ChemComm
www.rsc.org/chemcomm

"The splice is right": how protein splicing is opening new doors in protein science

Edmund C. Schwartz, Tom W. Muir* and Amy B. Tyszkiewicz

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; Tel: 212 327-7368

In the decade since the discovery of protein splicing, this natural process has been exploited to develop novel techniques that have expanded the scope of protein science. These techniques are being adapted for *in-vivo* use, facilitating the study of proteins in their natural environment.

AT A CERTAIN LEVEL, all science is driven by technology, and improvements in that technology change previously unattainable goals into routine experiments, as well as inspire entirely novel directions of study. The field of protein science is no exception. Protein scientists know the pain of having great ideas that were brainstormed in laboratory meetings eliminated by the reality of what is technically feasible at the time. The family of molecules known as proteins is both complex and diverse, making their study extremely difficult. Determining what a particular protein does requires that one be able to isolate it from all of the other proteins in the cell, either by synthesizing it in the laboratory or by purifying it from a population of cells. Neither of these tasks is trivial. Indeed, until the widespread use of macromolecular mass spectrometry, even determining if the protein synthesized or purified was the one desired was a task in and of itself. Isolating and studying the protein on the benchtop isn't enough, however. To truly understand a protein's function, one needs to be able to selectively affect and sort out the effects of a single protein when it is in an *in vivo* context, surrounded by literally thousands of other proteins, all chugging away at their own tasks. If isolating a protein in a test tube is difficult, isolating its effects while it's still in the cell is a truly Herculean task

We believe that protein splicing (Fig. 1) and its associated technologies, such as expressed protein ligation (EPL) and protein trans-splicing, have the potential to have a profound impact on the field of protein science by aiding in both the

Edmund C. Schwartz (RIGHT) received a B. S. with highest distinction from the University of Virginia in 2002, with majors in chemistry and biology, and is currently a graduate fellow at the Rockefeller University. At UVA he was an Echols scholar and was recently named an Howard Hughes Medical Institute predoctoral fellow starting in July 2003

Tom W. Muir (CENTRE) received his B. Sc (Hons, 1st class) in Chemistry from the University of Edinburgh in 1989 and his Ph.D. in Chemistry from the same institute in 1993 under the direction of Professor Robert Ramage. After postdoctoral studies with Stephen B.H. Kent at the Scripps Research Institute, he joined the faculty at the Rockefeller University in New York City in 1996, where he is now Full Professor. He has won a number of honors for his research, including: the Burroughs-Wellcome Fund New Investigator Award, the Pew Award in the Biomedical Sciences, the Alfred P. Sloan Research Fellow Award and the Leonidas Zervas Award from the European Peptide Society

Amy B. Tyszkiewicz (LEFT) received a B. S. in chemistry from the Massachusetts Institute of Technology in 2002, and is pursuing graduate research at the Rockefeller University. She is a Rockefeller University Women and Science Graduate Fellow.

DOI: 10.1039/b304989m

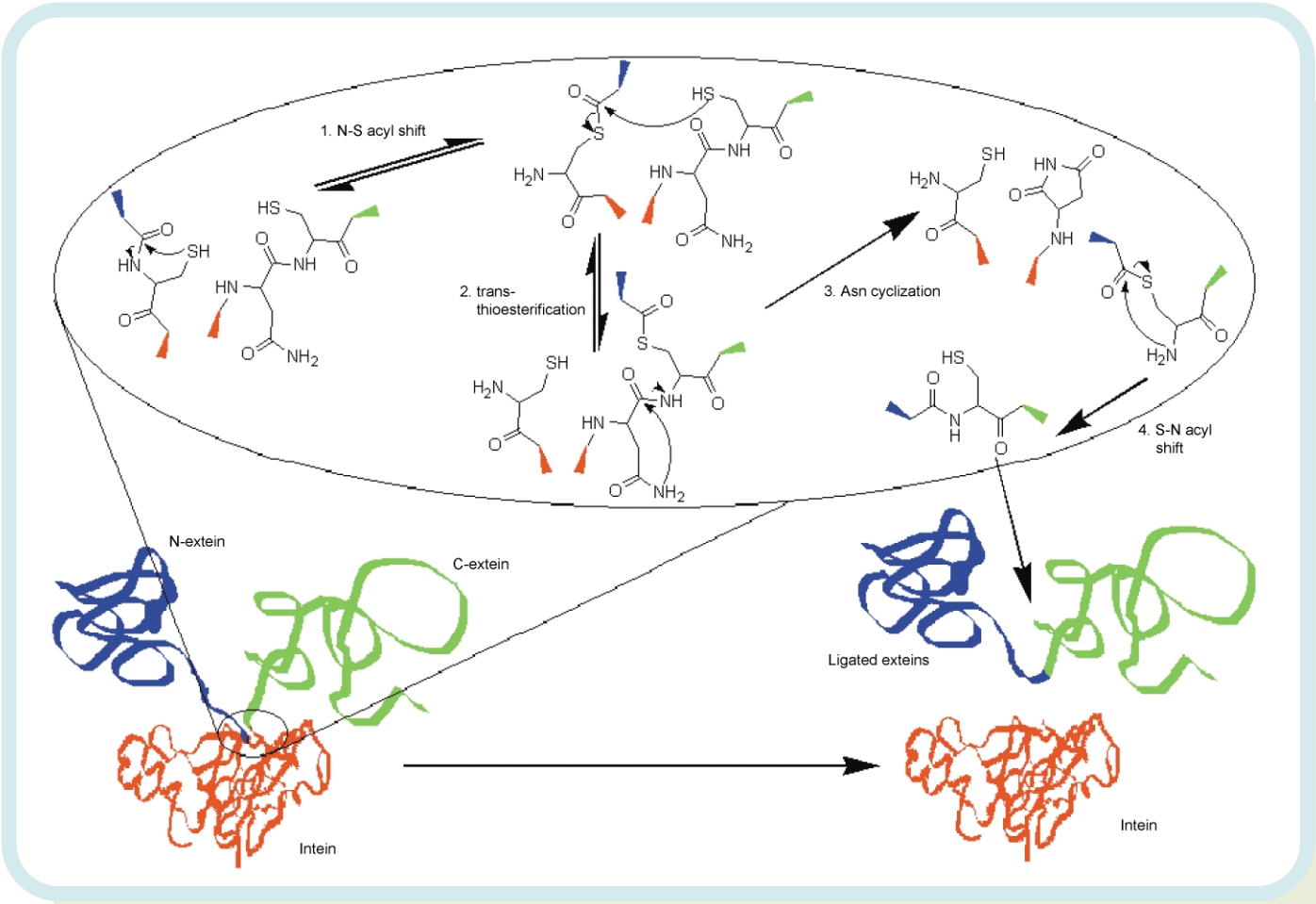

Fig. 1 The basic protein splicing mechanism. Note that the two exteins are joined together with a regular peptide bond.

synthesis and isolation of desired proteins as well as by allowing them to be regulated and detected in an *in-vivo* environment. By wedding benchtop chemical synthesis with molecular biology, expressed protein ligation is proving to be a significant breakthrough in building bigger, more complicated protein constructs, many of which were unattainable before the advent of this technology. Split intein-mediated protein trans-splicing is also opening the doors to novel experiments involving monitoring protein–protein interactions, regulation of protein function, as well as the synthesis of proteins that could not easily be achieved *via* other routes both *in vitro* and *in vivo*.

Protein splicing is a naturally occurring phenomenon in which an intervening protein sequence, or 'intein,' autocatalyzes its own removal from the flanking protein sequences, or 'exteins,' joining the two exteins with a native peptide bond, in a manner analogous to pre-mRNA splicing. Research has shown that inteins are promiscuous with respect to their flanking sequences, allowing their insertion into virtually any protein without affecting the splicing reaction. Following the discovery of the first intein in 1990,[1] the various steps and products of the autocatalytic reaction have been co-opted by researchers for a number of experimental purposes (for reviews, see references [2–4]).

The modification of inteins through site-directed mutagenesis has enabled the development of novel techniques to study protein function *in vitro*. Shortly after the elucidation of the basic splicing mechanism (Fig. 1), the autoprocessing function of inteins was harnessed in the form of fusion constructs to aid in protein purification. The mutation of residues in the active site of inteins allows the arrest of the splicing reaction at an intermediate step. The protein of interest is expressed as a fusion with an intein modified in this manner and an affinity domain. In one system, thiol-induced cleavage releases the protein from the intein, and consequently from the affinity domain, and allows easy elution of the protein off an affinity column.[5] A related system can also be applied to the overexpression and purification of cytotoxic proteins by separating the N- and C-termini of the protein with an intein containing a chitin binding domain that can reduce toxicity and allow reconstitution of the full-length protein *in vitro*.[6] If the thiols used to induce cleavage in intein-mediated purification are chosen properly, the protein generated will be in the form of a relatively stable α-thioester. This thioester is susceptible to attack by the N-terminal cysteine of another peptide or protein. The reaction results in the formation of a native peptide bond (Fig. 2). This technique, known as expressed protein ligation (EPL),[7] allows the chemical ligation of two or more polypeptides, at least one of which is recombinant in origin. This permits one to overcome the size limitations normally associated with total synthesis of peptides while retaining the diversity of structures available to organic chemistry. EPL has been used to incorporate, among other things, unnatural amino acids,[8] fluorescent labels,[9] post-translational modifications, such as phosphorylation[7a,10] and isotopic labels for NMR studies[11] into large, recombinant proteins. Theoretically, nearly anything that can be synthesized in an organic chemistry lab can be incorporated into a protein using EPL. For a more extensive review of EPL and its applications, see reference [4].

Inteins can also be modified for synthetic purposes by splitting them in such a manner that they can reconstitute the native intein fold *in trans*, *i.e.* intermolecularly, and participate in the standard splicing reaction—a technique known as protein trans-splicing.[2] Although the intein-mediated protein splicing reaction is spontaneous *in vivo*, artificially split inteins have to be reconstituted *in vitro*, most commonly by denaturation and renaturation. A recent report by Mootz and Muir showed a different method of

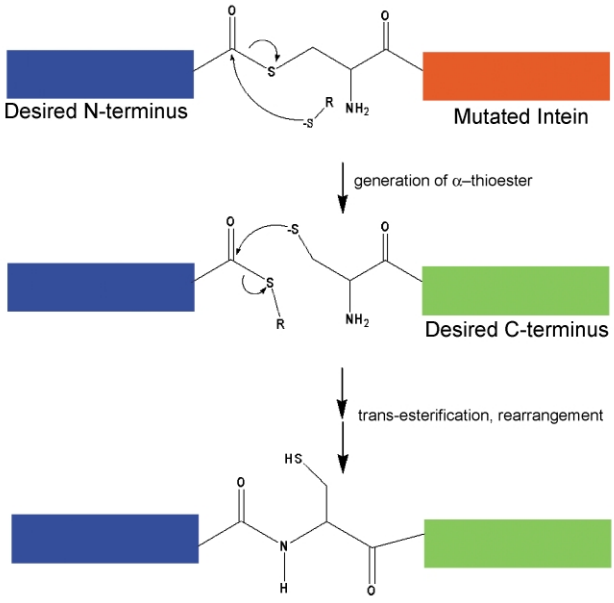

Fig. 2 A general expressed protein igation (EPL) scheme. In this case, the N-terminus is recombinantly expressed with a mutant intein that can carry out the initial N–S acyl transfer, but not the subsequent steps of the protein splicing reaction. The thioester is generated by the addition of thiols (HSR) to the reaction vessel. N-terminal building blocks for use in EPL can also be generated synthetically. C-terminal building blocks can be either recombinant or synthetic, as long as they contain an N-terminal cysteine.

reconstituting split inteins.[12] By fusing the halves of a split *Saccharomyces cerevisiae* VMA intein to the FKBP/FRB rapamycin-mediated heterodimerization system,[13] they were able to selectively induce trans-splicing by adding the small molecule rapamycin, without the need for denaturing conditions (Fig. 3A). Given the promiscuity of inteins with regard to the flanking exteins, this system theoretically allows selective regulation of virtually any protein with great temporal precision; for example, by ligating together two inactive pieces of a protein (to give a functional protein) in response to the cell-permeable rapamycin molecule. Indeed, recent data from our group indicates that this conditional protein trans-splicing system works extremely well *in vivo* (Mootz, *et al.* unpublished data).

The protein splicing reaction has also been exploited to allow chemical manipulations to be performed *in vivo*. Perhaps the most widely used example of this is in the generation of head-to-tail (backbone) cyclized peptides in cells. Several important naturally occurring antibiotics and immunosuppressants are based on cyclic peptides; therefore the ability to prepare this class of molecules using genetic approaches is an exciting development. Producing these cyclic peptides *in vivo* opens the possibility of creating large, combinatorial libraries and using diversity-based screening to discover new lead compounds. One method that has been used to cyclize peptides *in vivo* is intramolecular trans-splicing in which the C- and N-terminal halves of a split intein are expressed at the N- and C-termini, respectively, of the protein or peptide to be cyclized (Fig. 3B). The resulting intramolecular splicing results in a circular protein backbone.[14] Our group has used another technique to generate circular peptides *in vivo*. The protein is expressed as a fusion protein of the form Met-Cys-peptide-intein. Methionine-aminopeptidase, which is endogenous to the cell then removes the N-terminal methionine and exposes the cysteine which then takes part in an intramolecular ligation reaction with the α-thioester connecting the protein to the intein domain, also resulting in a cyclic peptide.[15] Split intein-mediated cyclization has been combined with the generation of peptide libraries to generate large numbers of cyclic peptides for screening.[16] In a recent application of this technique, a retroviral vector was used to deliver a library of random cyclic peptides into human B cells, and peptides that inhibit interleukin-4 signaling were identified.[17]

The many genome projects already completed and currently in progress can potentially reveal every protein present in

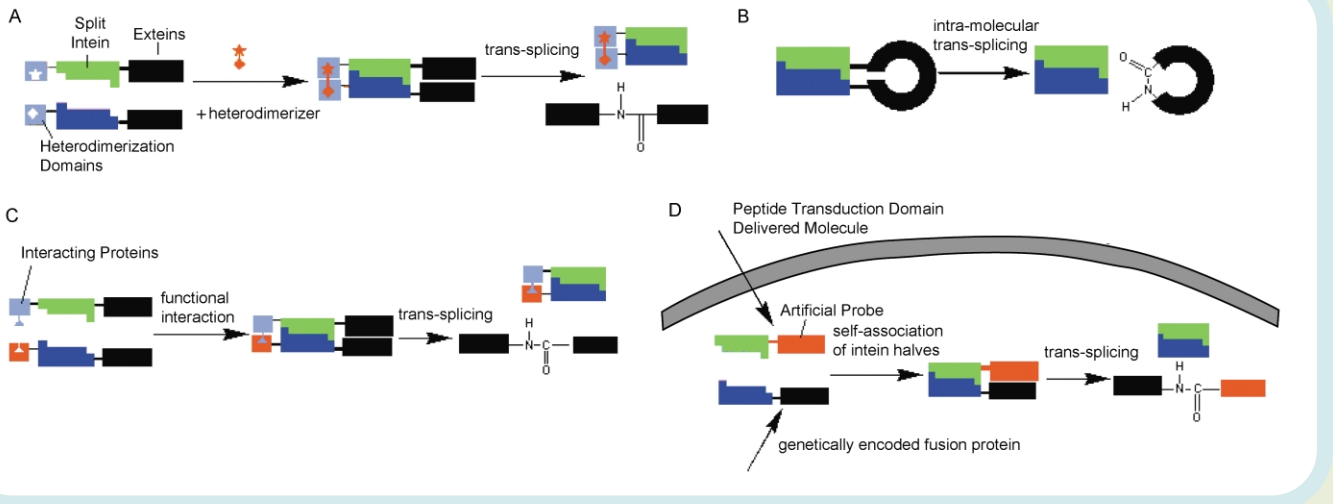

Fig. 3 Applications of protein trans-splicing. Natural and artificial split inteins have been utilized in a number of ways to accommodate protein synthesis. A) An artificially split intein in which the halves have low affinity for one another can be linked to a heterodimerization domain to allow small-molecule triggered splicing. B) A split and reversed intein can be used to generate cyclic peptides or proteins. C) An interaction between two proteins can be detected by utilizing protein trans-splicing to generate luciferase or GFP as a readout. D) A naturally split intein in which the halves have a high affinity for one another can allow *in vivo* protein semi-synthesis by delivering and targeting a synthetic molecule to a cellular protein.

an organism, but do not fully answer what each protein does, when and where each protein is expressed or what proteins interact with one another. The yeast two-hybrid system, in which a protein–protein interaction activates the transcription of a reporter gene, has been used to elucidate protein interactions, but is spatially limited to interactions that occur in the nucleus.[18] Umezawa *et al.* have developed a protein trans-splicing based interaction screen that can detect interactions anywhere in the cell (Fig. 3C). In it, putative interacting proteins are fused with two halves of a split intein and a split luciferase or GFP. If the two proteins are in close proximity, whether due to a functional interaction or to compartmentalization, the intein halves will be able to act *in trans*, leading to splicing of two halves of luciferase or GFP. The resulting fluorescence serves as an indicator of a protein interaction.[19] Recently, this technique has been used to identify mitochondrial proteins.[20]

The prospect of applying the myriad structures available through traditional chemical synthesis to *in-vivo* protein science is intriguing to say the least, whether in terms of introducing better probes and indicators into proteins, or in terms of actually engineering protein function *de novo*. This is made difficult by the harsh conditions often required for organic synthesis, which are incompatible with biological processes. A recent technique developed by Giriat and Muir allows an externally synthesized molecule to be delivered into a cell and ligated to a recombinant protein within the cell (Fig. 3D). The protein of interest is expressed as a fusion with one half of a naturally split intein. The externally synthesized molecule is coupled to the other half of the split intein and to a protein transduction domain peptide (PTD) which causes cells to internalize the construct. Once internalized, the two halves of the split intein reassociate, due to their high intrinsic affinity, and splice the internally expressed protein with the externally synthesized molecule.[21] In principle, this method will allow targeted cellular proteins to be elaborated with any number of artificial probes.

The discovery of protein splicing has catalyzed, in the span of barely over a decade, an expansion of the techniques available for the study of proteins. It has already aided in the purification and synthesis of a variety of proteins, as well as providing new tools for studying protein interactions and for the regulation of proteins. As expressed protein ligation gains in popularity and comes into more widespread use, we can expect more biochemical and biophysical problems to be solved, as previously synthetically unavailable proteins are made and studied. The continued development of *in-vivo* procedures will allow the study of proteins in their natural environment. Many of the experiments to date using these technologies have been proof-of-principle type experiments, and now that many of these principles have been proven, perhaps we can start to truly tap the potential that these approaches provide.

Acknowledgements

This work was supported by the N.I.H. and The Rockefeller University's Women in Science Fellowship Program. We thank Sarah C. Smith for the illustration in the Graphical Abstract.

Notes and references

1 (*a*) R. Hirata, Y. Ohsumi, A. Nakano, H. Kawasaki, K. Suzuki and Y. Anraku, *J. Biol. Chem.*, 1990, **265**, 6726; (*b*) P. M. Kane, C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebl and Tom H. Stevens, *Science*, 1990, **250**, 651.

2 H. Paulus, *Annu. Rev. Biochem.*, 2000, **69**, 447.

3 I. Giriat, T. W. Muir and F. B. Perler, *Genet. Eng.*, 2001, **23**, 171.

4 T. W. Muir, *Annu. Rev. Biochem*, 2003, **72**, 249.

5 (*a*) S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus and M.-Q. Xu, *Gene*, 1997, **192**, 271; (*b*) S. Chong, G. E. Montello, A. Zhang, E. J. Cantor, W. Liao, M.-Q. Xu and J. Benner, *Nucleic Acids Res.*, 1998, **26**, 5109.

6 W. Wu, D. W. Wood, G. Belfort, V. Derbyshire and M. Belfort, *Nucleic Acids Res.*, 2002, **30**, 4864.

7 (*a*) T. W. Muir, D. Sondhi and P. A. Cole, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 6705; (*b*) K. Severinov and T. W. Muir, *J. Biol. Chem.*, 1998, **273**, 16205; (*c*) T. C. Evans, J. Benner and M-Q. Xu, *Protein Sci.*, 1998, **7**, 2256.

8 (*a*) B. Ayers, U. K. Blaschke, J. A. Camarero, G. J. Cotton, M. Holford and T. W. Muir, *Biopolymers*, 2000, **51**, 343; (*b*) D. Wang and P. A. Cole, *J. Am. Chem. Soc.*, 2001, **123**, 8883.

9 (*a*) G. J. Cotton and T. W. Muir, *Chem. Biol.*, 2000, **7**, 254; (*b*) K. Alexandrov, I. Heinemann, T. Durek, V. Sidorovich, R. S. Goody and H. Waldmann, *J. Am. Chem. Soc.*, 2002, **124**, 5648.

10 M. Huse, M. N. Holford, J. Kuriyan and T. W. Muir, *J. Am. Chem. Soc.*, 2000, **122**, 8337.

11 (*a*) R. Xu, B. Ayers, D. Cowburn and T. W. Muir, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 388; (*b*) J. A. Camarero, A. Shekhtman, E. A. Campbell, M. Chlenov, T. M. Gruber, D. A. Bryant, S. A. Darst, D. Cowburn and T. W. Muir, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 8536; (*c*) J. J. Ottesen, U. K. Blaschke, D. Cowburn and T. W. Muir, *Biol Mag. Res.*, 2002, **20**, 35.

12 H. D. Mootz and T. W. Muir, *J. Am. Chem. Soc.*, 2002, **124**, 9044.

13 S. N. Ho, S. R. Biggar, D. M. Spencer, S. L. Schreiber and G. R. Crabtree, *Nature*, 1996, **382**, 822.

14 (*a*) C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon and S. J. Benkovic, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 13638; (*b*) T. C. Evans, D. Martin, R. Kolly, D. Panne, L. Sun, I. Ghosh, L. Chen, J. Benner, X.-Q. Liu and M.-Q. Xu, *J. Biol. Chem.*, 2000, **275**, 9091; (*c*) H. Iwai, A. Lingel and A. Plückthun, *J. Biol. Chem.*, 2001, **276**, 16548; (*d*) N. K. Williams, P. Prosselkov, E. Liepinsh, I. Line, A. Sharipo, D. R. Littler, P. M. G. Curmi, G. Otting and N. E. Dixon, *J. Biol. Chem.*, 2002, **277**, 7790.

15 J. A. Camarero, D. Fushman, D. Cowburn and T. W. Muir, *Bioorg. Med. Chem.*, 2001, **9**, 2479.

16 C. P. Scott, E. Abel-Santos, A. D. Jones and S. J. Benkovic, *Chem. Biol.*, 2001, **8**, 801.

17 T. M. Kinsella, C. T. Ohashi, A. G. Harder, G. C. Yam, W. Li, B. Peelle, E. S. Pali, M. K. Bennett, S. M. Molineaux, D. A. Anderson, E. S. Masuda and D. G. Payan, *J. Biol. Chem.*, 2002, **277**, 37512.

18 (*a*) S. Fields and O. Song, *Nature*, 1989, **340**, 245; (*b*) C. T. Chien, P. L. Bartel, R. Sternglanz and S. Fields, *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 9578; (*c*) R. L. Finley and R. Brent, *Proc Natl. Acad. Sci. USA*, 1994, **91**, 12980.

19 (*a*) T. Ozawa, A. Kaihara, M. Sato, K. Tachihara and Y. Umezawa, *Anal. Chem.*, 2001, **73**, 2516; (*b*) T. Ozawa, M. Takeuchi, A. Kaihara, M. Sato and Y. Umezawa, *Anal. Chem.*, 2001, **73**, 5866; (*c*) R. Paulmurugan, Y. Umezawa and S. S. Gambhir, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 15608.

20 T. Ozawa, Y. Sako, M. Sato, T. Kitamura and Y. Umezawa, *Nature Biotech.*, 2003, **21**, 287.

21 I. Giriat and T. W. Muir, *J. Am. Chem. Soc.*, 2003, **125**, 7180.