http://dx.doi.org/10.1038/nchembio.388

Jessani, N., Y. Liu, M. Humphrey, and B.F. Cravatt. 2002. Enzyme activity profiles of the secreted and membrane proteome that depict cancer cell invasiveness. *Proc. Natl. Acad. Sci. USA*. 99:10335–10340. http://dx.doi.org/10.1073/pnas.162187599

Lits, F.J. 1934. Contribution à l'étude des réactions cellulaires provoquées par la colchicine. *Comptes Rendus des Séances de la Société de Biologie Paris.* 115:1421–1423.

Liu, Y., M.P. Patricelli, and B.F. Cravatt. 1999. Activity-based protein profiling: the serine hydrolases. *Proc. Natl. Acad. Sci. USA*. 96:14694–14699. http://dx.doi.org/10.1073/pnas.96.26.14694

Mahrus, S., J.C. Trinidad, D.T. Barkan, A. Sali, A.L. Burlingame, and J.A. Wells. 2008. Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell*. 134:866–876. http://dx.doi.org/10.1016/j.cell.2008.08.012

Patterson, D.M., L.A. Nazarova, and J.A. Prescher. 2014. Finding the right (bioorthogonal) chemistry. *ACS Chem. Biol.* 9:592–605. http://dx.doi.org/10.1021/cb400828a

Paulsen, C.E., and K.S. Carroll. 2013. Cysteine-mediated redox signaling: chemistry, biology, and tools for discovery. *Chem. Rev.* 113:4633–4679. http://dx.doi.org/10.1021/cr300163e

Prescher, J.A., and C.R. Bertozzi. 2006. Chemical technologies for probing glycans. *Cell*. 126:851–854. http://dx.doi.org/10.1016/j.cell.2006.08.017

Rhee, H.W., P. Zou, N.D. Udeshi, J.D. Martell, V.K. Mootha, S.A. Carr, and A.Y. Ting. 2013. Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science*. 339:1328–1331. http://dx.doi.org/10.1126/science.1230593

Yu, M., G.C. Hon, K.E. Szulwach, C.X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, et al. 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*. 149:1368–1380. http://dx.doi.org/10.1016/j.cell.2012.04.027