DOI: 10.1002/cbic.200800319

Site-Specific Chemical Modification of Proteins with a Prelabelled Cysteine Tag Using the Artificially Split Mxe GyrA Intein

Thomas Kurpiers and Henning D. Mootz*[a]

The selective modification of proteins with a synthetic probe is of central interest for many aspects of protein chemistry. We have recently reported a new approach in which a short cysteine-containing tag (CysTag) fused to one part of a split intein is first modified with a sulfhydryl-reactive probe. In a second step, protein trans-splicing is used to link the labelled CysTag to a target protein that has been expressed in fusion with the complementary split intein fragment. Here, we present the generation and biochemical characterisation of the artificially split Mycobacterium xenopi *GyrA intein. We show that this split intein is active without a renaturation step and that it provides a significant improvement for the CysTag protein-labelling approach in terms of product yields and target protein tolerance. Two proteins with multiple cysteine residues, human growth hormone and a multidomain nonribosomal peptide synthetase, were site-specifically modified with high yields. Our approach combines the benefits of the plethora of commercially available cysteine-reactive probes with a straightforward route for their site-specific incorporation even into complex and cysteine-rich proteins.*

Introduction

The chemo- and regioselective introduction of synthetic moieties into proteins is of fundamental importance in basic protein research as well as in many medicinal and biotechnological applications. For example, appropriately placed biophysical probes, such as fluorophores and other spectroscopic labels, allow for the study of protein–protein interactions, protein folding and protein dynamics. Attachment to polymers or a solid phase can improve the properties of protein-based therapeutics and catalysts. Several approaches have been developed to prepare chemically modified proteins (for reviews see refs. [1]–[4]). Each of these techniques has certain advantages and limitations with respect to the nature and number of synthetic moieties, specificity and yield of their incorporation, and experimental complexity and cost. Another important criterion is the invasiveness of the method; that is, minimal addition of extra residues or auxiliary chemical moieties is desired to preserve the functional integrity of the protein under investigation.

Important recent techniques for protein modification include reprogramming of the cellular biosynthetic machinery,[1] chemical ligation reactions like native chemical ligation (NCL),[2] and expressed protein ligation (EPL)[3] as well as the enzymatic modification of fused proteins or peptide tags.[4] These techniques usually have high selectivity, but might require complex experimental setups or suffer from a limited choice of synthetic building blocks. In contrast, traditional bioconjugation approaches take advantage of the reactivity of functional groups, such as amino acid side chains within proteins.[5,6] Lysine and cysteine are most widely used for chemical modification. Several chemoselective reactions are known to address the amino or thiol moiety under mild conditions. Conveniently, a large number of useful bioconjugation reagents, for example, *N*-hydroxysuccinimide and maleimide or haloacetamide derivates, have been reported over the years and are commercially available. However, a general problem for reactions on amino acid side chains is their lack of regioselectivity. When more than one of the targeted amino acid is present in the protein of interest (POI), it is challenging or impossible to achieve a reaction exclusively at the desired residue. Differences in reactivity can stem from the electronic environment within the protein structure or spatial accessibility. Given the unique reactivity of the cysteine side chain, its selectivity towards certain functional groups, and its low abundance in proteins, this residue is in most cases the best choice for chemo- and regioselective protein modification. Although additional cysteines can be eliminated by site-directed mutagenesis, this is often undesirable and inapplicable to proteins that contain essential cysteines.

To expand the experimentally facile approach of cysteine bioconjugation to proteins that contain multiple and essential cysteine residues, we have recently reported a new general technique that circumvents the regioselectivity problem by incorporating a protein trans-splicing step in the bioconjugation procedure.[7] Our approach involves first the bioconjugation of a unique cysteine in a peptide sequence fused to an auxiliary split intein (for reviews on inteins and protein splicing see ref. [8]–[11]). In a second step, this cysteine-labelled auxiliary protein is incubated with the POI fused to the complementary

[a] *T. Kurpiers, Prof. Dr. H. D. Mootz*
Technische Universität Dortmund, Fakultät Chemie-Chemische Biologie
Otto-Hahn-Strasse 6, 44229 Dortmund (Germany)
Fax: (+49) 231-755-5159
E-mail: henning.mootz@tu-dortmund.de

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

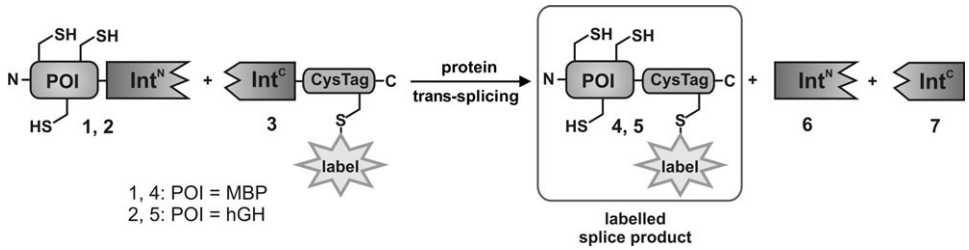

Figure 1. Site-specific chemical modification of proteins. The principle of the CysTag labelling approach is shown. A prelabelled peptide sequence with a unique cysteine (CysTag) is linked to a POI through protein trans-splicing. This reaction is catalysed by N- and C-terminal intein fragments (Int^N and Int^C). CysTag labelling is performed prior to the protein splicing reaction such that additional cysteines in the POI are not affected. For reasons of clarity, additional fusions of FKBP–His_6 to the C terminus of Int^N and of MBP to the N terminus of Int^C are omitted in this illustration.

split intein fragment (Figure 1). The chemically modified peptide sequence is linked to the POI through a peptide bond by protein trans-splicing. In this way, a labelled cysteine tag (CysTag) can be regio- and chemoselectively attached to a POI. This technique comprises some unique features: 1) native cysteines in the POI are not affected; 2) low concentrations of the POI–intein fusion protein are sufficient for effective labelling due to the affinity of the split intein pieces; 3) the protein trans-splicing reaction is highly selective and enables specific labelling of crude proteins in complex mixtures, such as cell lysates.[7] For this approach to be successful, the split intein auxiliary itself must be free of cysteine residues. This is especially noteworthy, since the best characterised split intein, the natively split Ssp DnaE intein from *Synechocystis* spp. PCC6803, has a catalytic cysteine in each half of the intein.[12] We chose the artificially split Ssp DnaB intein,[7] which contains a serine residue at the (+1) position of the C-terminal intein fragment (Int^C). This artificially split intein has the additional advantage that the two intein fragments are spontaneously active in protein trans-splicing when mixed under native conditions,[13] without the need for a renaturation step, in contrast to all other artificially split inteins reported so far.[14,15] However, one would expect the individual fragments of artificially split proteins to display more unfolded regions and show a higher tendency towards insolubility, which are potential disadvantages. Even though the intein fold seems to be quite tolerant to this kind of backbone engineering, these considerations might also apply to artificially split inteins. In fact, when fused to more delicate proteins of interest, the split DnaB intein described in our first report was found to either affect the solubility of the fusion protein or to show reduced activity in protein trans-splicing. Therefore, we were interested in identifying a more robust split intein for the CysTag labelling approach.

In this work, we report the generation, biochemical characterisation and protein modification ability of a new artificially split intein based on the *Mycobacterium xenopi* gyrase subunit A (Mxe GyrA) intein. The split Mxe GyrA intein was found to be highly active in protein trans-splicing without the requirement for a renaturation step. It proved beneficial in terms of solubility and robustness in protein trans-splicing when fused to various proteins of interest. We report the intein-mediated regioselective cysteine modification of the protein therapeutic human growth hormone (hGH) and of a 123 kDa multidomain nonribosomal peptide synthetase. Thus, in combination with cysteine-selective labelling reagents, the presented split intein provides a general and easy-to-perform approach for the chemo- and regioselective chemical modification of proteins.

Results and Discussion

Rationale for the choice of splitting the Mxe GyrA intein

In order to create a new artificially split intein with improved properties we hypothesised that artificially split native mini-inteins should prove beneficial in terms of solubility and expression levels. Native mini-inteins are relatively rare, because most inteins ("maxi-inteins") contain an inserted homing endonuclease domain. This domain can be excised on the DNA level to give an artificial mini-intein, as reported, for example, for the Mtu RecA,[16] Sce VMA[17,18] and Ssp DnaB inteins.[19] Compared to a native mini-intein, exposed hydrophobic patches and constraints from the introduced linker sequence might arise at the former intein–endonuclease junction. We therefore turned to the native mini-intein of the *M. xenopi* gyrase A (Mxe GyrA). Here, the N- and C-terminal intein regions (Int^N and Int^C, respectively) are connected by a short flexible linker region (see Figure 2 and Figure S1 in the Supporting Information for an illustration of the three-dimensional structure of the Mxe GyrA intein).[20] Earlier work showed that an insertion in this unstructured flexible loop region was tolerated.[21] Importantly, the C-terminal nucleophile essential for the protein splicing pathway is a threonine residue and thus qualifies the GyrA intein for our cysteine-labelling approach outlined above. Furthermore, a number of beneficial properties were reported for the GyrA intein. It is part of a commercial vector used in numerous laboratories for the tag-free purification of proteins and preparation of protein thioesters.[22] Additionally, the GyrA intein exhibits great promiscuity towards the N-extein sequence, and intein fusion proteins can be refolded from inclusion bodies.[23]

The artificially split GyrA intein is robust in protein trans-splicing with purified proteins

To generate the two intein halves necessary for protein trans-splicing we split the GyrA intein between residues Arg119 and Gly120 on the DNA level (Figure 2), and thereby created a 119 amino acid (aa) N-terminal intein fragment (Int^N) and a 79 aa C-terminal intein fragment (Int^C). The first two adjacent residues of the natural N-terminal extein sequence (positions −1 and −2) and the three residues of the natural C-terminal extein sequence (positions +1 to +3) were retained in our

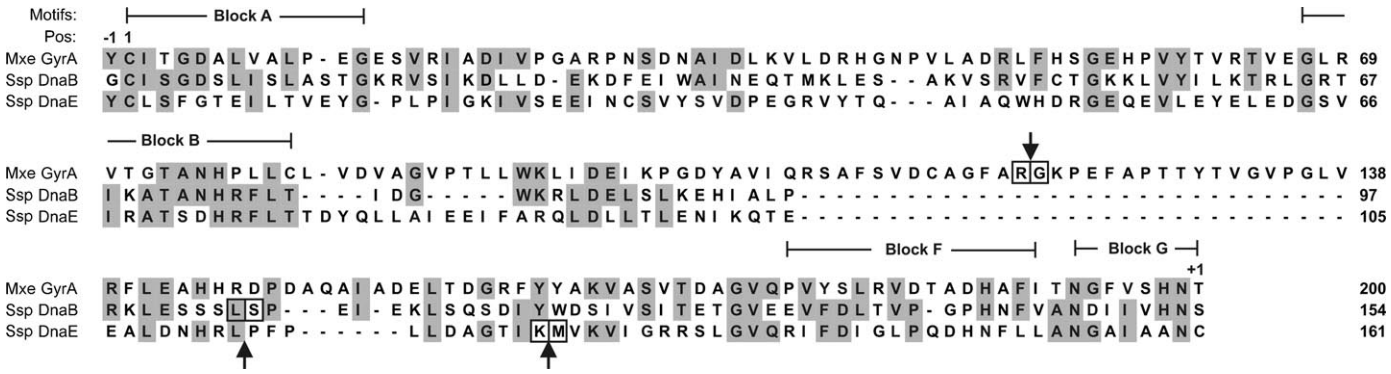

Figure 2. Sequence alignment of split inteins. Shown are the sequences of the herein described Mxe GyrA intein, the artificially split Ssp DnaB intein without the endonuclease domain and the naturally occurring split Ssp DnaE intein. Conserved sequence motifs are indicated and split positions are highlighted by arrows and boxes.

constructs. The resulting Int^C is free of cysteines. The single cysteine for protein labelling was added in a short peptide sequence to the C-terminal extein (Ext^C) sequence (T[+1]EAGS**C**S; CysTag). To introduce handles for affinity purification and extein sequences to follow the splicing reaction, the Int^N and Int^C halves were each fused on the DNA level to maltose binding protein (MBP) and a hexa-histidine tag (His_6) at their N and C termini, respectively. Additionally, to prevent the larger Int^N fragment from potential aggregation and minimize exposure of hydrophobic regions, which might cause misfolding, a FKBP domain was added at the C terminus. The initial constructs thus comprised the sequences MBP-$GyrA^N$-FKBP-His_6 (**1**) and MBP-$GyrA^C$-CysTag-His_6 (**3**; Figure 1).These proteins were produced in *Escherichia coli* and purified from the soluble fraction as described in the Experimental Section. Protein splicing activity was analysed by mixing the complementary proteins at equimolar concentrations and monitoring the formation of the reaction products by SDS-PAGE. Figure 3A shows that the expected splice product band (**4**) and the bands that correspond to the other excised intein fragments **6** ($GyrA^N$-FKBP-His_6, 27.2 kDa) and **7** (MBP-$GyrA^C$, 51.9 kDa) of the reaction were observed and increased in intensity over time. Interestingly, an additional protein band was observed at ~100 kDa (Figure 3A, marked with "#"), which occurred almost immediately after both proteins were mixed, reached a maximum after 30 min and then disappeared completely after 2 h. Because of its transient nature, we assume that this band corresponds to a branched intermediate (calcd mass = 96.6 kDa; see Figure S2 for a schematic illustration of the mechanism of protein trans-splicing and branched-intermediate formation). This interpretation is consistent with the detection of this band by an anti-His antibody (data not shown), the incorporation of the fluorescent signal when a fluorophore-labelled CysTag was used in the splicing reaction (see below) and its stability against boiling for 10 min in denaturing SDS-PAGE loading buffer, which argued against a noncovalent complex of the intein fragments. Furthermore, after excision of the protein band from the gel, tryptic digestion, and MALDI-TOF MS analysis, we identified several tryptic fragments from all parts of the two proteins **1** and **3** except for those that corresponded to the $GyrA^N$ and FKBP fragments (data not shown). The latter fragments would be present in a noncovalent complex but not in the branched intermediate. A branched intermediate was previously detected for the Psp Pol intein from *Pyrococcus* sp. GB-D.[24]

Densitometric analysis of the splice reaction revealed that splice product **4** was formed with a yield of up to 70% (Figure 3B). When one of the proteins was used in more than twofold excess, almost complete consumption of the complementary protein was observed (data not shown). The splice reaction was essentially complete after 4 h. Because of the formation of the intermediate we could not fit the data to simple first order kinetics. The temperature profile of the reaction shows that the highest yields were obtained between 20–30 °C, while significant activity was also observed at temperatures as low as 12 °C (Figure 3C). Even at 4 °C almost the same high yields could be obtained when the reaction was allowed to proceed for more than 24 h (data not shown).

A comparison to the previously reported split Ssp DnaB intein revealed the superior properties of the GyrA intein: although the splice product is formed faster in case of the DnaB intein, the product yields reached only 40–50% (Figure 3B). The temperature profile of both inteins is similar, but for the GyrA intein higher yields were generally obtained (Figure 3C). Moreover, we found that the split GyrA and DnaB inteins are orthogonal. No product formation was observed when the split intein constructs were incubated in either of the two heterologous combinations, but both pair-wise splice products were identified when all four constructs were combined (Figure S3). These findings are particularly interesting with respect to the overall charges of the Int^N and Int^C fragments. For the natively split Ssp DnaE intein electrostatic attractions were suggested to be important for fast association.[25] A similar explanation could hold true for the artificially split DnaB intein. Isoelectric points (pI) were calculated to be 9.5 and 3.8 for the $DnaB^N$ and $DnaB^C$ fragments, respectively, resulting in opposite charges at pH 7.[13] However, for the $GyrA^N$ and $GyrA^C$ intein fragments pI values of 5.5 and 5.4 were calculated, respectively; this indicates that other parameters, like local charges or hydrophobic interactions, must account for the initial association.

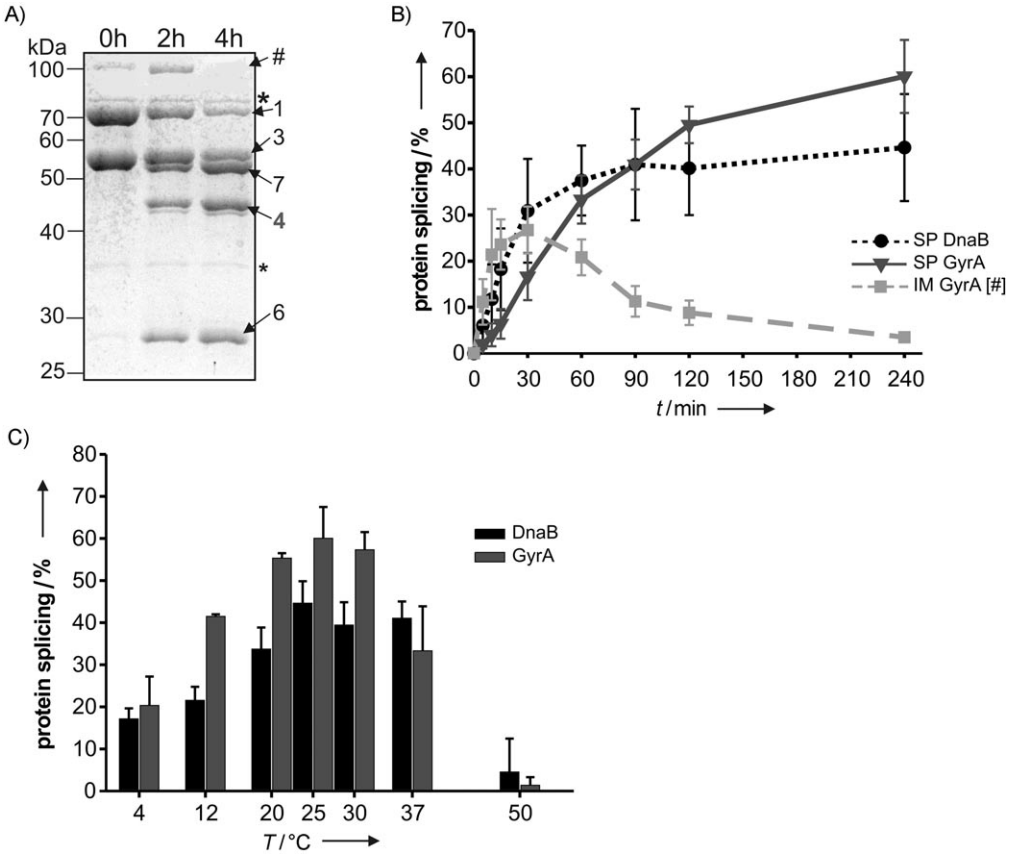

Figure 3. Biochemical characterisation of the artificially split Mxe GyrA intein. A) Analysis of the protein trans-splicing reaction of MBP-GyrA^{N}-FKBP-His$_6$ (protein **1**) with unlabelled MBP–GyrA^{C}–CysTag (protein **3**) on a SDS-PAGE gel stained with Coomassie brilliant blue. The splice product MBP–CysTag (**4**) as well as the other expected products **6** (GyrA^{N}-FKBP-His$_6$) and **7** (MBP–GyrA^{C}) of the reaction are indicated. The band indicated with "#" corresponds to the assumed branched intermediate; "*": protein contaminants B) Time-course of the formation of splice product (SP) and the branched intermediate (IM) determined by densitometric analysis of data as presented in A). For comparison, the respective data for a protein trans-splicing reaction of the split Ssp DnaB intein are shown.[13] C) Temperature dependence of the protein trans-splicing reaction of the split Mxe GyrA intein after 4 h and comparison to the split Ssp DnaB intein.

Exploring the split Mxe GyrA intein for regioselective cysteine bioconjugation of target proteins

We next examined the accessibility of the introduced thiol group in the CysTag sequence of protein **3** to modification reagents and the compatibility of this modification with splicing activity of the split GyrA intein. We labelled protein **3** with various reagents and observed the expected transfer of the labelled CysTag to the Ext^{N} sequence in the splicing reaction with protein **1** (Figure 4A and B for protein **3** labelled with iodoacteamide–fluorescein; **3-Fl**). The yields of the splicing reactions with modified protein **3-Fl** were indistinguishable from those obtained with unmodified protein **3**. In addition to the previously reported protein modifications,[7] we also tested the suitability of our approach for the modification of proteins with polyethylene glycol (PEG) moieties. Protein **3** was treated with a PEG5000 maleimide derivative with a yield of about 75% (Figure 4C). In a splicing reaction with **1**, the pegylated splice product **4-PEG** was observed at a size of 50 kDa along with an almost complete consumption of pegylated starting material (Figure 4D).

Preparation of labelled human growth hormone

To further evaluate the potential of the split GyrA intein for the CysTag protein labelling methodology we focussed on the modification of more challenging protein targets. Human growth hormone (hGH) is a commercially interesting protein therapeutic. It is a single chain polypeptide of 191 aa with four cysteine residues that form two disulfide bridges, which connect Cys53 and Cys165 as well as Cys182 and Cys189.[26,27] These cysteines prohibit the incorporation of synthetic labels through conventional cysteine bioconjugation, because under the reducing conditions required for the reaction they would be also converted into the free thiol form and chemically modified. Our CysTag approach introduces an additional, premodified cysteine and therefore circumvents this problem. Due to the poor solubility of hGH and high expression-level demands, the recombinant protein is usually directed into inclusion bodies when it is expressed in *E. coli*. It was therefore not surprising that a hGH–Int^{N} fusion protein (construct **2**, hGH-GyrA^{N}-FKBP-His$_6$, 50.1 kDa) was found to be insoluble after lysis of the *E. coli* expression host (data not shown). An attempt to refold the protein by renaturation from 8 M urea by

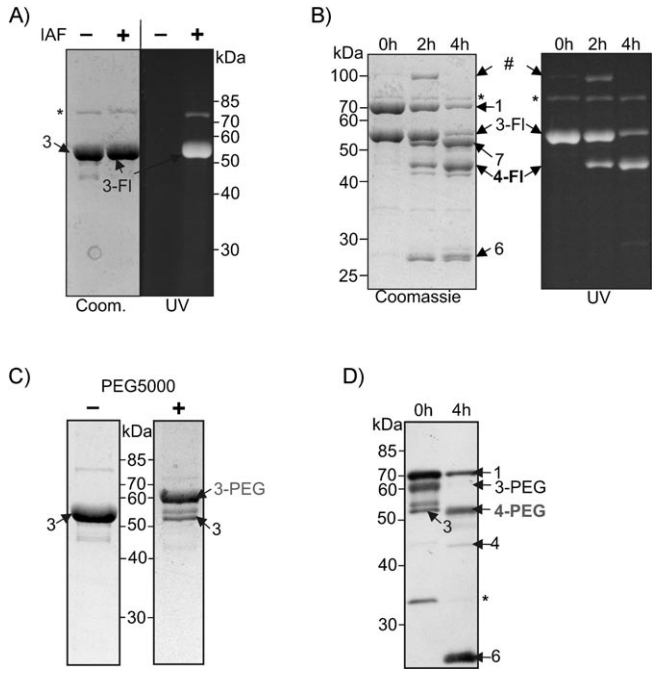

Figure 4. Chemical labelling of the CysTag and transfer by protein trans-splicing by using the artificially split Mxe GyrA intein. A) Modification of the CysTag protein **3** with 5-iodoacetamido-fluorescein (IAF). Shown is the analysis of the protein by SDS-PAGE stained with Coomassie brilliant blue (left, Coom.) and under UV illumination (right). B) Analysis of the protein trans-splicing reaction of protein **1** with **3-FI** is shown on the Coomassie brilliant blue stained (left) and UV illuminated (right) SDS-PAGE gel; "*": denotes a minor protein impurity also modified with fluorescein, that can also be seen in A). C) Modification of the CysTag protein **3** with polyethylene glycol 5000 maleimide (PEG) is visualized on a SDS-PAGE gel. D) Analysis of the protein trans-splicing reaction of protein **1** with **3-PEG** monitored with Western blot analysis by using an anti-His antibody. After 4 h the protein **3-PEG** was completely consumed and splice product MBP-CysTag-PEG5000, **4-PEG**, was formed. A small amount of unlabelled splice product MPB–CysTag (**4**) was also visible as a result of some remaining unlabelled protein **3** in the reaction mixture; "*": minor protein impurity also recognised by the antibody.

using a simple dialysis protocol failed. We therefore added the complementary CysTag intein fragment **3** to construct **2** in 8 M urea before dialysis and subjected the entire reaction mixture to dialysis against protein splicing buffer at 4 °C. After the last buffer exchange, the reaction mixture was incubated for an additional 2 h at 25 °C and then analysed by SDS-PAGE. Figure 5A shows that the expected splice product, hGH–CysTag, was formed under these conditions independently of whether the unmodified Int^C–CysTag protein **3** or fluorescent protein **3-FI** was used. Since the protein hGH–Int^N (**2**) was used in a twofold excess in these reactions, the complementary protein **3** was almost completely consumed in the reaction. To further verify the correct identity of the splice product, the band assigned to the fluorophore-labelled hGH–CysTag (**5-FI**) was excised from the gel, digested with trypsin and analysed by MALDI-TOF mass spectrometry. Figure 5B shows the masses of trypsin digest fragments. Most predicted fragments could be unambiguously identified, together with the fragment that contained the splice junction and the fluorescent label (peptide YTEAGSC(FI)SHHHHHH, $[M+H]^+_{obs}$ = 2026.878 *m/z*, $[M+H]^+_{calcd}$ = 2027.017 *m/z*). Similar experiments with the split DnaB intein were unsuccessful, because only negligible amounts of splice product could be obtained (<10%). Thus, these results demonstrate that poorly soluble proteins that require purification under denaturing conditions can still be submitted to our CysTag approach with the new split GyrA intein in combination with a renaturation protocol.

Preparation of a labelled 123 kDa nonribosomal peptide synthetase (NRPS)

We were also interested in the regioselective labelling of a multidomain nonribosomal peptide synthetase (NRPS). These enzymes produce a huge collection of peptidic secondary metabolites with broad structural diversity and biological activity (for example, cyclosporin and vancomycin) through a protein template mechanism.[28] NRPSs can be chemoenzymatically labelled on their peptidyl-carrier protein (PCP) domain by using a suitable CoA derivative and a 4′-phosphopantetheinyl transferase.[29] However, this procedure blocks the substrate binding site on the 4′-phosphopantetheine prosthetic group (Ppant) and therefore renders the enzyme catalytically inactive. We focused on TycA, which represents the first module of the protein assembly line for the cyclic decapeptide antibiotic tyrocidine A.[30] TycA is composed of three semi-autonomous domains, which are responsible for the activation (adenylation, A-domain) and binding (peptidyl carrier, PCP-domain) of phenylalanine, and its epimerization into the D-stereoisomer (epimerization, E-domain). The enzyme is 1088 aa long (122.7 kDa) and contains a total of five cysteines that prohibit specific, single-

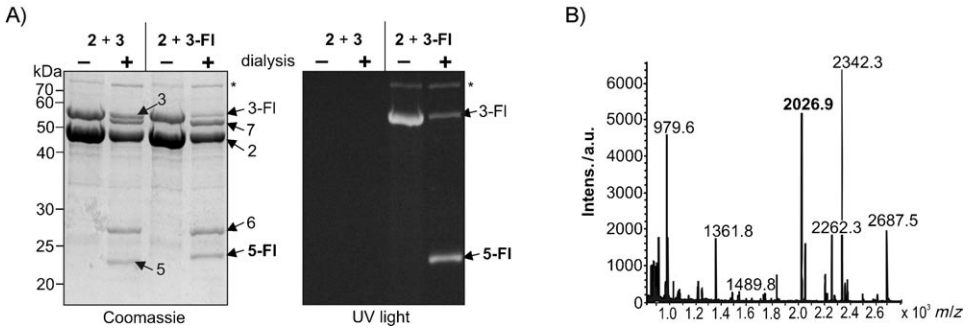

Figure 5. Preparation of fluorescein-labelled hGH. A) SDS-PAGE gel analysis of the reaction of hGH-$GyrA^N$-FKBP-His_6 (**2**) with complementary Int^C constructs **3** or **3-FI** by using Coomassie brilliant blue staining (left) or UV illumination (right). The reactions were performed by mixing **2** in buffer containing urea (8 M) with **3** or **3-FI**, and subsequent removal of the denaturant by dialysis. Samples taken before (−) and after (+) dialysis are indicated. B) Confirmation of the identity of the splice product hGH-CysTag-fluorescein (**5-FI**). The splice product band marked with **5-FI** was excised, digested with trypsin and analysed by MALDI-TOF mass spectrometry. All signals could be assigned to tryptic fragments of the expected splice product including the fragment that contained the splice junction (YTEAGSC(FI)SHHHHHH = calcd $[M^+]$ = 2027.017 *m/z*; obs $[M^+]$ = 2026.878 *m/z*). For the assignment of the other fragments see the Experimental Section.

site labelling with conventional cysteine bioconjugation methods. For these reasons, TycA is an ideal yet demanding target to show the strengths of selective labelling through the CysTag-intein approach. However, in this case the previously reported split DnaB intein also proved to be inappropriate since only very low splicing yields of $<10\%$ could be obtained (data not shown).

To test the new split GyrA intein, fusion protein **8** (Figure 6A; Strep-TycA-GyrA^N-FKBP, $M_W = 152$ kDa) was generated and purified from the soluble fraction after lysis of the *E. coli* expression host. Note that **8** harbours an N-terminal strep-tag II sequence but no C-terminal His_6-tag, whereas the complementary CysTag-intein construct **3** only has a C-terminal His_6-tag. As a consequence of this arrangement, only the desired splice product **9** possessed both affinity tags (Figure 6A). This allowed the straightforward purification of labelled TycA from the reaction mixture with two consecutive chromatography steps. Purified **8** and fluorescein-labelled **3-Fl** were mixed in an equimolar ratio at a concentration of 10 μM. Splice product formation proceeded to $>60\%$ yield after 4 h as visualized by SDS-PAGE (Figure 6B). Subsequently, the reaction mixture was applied to a Ni^{2+}-NTA and a strep-tactin matrix (Figure 6C). The splice product **9-Fl** could be bound and eluted (Figure 6C, lane 2) from the Ni^{2+}-NTA column and was essentially pure after binding and elution from the strep-tactin matrix (Figure 6C, lanes 5 and 6).

Enzymatic activity of selectively labelled **9-Fl** was assayed by D-Phe–L-Pro dipeptide formation of TycA with TycB1, which represents the next module in the tyrocidine NRPS assembly line. In this assay condensation of D-Phe with L-Pro results in the linear dipeptide bound at the Ppant moiety of TycB1. Due to the high Xaa–Pro peptide bond content in the *cis* conformation, this peptide rapidly cyclises to give the D-Phe–L-Pro-diketopiperazine (dFP–DKP).[31] In this way, the enzyme is regenerated for the next elongation cycle. Module TycB1, which consists of a condensation, an adenylation, and a PCP domain, as well as a TycA control for comparison were expressed and purified as described elsewhere.[32] Post-translational modification of the apo-enzyme with the prosthetic group Ppant is required for enzymatic activity. Thus, holo-enzymes of TycA, TycB1 and fluorescein-labelled TycA (**9-Fl**) were produced by incubation of the apo-enzymes with coenzyme A and phosphopantetheinyl transferase Sfp for 1 h at 37 °C. In the dFP–DKP formation assays, holo-TycB1 was incubated either with fluorescein-labelled holo-TycA (**holo-9-Fl**) or with a recombinantly produced positive control holo-TycA and with the substrates L-Phe, L-Pro

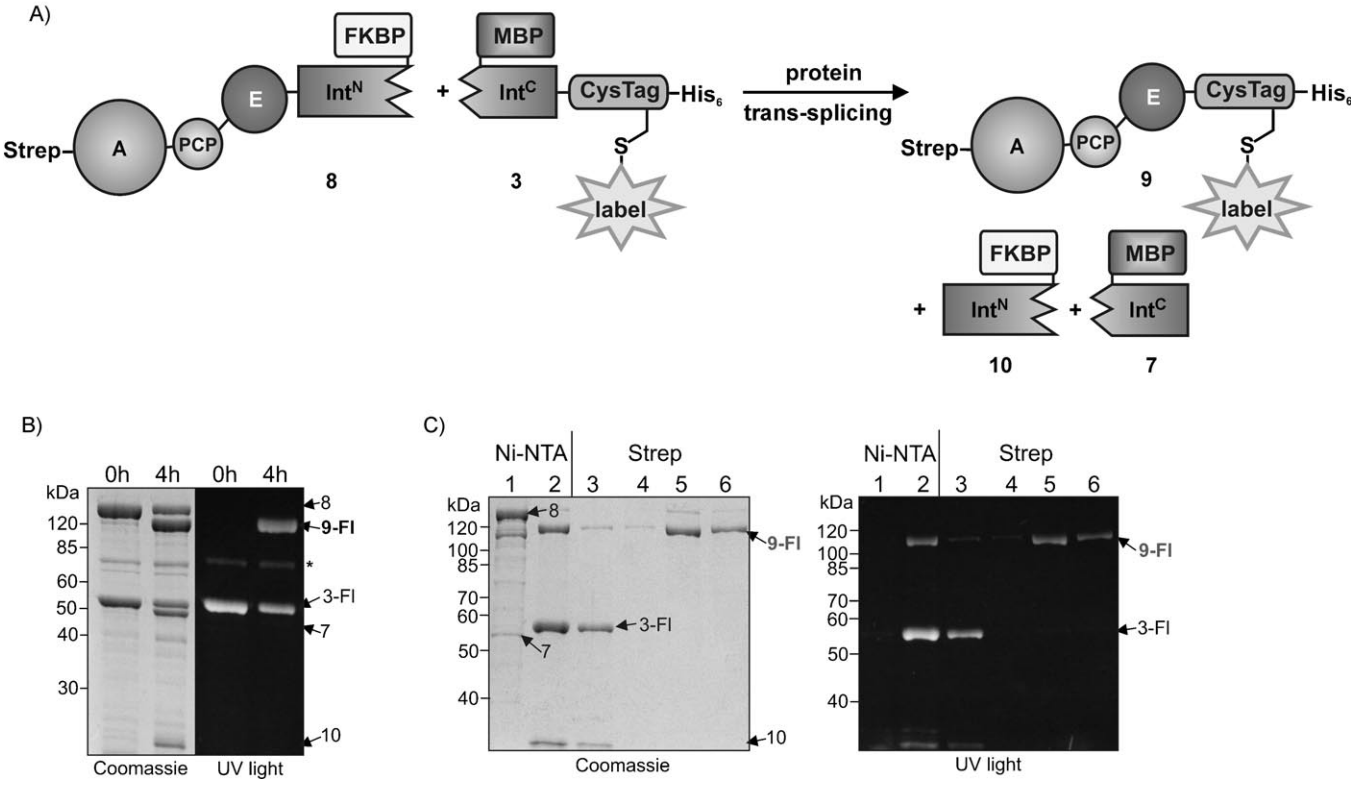

Figure 6. Preparation of a fluorescein-labelled multidomain NRPS. A) Schematic presentation of the reaction to generate labelled TycA, depicting the double-tag purification strategy. TycA consists of a phenylalanine-specific adenylation domain (A), a peptidyl-carrier protein (PCP) and an epimerisation domain (E). B) Analysis of the protein trans-splicing reaction between TycA-GyrA^N-FKBP (**8**) and complementary GyrA^C protein **3-Fl** on a Coomassie brilliant blue stained (left) or UV illuminated (right) SDS-PAGE gel. The splice product TycA-CysTag-fluorescein (**9-Fl**) was formed in high yields. C) Purification of **9-Fl** from the reaction mixture by two consecutive steps by using Ni^{2+}-NTA and a strep-tactin affinity matrix to give essentially pure protein (lanes 5 and 6). Lane 1: flow-through of the Ni^{2+}-NTA column; lane 2: elution fraction of the Ni^{2+}-NTA column; lanes 3 and 4: flow-through and washing fractions of the strep-tactin column; lanes 5 and 6: fractions eluted from the strep-tactin column with desthiobiotin. The fraction shown in lane 6 was used for the enzyme assay described in Figure 7.

and ATP. Figure 7 shows the HPLC analysis of the reaction mixtures. A peak at 18.7 min corresponded to the dFP–DKP product, as confirmed by mass spectrometry ($[M+H]^+_{obs} = 245$;

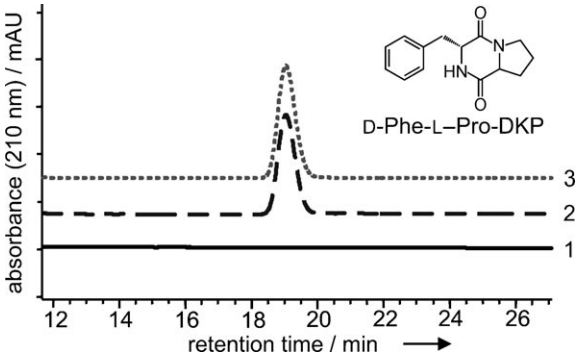

Figure 7. Enzymatic assay for the TycA-NRPS. Fluorescein-labelled apo-TycA (**9-Fl**) was converted into the holo form by incubation with Sfp and coenzyme A (data not shown). Upon incubation of **holo-9-Fl** with holo-TycB1 in the presence of substrates L-Phe, L-Pro and ATP, the cyclic dipeptide D-Phe-Pro-diketopiperazine (DKP) was formed. The product was detected by reversed phase HPLC at a retention time of 18.7 min (see trace 3). The product identity was further confirmed by MALDI-MS (not shown). A control reaction with recombinant control protein holo-TycA was performed under identical conditions (trace 2). For a negative control ATP was omitted (trace 1).

$[M+H]^+_{calcd} = 245$; data not shown). Indeed, integration of the product peaks from the two reactions revealed that the amount of product formed was indistinguishable, regardless of whether labelled enzyme obtained from the split-intein CysTag approach or recombinant control protein was used. In negative control reactions, in which either ATP or one of the amino acids or enzymes were omitted, dFP–DKP formation was not detected. These results demonstrate that the labelled TycA protein (**9-Fl**) was completely intact and enzymatically active. Fusion of the split intein fragment and the subsequent protein splicing reaction to append the prelabelled CysTag peptide sequence is thus an efficient and suitable approach for the selective chemical modification of large and complex proteins.

Conclusions

There is great demand for technologies that allow the selective incorporation of synthetic moieties into proteins. Some of the very powerful existing approaches, like chemical ligation and the tRNA suppressor strategies, are experimentally demanding and therefore difficult to perform in nonspecialised laboratories. Also, most of the required reagents have to be synthesised individually because they are not commercially available. In contrast, a huge portfolio of reagents with useful labels is commercially available for traditional amino acid side chain conjugation. To expand the scope of potential applications of these reagents we have recently developed a strategy in which a premodified cysteine-containing peptide sequence (the CysTag) is appended to a POI by protein trans-splicing (Figure 1). This allows the site-selective modification of a protein that contains multiple cysteine residues by using straightforward biochemical methodology.

In this work, we have shown that the split Mxe GyrA intein is a superior tool for protein trans-splicing and is compatible with the CysTag approach. We report for the first time an artificially split version of this intein that is highly active without the need for a denaturation step. Both the protein therapeutic hGH, which is difficult to express and insoluble, and a 123 kDa multifunctional NRPS were successfully labelled with high yields. The split GyrA intein will also be useful for the segmental isotopic labelling of proteins[9] and for further investigation of the mechanism of protein splicing. Particularly interesting is the rapid accumulation of the branched intermediate; this indicates that asparagine cyclisation is the rate-determining step in this case.

It is noteworthy that the fusion intein fragments excise themselves during the protein trans-splicing reaction. Thus, the sequence ultimately added to the POI is only a short peptide tag. Such a minimal extension can be advantageous for the positioning of chemical cross-linkers or fluorophores for inter- and intramolecular FRET studies. These potential applications can hardly be realised with the same level of spatial control with alternative techniques, such as alkyl guanine transferase (AGT), acyl- or peptidyl-carrier protein (ACP/PCP), that are geared to the labelling of a larger fusion protein. Furthermore, using our described CysTag approach, it is also conceivable that the cysteine-containing peptide sequence represents a larger part of the POI. If this is the case, the label can be introduced at an internal position (S. Brenzel, H.M., unpublished results). Finally, the approach might also be extended to other amino acid side chains or peptide motifs.

We believe that the herein described CysTag approach for selective protein labelling technology will be a powerful and easy to perform methodology to incorporate diverse synthetic moieties into a wide array of proteins.

Experimental Section

General: 5-Iodoacetamido-fluorescein (IAF) was purchased from Molecular Probes and polyethylene glycol 5000 maleimide (PEG) was from IRIS biotechnology. Ampicillin was used at a concentration of 100 $\mu g\,mL^{-1}$. Standard protocols were applied for DNA cloning. PCR was performed with Phusion polymerase (Finnzymes). Oligonucleotides were purchased from Operon (Cologne, Germany). All plasmids were verified by DNA sequencing. Standard chemicals were from Sigma–Aldrich (Munich, Germany), Roth (Karlsruhe, Germany) or AppliChem (Darmstadt, Germany).

Plasmid construction

Plasmids pAI13, pTK118, pTK120 and pTK130 for the expression of proteins **1**, **2**, **3** and **8**, respectively, were prepared as follows.

pAI13: The $GyrA^N$ encoding fragment was amplified with PCR by using plasmid pTXB1 (NEB) as a template and oligonucleotides 5′-ATA GAA TTC CGC TAC TGC ATC ACG GGA GAT GCA CTA GTT GCC-3′ and 5′-ATA TCT AGA GCG GGC AAA ACC TGC ACA GTC GAC GC-3′ (restriction sites underlined) and ligated into pSB13.[13]

pTK120: The $GyrA^C$ coding region was amplified with PCR by using the same pTXB1 template vector DNA and oligonucleotides 5′-ATA

ACT AGT GGA AAA CCC GAA TTT GCG CCC-3′ and 5′-TTT GGA TCC GGC TTC AGT ATT GTG GCT GAC GAA CCC-3′ and ligated into pSB13. The $DnaB^N$–FKBP coding region was replaced by the coding region for FRB from plasmid pSB15, which could subsequently be deleted by using restriction enzymes XbaI and SpeI followed by religation. The cysteine was introduced by site-directed mutagenesis with oligonucleotides 5′-CAA TAC TGA AGC CGG ATC CTG CGG TCA TCA CCA TCA CCA TCA C-3′ and the reverse complement. Sequencing revealed an additional point mutation, which led to the final Cys-tag with the amino acid sequence TEAGS**C**SHHHHHH.

pTK118: The DNA fragment encoding hGH was amplified from the cDNA by using oligonucleotides 5′-ATA CCA TGG GAT TCC CAA CCA TTC CCT TAA GC-3′ and 5′-ATA GAA TTC GAA GCC ACA GCT GCC CTC CAC-3′ and ligated into pAI13.

pTK130: TycA was PCR amplified from chromosomal DNA of *Bacillus brevis* ATCC8185 with oligonucleotides 5′-AAT GCT AGC ATG TTA GCA AAT CAG GCC-3′ and 5′-TTA GAA TTC GCG CAG TGT ATT TGC AAG-3′ and placed—by using restriction enzymes NheI and EcoRI—in a plasmid that encoded the N-terminal strep-tag II sequence. The strep-tag II-TycA encoding region was excised with restriction enzymes NcoI and EcoRI, and ligated into pAI13 to give pTK117. The His_6-tag was deleted by site-directed mutagenesis of pTK117 by the QuikChange protocol (Stratagene) and by using oligonucleotide 5′-CTA GTT ATG GAT CCA GAT AAC TAG TCA CCA TCA CCA TCA C-3′ and the reverse complement.

Protein expression and purification: *E. coli* BL21 cells were transformed with the respective expression plasmids and grown in LB medium containing ampicillin at 37 °C to an OD_{600} of 0.7. At this point the temperature was lowered to 30 °C or 25 °C and expression was induced by the addition of isopropyl-β-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM. After 3–5 h the cells were harvested by centrifugation, resuspended in Ni^{2+}-NTA loading buffer (50 mM Tris/HCl, pH 8.0, 300 mM NaCl, 1 mM EDTA) for proteins **1**, **2** and **3** or strep-tag loading buffer (100 mM Tris/HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA) for protein **8** and frozen at −80 °C. For protein purification the cell suspension was thawed on ice and lysed by two passages through an emulsifier (Avestin EmulsiFlex C5). Insoluble material was separated from the soluble supernatant by centrifugation at 30 000 *g* for 30 min. For proteins **1** and **3** the soluble fractions were applied to a Ni^{2+}-NTA affinity column and purified as described previously.[7] Protein **8** was loaded on a strep-tactin column. After being washed with loading buffer, **8** was eluted with strep-buffer containing desthiobiotin (2.5 mM). Protein **2** was purified from the insoluble fraction of the *E. coli* lysate under denaturing conditions by solubilization in buffer containing urea (8 M) and subsequent Ni^{2+}-NTA-chromatography. Protein concentrations were determined by using the calculated molecular extinction coefficient at 280 nm.

CysTag labelling and protein trans-splicing: Int^C constructs labelled at the CysTag were prepared as previously described.[7] Shortly, the single cysteine in purified protein **3** (MBP-$GyrA^C$-CysTag-His_6, 53.3 kDa) was reduced with DTT prior to the modification reaction. After being quenched with excess DTT, labelled protein could be used for subsequent splicing reactions. Proteins containing the complementary intein fragments were mixed in equimolar concentrations (4 μM each) in splice buffer (50 mM Tris, 300 mM NaCl, 1 mM EDTA, 2 mM DTT, pH 7.0) when not stated otherwise. At indicated time points samples were taken, mixed with 4× SDS-PAGE loading buffer to quench the reaction and boiled before being applied to a gel. For quantitative analysis, gels were stained with Coomassie brillant blue and relative intensities of protein bands were densitometrically determined by using the program "Scion Image" (http://www.scioncorp.com). Measurements at different temperatures were performed by mixing the Int^N construct with DTT and splice buffer, and preincubating this mixture at the respective temperature for 5 min prior to starting the splicing reaction by adding the complementary Int^C construct. The reactions were allowed to proceed for 4 h before being quenched and analysed. Time- and temperature-dependent measurements were performed in duplicate.

Construct **2** (in buffer containing 8 M urea) was mixed in a twofold excess with labelled or unlabelled protein **3** (in buffer without urea) in a final volume of 200 μL; this resulted in a final concentration of ~6 M urea. DTT was added to a final concentration of 2 mM and the reaction mixture was transferred into a dialysis tube (molecular weight cut-off of 3.5 kDa). Next, the reaction was dialysed at 4 °C against splice buffer (50 mM Tris/HCl, pH 7.0, 300 mM NaCl, 2 mM DTT, 1 mM EDTA), overnight, and two more times for 30 min, followed by a final dialysis step for 30 min at room temperature. The reaction mixture was then incubated for an additional 2 h at 25 °C before being analysed on a SDS-PAGE gel. The fluorescent band of splice product **5-Fl** was excised, digested with trypsin and analysed by MALDI-TOF mass spectrometry. The resulting spectrum is shown in Figure 5B. Assigned fragments are as follows: obs: 2687.466 = YSFLQNPQTSLC(PAM)FSESIPTPSNR ($hGH_{44–66}$, calcd: 2687.277); obs: 2342.297 = LHQLAFDTYQEFEEAYIPK ($hGH_{22–40}$, calcd: 2342.134); obs: 2262.276 = SVFANSLVYGASDSNVYDLLK ($hGH_{97–117}$, calcd: 2262.1288); obs: 2026.878 = YTEAGSC(Fl)SHHHHHH (splice junction with fluorescein modification, calcd: 2027,017); obs: 1489.779 = FDTNSHNDDALLK ($hGH_{148–160}$, calcd: 1489.692); obs: 1361.794 = DLEEGIQTLMGR ($hGH_{118–129}$, calcd: 1361.673); obs: 979.583 = LFDNAMLR ($hGH_{11–18}$, calcd: 979.503); PAM: acrylamide adduct, Fl: fluorescein-modified.

Labelled TycA (**9-Fl**) was prepared by mixing equimolar amounts (final concentration 4 μM for each protein) of proteins **8** and **3-Fl** in a total volume of 1 mL splice buffer. The reaction mixture was incubated at 25 °C for 12 h and then applied to a Ni^{2+}-NTA-column. Unreacted protein **9** was found in the flow-through, because it was the only protein without a His_6-tag in the reaction mixture. After three washing steps with Ni^{2+}-NTA buffer containing 5 mM, 20 mM and 40 mM imidazole, proteins were eluted with buffer containing 250 mM imidazole. Elution fractions were pooled and dialysed three times against strep-tag buffer (100 mM Tris/HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA) before being loaded on a strep-tactin column. After being washed with 10 column volumes of loading buffer and all other proteins were removed, **9-Fl** was eluted with strep-tag buffer containing desthiobiotin, and subsequently dialysed against assay buffer (50 mM HEPES, pH 8.0, 100 mM NaCl, 1 mM EDTA, 2 mM DTT, 10 mM $MgCl_2$).

Enzymatic assay for TycA: Apo proteins **9-Fl**, TycB1-His_6 and the TycA-His_6 control (1 μM each) were converted into their holo form by incubation with coenzyme A (100 μM) and recombinant Sfp (25 nM) for 1 h at 37 °C. For dFP–DKP formation **holo-9-Fl** was incubated with holo-TycB1 (500 nM for each protein) in the presence of 100 μM L-Phe and 100 μM L-Pro at 37 °C for 2 h in a final volume of 200 μL. The reaction was started by adding ATP to a final concentration of 2 mM. For the TycA–His_6 control the same conditions were applied. Negative controls were performed by omitting either ATP, an amino acid or an enzyme. Each reaction mixture was then quenched by the addition of methanol (1 mL), precipitated proteins were separated by centrifugation (16 000 *g*, 15 min) and supernatants were transferred to a fresh tube. The solvent was removed under vacuum, the residue was dissolved in methanol

(100 μL, 30%, *v/v*) and used for HPLC analysis. Separation of the reaction products was achieved by using a EC 150/4.6 Nucleodur C18 column by applying an isocratic method at a flow rate of 0.6 mL min^{-1} with 30% buffer B (0.045% formic acid/methanol, *v/v*; buffer A: 0.05% formic acid/water, *v/v*) as previously described.[33] The products were identified by detection at 210 nm, collected, and used for a MALDI-TOF analysis.

Abbreviations

CysTag: cysteine tag, dFP-DKP: D-Phe–L-Pro-diketopiperazine, $DnaB^C$/$DnaB^N$: C- and N-terminal fragment of the split Ssp DnaB intein, Ext^C/Ext^N: C- and N-extein, FKBP: FK506 binding protein, Fl: fluorescein label, $GyrA^C$/$GyrA^N$: C- and N-terminal fragment of the split Mxe GyrA intein, hGH: human growth hormone, His_6: hexahistidine tag, Int^C/Int^N: C- and N-terminal split intein fragment, MBP: maltose binding protein, NRPS: nonribosomal peptide synthetase, PEG: poly(ethylene glycol) (and -maleimide label), POI: protein of interest, Ppant: 4′-phosphopantetheine, Strep: strep-tag II, TycA: tyrocidine synthetase A.

Acknowledgements

We are grateful to A. Inkemann for cloning initial GyrA constructs and C. Ludwig for construction of a strep-tag II plasmid. We thank J. Zettler for help with mass analysis. Funding was provided by the DFG and the Fonds der Chemischen Industrie.

Keywords: cysteine bioconjugation · fluorescent probes · intein · protein modifications · protein splicing

[1] J. Xie, P. G. Schultz, *Nat. Rev. Mol. Cell. Biol.* **2006**, *7*, 775–782.
[2] P. E. Dawson, S. B. Kent, *Annu. Rev. Biochem.* **2000**, *69*, 923–960.
[3] T. W. Muir, *Annu. Rev. Biochem.* **2003**, *72*, 249–289.
[4] N. Johnsson, K. Johnsson, *ACS Chem. Biol.* **2007**, *2*, 31–38.
[5] *Bioconjugate Techniques* (Ed.: G. T. Hermanson) Academic Press, London, **1996**.
[6] M. B. Francis in *Chemical Biology: From Small Molecules to System Biology and Drug Design in Chemical Biology* (Eds.: S. L. Schreiber, T. M. Kapoor, G. Wess) Wiley, **2007**, pp. 593–634.
[7] a) T. Kurpiers, H. D. Mootz, *Angew. Chem.* **2007**, *119*, 5327–5330; *Angew. Chem. Int. Ed.* **2007**, *46*, 5234–5237.
[8] H. Paulus, *Annu. Rev. Biochem.* **2000**, *69*, 447–496.
[9] V. Muralidharan, T. W. Muir, *Nat. Methods* **2006**, *3*, 429–438.
[10] C. J. Noren, J. Wang, F. B. Perler, *Angew. Chem.* **2000**, *112*, 458–476; *Angew. Chem. Int. Ed.* **2000**, *39*, 450–466.
[11] I. Giriat, T. W. Muir, F. B. Perler, *Genet. Eng.* **2002**, *23*, 171–199.
[12] H. Wu, Z. Hu, X. Q. Liu, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9226–9231.
[13] S. Brenzel, T. Kurpiers, H. D. Mootz, *Biochemistry* **2006**, *45*, 1571–1578.
[14] K. V. Mills, B. M. Lew, S. Jiang, H. Paulus, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 3543–3548.
[15] M. W. Southworth, E. Adam, D. Panne, R. Byer, R. Kautz, F. B. Perler, *Embo J.* **1998**, *17*, 918–926.
[16] E. O. Davis, P. J. Jenner, P. C. Brooks, M. J. Colston, S. G. Sedgwick, *Cell* **1992**, *71*, 201–210.
[17] S. Chong, M. Q. Xu, *J. Biol. Chem.* **1997**, *272*, 15587–15590.
[18] H. D. Mootz, T. W. Muir, *J. Am. Chem. Soc.* **2002**, *124*, 9044–9045.
[19] H. Wu, M. Q. Xu, X. Q. Liu, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **1998**, *1387*, 422–432.
[20] T. Klabunde, S. Sharma, A. Telenti, W. R. Jacobs, Jr., J. C. Sacchettini, *Nat. Struct. Biol.* **1998**, *5*, 31–36.
[21] M. Fitzsimons Hall, C. J. Noren, F. B. Perler, I. Schildkraut, *J. Mol. Biol.* **2002**, *323*, 173–179.
[22] T. C. Evans, Jr., J. Benner, M. Q. Xu, *Protein Sci.* **1998**, *7*, 2256–2264.
[23] F. I. Valiyaveetil, R. MacKinnon, T. W. Muir, *J. Am. Chem. Soc.* **2002**, *124*, 9113–9120.
[24] M. Q. Xu, F. B. Perler, *EMBO J.* **1996**, *15*, 5146–5153.
[25] J. Shi, T. W. Muir, *J. Am. Chem. Soc.* **2005**, *127*, 6198–6206.
[26] A. M. de Vos, M. Ultsch, A. A. Kossiakoff, *Science* **1992**, *255*, 306–312.
[27] R. K. Chawla, J. S. Parks, D. Rudman, *Annu. Rev. Med.* **1983**, *34*, 519–547.
[28] J. Grunewald, M. A. Marahiel, *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 121–146.
[29] N. George, H. Pick, H. Vogel, N. Johnsson, K. Johnsson, *J. Am. Chem. Soc.* **2004**, *126*, 8896–8897.
[30] H. D. Mootz, M. A. Marahiel, *J. Bacteriol.* **1997**, *179*, 6843–6850.
[31] T. Stachelhaus, H. D. Mootz, V. Bergendahl, M. A. Marahiel, *J. Biol. Chem.* **1998**, *273*, 22773–22781.
[32] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5848–5853.
[33] U. Linne, D. B. Stein, H. D. Mootz, M. A. Marahiel, *Biochemistry* **2003**, *42*, 5114–5124.

Received: May 9, 2008
Published online on August 29, 2008

