Protein Modifications

DOI: 10.1002/anie.200700719

Regioselective Cysteine Bioconjugation by Appending a Labeled Cystein Tag to a Protein by Using Protein Splicing in *trans***

*Thomas Kurpiers and Henning D. Mootz**

The preparation of chemically modified proteins is key for many aspects of basic research and applied technologies. For example, biophysical probes can be introduced to study protein structure and function, attached polymers can increase the stability and biocompatibility of pharmaceutically relevant proteins, and tethering to solid supports is exploited in the design of protein chips and immobilized enzyme catalysts.[1,2]

Bioconjugate techniques take advantage of the reactivity of functional groups in proteins and provide the easiest and probably most widely used route to chemical modification of proteins.[3] Cysteine can be considered as the most important target for bioconjugation among the proteinogenic amino acids; this can be explained by the unique chemical reactivity of the sulfhydryl moiety and the low abundance of cysteines in proteins. The cysteine side chain facilitates chemoselective modification with a variety of functional groups, including, for example, maleimides, haloacetamides, and thiols.[3] Conveniently, a large number of synthetic probes linked to these functional moieties are commercially available. However, cysteine modification is strictly regioselective only in proteins with a single cysteine residue. Thus, the approach can be severely limited or even impractical for a protein of interest (POI) that contains an essential or multiple cysteines. We hypothesized that the problem of regioselectivity can be circumvented when the peptide backbone of the POI is assembled from two parts by protein *trans*-splicing using a split intein[4,5] and prior incubation of only one part with the cysteine reactive probe (Figure 1). Cysteines in the other part of the POI would then remain unaffected. Another advantageous feature of this general concept would result from the specific interaction between the two split intein halves that should facilitate the protein *trans*-splicing reaction in complex mixtures such as a cell lysate or a living cell.

Our experimental outline to establish regioselective cysteine modification included splicing a short peptide tag with a labeled cysteine (Cys-tag) to a POI in order to attach

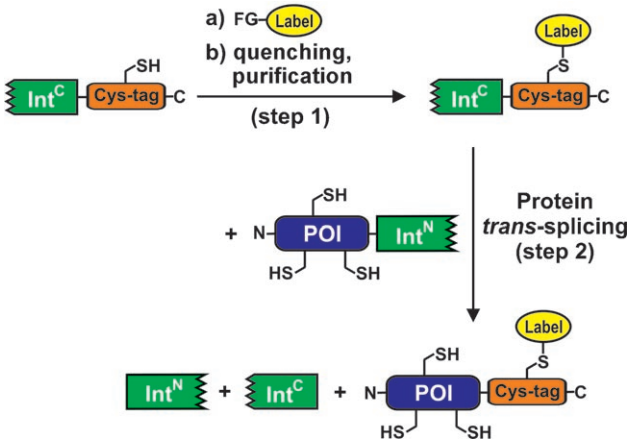

Figure 1. Principle of the approach. A peptide sequence with a single cysteine (Cys-tag) is expressed in fusion with the C-terminal fragment of a split intein and chemically modified. In a second step the labeled Cys-tag is linked to a protein of interest (POI) by protein *trans*-splicing. FG = functional group, for example, haloacetamide, maleimide.

various probes to its C terminus (Figure 1). One prerequisite for this strategy is the absence of a cysteine in the respective intein fragment. However, most inteins, including the well-characterized natively split *Ssp* DnaE intein,[6] as well as the artificially split *Sce* VMA[7] and *Mtu* RecA inteins,[8] employ a catalytically active cysteine residue at each splice junction. These cysteines are involved in the peptide-bond breaking and forming steps and thus preclude the use of the respective inteins for the envisioned approach. Very recently, we have described the artificially split *Ssp* DnaB intein for biochemical applications using purified proteins.[9] This intein has a serine at the C-terminal splice junction and also the remaining sequence of the C-terminal fragment ($DnaB^C$) is free of cysteines, making it the ideal candidate for our approach.[10] Furthermore, the DnaB intein is the only engineered intein described so far that spontaneously associates into its active form from the $intein^N$ and $intein^C$ fragments under native conditions without the requirement for a renaturation step.[9]

We introduced the cysteine of the Cys-tag peptide by site-directed mutagenesis into the extein sequence of the previously described construct MBP-$DnaB^C$-His_6 (**4**)[9] to give MBP-$DnaB^C$-Cys-His_6 (**2**; 50.0 kDa, sequence of the C-extein: SIRS**C**GHHHHHH; see Figure 2). The maltose-binding protein (MBP) at the N terminus of these constructs was included to increase expression yields, while the His_6 tag facilitated simple purification by affinity chromatography. Construct **2** was expressed in *E. coli* and purified from the soluble fraction. Modification of the single cysteine residue was carried out using standard protocols with the labeling reagent 5-(iodoacetamido)fluorescein (5-IAF; see the Sup-

[*] T. Kurpiers, Prof. Dr. H. D. Mootz
Universität Dortmund, Fachbereich Chemie – Chemische Biologie
Otto-Hahn-Strasse 6, 44227 Dortmund (Germany)
Fax: (+49) 231-755-5159
E-mail: Henning.Mootz@uni-dortmund.de

[**] We are grateful to S. Brenzel for construction of plasmid pSB34 and C. Ludwig for construction of plasmid pCL39. We thank U. Linne for help with mass spectrometry and M. Marahiel for support. We thank F. Bernhard for providing plasmid pAT37. Funding was provided by the DFG (Emmy Noether program) and the Fonds der Chemischen Industrie.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

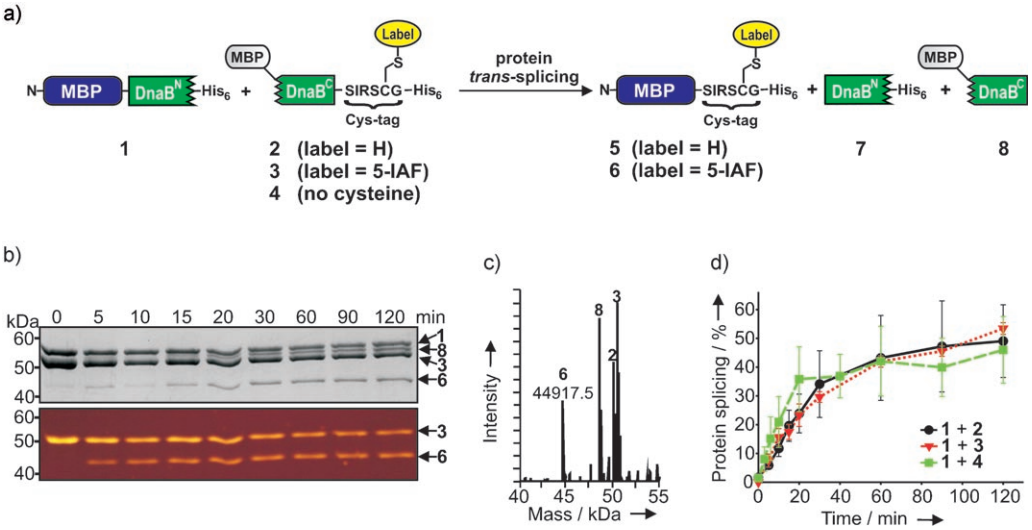

Figure 2. Proof-of-concept for protein *trans*-splicing reactions. a) Schematic illustration of the reactions. b) Coomassie brilliant blue stained (top) or UV-illuminated (bottom) SDS-PAGE gel of reaction between **1** and **3**. By-product **7** cannot be seen on this gel because of its small size (13.1 kDa). c) ESI-TOF mass spectrum of the reaction mixture after 2 h. Observed molecular masses [Da] (calculated values in brackets): **6** 44917.5 (44917.8); **8** 48613.4 (48613.7), **3** 50428.5 (50428.6), **2** 50040.5 (50040.5). d) Time courses of the indicated protein *trans*-splicing reactions. Kinetic data for the reaction of **1** and **4** was taken from.[9]

porting Information). Excess reagent was quenched with 1,4-dithiothreitol (DTT) and optionally separated by gel filtration. Conversion of **2** to the modified protein MBP-DnaB^C-Cys(5-IAF)-His$_6$ (**3**) was monitored by a band shift on a Coomassie brilliant blue stained SDS-PAGE gel and by visualizing the product band with UV light (see the Supporting Information). Stoichiometric addition of a single 5-IAF fluorophore was also confirmed by ESI-TOF mass spectrometry (see Figure 2). From these data the yield of the conversion to the labeled adduct was determined to be $\geq 90\,\%$, as expected for this type of reaction and implying that the single cysteine residue in construct **2** was freely accessible.

Protein *trans*-splicing of modified protein **3** was investigated by incubation with MBP-DnaB^N-His$_6$ (construct **1**, 56.4 kDa), which contained the complementary intein fragment (DnaB^N) and MBP as the POI (Figure 2a). Construct **1** was obtained from the soluble fraction after expression in *E. coli*.[9] Figure 2b shows that the protein *trans*-splicing reaction at 25 °C and a concentration of 2 μM for each protein proceeded as expected: the time-dependent appearance of the band on the SDS-PAGE gel corresponded to the splice product MBP-Cys(5-IAF)-His$_6$ (**6**). Also, the by-products **7** and **8** could be identified (**7** is not visible on the gels in Figure 2b owing to its size of 13 kDa). Analysis of a reaction mixture by ESI-TOF-MS further confirmed the identity of all predicted splice products (Figure 2c). Protein *trans*-splicing proceeded to a yield of about 40–50 % after 2 h, which was very similar to the control reactions with parent construct **4** (Figure 2d) and is often observed for split inteins reconstituted in vitro.[6,9] Also the control reaction of unmodified Cys-tag construct **2** with **1** was virtually indistinguishable (Figure 2d), which suggests that neither the fluorophore-labeled nor the free cysteine itself in close proximity to the active site nucleophile of the intein had an effect on its activity. In accordance with the model that intein fragment association is fast compared to the rate-determining step of intein folding and/or protein *trans*-splicing,[11] the reaction kinetics could be fitted to first-order functions with rate constants of $(5.5 \pm 0.4) \times 10^{-4}\,s^{-1}$ (**1** + **2**) and $(4.7 \pm 0.4) \times 10^{-4}\,s^{-1}$ (**1** + **3**). This compares well with the previously reported rate constant of $(9.9 \pm 0.8) \times 10^{-4}\,s^{-1}$ for (**1** + **4**).[9]

To study the generality of our approach with regard to the attached chemical modification, we tested various labeling reagents. Similar results were obtained when **2** was modified with maleimide-coupled fluorescein, Alexa Fluor 546, Texas Red, and with biotin (see the Supporting Information). This indicates that the intein's activity is largely independent of the nature of the attached label and that most, if not all, such groups can be regioselectively transferred to the C terminus of a POI.

We then turned to investigate the scope of our approach in terms of the nature of the POI. One strength of the approach is that proteins containing multiple cysteine residues can be labeled regioselectively with conventional Cys-reactive probes. Therefore, the proteins β-lactamase (Bla) and thioredoxin (Trx) were chosen as N-exteins. β-Lactamase contains two cysteines that form a disulfide bridge[12] and that would react with sulfhydryl-directed probes under the reductive conditions of the bioconjugation procedure. Thioredoxin employs two cysteine residues for its catalytic cycle.[13] Both proteins were prepared as fusion proteins with the N-terminal intein fragment to give constructs Bla-DnaB^N-His$_6$ (**9**, 43.5 kDa; see Figure 3) and StreptagII-Trx-DnaB^N-His$_6$ (26.8 kDa, see the Supporting Information). For the following protein *trans*-splicing experiments we also developed a simplified purification strategy for the isolation of the splice product from the reaction mixture. To this end, the new construct His$_{10}$-gpD-DnaB^C-Cys (**10**, 19.3 kDa) with the C-terminal intein fragment was made (Figure 3a), in which the C-terminal His-tag was moved to the N terminus of the

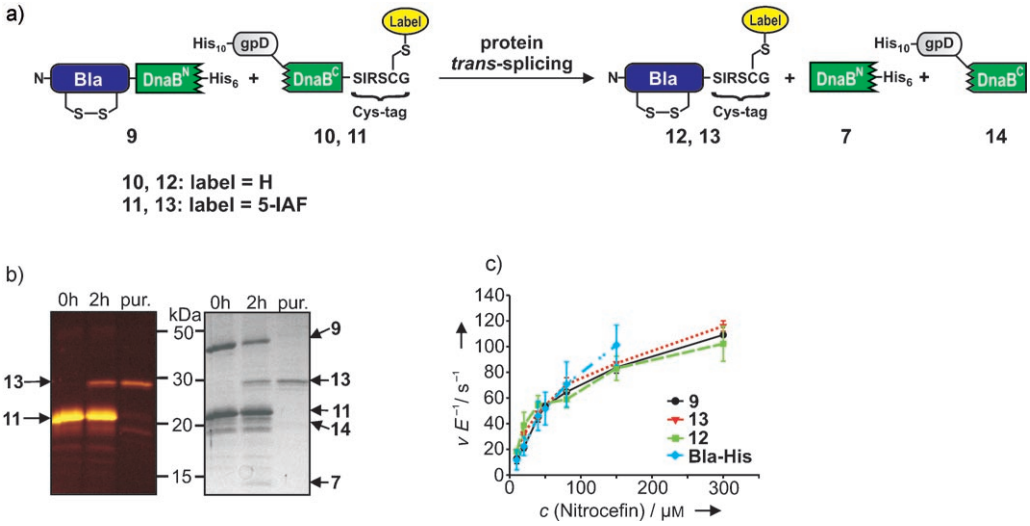

Figure 3. Preparation of regioselectively fluorescein-labeled and catalytically active β-lactamase (Bla). a) Reaction scheme. b) Analysis of the protein *trans*-splicing reaction between **9** and **11** and the purified splice product **13** on a Coomassie brilliant blue stained (right) or UV-illuminated (left) SDS-PAGE gel. Purified **13** was obtained in the flow-through of Ni^{2+}–NTA purification. c) Enzymatic activity of labeled protein **13** monitored by nitrocefin hydrolysis and compared in control experiments to unlabeled protein **12**, fusion protein **9**, and recombinant control protein Bla-His_6; the Michaelis–Menten plots are shown. E = standardized enzyme concentration.

protein. As a consequence, the desired splice product was the only protein in the reaction mixture without a His-tag and could be collected from the flow-through of a Ni^{2+}–NTA affinity column. Additionally, MBP was substituted by the smaller bacteriophage lambda head protein gpD.[14] The resulting construct **10** harbored the cysteine as the second to last amino acid in its Cys-tag sequence SIRS**C**G. Construct **10** could be obtained as a highly soluble protein from expression in *E. coli* and could be efficiently labeled with various reagents at its single cysteine residue.

For the preparation of C-terminally modified β-lactamase, construct **9** (8 μM concentration) was mixed at a ratio of about 1:2 with the fluorescein-modified construct His_{10}-gpD-$DnaB^C$-Cys(5-IAF) (**11**). The mixture was then incubated for 2 h at 25 °C. Figure 3b shows that the splice product Bla-Cys(5-IAF) (**13**, 31.2 kDa) was formed with a yield of about 40 % with regard to starting material **9**. To isolate purified splice product **13**, a reaction at larger scale (1 mL total volume) was applied onto a Ni^{2+}-NTA column and pure fractions of **13** were obtained in the flow-through (right lanes in Figure 3b). Unlabeled splice product Bla-Cys (**12**) was prepared as a control in the same manner using unmodified construct **10** in the protein *trans*-splicing reaction. β-Lactamase activity of the pure splice products **12** and **13** was assayed photometrically by nitrocefin hydrolysis:[12] $DnaB^N$-fusion protein **9** and recombinant Bla-His_6[15] were included as further controls. The k_{cat}/K_M values obtained from the measurements shown in Figure 3c were very similar: 1.6 $s^{-1}\,\mu M^{-1}$ (**13**), 2.3 $s^{-1}\,\mu M^{-1}$ (**12**), 1.4 $s^{-1}\,\mu M^{-1}$ (**9**), and 1.3 $s^{-1}\,\mu M^{-1}$ (Bla-His_6).

Likewise, preparation of thioredoxin regioselectively modified with fluorescein at its C terminus using the same procedure of protein *trans*-splicing and purification yielded a catalytically active protein (see the Supporting Information). These results are consistent with the notion that expression as an $intein^N$ fragment fusion protein and subsequent protein *trans*-splicing does not affect the integrity of the POI and therefore proved the utility of the approach.

Another advantage of our method results from the selectivity of the protein *trans*-splicing reaction mediated by the complementary split intein fragments. Only proteins or polypeptides expressed in fusion with complementary intein fragments can react with each other, as has been shown by co-expression[10] and protein transduction of one construct into cells expressing the corresponding partner protein.[16] To test if the described split DnaB intein constructs fulfilled the criteria of stability and selectivity for efficient modification of a POI in a cell extract, fluorescein-labeled **3** (10 μM) was added in an approximately equimolar ratio to lysed *E. coli* cells that had either expressed or not expressed the counterpart protein **1**. The reaction was stopped after 0 h or 1 h by adding SDS-PAGE loading buffer and analyzed by SDS-PAGE (see Figure 4). Even in the presence of a very high amount of *E. coli* proteins as potential unspecific interaction partners, **3** was not degraded and did not react with any other proteins in the absence of **1**. Only when expressed protein **1** was present in the cell lysate, was the expected fluorescein-labeled splice product **6** was formed specifically in very good yields of more than 80 %. The integrity of the splice product **6** was further confirmed by an analytical Ni^{2+}–NTA affinity purification, by which the other His-tag-containing proteins **1** and **3** were co-purified (Figure 4). Thus, our approach can also be used to selectively label proteins in cell lysates. This could be useful to either bypass one purification step of the POI or to study labeled proteins directly in a cell lysate or other complex mixtures.

We have shown here that a short, prelabeled Cys-tag can be spliced to a POI. Alternatively, splitting the POI at an appropriate site will also allow a cysteine residue located further in the interior of the protein with respect to the

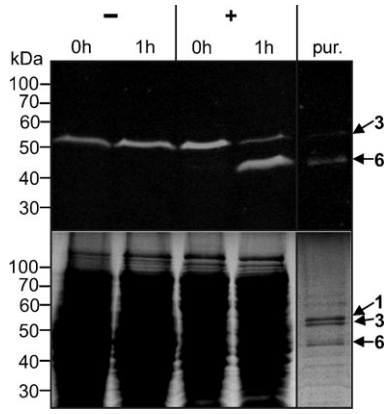

Figure 4. Protein labeling in a cell lysate. Fluorescein-labeled Cys-tag reagent **3** was added to lysates of *E. coli* cells either expressing (+) or not expressing (−) construct **1**. The Coomassie brilliant blue stained (bottom) and UV-illuminated (top) SDS-PAGE gel is shown. The labeled splice product **6** could be purified by analytical Ni^{2+}–NTA chromatography (see right lane). Under these conditions all other proteins containing a His-tag are co-purified as indicated.

primary sequence to be addressed. This generally applicable technique overcomes previous limitations of cysteine bioconjugation. It will be useful for biochemical investigations and manipulations in which regioselective attachment of a chemical moiety to a protein containing more than one cysteine is required. Another specific application would be to incorporate a second chemical group to another cysteine in the protein after the protein splicing step, for example to prepare a doubly labeled protein for intramolecular FRET measurements. While we have demonstrated the feasibility of this method for C-terminally modified proteins, it could also be configured in a way to obtain proteins labeled at their N terminus. Suitable split inteins would require a non-cysteine residue at the first position of the intein and ideally no further cysteines in the N-terminal fragment (the *Ssp* DnaB intein used in this study has a cysteine at this position). For example, the artificially split *Psp* Pol-1 intein[17] and the natively split *Neq* Pol intein[18] are suitable candidates because they operate with serine and threonine residues at both splice junctions.

We believe that our approach provides a valuable alternative to native chemical ligation (NCL) and expressed protein ligation (EPL) for many applications.[19,20] Most importantly, the requirement for the thioester and cysteine groups at the termini is avoided, and the reactions can be carried out at low micromolar protein concentrations. Furthermore, the attachment of a chemical label to a protein will be easier and faster in many cases and for many laboratories because of the wealth of well-established and commercially available cysteine bioconjugation reagents. The potential for performing the labeling reaction in complex protein mixtures further expands the scope of this technique with possible applications in protein purification and, for example, chemical cross-linking in cell extracts.

Received: February 15, 2007
Published online: May 30, 2007

Keywords: bioconjugation · fluorophores · intein · protein modifications · protein splicing

[1] L. Wang, P. G. Schultz, *Angew. Chem.* **2005**, *117*, 34; *Angew. Chem. Int. Ed.* **2005**, *44*, 34.
[2] M. E. Hahn, T. W. Muir, *Trends Biochem. Sci.* **2005**, *30*, 26.
[3] G. T. Hermanson, *Bioconjugate Techniques*, Academic Press, London, **1996**.
[4] C. J. Noren, J. Wang, F. B. Perler, *Angew. Chem.* **2000**, *112*, 458; *Angew. Chem. Int. Ed.* **2000**, *39*, 450.
[5] H. Paulus, *Annu. Rev. Biochem.* **2000**, *69*, 447.
[6] D. D. Martin, M. Q. Xu, T. C. Evans, Jr., *Biochemistry* **2001**, *40*, 1393.
[7] H. D. Mootz, T. W. Muir, *J. Am. Chem. Soc.* **2002**, *124*, 9044.
[8] K. V. Mills, B. M. Lew, S. Jiang, H. Paulus, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3543.
[9] S. Brenzel, T. Kurpiers, H. D. Mootz, *Biochemistry* **2006**, *45*, 1571.
[10] H. Wu, M. Q. Xu, X. Q. Liu, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **1998**, *1387*, 422.
[11] T. J. T. Evans, M. Q. Xu, *Chem. Rev.* **2002**, *102*, 4869.
[12] C. H. O'Callaghan, A. Morris, S. M. Kirby, A. H. Shingler, *Antimicrob. Agents Chemother.* **1972**, *1*, 283.
[13] A. Holmgren, *J. Biol. Chem.* **1979**, *254*, 9627.
[14] P. Forrer, R. Jaussi, *Gene* **1998**, *224*, 45.
[15] C. Ludwig, M. Pfeiff, U. Linne, H. D. Mootz, *Angew. Chem.* **2006**, *118*, 5343; *Angew. Chem. Int. Ed.* **2006**, *45*, 5218.
[16] I. Giriat, T. W. Muir, *J. Am. Chem. Soc.* **2003**, *125*, 7180.
[17] M. W. Southworth, E. Adam, D. Panne, R. Byer, R. Kautz, F. B. Perler, *EMBO J.* **1998**, *17*, 918.
[18] J. J. Choi, K. H. Nam, B. Min, S. J. Kim, D. Soll, S. T. Kwon, *J. Mol. Biol.* **2006**, *356*, 1093.
[19] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776.
[20] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705.

