Insertion of a Synthetic Peptide into a Recombinant Protein Framework: A Protein Biosensor

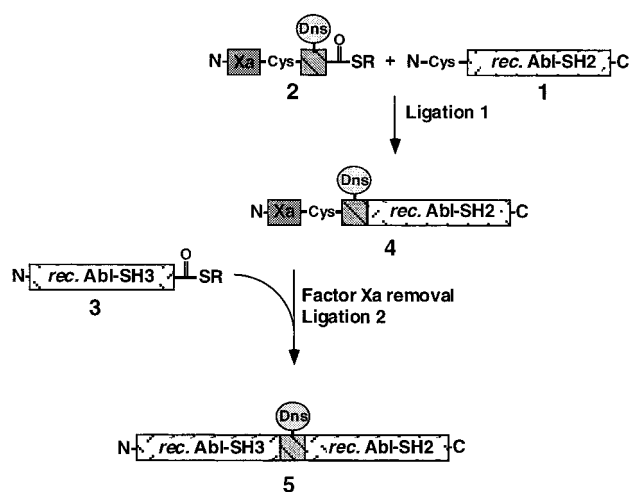Graham J. Cotton,*,[†] Brenda Ayers,[†] Rong Xu,[‡] and Tom W. Muir*,[†]

Laboratories of Synthetic Protein Chemistry and Physical Biochemistry Rockefeller University 1230 York Avenue, New York, New York 10021

The ability to insert synthetic peptides into recombinantly expressed proteins would open powerful new avenues in protein engineering and would complement existing approaches for introducing unnatural amino acids and biophysical probes into large proteins.[1] One exciting application of such a technology would be in the generation of fluorescence-based protein biosensors for investigating molecular processes;[2] in principle, an appropriate fluorophore(s) could be selectively introduced into a protein such that its fluorescence properties are dependent on the "functional state" of a process under investigation. We report here a general strategy for site specifically incorporating a synthetic peptide sequence within the backbone of a target recombinant protein. To illustrate this methodology, a synthetic tripeptide containing the environmentally sensitive 5-(dimethylamino)-naphthalene-1-sulfonamide (Dns) fluorophore[3] was inserted between the recombinantly derived Src homology 3 (SH3) and SH2 domains of the Abelson nonreceptor protein tyrosine kinase (Abl).[4] The resulting semisynthetic construct was found to biosense for high affinity bidentate interactions with the Abl-SH(32) domain pair.

Recently we described an approach, expressed protein ligation,[1h–k] which allows synthetic peptides to be chemically ligated via an amide bond[5] to the C-terminus of recombinant protein α thioesters, themselves generated by chemically intercepting a protein-splicing reaction. The biosynthetic strategy outlined in Figure 1 extends this approach by allowing a peptide insert, containing both an N-terminal cysteine residue and an αthioester moiety, to be inserted between two recombinant proteins containing complementary reactive groups at their N- or C-termini. Previously, we demonstrated the feasibility of such a process using

---

*To whom correspondence should be addressed: G. J. Cotton (cottong@rockvax.rockefeller.edu) or T. W. Muir (tmuir@rockvax.rockefeller.edu), Laboratory of Synthetic Protein Chemistry, Rockefeller University, 1230 York Avenue, New York, NY 10021. Phone: 212-327-7368. Fax: 212-327-7358.

[†] Laboratory of Synthetic Protein Chemistry.
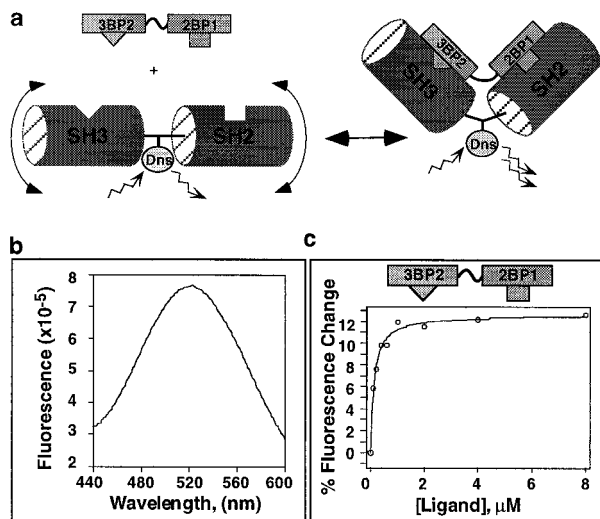
[‡] Laboratory of Physical Biochemistry.

(1) (a) Cornish, V. W.; Mendel, D.; Schultz, P. G. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 621–633. (b) Cornish, V. W.; Hahn, K. M.; Schultz, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 8150–8151. (c) Hubbell, W. L.; Mchaourab, H. S.; Altenbach, C.; Lietzow, M. A. *Structure* **1996**, *4*, 779–783. (d) Chen, Y.; Ebright, Y. W.; Ebright, R. H. *Science* **1994**, *265*, 90–92. (e) Budisa, N.; Minks, C.; Medrano, F. J.; Lutz, J.; Huber, R.; Moroder, L. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 455–459. (f) Muir, T. W. *Structure* **1995**, *3*, 649–652. (g) Wilken, J.; Kent, S. B. H. *Curr. Opin. Biotechnol.* **1998**, *9*, 412–426. (h) Muir, T. W.; Sondhi, D.; Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 6705–6710. (i) Severinov, K.; Muir, T. W. *J. Biol. Chem.* **1998**, *273*, 16205–16209. (j) Evans, T. C., Jr.; Benner, J.; Xu, M.-Q. *Protein Sci.* **1998**, *7*, 2256–2264. (k) Xu, R.; Ayers, B.; Cowburn, D.; Muir, T. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, in press.

(2) For recent reviews, see; (a) Tsien, R. Y.; Miyawaki, A. *Science* **1998**, *280*, 1954–5 (b) Hellinga, H. W.; Marvin, J. S. *Trends Biotechnol.* **1998**, *16*, 183–189. (c) Giuliano, K. A.; Taylor, D. L. *Trends Biotechnol.* **1998**, *16*, 135–40.

(3) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*; Plenum Press: New York, 1983; pp 187–214.

(4) (a) Overduin, M.; Rios, C. B.; Mayer, B. J.; Baltimore, D.; Cowburn, D. *Cell* **1992**, *70*, 697–704. (b) Gosser, Y. Q.; Zheng, J.; Overduin, M.; Mayer, B. J.; Cowburn, D. *Structure* **1995**, *3*, 1075–1086. (c) Nam, H.-J.; Hasser, W. G.; Roberts, T. M.; Frederick, C. A. *Structure* **1996**, *4*, 1105–1114.

(5) (a) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776–779. (b) Hackeng, T. M.; Mounier, C. M.; Bon, C.; Dawson, P. E.; Griffin, J. H.; Kent, S. B. H. *Proc. Natl. Acad. Sci. U.S.A.*, **1997**, *94*, 7845–7850. (c) Dawson, P. E.; Churchill, M. J.; Ghadiri, M. R.; Kent, S. B. H. *J. Am. Chem. Soc.* **1997**, *119*, 4325–4329. (d) Muir, T. W.; Dawson, P. E.; Kent, S. B. H. *Methods Enzymol.* **1997**, *289*, 266–298.

Figure 1. Sequential ligation strategy. The first step involves native chemical ligation of the synthetic peptide RGKIEGR-CK(Dns)G-propionamide α thioester[11] to the recombinant N-terminal Cys fragment, Abl-[C[121]]SH2.[12] After purification the pro-sequence RGKIEGR (Xa) is removed from the ligation product by treatment with factor Xa. This exposes a N-terminal cysteine residue which is then reacted with the recombinant thioester fragment, Abl-[G[120]]SH3-ethyl α thioester,[13] in a second native chemical ligation step.

a synthetic peptide model system.[6] Key to this sequential ligation process was the reversible N^α-protection of the cysteine residue in the central peptide which allows the two ligation reactions to be performed in a controlled, directed fashion.[7] However, the base-labile 2-(methylsulfonyl)methyloxycarbonyl (Msc) N^α-protecting group used in these model studies was deemed to be incompatible with the use of recombinant proteins due to the harsh conditions required for its removal (pH 13, for 1–2 min). To resolve this problem, a mild enzymatic N^α-protection strategy was employed which involved appending the short propeptide, RGKIEGR, to the N-terminal Cys of the central peptide. Model studies indicated that this peptide-protecting group prevented self-ligation yet could be cleanly removed at pH 7.5 by treatment with the protease factor Xa.[8] Note, this enzymatic N^α-protection strategy is compatible with both the chemical synthesis and biosynthesis of peptides and proteins.

The applicability of our biosynthetic insertion approach was demonstrated by the regioselective incorporation of the fluorescent peptide sequence, CK(Dns)G, between the SH3 and SH2 domains of Abl (Figure 1). The two SH domains of Abl are adjacent in the full length protein and are critical for its regulation.[9] We anticipated that introduction of the environmentally sensitive dansyl group between these two regulatory domains would allow any changes that occur in their relative motions upon ligand binding to be monitored spectrofluorometrically (Figure 2a).[10]

In the first step of our insertion process, the protected synthetic peptide insert, RGKIEGR-CK(Dns)G-propionamide α thioester (**2**),[11] was chemoselectively ligated to the N-terminal cysteine-containing

(6) Camarero, J. A.; Cotton, G. J.; Adeva, A.; Muir, T. W. *J. Pept. Res.* **1998**, *51*, 303–316.

(7) Without N^α protection, a peptide containing both an N-terminal cysteine residue and an α thioester group undergoes uncontrolled intra- and inter-molecular self-ligation. See ref 6 and (a) Camarero, J. A.; Muir, T. W. *J. Chem. Soc., Chem. Commun.* **1997**, 1369–1370. (b) Zhang, L.; Tam, J. P. *J. Am. Chem. Soc.* **1997**, *119*, 2363–2370. (c) Camarero, J. A.; Pavel, J.; Muir, T. W. *Angew. Chem., Int. Ed.* **1998**, 37, 347–349.

(8) Previous studies have shown that recombinant N-terminal cysteine proteins can be generated by cleavage of an appropriate fusion precursor with the protease factor Xa: Erlanson, D. A.; Chytil, M.; Verdine, G. L. *Chem. Biol.* **1996**, *3*, 981–991.

(9) (a) Mayer, B. J.; Baltimore, D. *Mol. Cell. Biol.* **1994**, *14*, 2883–2894. (b) Wang, J. Y. J. *Curr. Opin. Gen. Dev.* **1993**, 35–43. (c) Pendergast, A. M. *Curr. Opin. Cell. Biol.* **1996**, *8*, 174–181.

Figure 2. (a) Anticipated change in the dansyl emission spectra of Abl-SH3−CK(Dns)G-SH2 upon binding the consolidated ligand 3BP2-2BP1.[16,17] Binding of the bidentate ligand constrains the interdomain motions which is expected to perturb the microenvironment of the fluorophore. (b) Fluorescence emission spectrum of the final ligation product (**5**), indicating the presence of the dansyl group within the construct. Excitation = 340 nm. (c) Change in the emission intensity (background corrected) of the construct at 530 nm on addition of increasing concentrations of the Abl-SH(32) consolidated ligand 3BP2-2BP1. All experiments were conducted in 10 mM DTT, 150 mM GdmCl, 140 mM NaCl, 25 mM phosphate pH 7.2, with excitation at 340 nm.

recombinant protein, Abl-[C121]SH2 (**1**).[12] This was achieved by simply stirring both reactants in 6 M GdmCl, 140 mM NaCl, 200 mM phosphate buffer pH 7.2 containing 1.5% v/v of both thiophenol and benzylmercaptan. After 96 h, the desired ligation product RGKIEGR-CK(Dns)G-Abl-[C121]SH2 (**4**) was purified by reverse phase HPLC and concentrated into 0.03 M GdmCl, 100 mM NaCl, 5 mM phosphate buffer pH 7.5. The leader sequence was then cleanly removed by overnight treatment with factor Xa at 25 °C. The cleavage product, CK(Dns)G-Abl-[C121]-SH2, thus contained the requisite free N-terminal cysteine residue necessary for the second ligation reaction. This final step was facilitated *in situ* by concentrating the factor Xa cleavage mix into the ligation buffer (6 M GdmCl, 140 mM NaCl, 200 mM phosphate buffer pH 7.2 containing 1.5% thiophenol and benzylmercaptan) and then directly adding this to a molar excess of purified recombinant Abl-(G120)SH3-ethyl$^\alpha$thioester (**3**).[13] After the mixture stirred for 4 days at room temperature, the desired insertion product, [Abl-(G120)SH3]-CK(Dns)G-[(C121)SH2] (**5**), was purified by HPLC and characterized by fluorescence spectroscopy (Figure 2b) and electrospray mass spectrometry.[14] Digestion of this protein with the enzyme Lys-C produced the expected fragment corresponding to residues SH3(106−120)-CK-(Dns)-SH2(121−124) of the construct, confirming that the dansyl-

peptide has been inserted between the two recombinant domains.

The purified, refolded insertion product **5** was used to investigate whether ligands for Abl-SH(32) induce a significant fluorescence change upon binding.[15] Known peptide ligands for Abl-SH2, Abl-SH3, and Abl-SH(32) were titrated against construct **5** and the fluorescence emission intensity was monitored at 530 nm. As illustrated in Figure 2c, addition of the consolidated ligand 3BP2-2BP1,[16] a branched peptide which coordinates both domains of Abl-SH(32),[17] causes a sizable increase in the fluorescence intensity. The data fits a two-state binding isotherm, yielding a K_d of 0.123 ± 0.017 μM, a value consistent with that previously reported for recombinant Abl-SH(32).[17] Ligands for the individual SH2 and SH3 domains, 2BP1 and 3BP2 respectively, also produced the expected binding isotherms (data not shown), although the maximum fluorescence change was significantly smaller (∼4% at saturation) as compared to the bidentate ligand. The fluorescence data obtained using construct **5** is consistent with a model whereby the relative motions of the SH3 and SH2 domains are constrained when the protein is complexed to the bidentate ligand.[18] Importantly, at low concentrations of ligand (e.g., 0.5 μM or below) the change in fluorescence is far greater (at least 6−10-fold) for the bidentate interaction compared to its monovalent counterparts. This ability to distinguish between these different interactions at low ligand concentrations indicates that the construct **5** is a specific protein-biosensor for high affinity bidentate interactions with the SH(32) region of Abl. In future studies this reagent will be used to identify novel high affinity bidentate ligands for Abl-SH(32), through the *in vitro* screening of combinatorial peptide libraries and to biochemically characterize protein−protein interactions which potentially regulate Abl function.

The peptide-insertion strategy described above is expected to be a generally applicable method for protein engineering, enabling a molecule to be subjected to a wide variety of specific biochemical transformations (including backbone mutations) regardless of its size or primary sequence. All of the steps in the procedure are performed at ca. physiological pH and can be accomplished under native conditions if the final product cannot be refolded from its denatured state. The development of the "proteolytic cleavage" method of N-terminal protection means the insert can, in principle, be another recombinant fragment, permitting the combinatorial recombination of numerous protein domains. In this example the dansyl fluorophore was selectively incorporated between the SH3 and SH2 domains of Abl to generate a biosensor which reports bidentate interactions with this regulatory region of the protein. The ability of such reagents to report specific biological interactions may be extremely useful for investigating molecular recognition processes both *in vitro* and *in vivo*.

(10) (a) Walkup, G. K.; Imperiali, B. *J. Am. Chem. Soc.* **1997**, *119*, 3443−3450. (b) Marvin, J. S.; Hellinga, H. W. *J. Am. Chem. Soc.* **1998**, *120*, 7−11 and references therein.

(11) The peptide was assembled on a S-propionamide-derivatized 4-methylbenzhydrylamine (MBHA) resin[6] according to the in situ neutralization protocols for Boc solid-phase peptide synthesis: Schnölzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. H. *Int. J. Pept. Protein Res.* **1992**, 40, 180−193. The Dns group was incorporated directly as the α-(*t*-Boc)-ε-(4-dimethylaminonaphthalene-1-sulfonyl) derivative of lysine.

(12) The recombinant N-terminal cysteine protein Abl-(C121)SH2 was generated by factor Xa treatment of a precursor GST fusion protein containing the sequence Ile-Glu-Gly-Arg-Cys immediately N-terminal to the SH2 domain (see refs 1k and 8).

(13) Recombinant Abl-(G120)SH3-ethyl$^\alpha$thioester was produced by overnight treatment of the corresponding Abl-SH3-Intein-CBD fusion protein with 3% v/v ethanethiol in 200 mM NaCl and 200 mM phosphate buffer pH 6.0, and the desired product was purified by HPLC. The C-terminal residue N120 was mutated to Gly to improve the kinetics of the ligation reaction (see ref 1k).

(14) Electrospray mass spectrometry; observed mass =18604 ± 0.9 Da, expected (av isotope comp.) = 18602.9 Da. The overall yield for the two ligation steps and factor Xa removal of the N$^\alpha$-protecting group was 28%.

(15) The protein was refolded by rapid dilution (>10-fold) from 6 M GdmCl with 10 mM DTT, 140 mM NaCl, 25 mM phosphate pH 7.2.

(16) Consolidated ligand 3BP2-2BP1 is the branched peptide α-[3BP2]-ε-[2BP1-(Gly)$_6$]-lysine; 3BP2 is PPAYPPPPVP; 2BP1 is PVY*ENV (Y* = phosphotyrosine).

(17) The interaction of the consolidated ligand 3BP2-2BP1 with recombinant Abl-SH(32) has a reported K_d of 0.249 ± 0.005 μM, whereas the individual components, 2BP1 and 3BP2, have K_d values of 2.35 ± 0.34 μM and 10.7 ± 0.3 μM, respectively: Cowburn, D.; Zheng, J.; Xu, Q.; Barany, G. *J. Biol. Chem.* **1995**, *270*, 26738−26741.

(18) Structural studies on fully recombinant Abl-SH(32) indicate that the relative orientation of the SH2 and SH3 domains is not restricted in solution (ref 4b).