AWARD ADDRESS

2005 Irving Sigal Young Investigator Award

TOM W. MUIR

Laboratory of Synthetic Protein Chemistry, The Rockefeller University, New York, New York 10021, USA

It is a terrific honor to receive the Irving Sigal Young Investigator Award. As I look back at the past recipients of this award, I am truly humbled, and I can only hope that being added to this list will in some way catapult my future research endeavors to something approaching the level of these great protein scientists. Such aspirations aside, I would like to thank the Protein Society for recognizing our work over the last several years and Merck Research Laboratories for sponsoring this award.

My research interests revolve around the generation and uses of (semi)synthetic proteins. The idea of using organic chemistry to make proteins is certainly not new; peptide and protein chemists have been pursuing this goal for a century. However, it is only in the last 10–15 years that the routine chemical synthesis of proteins has been possible (Dawson and Kent 2000). Thus, I have had the very good fortune to “grow up” in this area at an opportune time. I have also had the great privilege to train under two of the world’s leaders in the field of synthetic protein chemistry, namely Professor Robert Ramage, with whom I worked as a graduate student at the University of Edinburgh, and Professor Stephen B.H. Kent, with whom I worked as a post-doctoral fellow at the Scripps Research Institute. I learned a great deal from these mentors, too much to do justice to in the limited space available here, but I guess what I am most grateful for is that they were able to infect me with their enthusiasm for science and, specifically, for their absolute belief that the ability to make proteins chemically would represent an important vehicle for an expanded dialogue between chemists and biologists.

The transition from peptide synthesis to protein synthesis has not been straightforward. Classically, two different approaches were taken to this problem. The first involved essentially a linear extrapolation of the

stepwise solid-phase peptide synthesis (SPPS) strategies used to make small peptides. SPPS has been subject to almost constant refinement since its inception, thereby allowing increasingly longer polypeptides to be assembled. Nonetheless, it still remains exceedingly difficult to synthesize peptides longer than ~50 amino acid residues in length, at least in an efficient fashion (Kent 1988). Fragment condensation approaches sidestep this ceiling by employing a convergent synthetic approach in which multiple, smaller peptides are condensed together to give the final polypeptide. Standard peptide coupling procedures are typically used in the condensation steps, which means that the peptide fragments need to be fully protected to ensure regioselectivity, resulting in progressively intractable solubility problems. Thus, while stepwise SPPS and fragment condensation have both yielded notable successes, neither represented a general solution to the problem of how to synthesize proteins.

In the early 1990s, the field of synthetic protein chemistry received a much-needed shot in the arm with the introduction of the chemical ligation approach. As with any breakthrough, chemical ligation did not emerge out of thin air, but was conceived against a growing realization by the field of the need to dispense with maximal protection strategies in fragment condensation. Taking this line of thought to its logical conclusion, Schnölzer and Kent demonstrated in 1992 that fully unprotected peptides could be regioselectively ligated together in water by using a chemoselective thioester-generating coupling reaction (Schnölzer and Kent 1992). This seminal study opened the floodgates, and soon a plethora of different ligation chemistries were developed for the purposes of hooking together synthetic peptides. The most widely used of these is the native chemical ligation (NCL) strategy that results in the generation of a normal peptide bond at the ligation junction (Dawson et al. 1994). NCL is compatible with all the natural amino acids in their unprotected form and requires only that the two unprotected peptide segments contain the necessary reactive handles for ligation, namely, an N-terminal Cys residue in one fragment and an α -thioester group in

Reprint requests to: Tom W. Muir, Laboratory of Synthetic Protein Chemistry, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA; e-mail: muirt@mail.rockefeller.edu; fax: (212) 327-7358.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.051848105>.

the other. The NCL approach has been widely used by biological chemists to prepare small to moderately sized proteins.

Upon completing my postdoctoral studies, I was fortunate enough to be offered a junior faculty position at the Rockefeller University. Surrounded by some of the world's great biologists, I soon came to realize that most of the proteins involved in say, cellular signaling, were going to be out of reach to total synthesis, even using NCL. At the same time, I remained convinced that synthesis could be helpful in elucidating the biological function and regulation of proteins of this type. After some thrashing around, my group and I decided to take a semisynthetic approach to the problem. The idea of hooking together synthetic peptides and naturally derived protein fragments (i.e., semisynthesis) was by this point well established. Indeed, some of the first reports of chemoselective ligation reactions involved linking together modified recombinant protein fragments (Gaertner et al. 1992). Moreover, in a very influential study, Verdine and coworkers had already confirmed that NCL was compatible with semisynthesis; in this case, a synthetic probe was ligated to the N terminus of a recombinant protein bearing an N-terminal Cys (Erlandson et al. 1996). The remaining piece of the puzzle was how to prepare recombinant protein α -thioesters, which was needed if NCL was to be fully integrated with semisynthesis. As it turned out, this problem had already been solved by researchers at New England Biolabs, who had generated a mutant

self-cleaving intein for the purposes of recombinant protein purification (Chong et al. 1997). Inteins are a family of autocatalytic proteins that excise themselves from a host protein (termed an extein) in a post-translational process known as protein splicing (by analogy with RNA splicing). Biochemical studies had revealed the basic chemical steps in protein splicing, and had identified the key conserved residues within the intein family required for these steps. This information allowed the design of an intein point-mutant that leads to the accumulation of a protein splicing intermediate in which the N-extein polypeptide is linked to the intein via a labile α -thioester bond. Importantly, these mutant inteins appeared to be rather promiscuous with respect to the nature of the N-extein, thereby providing a potentially general route to recombinant protein α -thioesters.

Working in collaboration with Phil Cole's group, at that time also at Rockefeller, we quickly found that recombinant protein α -thioesters could be isolated by thiolysis of the corresponding intein fusions using small molecules (Muir et al. 1998). Once generated, these reactive building blocks could then be used in standard NCL reactions with suitable synthetic peptides (Fig. 1). Expressed protein ligation (EPL), as this technology was termed, has turned out to be a remarkably general approach for the semisynthesis of proteins (Muir 2003). Indeed, EPL has been widely used and applied to many classes of protein, including kinases (as in the original study), phosphatases, GTPases, polymerases, transcription factors, antibodies, ion channels, and so on. As was

Figure 1. Principle of EPL in which synthetic and recombinant proteins are chemically ligated together using the NCL reaction. SPPS, solid phase peptide synthesis.

originally hoped, synthetic access to these molecules has facilitated many aspects of function and regulation to be studied through the site-specific incorporation of biochemical and biophysical probes (Fig. 2). In terms of our own work, we have primarily focused on using EPL to study the regulation of proteins involved in cellular signaling. This is best illustrated by our work on the TGF- β signal transduction pathway.

TGF- β signaling is the quintessential growth inhibitory cytokine signaling pathway, and is involved in a myriad of normal and disease processes, including development, tissue homeostasis, and cancer. This signaling pathway, as with many others, is regulated by reversible protein phosphorylation events. Using EPL, we have been able to probe the mechanisms of phosphoregulation of TGF- β signaling in some detail. This has revealed, for example, how hyperphosphorylation of a regulatory region in the type I TGF- β receptor Ser/Thr kinase (T β RI) switches the protein from an inhibited state into an activated form capable of binding and doubly phosphorylating its substrate, Smad2 (Huse et

al. 2001). In subsequent studies, EPL was used to prepare phospho-variants of Smad2 (Ottesen et al. 2004). These were used to show that T β RI kinase phosphorylates its substrate in an ordered, possibly processive, fashion. This information has turned out to be useful in the design of optimal peptide substrates of T β RI for use in medicinal chemistry initiatives.

There is, of course, a limit to the type of questions one can ask about a signaling network using a reconstituted biochemical system. With this in mind, we have begun to think about how semisynthesis might be used to probe cell signaling *in vivo*. Again, TGF- β signaling has been the initial focus and, specifically, we have developed two strategies that allow the activity of Smad2 to be controlled with light. In one approach, phosphorylated Smad2 was modified with a fluorophore and a photocleavable moiety that acted as both a caging group and a fluorescence quenching group (Pellois et al. 2004). In the caged state, the protein formed a nonfluorescent heterodimer with the regulatory protein SARA. Irradiation with UV light led to photocleavage of the caging group

Figure 2. Examples of biochemical and biophysical probes incorporated into proteins using EPL. The approach has also been used to ligate proteins to DNA and surfaces and to circularize recombinant proteins.

and produced a fluorescent homotrimer of the protein. These *in vitro* experiments demonstrated that a photochemical trigger mimicking the critical biochemical event of Smad2 phosphorylation in the TGF- β signaling pathway could be achieved and that fluorescence could be used as a coupled read-out of protein activity. Very recently, we have shown that this type of strategy can be used to trigger and simultaneously monitor Smad2 nuclear import in real time in living mammalian cells (J.-P. Pellois and T.W. Muir, unpubl.). In a second approach, we prepared an analog of Smad2 directly caged on the two activating phosphoserine residues (Hahn and Muir 2004). EPL was used to chemically ligate a recombinant fragment of the protein to a synthetic segment bearing two caged phosphate moieties. Biochemical and cell biological experiments demonstrated that the resulting caged semisynthetic protein was fully activated upon irradiation with UV light. This sets the stage for a detailed kinetic analysis of Smad2 function in living cells.

One of the drawbacks of using EPL to make caged-Smad2 analogs is that the semisynthetic proteins must be microinjected into cells. This is a tedious business, and limits the approach to questions that can be addressed at the single cell level. To solve this problem, we developed an approach that allows ligation of synthetic molecules to target proteins inside living cells (Giriati and Muir 2003). This technology makes use of the *Ssp* DnaE intein, which is naturally split into two pieces. Recent biophysical studies from our lab indicate that that N- and C-terminal halves of this intein interact tightly and with rapid association kinetics (Shi and Muir 2005). We reasoned that DnaE intein complementation and subsequent *trans*-splicing could be used to effect semisynthesis in cells. In this scheme, a cellular protein is genetically tagged with one half of the split intein. The complementary half is linked *in vitro* to the synthetic probe, and this fusion is delivered into cells using a transduction peptide. Association of the intein halves in the cytosol triggers protein *trans*-splicing, resulting in the ligation of the probe to the target protein through a peptide bond. Importantly, the process is traceless since the intein is removed from the protein of interest and replaced by the probe. In proof-of-principle studies, this *in vivo* semisynthesis strategy was shown to be specific and applicable to both cytosolic and integral membrane proteins (Giriati and Muir 2003). This technology adds to the growing number of approaches being developed that allow recombinant fusion proteins to be modified with small molecule probes in cells (Hahn and Muir 2005).

Protein splicing results in a dramatic change in the primary structure and, by extension, function of a protein. Thus, the ability to control the activity of an intein might provide a useful way of controlling protein function at the post-translational level. To that end, we re-

cently developed a conditional protein splicing (CPS) reaction that is triggered by the addition of a small molecule (Mootz and Muir 2002). CPS works by fusing the two halves of an artificially split yeast intein to FKBP and FRB heterodimerizer domains. The intein halves have very low affinity for one another; however, upon addition of the small molecule rapamycin and subsequent assembly of the ternary complex, the intein reconstitutes and splicing is triggered. CPS has been shown to work in mammalian cells, allowing two genetically encoded polypeptides of choice to be spliced together in a dosable and temporally controlled manner (Mootz et al. 2003). In principle, CPS can be used to control the activity of a protein in a number of ways. For example, the target protein could be split into two inactive fragments and CPS used to assemble the functional protein. Alternatively, it should be possible to exploit the bond-breaking steps in splicing by appending a *cis*-acting regulatory element to an intact protein such that it would be cleaved off following splicing. As an example of this, we recently generated an autoregulated version of a protein kinase in which a pseudosubstrate inhibitor was linked to the kinase via one of the CPS elements (Mootz et al. 2004). Induction of splicing with rapamycin resulted in removal of the intrasteric inhibition with concomitant activation of the kinase activity. It should be stressed that CPS is still a fairly new technique, and hence, its scope and limitations remain to be fully defined. Nonetheless, the approach appears to allow the levels of a spliced protein to be controlled with some precision in a cell (and perhaps even an animal). This feature could have important implications for studying kinetic and thermodynamic aspects of protein function *in vivo*.

As noted at the beginning of this article, I have had the opportunity to work in the area of synthetic protein chemistry during a period of rapid progress. The last decade has seen the development of semisynthetic (Muir 2003) and biosynthetic (Wang and Schultz 2005) strategies that allow the primary structure of proteins to be chemically tailored in quite remarkable ways. Thus, I believe there is (or should be) less of an emphasis now on how we might make proteins and more of an emphasis on what we might do with proteins now that we *can* make them. Many areas stand to benefit from synthetic access to proteins. Most applications to date have been of a biochemical or biophysical nature. These studies illustrate the power of synthetic protein chemistry for probing structure-activity relationships in proteins. Nonetheless, there are other areas such as cell biology and nanobiotechnology where synthetic protein chemistry could make an impact. Indeed, I think it is safe to say that we have only scratched the surface in terms of what is possible.

Thus, we can expect the unexpected now that synthetic chemistry has been unleashed on the world of proteins.

Acknowledgments

I thank the many members of my research group, past and present, for making the journey so far both stimulating and enjoyable. I also thank my collaborators for guiding me to interesting problems and for teaching me that chemical biology is a collaborative enterprise. Finally, I acknowledge the support of the NIH for financing our research.

References

- Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., Perler, F.B., Benner, J., Kucera, R.B., Hirvonen, C.A., et al. 1997. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**: 271–281.
- Dawson, P.E. and Kent, S.B.H. 2000. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **69**: 923–960.
- Dawson, P.E., Muir, T.W., Clark-Lewis, I., and Kent, S.B.H. 1994. Synthesis of proteins by native chemical ligation. *Science* **266**: 776–779.
- Erlanson, D.A., Chytil, M., and Verdine, G.L. 1996. The leucine zipper domain controls the orientation of AP-1 in the NFAT-AP-1-DNA complex. *Chem. Biol.* **3**: 981–991.
- Gaertner, H.F., Rose, K., Cotton, R., Timms, D., Camble, R., and Offord, R.E. 1992. Construction of protein analogues by site-specific condensation of unprotected peptides. *Bioconjugate Chem.* **3**: 262–268.
- Giriat, I. and Muir, T.W. 2003. Protein semi-synthesis in living cells. *J. Am. Chem. Soc.* **125**: 7180–7181.
- Hahn, M.E. and Muir, T.W. 2004. Photocontrol of Smad2, a multi-phosphorylated cell signaling protein through caging of activating phosphoserines. *Angew. Chem.* **43**: 5800–5803.
- . 2005. Manipulating proteins with chemistry: A cross-section of chemical biology. *Trends Biochem. Sci.* **30**: 26–34.
- Huse, M., Muir, T.W., Xu, L., Chen, Y.-G., Kuriyan, J., and Massague, J. 2001. The TGF β receptor activation process: An inhibitor to substrate-binding switch. *Mol. Cell* **6**: 671–682.
- Kent, S.B.H. 1988. Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.* **57**: 957–989.
- Mootz, H.D. and Muir, T.W. 2002. Protein splicing triggered by a small molecule. *J. Am. Chem. Soc.* **124**: 9044–9045.
- Mootz, H.D., Blum, E.S., Tyszkiewicz, A.B., and Muir, T.W. 2003. Conditional protein splicing: A new tool to control protein structure and function in vitro and in vivo. *J. Am. Chem. Soc.* **125**: 10561–10569.
- Mootz, H.D., Blum, E.S., and Muir, T.W. 2004. Activation of an auto-regulated protein kinase by conditional protein splicing. *Angew. Chem.* **43**: 5189–5192.
- Muir, T.W. 2003. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **72**: 249–289.
- Muir, T.W., Sondhi, D., and Cole, P.A. 1998. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci.* **95**: 6705–6710.
- Ottesen, J.J., Huse, M., Sekedat, M.D., and Muir, T.W. 2004. Semisynthesis of phosphovariants of Smad2 reveals a substrate preference of the activated T β R1 kinase. *Biochemistry* **43**: 5698–5706.
- Pellois, J.-P., Hahn, M.E., and Muir, T.W. 2004. Simultaneous triggering of protein activity and fluorescence. *J. Am. Chem. Soc.* **126**: 7170–7171.
- Schnölzer, M. and Kent, S.B.H. 1992. Constructing proteins by dovetailing unprotected synthetic peptides: Backbone-engineered HIV protease. *Science* **256**: 221–225.
- Shi, J. and Muir, T.W. 2005. Development of a tandem protein *trans*-splicing system based on native and engineered split inteins. *J. Am. Chem. Soc.* **127**: 6198–6206.
- Wang, L. and Schultz, P.G. 2005. Expanding the genetic code. *Angew. Chem.* **44**: 34–66.