

# Challenges for the ‘chemical-systems’ biologist

Gabriel M Simon & Benjamin F Cravatt

**As the field of chemical biology matures, its practitioners are tackling ever more sophisticated biological problems. Chemical approaches, both synthetic and analytical, provide researchers with powerful new technologies to perturb, dissect and even reconstruct complex biological systems. Here we discuss the special challenges and opportunities confronted at the burgeoning interface of chemical and systems biology.**

What’s in a name? Depending on one’s perspective, perhaps everything or nothing. By naming or categorizing fields of science, we achieve certain desirable objectives (for example, uniting groups of scientists with common research interests), but these benefits come at the cost of creating potentially artificial chasms between related disciplines. From journal to departmental titles, simple categorizations can result in groups of scientists with common interests reading distinct sets of literature or being divided into separate buildings, substantially raising the activation barrier for conversation and collaboration. One might argue that specialization and scientific division is an inevitable product of our increasing sophistication as researchers. However, the past decade has witnessed a growing movement to ‘de-differentiate’ research in the life sciences, engendering “hybrid” disciplines that reflect a desire from within the scientific community for a resurgence of integrative approaches. Chief among these emerging (or reemerging) cross-disciplinary endeavors are chemical biology and systems biology.

Are there features that distinguish chemical biology from the more classical fields of biochemistry and pharmacology? These disciplines certainly share common scientific goals—namely to understand and exert chemical control over the function of biological

macromolecules. The ‘new’ distinguishing characteristic of chemical biology might instead be viewed as one of methodology. A conceptually similar distinction can be illustrated by reflecting on the historical emergence of the field of molecular biology. Molecular biology is defined by a set of technologies such as gene amplification, subcloning and sequencing that, more than any particular research objective, distinguishes the molecular biologist from other types of scientist. Similarly, by developing and implementing new analytical and synthetic methods, chemical biologists have dramatically expanded the scale and scope of biological problems that can be tackled with chemical techniques.

Similar questions could be asked of systems biology. How does this field, for instance, differ from physiology? We would argue here, too, that the distinction is less about scientific objectives, where both fields aim to generate a holistic understanding of how molecules and pathways interact to form complex life processes, than the approaches that are taken in pursuit of these goals. Systems biology is a discipline that has emerged from the concordant development of large-scale profiling technologies and sophisticated computational tools to analyze huge amounts of data. The discovery of the genetic code was singularly transformative because it made clear that biology is essentially an information science<sup>1</sup>. It is no coincidence that the development of powerful computers and efficient software went hand-in-hand with cloning and sequencing technologies, culminating in the unraveling of the human genome in 2001. Now, less than a decade later, genomes of new organisms are

completed weekly, providing an ever-growing database on which many large-scale experimental approaches, including DNA microarrays and mass spectrometry-based proteomics, have been founded. It is becoming increasingly clear that organizing and interpreting the enormous volumes of data generated by genome-scale science is a challenge even greater than the sequencing itself.

Viewing chemical and systems biology as methodological rather than conceptual departures from more traditional disciplines underscores the technical prowess and versatility of these emerging fields and their potential for cross-fertilization. It is thus perhaps not surprising that new departments have formed with the expressed goal of fostering interactions between chemical and systems biology (for example, Stanford University’s Department of Chemical and Systems Biology and The Scripps Research Institute’s Department of Chemical Physiology). As cohorts of scientists are built with a commitment to working at this interface, a key question surfaces: what important biological problems are these bands of interdisciplinary researchers most uniquely suited to solve? Here, we offer possible answers to this question by highlighting several research areas where advances in our understanding of ‘systems-level’ biological problems have been achieved using chemical approaches.

## Chemical approaches for the global analysis of PTMs

Genome sequencing projects have revealed that mammals contain surprisingly few genes (~20,000), on par with the number found in the comparatively simpler round worm.

*Gabriel M. Simon & Benjamin F. Cravatt are at The Skaggs Institute for Chemical Biology and the Department of Chemical Physiology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. e-mail: cravatt@scripps.edu*

Concomitant with this discovery was the realization that gene products (proteins) are found in a dizzying variety of modified forms *in vivo*, leading some to speculate that the human proteome may be comprised of in excess of 1,000,000 distinct proteins. Post-translational modifications (PTMs) alter the chemical state of proteins in often subtle ways that are not easily detected by standard gene or protein profiling techniques. Furthermore, PTMs are usually dynamic, with the enzymes that add and remove them having many protein substrates in cells and tissues. This combination of factors highlights the remarkable chemical complexity of mammalian proteomes and suggests that a comprehensive understanding of any PTM will require the ability to perform dynamic, systems-level analyses of cells or organisms. Here, proteomic researchers have enjoyed much success by creating large-scale profiling technologies founded on the principles of chemistry.

A central challenge for the global analysis of PTMs is the substoichiometric level at which the modifications are often found, which is further compounded by their chemical and enzymatic lability. Key to identifying low-abundance, modified forms of proteins from within the sea of unmodified proteins has been their covalent capture and enrichment. Chemical biologists have pioneered methods to achieve this goal for multiple PTMs, including phosphorylation and glycosylation. By engineering kinases and glycosyltransferases that are capable of transferring chemical affinity tags onto phosphorylated and glycosylated proteins, respectively, the Shokat<sup>2</sup> and Hsieh-Wilson<sup>3</sup> groups have provided general strategies to label, enrich and profile the phospho- and O-GlcNAc proteomes with unprecedented breadth and depth (Fig. 1). In both cases, advanced analytical methods

that exploit the awesome power of mass spectrometry (MS) were used to fully inventory the captured proteins. These approaches further offer a way to assign endogenous substrates to individual members of huge enzyme classes, such as the kinases—a problem that has been challenging to address with genetic approaches owing to overlapping and/or compensatory enzyme activities.

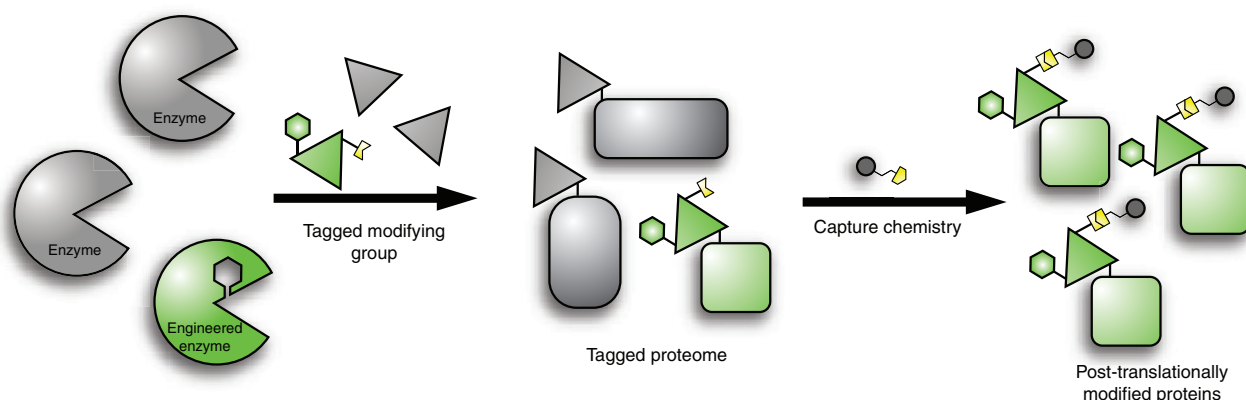
Despite these successes, we still lack general profiling methods for many PTMs, not to mention techniques to discover new PTMs and connect these events to their cognate enzymes. Will the integrated application of synthetic chemistry, protein engineering and advanced MS-based analytical methods prove to be a general strategy for mapping the diversity of PTM systems operational in eukaryotic cells? We suspect that this will be the case, although each PTM offers its own unique challenges. Take for instance lysine acetylation and lysine/arginine methylation. The ability to identify these modifications by MS is not prohibitively difficult; what is lacking is a selective capture or enrichment strategy for these modifications. It has proven difficult to develop selective antibodies against acetyl- or methyllysines. Could researchers engineer acetyltransferases or methyltransferases that accept modified acetyl-CoA or AdoMet groups to permit selective capture and identification of the entire substrate repertoire of individual transferases? Indeed, certain methyltransferases have been shown to accept alkyne derivatives of AdoMet<sup>4</sup>, which opens up the exciting possibility of transferring alkyne-tagged groups to lysines or arginines for capture via click chemistry<sup>5</sup> as a strategy for substrate identification. When framed in this context, it becomes clear that the interplay among bio-orthogonal chemical capture technologies, protein engineering and high-capacity

analytical chemistry platforms is a major driving force for much of our modern understanding of PTMs and their role in biology.

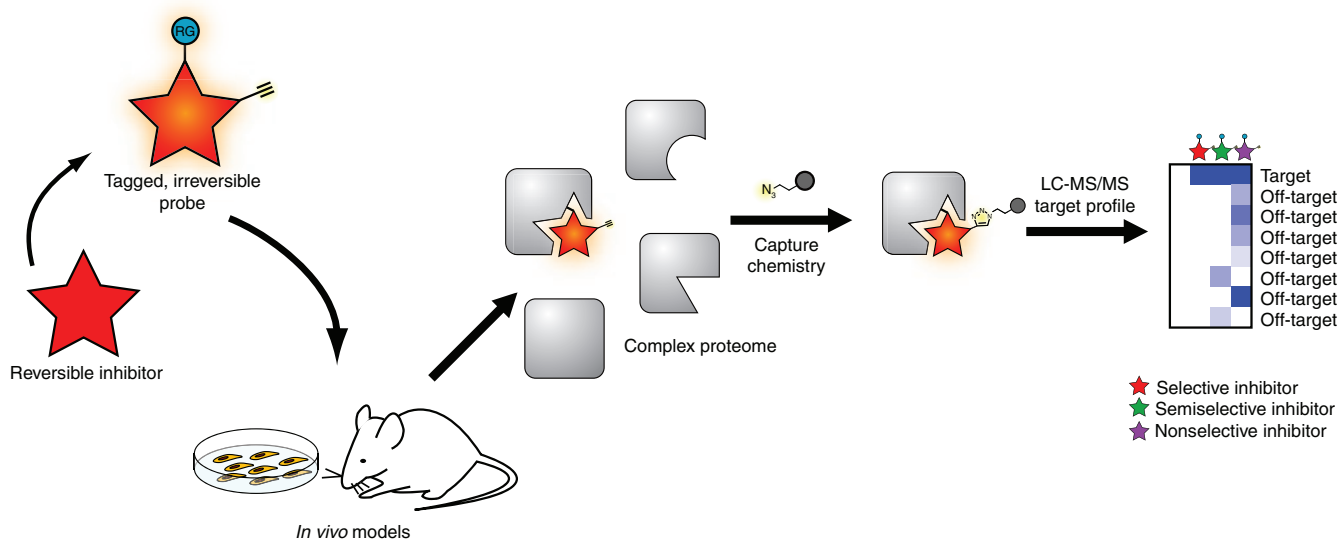
### Chemical approaches for the global analysis of protein activity

Chemical probes that contain a reactive group that covalently modifies the active sites of enzymes have engendered the field of activity-based protein profiling (ABPP), in which differences in protein activity, rather than abundance, can be measured<sup>6</sup>. The interplay of ABPP with systems biology was founded on the premise that probes targeting a broad swath of enzymes could be created by exploiting common mechanistic and/or other active site features. The ability to profile many enzyme activities in parallel, first by gel<sup>6,7</sup> and later by MS<sup>6,8</sup>, has provided a global view of the functional state of the proteome in a variety of physiological and pathological settings.

Early studies with activity-based probes taught us that valuable biological information can be garnered with promiscuous chemical reagents that target many enzymes from a given class. These probes were founded on a rich history of knowledge on affinity labels for specific enzyme classes. This concept has been expanded by our lab with the development of general electrophile probes that label enzymes from several mechanistically distinct classes<sup>9</sup>. These electrophilic probes can be thought of as minimalist 'reactivity probes' because their labeling is contingent largely on the presence of activated nucleophiles in the proteome, rather than a specific binding pocket or catalytic mechanism. Initial reports on these reactivity probes, which bear mild carbon electrophiles such as Michael acceptors, chloroacetamides and sulfonate esters, have focused on obtaining inventories of their protein targets and



**Figure 1** Chemical strategies to profile post-translationally modified proteins with engineered enzymes. An enzyme is engineered to accept and transfer a tagged group (for example, a phosphate or sugar) to its natural substrate within a complex proteome. Subsequently, the tagged protein targets can be covalently captured and enriched for identification. This approach has been used by the Shokat<sup>2</sup> and Hsieh-Wilson<sup>3</sup> groups to profile the phosphorylated and O-GlcNAc-modified proteomes, respectively.



**Figure 2** Chemoproteomic strategy to characterize the selectivity of bioactive small molecules *in vivo*. Covalent inhibitors<sup>25</sup> (or reversibly binding small molecules<sup>12</sup> that have been derivatized with a reactive group, RG) are converted into activity-based probes via addition of a bio-orthogonal 'handle'. These probes are added to living systems (cells or animals) and given time to react with protein targets. Probe-labeled proteins are then captured and identified from proteomes using bio-orthogonal chemistry and LC-MS-based proteomic methods, respectively. Using this approach, selective small molecules (red) can be distinguished from compounds with variable degrees of off-target reactivity (green, purple).

specific sites of modification in proteomes, but the potential for greater systems-level application is clear. A logical extension of this work would be to attach such tempered carbon electrophiles, or other tunable reactive elements<sup>10,11</sup>, to reversibly binding molecules to impart upon them the ability to covalently modify their protein targets (Fig. 2). Successful examples of such a strategy have already begun to emerge in the literature, resulting for instance in the creation of irreversible inhibitors of Rsk kinases<sup>12</sup>.

Projecting forward, incorporation of reactive elements into bioactive compounds of ill-defined mechanism could facilitate characterization of their protein targets in proteomes or even living systems<sup>13,14</sup>. These studies might in turn reveal that well-designed covalent inhibitors display a surprisingly high level of selectivity in the proteome<sup>12,15</sup>. In this regard, it is interesting to note that the pharmaceutical industry has historically resisted the idea of purposefully designing irreversible enzyme inhibitors as drugs, owing to worries that they might covalently modify other proteins to catastrophic effect<sup>16</sup>. We would posit, however, that modern chemical and systems biology has provided all of the methods required to experimentally address this concern in relevant model systems. Covalent inhibitors can be readily modified with clickable tags that are so sterically small as to make the resulting activity probes almost indistinguishable from the original agent. These probes can then be applied to any living system (cells or whole ani-

mals) across a broad dose range to fully inventory their on- and off-targets (Fig. 2). Indeed, one might even conclude that it is now more straightforward to define the proteome-wide specificity of irreversible inhibitors than reversibly acting compounds.

### Chemical genetics: pharmacology meets systems biology

In the 1970s and 1980s, a tremendous amount of biological discovery was driven by forward genetic screening. Yeast were ideal for these studies owing to their facile genetics and clonal growth. Forward genetics in higher eukaryotes is also possible, but is more difficult owing to longer lifespans, diploid genomes and difficulty isolating clones. Forward chemical genetics has, to some degree, picked up where traditional forward genetics left off, using a similar theoretical framework, but employing small-molecule libraries to interrogate phenotype and protein-capture technology to identify targets<sup>17</sup>. The peculiar challenges associated with high-throughput screening of small molecules, not the least of which is target identification<sup>18</sup>, have prompted many researchers to turn to screening efforts that use molecular biology techniques such as RNA interference (RNAi) or complementary DNA overexpression. These methods are certainly valuable and work well to complement and substantiate targets of small molecules. However, they also raise a more general question: if phenotypes can be discovered by direct manipulation of mRNA, what value does chemical screening hold for discovering

new features of biological systems? An excellent recent commentary has addressed precisely this issue<sup>19</sup> and raises the point that protein depletion, in contrast to chemical inhibition, can lead to discordant phenotypes when proteins perform multiple functions (for instance, when a protein has both catalytic and scaffolding activities that are simultaneously disrupted by depletion). Here, we would like to highlight another often-overlooked special feature of bioactive small molecules: their ability to perturb multiple targets in a cellular system.

Inhibitors with high target selectivity are typically thought of as superior to those that perturb multiple proteins, and for good reason: off-target effects of pharmaceuticals are often responsible for toxicity and unexpected or misleading biological results. That said, sometimes it is precisely the plurality of targets that is responsible for the efficacy of an inhibitor, and, in this respect, small molecules can reveal the interplay of biochemical pathways in complex biological systems that would otherwise be missed with selective agents such as RNAi. A recent and compelling example is the drug imatinib (Gleevec), which is used to treat chronic myelogenous leukemia (CML). The intended target of imatinib is the fusion protein BCR-ABL, which is a constitutively active tyrosine kinase that drives proliferation of CML cells. Despite its relatively high selectivity for BCR-ABL over other protein kinases, imatinib also inhibits other tyrosine kinases, including c-Kit. An analog-sensitive (*as*)-BCR-ABL was engineered that allowed selective inhibition of

this kinase without concomitant inhibition of c-Kit and was used to show that simultaneous inhibition of both kinases was, indeed, required for the potent cytotoxic effects of imatinib on leukemia cells<sup>20</sup>. Thus, the clinical efficacy of imatinib, lauded as a paragon of rational drug design, is likely due to unintentional blockade of multiple targets—a finding that was not fully appreciated until researchers studied this drug with integrated chemical and systems biology methods. This study and others<sup>21</sup> demonstrate that small molecules are often uniquely suited to drive discovery in complex biological systems that may be too robust to be perturbed by single target-specific approaches.

### Summary and outlook

The fields of chemical and systems biology are poised to tackle complex biological problems owing to the advent of a number of enabling 'bridge' technologies. Chemical capture techniques to enrich specific classes of proteins have provided the first global portraits of protein PTMs and activity in native biological systems. The utility of chemical probes has been augmented by advances in bio-orthogonal and analytical chemistry that improve methodological versatility, resolution and sensitivity. Furthermore, incorporating reactive tags into small molecules to comprehensively identify their interacting proteins offers a potentially universal strategy to determine mechanism of action directly in living systems and might one day even make a case for the preferential adoption of covalent agents as pharmacological probes and drugs.

Looking forward, a number of challenging problems remain for the ambitious 'chemical systems' biologist. Innovative informatic strategies are required to integrate and interpret the huge volumes of data produced by large-scale experimental approaches so that coherent descriptions of complex biochemical processes can emerge. A glimpse of what the future may hold can be found in recent studies where advanced protein engineering<sup>22</sup> and analytical/bioinformatic<sup>23</sup> platforms were used to generate complementary proteome-wide profiles of proteolytic events in apoptotic cells. The unification of such experimental and computational platforms promises to provide the first holistic view of proteolysis. On the chemical side, facile approaches are still needed to generate probes to profile and perturb biomolecular phenomena, such as protein-protein interactions, that have historically been viewed as "undruggable," although some headway in this area is being made<sup>24</sup>. Ultimately, scientists working at the interface of chemical and systems biology should aspire to create a complete tool kit to characterize and manipulate the functional state of any protein (or set of related proteins) in native biological systems. The proteome-wide implementation of these methods has the potential to provide the fundamental molecular information required to construct the biochemical networks of life. As an added bonus, the very chemical tools that are used to interrogate life systems should also facilitate drug development efforts to treat these systems when they go awry.

1. Crick, F.H., Barnett, L., Brenner, S. & Watts-Tobin, R.J. *Nature* **192**, 1227–1232 (1961).
2. Blethrow, J.D., Glavy, J.S., Morgan, D.O. & Shokat, K.M. *Proc. Natl. Acad. Sci. USA* **105**, 1442–1447 (2008).
3. Khidekel, N., Ficarro, S.B., Peters, E.C. & Hsieh-Wilson, L.C. *Proc. Natl. Acad. Sci. USA* **101**, 13132–13137 (2004).
4. Dalhoff, C., Lukinavicius, G., Klimasauskas, S. & Weinhold, E. *Nat. Chem. Biol.* **2**, 31–32 (2006).
5. Kolb, H.C., Finn, M.G. & Sharpless, K.B. *Angew. Chem. Int. Ed.* **40**, 2004–2021 (2001).
6. Cravatt, B.F., Wright, A.T. & Kozarich, J.W. *Annu. Rev. Biochem.* **77**, 383–414 (2008).
7. Greenbaum, D. *et al. Mol. Cell. Proteomics* **1**, 60–68 (2002).
8. Jessani, N. *et al. Nat. Methods* **2**, 691–697 (2005).
9. Weerapana, E., Simon, G.M. & Cravatt, B.F. *Nat. Chem. Biol.* **4**, 405–407 (2008).
10. Lee, J., Yu, P., Xiao, X. & Kodadek, T. *Mol. Biosyst.* **4**, 59–65 (2008).
11. Salisbury, C.M. & Cravatt, B.F. *Proc. Natl. Acad. Sci. USA* **104**, 1171–1176 (2007).
12. Cohen, M.S., Hadjivassiliou, H. & Taunton, J. *Nat. Chem. Biol.* **3**, 156–160 (2007).
13. Evans, M.J., Saghatelian, A., Sorensen, E.J. & Cravatt, B.F. *Nat. Biotechnol.* **23**, 1303–1307 (2005).
14. Arastu-Kapur, S. *et al. Nat. Chem. Biol.* **4**, 203–213 (2008).
15. Ahn, K. *et al. Biochemistry* **46**, 13019–13030 (2007).
16. Robertson, J.G. *Biochemistry* **44**, 5561–5571 (2005).
17. Knight, Z.A. & Shokat, K.M. *Cell* **128**, 425–430 (2007).
18. Burdine, L. & Kodadek, T. *Chem. Biol.* **11**, 593–597 (2004).
19. Weiss, W.A., Taylor, S.S. & Shokat, K.M. *Nat. Chem. Biol.* **3**, 739–744 (2007).
20. Wong, S. *et al. Proc. Natl. Acad. Sci. USA* **101**, 17456–17461 (2004).
21. Chen, S. *et al. Proc. Natl. Acad. Sci. USA* **103**, 17266–17271 (2006).
22. Mahrus, S. *et al. Cell* **134**, 866–876 (2008).
23. Dix, M.M., Simon, G.M. & Cravatt, B.F. *Cell* **134**, 679–691 (2008).
24. Wells, J.A. & McClendon, C.L. *Nature* **450**, 1001–1009 (2007).
25. Alexander, J.P. & Cravatt, B.F. *Chem. Biol.* **12**, 1179–1187 (2005).

