Review

Semi-Synthesis of Labeled Proteins for Spectroscopic Applications

Lucia De Rosa [1], Anna Russomanno [1], Alessandra Romanelli [2] and Luca Domenico D'Andrea [1],*

[1]   Istituto di Biostrutture e Bioimmagini, CNR, Via Mezzocannone 16, Napoli 80134, Italy;
    E-Mails: lucia.derosa@unina.it (L.D.R.); anna.russomanno@gmail.com (A.R.)
[2]   Dipartimento delle Scienze Biologiche, Università di Napoli "Federico II", Via Mezzocannone 16,
    Napoli 80134, Italy; E-Mail: alessandra.romanelli@unina.it

* Author to whom correspondence should be addressed; E-Mail: ldandrea@unina.it;
    Tel.: +39-081-253-6679; Fax: +39-081-253-4574.

Abstract: Since the introduction of SPPS by Merrifield in the 60s, peptide chemists have considered the possibility of preparing large proteins. The introduction of native chemical ligation in the 90s and then of expressed protein ligation have opened the way to the preparation of synthetic proteins without size limitations. This review focuses on semi-synthetic strategies useful to prepare proteins decorated with spectroscopic probes, like fluorescent labels and stable isotopes, and their biophysical applications. We show that expressed protein ligation, combining the advantages of organic chemistry with the easy and size limitless recombinant protein expression, is an excellent strategy for the chemical synthesis of labeled proteins, enabling a single protein to be functionalized at one or even more distinct positions with different probes.

1. Introduction

Protein structure, folding dynamics, function and interactions with other macromolecules can be widely explored and characterized by biophysical techniques such as fluorescence and NMR spectroscopies. Spectroscopic techniques rely tightly on protein labeling strategies by which the chemical structure of a protein is modified through the introduction of biophysical probes, such as fluorophores or isotopes.

A wide collection of protein labeling approaches have been developed in recent years [1–4], leading to great discoveries and innovations. In particular, the introduction of native chemical ligation (NCL) methodologies for chemical synthesis of proteins marked a breakthrough in protein and peptide chemistry, with a strong impact on chemical biology and biophysical applications [5,6]. Here, we review the semi-synthetic strategies employed for the preparation of labeled proteins and their spectroscopic applications.

1.1. Total Synthesis of Proteins: SPPS and Chemical Ligation

The great advantage of chemical protein synthesis and semi-synthesis over traditional recombinant protein expression refers to the precise control over the kind and number of modifications that can be introduced into a protein molecule, enabling one to realize any desired change of its covalent structure with surgical precision. Since the introduction in 1963 of the stepwise solid-phase peptide synthesis (SPPS) by Merrifield [7], chemical synthesis of proteins greatly expanded and culminated with the development of chemical ligation approaches [8]. Chemical ligation methods allow one to overcome the size-limit of SPPS, mainly due to incomplete coupling and deprotection reactions (Figure 1), affording the synthesis of large proteins by stitching together short synthetic segments to give a polypeptide chain of any desired length and harboring a plethora of possible chemical modifications.

Figure 1. Dependence of stepwise SPPS yield on coupling efficiency (**A**) and number of coupling steps (**B**).
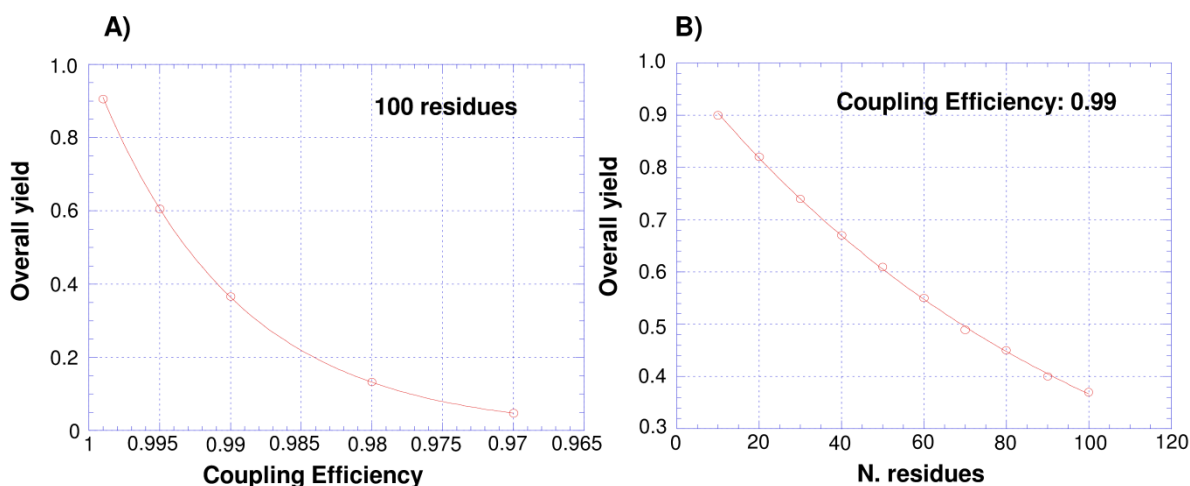
In this scope, ligation strategies are based on chemoselective reactions between two functional groups respectively placed at N- and C-terminus of two contiguous peptide segments which react by forming a stable chemical bond. The type of chemical bond made at the site of junction between two peptide fragments depends on the reactive groups employed (Figure 2). For instance, the reaction between an hydrazide or an aminooxy group with an aldehyde or a ketone gives rise to hydrazones or oximes (Figure 2A,B); a pseudoproline analog results from the reaction between a peptide bearing a C-terminal glycolaldehyde ester and a second peptide starting with Cys, Ser or Thr [9,10] (Figure 2C); a peptide-thiocarboxylate reacts with a N^α-bromoacetyl-peptide to give a thioester-linked polypeptide product (Figure 2D) [8]; the reaction between a thiol and a bromoacetyl group yields a polypeptide

joined by thioether linkage (Figure 2E) [11]; thiol and maleimide groups react through a Michael addition (Figure 2F) [12]; Cu(I)-catalyzed azide-alkine [3+2] cycloaddition leads to the formation of a triazole ring (Figure 2G) [13,14]; Diels-Alder cycloaddition reaction, which involves a diene and a dienophile, yields a six member carbocycle (Figure 2H) [15].

Figure 2. Schematic representation of chemoselective reactions useful for non-native chemical ligation. (**A**) and (**B**) R = H, CH$_3$; (**C**) R = SH, OH.
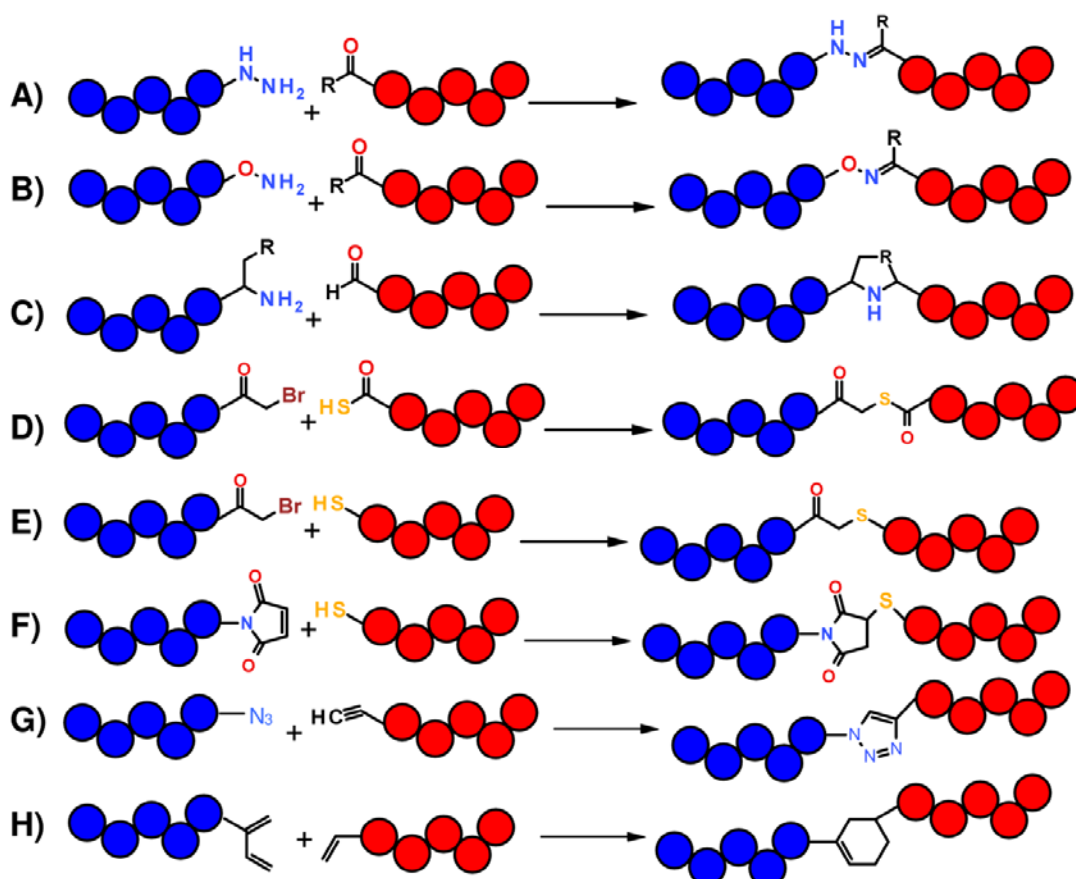
Even though numerous proteins have been successfully prepared by chemical ligation strategies yielding a non-native bond at the junction sites [16,17], such unnatural structures are not always compatible with the assumption of correct protein folding. Thus, great efforts were directed towards the development of traceless chemical ligation approaches, ending up in the introduction of the NCL by Kent and coworkers [18]. This method leads to the formation of a single polypeptide chain bearing a native peptide bond at the ligation site after coupling of two peptide segments, one containing a carboxy-terminal α-thioester group and the other bearing a 1,2 aminothiol, like an amino-terminal Cys residue (Figure 3). The reaction occurs, even under mild conditions (aqueous solution at pH around neutrality), through a trans-thioesterification reaction which leads to the formation of a thioester intermediate rapidly evolving toward the desired stable amide-linked product through a spontaneous intramolecular S-to-N acyl shift. This reaction is orthogonal with respect to functional groups present in a protein and also proceeds selectively in the presence of internal Cys residues. Another powerful chemical ligation method is the traceless Staudinger ligation, which relies on the selective reaction between a phosphinothioester and an azide to form an amide bond [19]. This method does not

necessitate of a Cys residue at the junction between two peptide fragments and, thus, appears as a strategy complementary to NCL.

Figure 3. Mechanism of the NCL reaction.
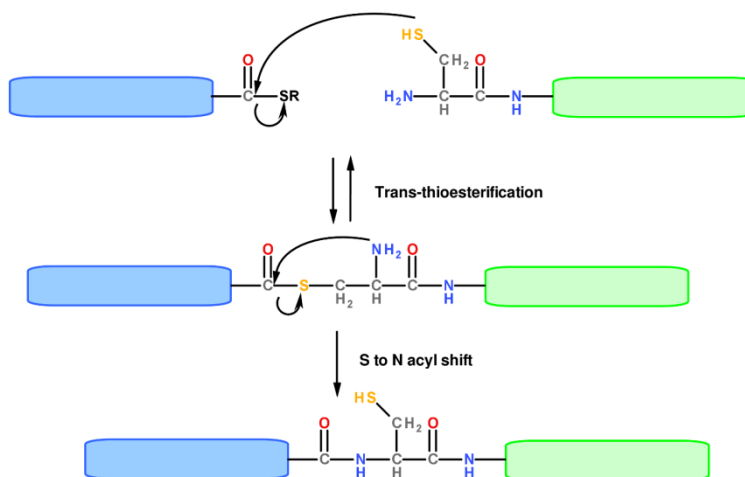
The efficacy of the NCL has been significantly broadened by the introduction of a "recombinant" version, named Expressed Protein Ligation (EPL) [20,21]. In this case one or more of the peptide building blocks are prepared by recombinant DNA technology and ligated through a NCL reaction between a carboxy-terminal α-thioester and 1,2 aminothiol groups. For this reason, EPL is referred to as a semi-synthetic approach. To achieve C-terminal thioester or N-terminal Cys containing protein segments, EPL exploits a class of proteins called inteins, which are auto-cleavable elements able to catalyze their self-removal from flanking polypeptides [21,22]. EPL, as well as NCL, has been employed for the introduction into proteins of unnatural amino acids, post-translational modifications [6,20,23,24] and covalent dimers formation [25,26]. Furthermore, EPL, combining the advantages of organic chemistry with the easy and size limitless recombinant protein expression, allows for the preparation of modified large proteins.

1.2. Protein Synthesis via Intein Chemistry (EPL and PTS)

The mechanism by which an intein is removed from a pre-mature polypeptide chain is known as protein splicing (Figure 4). Protein splicing proceeds through an N→S (or N→O) acyl shift in which the *N*-extein is transferred to the side chain -SH or -OH of a Cys/Ser residue, located at N-terminus of the intein. The N-extein is then transferred to a second Cys/Ser residue located at the N-terminus of the C-extein through a trans(thio)esterification reaction. A cyclization reaction involving a conserved Asn residue at the intein C-terminus leads to intein excision and to the formation of a new peptide bond between the two exteins, giving the mature polypeptide chain. A wide number of genetically engineered inteins able to catalyze their self-cleavage from only a N- or C-extein have been obtained, providing a powerful tool to prepare recombinant C-terminal thioester or N-terminal cysteinyl polypeptides required to perform EPL.

Figure 4. Mechanism of the intein protein splicing reaction.
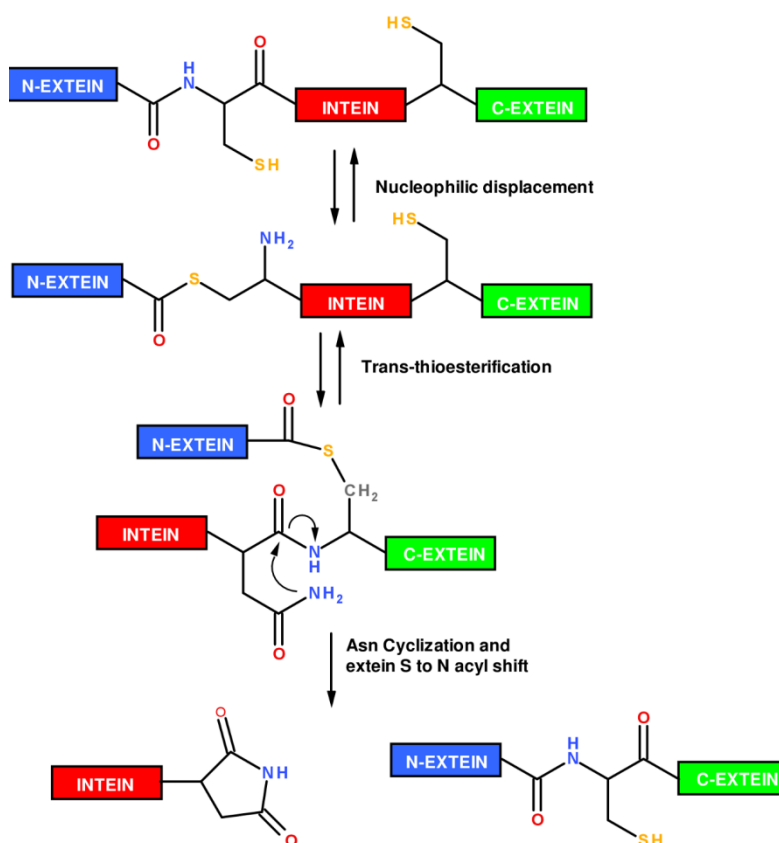
To prepare C-terminal thioester proteins, inteins harboring the mutation of the C-terminal Asn into an Ala have been designed [22]. Such mutant inteins are unable to evolve from the thioester intermediate to the final spliced product. In this case, the thioester intermediate can be cleaved off by addition of thiols releasing the N-extein as C-terminal thioester protein (Figure 5A). Several *E. coli* expression vectors are commercially available that allow the recombinant expression of fusion constructs with engineered inteins, such as the *Saccharomyces cerevisiae* vacuolar ATPase subunit (*Sce* VMA) intein, *Methanobacterium thermo-autotrophicum* ribonucleoside diphosphate reductase (*Mth* RIR1) intein or the *Mycobacterium xenopi* DNA gyrase A (*Mxe* GyrA) intein [27]. The *Mxe* GyrA intein [28] is the most widely used as *Mxe* GyrA intein is small (198 amino acids), expresses at high level in bacteria, is able to efficiently refold from inclusion bodies and is cleaved by a variety of thiols, even in the presence of low concentration of denaturants (*i.e.*, 2 M urea), detergents or organic solvents [29]. Special care has to be devoted to the choice of the -1 residue at the junction extein-intein, as the efficiency of splicing is intein and N-extein sequence dependent. Furthermore, the extein sequence modulates the kinetic of protein splicing. Notably, a -1 Asp residue induces high levels of premature cleavage *in vivo*, whereas a -1 Pro should be avoided as completely inhibits the cleavage [20,27].

The thiol sodium 2-mercaptoethanesulfonate (MESNA) is often employed to mediate intein thiolysis. In fact, while thioalkyl-esters, obtained, for example, from ethanedithiol (EDT) and ethanethiol (ET), are quite stable, but not reactive enough in NCL, whereas thioaryl-esters, which result from the use of benzylmercaptan and thiophenol as splicing inducing agents, are much more reactive but even more susceptible to hydrolysis. Usually, an alkyl-thioester is preferred during the preparation and purification steps, while during NCL, the addition of aryl-thiols to the reaction mixture

allows for conversion of the thioester species into a more reactive thioaryl-ester through trans-thioesterification. The thiol MESNA allows a good compromise between stability and reactivity of the thioester protein obtained. Furthermore, MESNA is a non-malodourous thiol, facilitating thioester protein handling and purification. A complete characterization of the reactivity of a wide series of thiols was performed by Johnson and Kent [30]. They also reported another useful non-maleodorous thiol, the *p*-mercaptophenylacetic acid (MPAA), which can be added to the NCL mixture to promote the *in-situ* generation of a very reactive thioester, resulting in a sensible improvement in the efficiency both in terms of yield and time of reaction.

Figure 5. Schematic representation of reactions performed by engineered inteins yielding α-thioester proteins (**A**) or N-terminal cysteinyl-proteins (**B**).
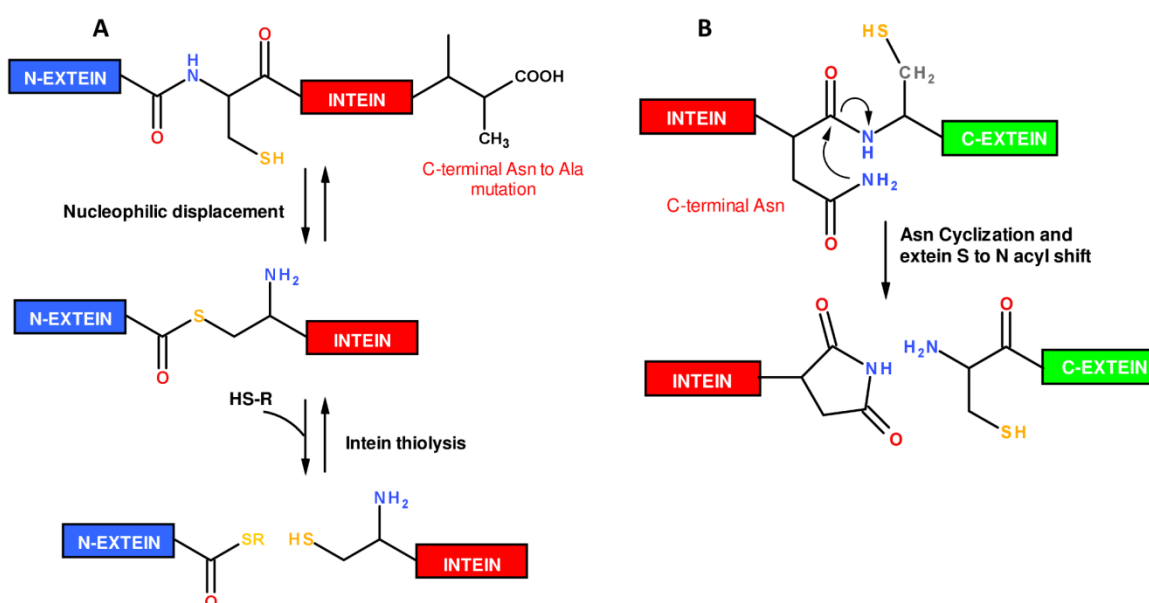

Protein splicing can also be exploited to prepare amino-terminal Cys proteins. In this case, inteins have been mutated in order to induce the cleavage at the C-terminal splice junction, between the intein and a C-extein starting with Cys, through pH and temperature changes (Figure 5B). Splicing of such modified inteins allows for the release of an amino-terminal cysteinyl protein. The *Mxe* GyrA intein was also adapted to this application [27], as well as the *Synechocystis* sp. PCC6803 DnaB helicase (*Ssp* DnaB) intein [31]. A drawback of this intein-based approach refers to the spontaneous cleavage at the intein-extein junction which can occur during expression and purification of the fusion protein. Finally, due to their self-removing nature, inteins have been exploited as auto-cleavable fusion partner in protease-free purification schemes [32–34].

Similarly to EPL, another process called Protein Trans-splicing (PTS) can be employed for the site-specific labeling of a protein. Such an approach relies on the use of a particular class of inteins which are naturally split or can be split into two pieces and, upon mixing, reassemble into a functional intein able to splice. Ligation of the proteins, which can be decorated with functional probes, fused to the split inteins occurs after the intein has been reconstituted (Figure 6) [35]. Natural or artificially split inteins have been examined to identify different possible sites of splitting. Interestingly, such investigations showed that in some inteins, such as *Mxe* GyrA, *Ssp* DnaB and *Ssp* GyrB inteins, the

splitting site can be shifted very close to intein termini [36–39], rendering N-terminal or C-terminal split intein fragments so short that the preparation of this segments and their exteins can be accomplished by SPPS, expanding the repertoire of possible chemical modifications that can be introduced into a protein.

Figure 6. Schematic representation of the protein trans splicing reaction. Target proteins (yellow and gray) are fused to N- (IntN, red) and C-terminal (IntC, green) split intein fragments. After reconstitution, intein splices releasing the two target proteins linked through an amide bond.
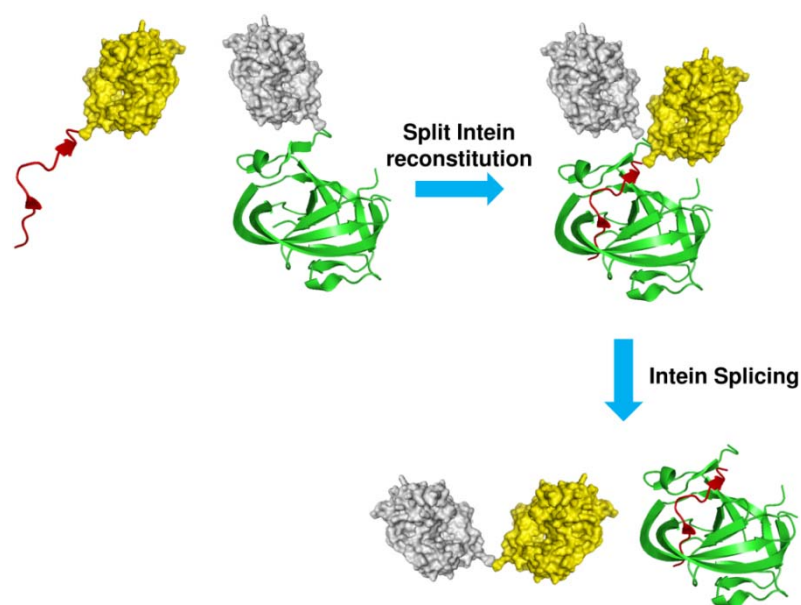
2. Fluorescent Labeling

Tagging proteins with fluorescent probes provides a tool to study proteins both *in vivo* and *in vitro* by confocal microscopy and fluorescence techniques, such as Förster resonance energy transfer (FRET), fluorescence correlation spectroscopy (FCS) and fluorescence polarization (FP). Genetic fusion to green fluorescent protein (GFP) and its color burst of spectral variants has allowed elucidation of the function of many proteins by mapping their intracellular localization and trafficking [40–44]. However, due to their large size, fluorescent fusion partners may interfere with protein folding and function. Moreover, fluorescent proteins quantum yield and photostability are quite limited [45]. Thus, it could be convenient to replace fluorescent partners with organic dyes [1,46]. Conventional methods for protein derivatization with fluorophores exploit the reactivity of native or engineered Lys or Cys side chains, as numerous dyes are commercially available as *N*-hydroxysuccinimide esters, isothiocyanates, iodoacetamides, vinyl sulfonones, maleimides and bromo-maleimides that react selectively with amino or thiol group [47,48]. However, this approach could be hampered by the presence of more than one Cys or Lys in the protein sequence, resulting in multiple labeling and heterogeneous protein preparations. Furthermore, many of such approaches suffer of a poor versatility, being tailored on a specific protein target [49,50]. Other approaches described are limited to the labeling of extra-protein sequence signal peptides, such as the reaction with FlASH, CrAsH or ReASH (biarsenical derivatives of fluorescein and resorufin) [51–53], which bind non-covalently to a short

hexapeptide containing a tetracysteine motif (CCXXCC, were X stands for any amino acid, preferentially Pro or Gly), rhodamine-derived bisboronic acid (RhoBo) dye, which reacts with a tetraserine tag sequence (SSPGSS) [54], nickel-nitriloacetic acid derivatized chromophores [55,56], which interact with hexa or decahistidine tags, fluorophores complexed to Zn^{2+}-2,2'-dipicoylamine based on L-tyrosine scaffold (Zn^{2+}-DpaTyr), which bind to oligoaspartic acid sequence [57]. Enzymatic approaches for protein labeling were also described, such as the SNAP-tag [58], CLIP-tag [59] and Halo-tag [60] technologies. Other enzymatic strategies require the fusion of the target protein to a short signal peptide, such as the reaction catalyzed by sortase [61], transglutaminase [62], biotin-ligase [63], phosphopantetheine transferase [64], lipoic acid ligase, myristoyl-CoA:protein *N*-myristoyltransferase [65]. Although of great originality and inventiveness, the above mentioned labeling procedures are limited to N- or C-protein termini and generally do not allow incorporation of multiple, different molecular probes into a protein. The drawbacks of these fluorescent labeling strategies are successfully overcome using chemical ligation approaches, which do not require the addition to the protein of extra-sequence signal peptides or labeling domains and ensure open access to any site of any protein sequence with unique level of specificity.

2.1. Protein Labeling with Fluorescent Probes for FRET Studies

A large number of fluorescent proteins have been prepared by EPL and employed for the conduction of structural and functional studies by FRET. FRET can be used to monitor dynamic processes involving protein structural changes, protein interaction or oligomerization through the variation of the spatial distance between two fluorophores. Relevant examples were reported by Muir, who pioneered the field with outstanding works [66–69]. Of great interest is his description of a solid-phase expressed protein ligation (SPPL) strategy that enabled the semi-synthesis of a large protein on a solid support. The designed approach was exploited to label the N and C termini of the Crk-II protein with the fluorescein and tetramethylrhodamine FRET pair. Crk-II was expressed in bacteria as fusion construct with the yeast vacuolar membrane ATPase (VMA) intein and with the affinity tag chitin binding domain (CBD). The fusion construct harbored an N-terminal Cys masked by the proteolysis site recognized by the factor Xa protease. The chimeric protein Xa site-Crk-II-intein-CBD was purified from cytosolic extract using a chitin affinity resin. The protein was incubated on the solid matrix with a synthetic peptide containing an N-terminal Cys residue and the probes fluorescein (Fl) and biotin, separated by the PreScission protease cleavage site (PS). The peptide reacted in an one pot reaction with the C-terminal thioester derivative of Crk-II protein, released from VMA intein splicing, affording the C-terminal Crk-II labeling with the first dye (Fl). The biotin handle was exploited to isolate the EPL product by an affinity chromatography step on an avidin-resin. Once bound to the resin, the Xa site-Crk-II-Fl PS biotin was incubated with factor Xa protease, in order to expose the N-terminal Cys while leaving the fusion construct attached to the resin. The protein underwent a second ligation reaction with a synthethic C-terminal-thioester peptide labeled with tetramethylrhodamine (Rd), affording the introduction of the second label. The doubly-labeled construct was cleaved from the resin upon incubation with biotin or with the PreScission protease. The latter method allowed the removing of the affinity biotin handle, giving the doubly-labeled species Rd-Crk-II-Fl. FRET experiments performed on dual labeled Crk-II construct before and after its phosphorylation by the

nonreceptor protein tyrosine kinase c-Abl resulted in a variation of FRET efficiency between the dyes pair, suggesting conformational changes upon phosphorylation [67]. In a successive work, a doubly labeled truncated version of Crk-II that worked as fluorescent biosensor was synthesized, enabling real-time monitoring of c-Abl kinase activity and provided a rapid tool for screening potential c-Abl kinase inhibitors [68]. A combination of EPL and selective Cys-labeling was employed by the Ebright's group to introduce the FRET pair tetramethylrhodamine and fluorescein into different subunits of the *E. coli* RNA polymerase holoenzyme [70,71]. FRET studies led to the elaboration of a structural model for the holoenzyme complex. More recently, Xie *et al.* used EPL to label the histone acetyltransferases (HATs) PCAF and p300 with Dabcyl as FRET acceptor, while HAT substrate analogues were labeled with the acceptor dye methoxycoumarin through SPPS. The molecules were subjected to FRET and fluorescence anisotropy assays to detect HATs inhibitors [72]. A FRET based approach to observe protein oligomerization was described by Scheibner *et al.* [73]. A synthetic dipeptide Cys-Lys(ε-fluorescein) (donor) or Cys-Lys(ε-rhodamine) (acceptor) was attached via EPL to the C-terminus of three recombinant proteins (glutathione S-transferase, SH2 domain phosphatase-1 and serotonin *N*-acetyltransferase) and a mixture of the two singly-labeled proteins, carrying the donor or the acceptor fluorophore, was analyzed by FRET. A similar labeling strategy was adopted in a study performed on two GTPases from the superfamily of Ras-like small GTPases, H-Ras and Ypt1, and two of their interacting partners, the Ras-binding domain (RBD) of c-Raf1 and MRS6. Target proteins were prepared as C-terminal thioester derivatives using intein fusion technology. Thioester proteins were reacted with synthetic decapeptides labeled in the solid phase with a flurophore and containing six His residues, the latter being useful to purify the ligation products. Labeled proteins were subjected to protein-protein and protein-nucleotide interaction studies by FRET and fluorescence cross-correlation spectroscopy (FCCS) [74]. Multicolor protein labeling is a useful way for the characterization of protein structure and folding by FRET studies. These studies require the attachment of a donor and an acceptor dye inside the same polypeptide chain at specific positions. An example of semi-synthetic protein doubly labeled with fluorescent dyes was reported by Yi *et al.* [75]. The described approach is limited to protein N- and C-termini and uses a combination of EPL and oxime ligation [76]. Rab7 GTPase protein was obtained as C-terminal thioester after intein splicing. The thioester protein was reacted with (bis)oxyamine moiety. The C-terminal oxamino-modified protein derivative selectively reacts with fluorophores containing a ketone functional group. To afford the second labeling, an N-terminal Cys residue was exposed by Tobacco Etch Virus (TEV) protease digestion of the protein which was then reacted by EPL with the second dye supplied as thioester derivative. The use of TEV protease to release an amino-terminal cysteinyl protein was introduced by Tolbert & Wong, who demonstrated that TEV protease could accept Cys in the P1 position of its proteolytic site rather than the usual Gly [77], enabling the use of TEV recognition site as "protecting group" for the N-terminal Cys of a recombinant protein in alternative to the factor Xa recognition site introduced by Muir [67]. In another example, Iakovenko and co-workers incorporated a fluorescent probe into a semisynthetic version of Rab7, a small GTPase [78]. The fluorophore was incorporated at the C-terminus of the protein, a region that is known to be post-translation prenylated by Rab geranylgeranyl transferase (RabGGTase). Using this approach, a library of 46 Rab7 analogs conjugated to different fluorophores was constructed and the molecules were used as sensors that report on the interaction of Rab7 with RabGGTase and the escort protein REP-1 [79]. Recently, we contributed to widen the applications of

EPL describing a general method for the incorporation of two molecular probes at different, specific positions along the protein framework [80]. The approach is schematically illustrated in Figure 7 and combines protein semi-synthesis by EPL with the labeling in solution of a convenient functional group (a thiol in our case).

Figure 7. Schematic procedure to prepare double/triple modified proteins combining EPL and chemoselective labeling in solution [80].
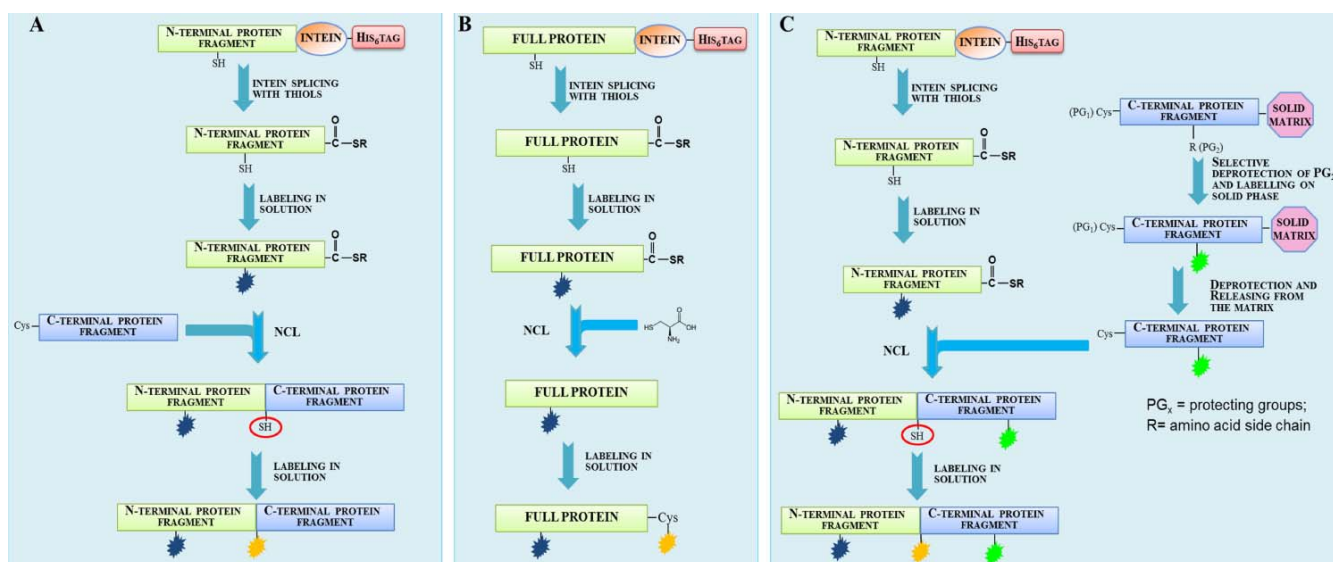
The protein of interest is split in two fragments (N- and C-terminal), possessing a single Cys residue. The N-terminal fragment fused to an intein is expressed in *E. coli* and spliced in presence of MESNA, affording the thioester derivative. Then the thioester protein is selectively labeled on its unique Cys residue, for example using a probe conjugated with the maleimide. Successively, the mono-labeled thioester fragment is reacted with an N-terminal Cys residue by a NCL reaction to afford full-length mono-labeled protein. Finally, the Cys residue involved in NCL reaction is exploited to introduce the second probe (Figure 7A). Unlike previously reported [81], both labeling reactions are performed in solution, enabling the use of mild excess of fluorophores and probes with a chemical structure that is not compatible with the harsh conditions of solid phase synthesis. A simpler version of our approach can be adopted when the second probe is located at the C-terminus (Figure 7B). In this case, the full-length protein containing a single Cys is prepared as thioester derivative and, after the first labeling in solution, is reacted with a Cys in an EPL reaction and then labeled with the second probe. As proof of concept, we successfully applied our synthetic approach to prepare four fluorescent variants of the repeat protein CTPR3 with different probes distances, confirming its efficacy and versatility. The protein variants were analyzed by CD and by ensemble-FRET. As the doubly-labeling strategy described requires RP-HPLC purification steps, the fluorescent proteins obtained are characterized by high purity and homogeneity, thus resulting amenable for single-molecule studies. This strategy could be also applied to prepare triple labeled proteins. In fact, the C-terminal fragment ending with a Cys residue could be chemically synthesized and selectively labeled on solid phase in a specific position. Then, it could be ligated to a labeled protein-thioester fragment to afford a doubly labeled species harboring a Cys residue exploitable for the introduction of the third probe (Figure 7C).

Semi-synthesis of fluorescent proteins can be also accomplished by PTS. Useful strategies have been described to tag the N- or C-terminus of a protein of interest with a chemical probe using specific split inteins characterized by a splitting site near to one end of the amino acid sequence, allowing for the preparation of the shortest domain of the intein by SPPS. For example, an engineered version of the *Ssp* DnaB intein was used by Ludwig *et al.* [38]. In this case the split site has been shifted at the amino acid 11, which is compatible with the chemical synthesis of the Int^N fragment and its extein. This approach was used to synthesize two fluorescein labeled proteins, the thioredoxin (Trx) and the β-lactamase (βLac). For the preparation of fluoresceinated Trx, the Int^C domain of *Ssp* DnaB intein (143 aa) was fused to Trx and an His$_6$tag (Int^C-Trx-His$_6$). The construct was mixed in equimolar amount with a synthetic Int^N domain (11 aa) carrying a peptide labeled with a fluorescein as N-extein (Fl-extein-Int^N). PTS yielded the fluorescent product Fl-Trx-His$_6$ with a reaction rate of about 17 folds lower than that reported for *Ssp* DnaB intein split at previously known site, probably due to the strong impact that the new splitting site has on the reassembling and refolding attitude of the *Ssp* DnaB intein. Despite the reduced rate of splicing, the final trans-splicing yield (40–45%) was similar to those observed for *Ssp* DnaB intein split at the canonical site. The use of a higher excess of the synthetic domain did not result in an improvement of the final yield. The second protein used as case-of-study, βLac, required a refolding step, as the construct Int^C-βLac-His$_6$ aggregated in inclusion bodies during expression in prokaryotic hosts. Refolding was performed in one step by dialysis and the soluble Int^C-βLac-His$_6$ was reacted with Fl-extein-Int^N, even in this case using equimolar amounts of the two reactants. PTS afforded Fl-βLac-His$_6$ with a 35% yield. Both Fl-Trx-His$_6$ and Fl-βLac-His$_6$ were subjected to enzymatic activity assays, demonstrating that the semi-synthetic origin of the enzymes did not affect their functionality. The utility of the split *Ssp* DnaB intein has also been improved through the selection by a directed evolution approach of a mutant which shows a lower restriction about sequences requirements of the flanking exteins and an higher rate of trans-splicing [82]. Furthermore, a split variant of the *Ssp* GyrB intein with the split site near its C-terminus was designed, providing an N-terminal intein domain of 150 aa and a C-terminal intein domain of 6 aa only. The use of such intein allows for the chemical modification of proteins in the C-terminal region, in a similar fashion to that described to label a protein N-terminal region with *Ssp* DnaB intein. Protein *trans*-splicing has also been exploited to introduce a FRET pair into a single protein [83]. The reported strategy combines PTS on recombinant fragments and Cys-labeling in solution. The protein of interest was split at a Ser residue in two fragments (Prot^N and Prot^C), each one bearing a single Cys residue along the amino acid sequence. Each fragment was expressed in fusion with a domain of *Npu* DnaE split intein (Int^N and Int^C) and a solubility enhancing partners. The Ser residue at which the protein of interest has been split become placed at the catalytic junction between Int^C-Prot^C. Usually, inteins prefer a Cys residue at this site (however *Npu* DnaE intein tolerates also this nucleophile) allowing to realize the protein doubly-labeling in a three steps scheme. First, Int^C-Prot^C construct was labeled at Cys with a thiol-reactive derivative of the first dye; the presence of a catalytic Ser instead of a Cys ensured to maintain intact the nucleophile at this position during Cys labeling. Then, the mono-labeled full protein was obtained through PTS by mixing Int^C-Prot^C and Prot^N-Int^N. The second dye was finally introduced onto the full protein by labeling the only remaining Cys residue. As Int^N domain also contains a Cys residue involved in the splicing reaction, to preserve intein functionality, the labeling with the second dye was performed on the full protein after ligation instead than on the intermediate construct Prot^N-Int^N. The

designed approach was validated through the synthesis of a doubly-labeled di-ubiquitin molecule, in which two ubiquitin units are joined by a linker sequence containing a Ser residue. Maltose binding protein was selected as enhancer of solubility and purification tag of the construct Int^C-Ub while Ub-Int^N was expressed in fusion with a His$_6$tag. Each unit of Ub bore a single Cys residue introduced by mutagenesis respectively at position 47 and 7. These sites were selected as they are exposed outside the Ub hydrophobic core. The doubly-labeled construct obtained was used to study the unfolding process by FRET. Recently, Muir's group performed a systematic study of 18 cyanobacterial split DnaE inteins and several of them resulted to be "ultrafast" inteins, able to catalyze protein trans-splicing in tens of seconds, as previously observed for the *Npu* DnaE intein [84]. They analyzed the effect caused by C-extein sequence variation on ultrafast inteins trans-splicing rate, observing different degree of tolerance for each intein and kind of mutation. The ultrafast inteins appear an attractive and efficient tool to prepare α-thioester proteins for EPL [85].

The general applicability of PTS approach is however limited by the instability of the split protein fragments which can aggregate in inclusion bodies and for whom refolding procedures are not always successful. Besides, many inteins are not completely promiscuous regarding their exteins, showing often strong restrictions on the sequences that they tolerate in the proximity of the junction site extein-intein, limiting the number of protein targets that can be prepared by PTS.

2.2. Protein Labeling with Fluorescent Probes for Fluorescence Microscopy in Living Cells

The ability to track the position and the movements of proteins inside a living cell is a key approach to describe cellular mechanisms and protein functions. In the last years, we assisted to an explosion of chemical biology tools to label proteins for microscopy applications. Beck-Sickinger and coworkers described the site-specific fluorescent modification by EPL of the SDF1α chemokine. SDF1α (1–49) fragment was expressed as *Mxe* GyrA intein-fusion protein. Fusion protein was purified on chitin resin, exploiting the presence of the CBD at the intein C-terminus; thioester SDF1α (1–49) was obtained on resin after incubation with MESNA. The SDF1α (50–68) C-terminal peptide, starting with a native Cys, was synthesized by SPPS and labeled on solid phase with carboxyfluorescein at the C-terminal Lys residue. It was selected for labeling as it was assumed it did not interfere with the correct protein folding. EPL between the two fragments lasted 24 h. Since SDF1α contains two disulfide bridges, after ligation the full product was refolded and oxidized using the cysteine/cystine redox system. The well folded product was purified by HPLC, as the retention time of the oxidized species shifted with respect to the linear compound. Fluorescence microscopy studies carried out on the labeled chemokine demonstrated that the semi-synthetic SDF1α is biologically functional molecule, able to induce chemotaxis and to be internalized upon specific binding to its receptor CXCR4, supporting the extension of such synthetic protocol to other chemokine and to the introduction of other modifications [86]. EPL has also been harnessed to control the activity of a post-translational modified protein inside living cells by the use of photocleavable caging groups. In this approach, EPL is used to synthesize the protein of interest modified with a fluorophore and a caging group which quenches the fluorophore and, contemporary, suppresses the protein activity. The caging group is photocleavable and thus it can be removed upon light exposure ensuring to switch on protein activity and dye fluorescence emission. For example, this strategy was applied to the protein Smad2, involved

in the transforming growth factor *α* (TGF-*α*) signaling pathway. The activation of Smad2 requires phosphorylation of two Ser residues placed at protein C-terminus. In the phopshorylated Smad2, the C-terminal carboxylate stabilizes a homotrimeric structure. The presence of a caging moiety at the C-terminus, able to disrupt the homotrimeric interactions, maintains the phosphorylated protein in an inactive state. UV irradiation induces the cleavage of the caging group and restores Smad2 activity. Smad2 analog harboring two phospho-Ser residues, a fluorescein as fluorescent probe, a dabcyl as quencher and the 4-[4-(1-hydroxyethyl)-2-methoxy-5-nitrophenoxy]butanoic acid as photo-cleavable caging group was prepared by semi-synthesis. Thioester Smad2 lacking the last five residues was prepared by intein-fusion technology, complexed with the membrane-anchored protein SARA and reacted with a synthetic peptide containing all the functional groups listed before [87]. In a subsequent work, the same concept and semi-synthetic strategy was applied to prepare caged phosphorylated and unphosphorylated Smad2, containing respectively the green fluorophore carboxyfluorescein and the red fluorophore tramethylrhodamine. Fluorescence appeared to be titratable by modulating the extent of UV-light exposition. The two proteins were co-injected into live cells and their fluorescence before and after photoactivation was monitored. This strategy enabled to track the phosphorylated and nonphosphorylated Smad2 inside the same single-living cell [88]. A C-terminal protein labeling was described by Chaisemartin *et al.* using a Cys analog linked to a N-1' fluorescent biotinyl derivative; the molecule was used to induce splicing of the fusion construct consisting of the intein *Mxe* GyrA and a scFv directed against the GTPase RhoB. Splicing afforded the fluorescent scFv which was purified to homogeneity exploiting the biotin hand [89], while fluorescent probe allowed the *in vivo* antibody detection. A "mirror" approach, useful to label protein N-terminus, has also been described and applied to the imaging in living cells [3]. In this work, the intein-mediated protein splicing allowed the *in vivo* generation of the target protein bearing an N-terminal Cys residue. EPL with membrane-permeant thioester-containing fluorophore allowed site-specific labeling of the protein. Other examples of N-terminal Cys containing proteins labeling with thioester molecules were also reported [90–92]. Protein labeling can be also accomplished *in vivo* by using protein trans-splicing, as firstly described by Giriat and Muir [93]. To validate their semi-synthetic approach, GFP was expressed in cells as fusion protein with the N-terminal fragment of *Ssp* DnaE split intein. The C-terminal *Ssp* DnaE intein fragment was instead fused to the FLAG epitope and supplied to the cell media. The latter construct was able to penetrate through cell wall thanks to a Protein Transduction Domain (PTD) peptide which served as a signal peptide and which was ligated to Int[C]-FLAG through a disulfide bridge. Western-blot analysis of the cell lysate with anti-GFP and anti-FLAG antibodies demonstrated the ability of the intein to reconstitute *in vivo* and splice giving the GFP-FLAG as product. The naturally occurring split intein *Npu* DnaE, characterized by the highest rate of trans-splicing reaction, has also been adopted to ligate an exogenous polypeptide to a membrane protein exposed by living cells, offering a useful way to modify a cell surface protein. The 36 amino acids C-terminal domain of the *Npu* DnaE intein (Int[C]) was expressed in eukaryotic cells as fusion with a transmembrane domain derived from PDGF receptor (TM), a soluble partner (Trx) and the fluorescent protein mCherry. Microscopy analysis on living cells showed that the construct Int[C]-Trx-TM-mCherry was localized on plasma membrane. Cells were incubated with a green fluorescent protein carrying the complementary moiety of the *Npu* DnaE intein (102 amino acids), the construct eGFP-Int[N], which was expressed in *E. coli* and purified. Specific co-localization of mCherry and eGFP proteins at cell surface was verified by confocal microscopy,

demonstrating that *Npu* DnaE intein was able to reconstitute into a functional splicing unit in a cellular context. *Npu* DnaE intein has two catalytic cysteines involved in PTS reaction that must be reduced to ensure successful trans-splicing. Authors observed that PTS was not compromised if the contruct eGFP-Int[N] was preincubated with DTT and dyalized against PBS before the incubation with cells. Direct incubation of cells with DTT should be avoided as thiols may damage cell membranes. The semi-synthetic approach was extended to the preparation of proteins attached to the membrane through a glycosylphosphatidylinositol (GPI) anchor. To this scope, the Int[C] domain was expressed endogenously in fusion with the GPI signaling sequence Gas1p, which triggers its attachment to a glycolipid anchor on the cell membrane. Incubation with eGFP-Int[N] allowed membrane labeling through a GPI-anchor [94].

3. Isotopic Labeling

Structural characterization by NMR spectroscopy may be not trivial if applied to large proteins, due to significant loss of spectral resolution and chemical shifts overlap. Both effects are progressively amplified as the protein size increases, hampering unambiguous signals assignment. The synthesis of large proteins by ligation methods, such as EPL, may resolve this issue. Through a synthetic approach, named segmental isotopic labeling, a selected portion of a protein, such as a single protein domain or region, can be specifically labeled with ^{13}C, ^{15}N and/or ^{2}H isotopes and analyzed by NMR spectroscopy in the context of the native full protein (Figure 8). Segmental labeling allows to reduce the complexity of the NMR spectrum as only the signals of the labeled region are revealed, facilitating the structural analysis of the labeled protein portion.

Figure 8. Semi-synthetic strategy for isotope labeling of an internal protein domain. The labeled protein domain is reported in yellow. PG is a protease recognition site which releases a N-terminal cysteinyl-protein.
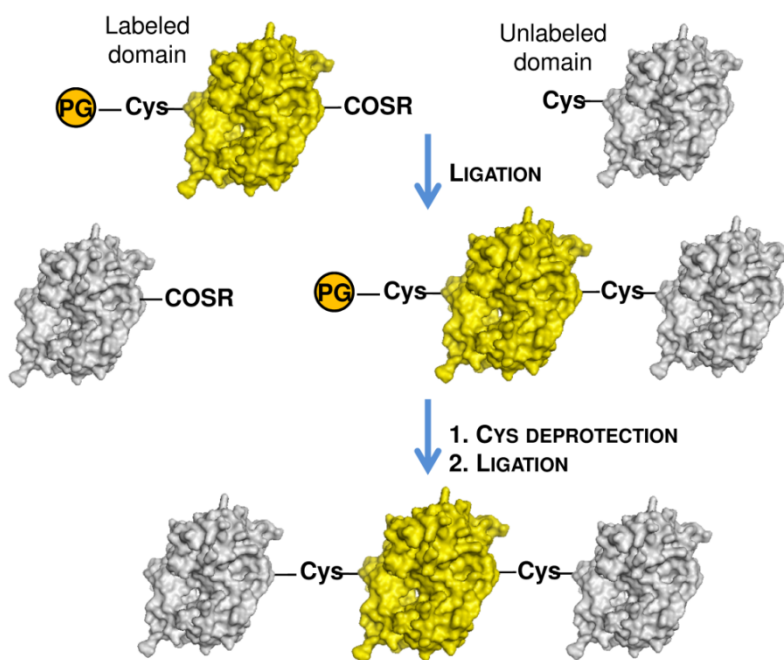
Muir's group reported the first example of segmental labeling by EPL [95,96]. In this work, EPL was used to prepare the Abelson protein tyrosine kinase-SH(3,2) domain pair, in which only one of the

two domains was labeled with ^{15}N. Unlabeled Abl-SH3 domain was prepared as C-terminal ethyl thioester using intein technology and reacted with N-terminal Cys containing (^{15}N)Abl-SH2. A combination of mass spectrometry and NMR spectroscopy techniques was used to confirm the identity and the correct structure of the ligation product, which also showed the appropriate ligand-binding properties. The same semi-synthetic procedure was adopted by Camarero *et al.* to isotopically label the C-terminal region of a sigma70-like subunit from *Thermotoga maritima* in order to study the mechanism of autoinhibition mediated by its own N-terminal 90 amino acids by NMR [97]. More recently, the Gierasch's group described the segmental isotopic labeling of the *E. coli* Hsp70 molecular chaperone DnaK [98]. One of its two constituent domains, the ATPase domain, was expressed as N-terminal fusion with the *Mxe* GyrA intein in minimal medium containing ^{2}H, ^{13}C and ^{15}N isotopes sources, generating a labeled C-terminal thioester derivative after incubation with the thiol MESNA. The complementary substrate-binding domain (SBD) was expressed as C-terminal fusion with *Ssp* DnaB intein to generate, after intein splicing, a N-terminal Cys residue. The two domains were then ligated, affording full DnaK bearing the ATPase domain labeled with NMR active isotopes. In order to optimize ligation yield, many conditions were tested, evaluating the effects of varying pH, temperature, reagents concentration and ratio, denaturant concentration. Interestingly, they found appropriate to carry out the ligation reaction at low urea concentration (2 M) to enhance ligation efficiency, while higher or lower denaturant concentration resulted in a decreased ligation yield. They interpreted these results hypothesizing that urea concentrations higher than 2 M disrupt specific domain-domain interactions which bring together the ligands and promote ligation. At lower urea concentrations, instead, domains interaction persists but ligation yield falls down due to poor accessibility of C-terminal thioester group and N-Cys of the two domains. The former effect does not affect ligation yield if the reaction is carried out using millimolar reagent concentrations and an excess of SBD domain.

Vitali *et al.* described an interesting segmental labeling reaction performed on resin, which requires the concerted use of two inteins, the *Mxe* GyrA and *Ssp* DnaB inteins. The synthetic procedure was adopted to prepare and characterize by NMR a series of segmentally labeled variants of the RNA recognition motifs (RRM3 and RRM4) of the polypyrimidine tract binding protein (PTB). The RRM3 motif was expressed in fusion with *Mxe* GyrA intein and a CBD as affinity partner, while RRM4 was fused at the C-terminus of the *Ssp* DnaB intein and at the N-terminus to a CBD. As both fusion proteins bore the same affinity tag, they were co-loaded onto a chitin resin and the splicing of each intein was induced. Firstly, *Mxe* GyrA intein was spliced upon incubation with MESNA, releasing a C-terminal α-thioester RRM3. Then, a temperature switch to 37 °C allowed *Ssp* DnaB intein splicing, leaving RRM4 domain with an amino-terminal Cys residue. Once released, the two domains promptly reacted, affording a unique polypeptide chain. The full-length product was finally purified on Ni^{2+}-resin, exploiting the His$_6$ tag placed upstream RRM3 as additional affinity moiety. The ability of the two domains to interact facilitated their successful ligation. Segmental labeling could simply be accomplished loading onto the chitin resin an isotopically labeled RRM-Intein-CBD construct. By the use of the described on column procedure, many segmentally-labeled variants of RRM34 were prepared carrying ^{15}N,^{13}C-labeled RRM3 and an unlabeled RRM4, unlabeled RRM3 and a ^{15}N,^{13}C-labeled RRM4 and a ^{15}N,^{13}C-labeled RRM3 and a ^{15}N-only-labeled RRM4 [99]. The same procedure was adopted and improved by Skrisovska and Allain, who described the characterization of two different multidomain proteins containing RNA recognition motifs (RRMs), heterogeneous nuclear ribonucleoprotein L and

Npl3p. In this work, the splicing-ligation on column procedure was extended to insoluble protein, adding a refolding step before protein binding on resin, and to non-interacting domains, for which full-length protein yield was enhanced by eluting from the resin and concentrating the non-ligated reactants. The proteins were obtained in high yields, allowing to characterize their interdomain interactions by NMR spectroscopy [100]. An elegant approach was also proposed by Zhao *et al.* that reported an efficient on-column EPL strategy for the semi-synthesis of human apolipoprotein E (apoE) triply-labeled with ^{2}H, ^{15}N and ^{13}C. The protein was expressed in two fragments. The N-terminal portion of ApoE was expressed as intein fusion construct and also bore a CBD as affinity tag. After binding on a chitin-resin, the fusion protein was incubated with the C-terminal ApoE fragment, also prepared by recombinant means and harboring a N-terminal Cys, to perform on-resin ligation reaction. Using this on-column ligation approach, once generated from intein thiolysis, the thioester species may readily react with the N-terminal Cys-fragment, strongly reducing the probability of hydrolysis of the thioester group and increasing the ligation yield. By this protocol, several variants of ApoE were prepared combining the two fragments labeled with three different NMR active nuclei, ^{2}H, ^{15}N or ^{13}C [101,102]. Another biological problem which was addressed by segmental labeling is the study of ubiquitin biology. The attachment of a ubiquitin (Ub) or of poly-Ub to a target protein may regulate a great number of cellular processes, such as protein proteasomal degradation, transcriptional activation, vesicular trafficking of proteins. Poly-Ub chains are made by linking to each other Ub monomers through an isopeptide bond between the ε-amino group of one of the seven Lys residues of a monomer with the C-terminus of the next one or head-to-tail. Depending on the Lys involved into poly-Ub assembling, poly-Ub chains may work as molecular signals for different biological processes. Probably, the specificity of each poly-Ub chains depends on the different conformation that each one assumes depending on the sites of ramification. The characterization of the interaction of different poly-Ub with their receptors requires the ability to prepare homogeneous preparation of poly-Ub with a defined linkage between each Ub monomer. EPL fulfills such request, being more selective and versatile than enzymatic methods. However, due to its homo-polymeric nature, NMR studies of poly-Ub chains represent a challenging task. Segmental labeling offers a solution to the problem, allowing to observe each single Ub monomer in the context of the branched poly-Ub. Castaneda *et al.* reported the semi-synthesis of a segmentally isotopic labeled Ub$_2$ chains. A Ub monomer carrying a δ-mercaptolysine at two positions (33 or 48) was prepared by chemical synthesis and reacted with a C-terminal thioester (^{15}N)-labeled Ub monomer, obtained by recombinant expression as intein fusion protein. EPL allowed the formation of an isopeptide bond between the two monomers. After ligation, the thiol group in δ-position of the Lys was removed by desulfurization, affording a native- like di-Ub molecule [103]. The variant in which the two Ub monomers are linked through Lys48 was used as control construct, as this poly-Ub has been widely characterized by NMR. Data obtained using the segmentally labeled synthetic K48-linked di-Ub demonstrated that it is structurally identical to the one assembled using enzyme, confirming that the protein chains obtained by EPL are ''native-like'' and paving the way to the structural, conformational and ligand-binding properties characterization by NMR of the K33-linked di-Ub variant. EPL is a useful approach even for the site-specific introduction of one or few stable isotopes within a protein. Romanelli *et al.* applied the segmental isotopic labeling to selectively dual label the scissile peptide bond at the N-extein/intein junction of *Mxe* GyrA intein with ^{13}C and ^{15}N nuclei. ^{15}N uniformly labeled *Mxe* GyrA intein was ligated to a synthetic pentapeptide

labeled with ^{13}C only on the C-terminal carboxylic group (Figure 9). On such construct they were able to measure the amide $^1J_{NC'}$ coupling constant which was found to be of 12.3 Hz. This result suggests that this amide bond is unusually polarized, because of non-planarity, allowing one to explain the extreme lability of the N-extein-intein bond which is broken in the first step of protein splicing. Additional studies demonstrated that a conserved His residue of intein block B is absolutely required to catalyze the first step of the splicing reaction and contributes to maintain the (-1) scissile bond in its unusual conformation [104]. Further investigations on *Mxe* GyrA intein splicing mechanism were carried out performing NMR studies on a semi-synthetic branched intermediate prepared ligating a recombinant 1–184 α-thioester intein fragment with a synthetic branched peptide reproducing the intein remaining portion (aa 185–198) with the C-extein and carrying the N-extein attached to the side chain of a Thr. Furthermore, the ^{13}C and ^{15}N isotopes were specifically incorporated into the scissile +1 peptide bond [105]. In another example, a series of ^{13}C-labeled amino acids were incorporated at the C-terminus of the α-subunit of a heterotrimeric G protein [106]. Using EPL, 9-mer peptides containing ^{13}C labels in Leu-348 (uniform), Gly-352 (alpha carbon), and Phe-354 (ring) were ligated to recombinant Gα subunit lacking the corresponding carboxyl-terminal residues. Analysis of the ^{13}C resonances indicated that the C-terminus of the Gα subunit is unstructured when the protein is bound to GDP, but adopts an ordered conformation upon activation by AlF$_4^-$.

Figure 9. Semi-synthesis of *Mxe* GyrA intein with the scissile (-1) peptide bond dual labeled with ^{13}C and ^{15}N. *Mxe* GyrA intein (green) was uniformly labeled with ^{15}N. Extein peptide (yellow) was synthesized by SPPS with a ^{13}C α-carboxylate.
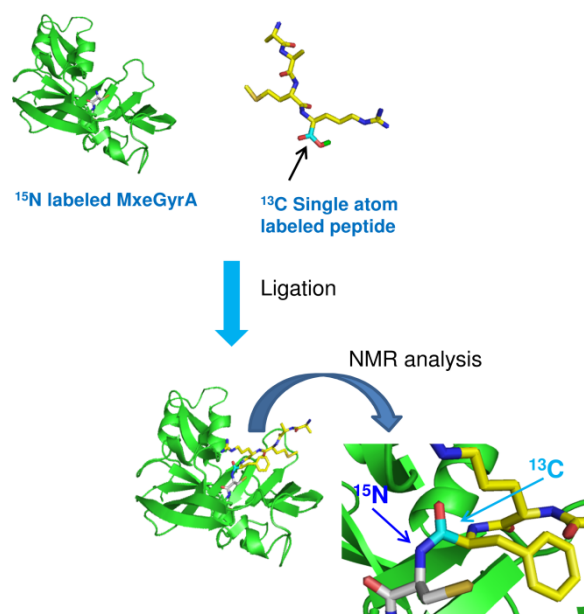
EPL allows even for the segmental labeling of a protein internal region. To this purpose EPL makes use of sequential-ligation tools, as firstly described by Muir's group who semi-synthesized the 304 amino acids Crk-II protein through the sequential assembling of three recombinant polypeptides. In such approach, the central segment bears both a C-terminal thioester and a "cryptic" N-terminal Cys masked by a protease cleavage site which prevents the self-reaction of the central polypeptide. After EPL between the central thioester segment and the C-terminal polypeptide, protease cleavage allows to

reveal the α-Cys in the intermediate protein. Reaction with N-terminal segment, harboring the C-terminal thioester function, may afford the full protein [107]. Zaragöz *et al.* applied EPL to the semi-synthesis of the dimeric protein Hsp90, evaluating the effect on ligation yield of varying critic parameters such as reactants concentration, pH, ligation site. They developed an EPL based protocol suitable for the preparation of segmentally labeled dimeric proteins [108]. Another application of segmental labeling refers to infrared spectroscopy. Moran *et al.* applied segmental labeling to the study of human γD-crystallin amyloid fibrils by two-dimensional IR spectroscopy. EPL was used to uniformly label with ^{13}C one of the two Greek key domains in order to individually resolve them in (2D) IR spectra of acid-induced amyloid fibrils [109,110]. Segmental labeling can also be accomplished by intein trans-splicing. This approach has been, recently, exploited by the Iwai's group which described a strategy to label an internal region of a protein of interest using the split intein *Npu* DnaE [111] and more recently the preparation of a segmentally labeled version of the multi-domain CheA protein from *E. coli* both by EPL and trans-splicing [112]. In the latter work, pTWIN vector was used to express the two domains of the 140 kDa protein CheA respectively as *Mxe* GyrA and *Ssp* DnaB fusion constructs. EPL between the two domains was performed in native conditions and thus gave a ligation yield of only 10%–40% due to the impossibility to reach high reactant concentrations under non-denaturing conditions. PTS performed using *Npu* DnaE intein gave better results because the trans-splicing reaction does not require high reactant concentrations as the two intein fragments show a very high affinity. Intein trans-splicing mediated segmental labeling was also applied to study F_1-ATPase, whose characterization by NMR is hampered by its amino acid length. A segmental labeling procedure based on the use of PI-*pfu*I intein was applied to label the β subunit of the protein and allowed to successfully assign the great majority of signals of the beta subunit monomer. The structures elaborated from the collected NMR data suggested that the subunit beta monomer assumes the open form in the absence of the nucleotide, while nucleotide binding induces a conformational change from the open to the closed form. The structural change of the beta subunit monomer induced by nucleotide binding triggers the rotation of F_1-ATPase [113]. In a subsequent work, the segmentally labeled F1-ATPase was studied to gain deeper insights into the rotation of F1-ATPase, driven by the open/close bending motion of the β subunit [114]. An interesting approach has been described by the Iwai's group to segmentally label a protein by *in vivo* PTS [115–117]. In this synthetic scheme the two protein fragments of a target protein are expressed in fusion with the N or C domain of a split intein. Each fusion construct is placed under the control of an inducible promoter, for instance the T7 and arabinose promoters respectively inducible by IPTG and arabinose. Bacterial cells co-transformed with both the recombinant plasmids are cultured and subjected to two sequential induction steps. First, labeled domain expression is induced in minimal medium containing isotopes source; then, cells are harvested by centrifugation, resuspended in rich culture medium and then expression of the complementary domain is induced. *In vivo* PTS affords segmentally labeled full-length protein target. Such approach has been used by Iwai's group to prepare a ^{15}N-segmentally labeled variant of the c-CRKII adaptor domain using the split intein *Npu* Dna E [115]. Segmental labeling by PTS has also been adopted to overcome the solubility limit of a protein sample for NMR studies. In fact, a common obstacle to NMR studies of proteins is the preparation of samples in soluble form at high concentration, usually in the millimolar range. To improve protein solubility, fusion of the target protein to a solubility enhancement tag (SET), such as the glutathione S-transferase, the maltose-binding protein or the thioredoxin, allows one

to overcome the problem and at the same time facilitates the purification procedures. However, many extra signals arise from the SET, complicating the NMR spectrum. Segmental isotopic labeling offers the right solution to the quest trough the so called "invisible SET'' approach, referring to the preparation of a SET fusion protein in which solely the target protein is isotopes enriched. Using *in vivo* intein trans-splicing, the non-isotope labeled SET protein GB1 (*Streptococcus* Protein G B1 domain) was fused to ^{15}N-labeled chitin binding domain (CBD), improving its solubility for a study by solution NMR spectroscopy [117]. Although of great originality, the shortcoming of such approach is that the target protein does not possess a stabilizing fusion partner during expression and thus it may aggregate into inclusion bodies before *in vivo* ligation, becoming unable to react with the SET. In order to overcome such limitation, Kobayashi *et al.* developed an alternative invisible SET approach, always based on intein trans-splicing mechanism. The target protein, the ribosome binding factor A (RbfA), was expressed in isotopes enriched medium as fusion construct with a tandem repeated N-terminal domains of the protein S from *Myxococcus xanthus*, previously reported to be a useful SET [118], and with the C-terminal domain of a split intein. The same SET protein was expressed fused to the split N-terminal domain of the intein. Intein trans-splicing between them allowed to swap the labeled SET from RbfA with an unlabeled one. Using this approach, RbfA soluble expression is ensured by fusion with protein S tag even during expression phase and NMR characterization is facilitated by suppressing SET signal trough segmental labeling [119].

4. Conclusions

Protein semi-synthesis by EPL is nowadays a mature technology, it has demonstrated its utility to solve numerous scientific issues, especially in chemical biology and biophysics. In particular, EPL is one of the favorite methods to prepare high pure and homogenous proteins modified with specific probes, as fluorescent dyes or isotopes. This aim is achieved developing several ingenious synthetic strategies, indicative of the flexibility of EPL technology. In fact, EPL combines the advantages of organic chemistry with the easy and size limitless of recombinant protein expression, enabling a single protein to be functionalized at one or even more specific positions with different probes irrespective of its size. In definitive, EPL is an excellent strategy for the chemical synthesis of labeled protein which can find application in wide variety of life-science laboratories.

References

1. Stephanopoulos, N.; Francis, M.B. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **2011**, *7*, 876–884.
2. Basle, E.; Joubert, N.; Pucheault, M. Protein chemical modification on endogenous amino acids. *Chem. Biol.* **2010**, *17*, 213–227.
3. Chattopadhaya, S.; Abu Bakar, F.B.; Yao, S.Q. Use of intein-mediated protein ligation strategies for the fabrication of functional protein arrays. *Meth. Enzymol.* **2009**, *462*, 195–223.
4. Hinner, M.J.; Johnsson, K. How to obtain labeled proteins and what to do with them. *Curr. Opin. Biotechnol.* **2010**, *21*, 766–776.
5. Kent, S.B. Total chemical synthesis of proteins. *Chem. Soc. Rev.* **2009**, *38*, 338–351.

6.   Muralidharan, V.; Muir, T.W. Protein ligation: An enabling technology for the biophysical analysis of proteins. *Nat. Methods* **2006**, *3*, 429–438.

7.   Merrifield, B.R. Solid Phase Peptide Synthesis. I. The synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.

8.   Schnolzer, M.; Kent, S.B. Constructing proteins by dovetailing unprotected synthetic peptides: Backbone-engineered HIV protease. *Science* **1992**, *256*, 221–225.

9.   Liu, C.F.; Tam, J.P. Peptide segment ligation strategy without use of protecting groups. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 6584–6588.

10.   Liu, C.F.; Tam, J.P. Chemical ligation approach to form a peptide bond between unprotected peptide segments. Concept and model study. *J. Am. Chem. Soc.* **1994**, *116*, 4149–4153.

11.   Robey, F.A.; Fields, R.L. Automated synthesis of *N*-bromoacetyl-modified peptides for the preparation of synthetic peptide polymers, peptide-protein conjugates, and cyclic peptides. *Anal. Biochem.* **1989**, *177*, 373–377.

12.   Smyth, D.G.; Blumenfeld, O.O.; Konigsberg, W. Reactions of *N*-Ethylmaleimide with Peptides and Amino Acids. *Biochem. J.* **1964**, *91*, 589–595.

13.   Rostovtsev, V.V.; Green, L.G.; Fokin, V.V.; Sharpless, K.B. A stepwise huisgen cycloaddition process: Copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 2596–2599.

14.   Tornoe, C.W.; Christensen, C.; Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-Triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **2002**, *67*, 3057–3064.

15.   De Araujo, A.D.; Palomo, J.M.; Cramer, J.; Seitz, O.; Alexandrov, K.; Waldmann, H. Diels-Alder ligation of peptides and proteins. *Chem. Eur. J.* **2006**, *12*, 6095–6109.

16.   Dawson, P.E.; Kent, S.B. Synthesis of native proteins by chemical ligation. *Annu. Rev. Biochem.* **2000**, *69*, 923–960.

17.   Dirksen, A.; Dawson, P.E. Expanding the scope of chemoselective peptide ligations in chemical biology. *Curr. Opin. Chem. Biol.* **2008**, *12*, 760–766.

18.   Dawson, P.E.; Muir, T.W.; Clark-Lewis, I.; Kent, S.B. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266*, 776–779.

19.   Nilsson, B.L.; Kiessling, L.L.; Raines, R.T. Staudinger ligation: A peptide from a thioester and azide. *Org. Lett.* **2000**, *2*, 1939–1941.

20.   Muir, T.W. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **2003**, *72*, 249–289.

21.   Muir, T.W.; Sondhi, D.; Cole, P.A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705–6710.

22.   Xu, M.Q.; Perler, F.B. The mechanism of protein splicing and its modulation by mutation. *EMBO J.* **1996**, *15*, 5146–5153.

23.   Flavell, R.R.; Muir, T.W. Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. *Acc. Chem. Res.* **2009**, *42*, 107–116.

24.   Cowburn, D.; Muir, T.W. Segmental isotopic labeling using expressed protein ligation. *Meth. Enzymol.* **2001**, *339*, 41–54.

25. Ziaco, B.; Pensato, S.; D'Andrea, L.D.; Benedetti, E.; Romanelli, A. Semisynthesis of dimeric proteins by expressed protein ligation. *Org. Lett.* **2008**, *10*, 1955–1958.

26. Xiao, J.; Hamilton, B.S.; Tolbert, T.J. Synthesis of N-terminally linked protein and peptide dimers by native chemical ligation. *Bioconjug. Chem.* **2010**, *21*, 1943–1947.

27. Southworth, M.W.; Amaya, K.; Evans, T.C.; Xu, M.Q.; Perler, F.B. Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein. *Biotechniques* **1999**, *27*, 110–114,116,118–120.

28. Telenti, A.; Southworth, M.; Alcaide, F.; Daugelat, S.; Jacobs, W.R., Jr.; Perler, F.B. The Mycobacterium xenopi GyrA protein splicing element: Characterization of a minimal intein. *J. Bacteriol.* **1997**, *179*, 6378–6382.

29. Valiyaveetil, F.I.; MacKinnon, R.; Muir, T.W. Semisynthesis and folding of the potassium channel KcsA. *J. Am. Chem. Soc.* **2002**, *124*, 9113–9120.

30. Johnson, E.C.; Kent, S.B. Insights into the mechanism and catalysis of the native chemical ligation reaction. *J. Am. Chem. Soc.* **2006**, *128*, 6640–6646.

31. Mathys, S.; Evans, T.C.; Chute, I.C.; Wu, H.; Chong, S.; Benner, J.; Liu, X.Q.; Xu, M.Q. Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: Facile production of protein building blocks for protein ligation. *Gene* **1999**, *231*, 1–13.

32. Sun, Z.; Chen, J.; Yao, H.; Liu, L.; Wang, J.; Zhang, J.; Liu, J.N. Use of Ssp dnaB derived mini-intein as a fusion partner for production of recombinant human brain natriuretic peptide in Escherichia coli. *Protein Expr. Purif.* **2005**, *43*, 26–32.

33. Esipov, R.S.; Stepanenko, V.N.; Chupova, L.A.; Boyarskikh, U.A.; Filipenko, M.L.; Miroshnikov, A.I. Production of recombinant human epidermal growth factor using Ssp dnaB mini-intein system. *Protein Expr. Purif.* **2008**, *61*, 1–6.

34. Yan, S.S.; Yan, J.; Shi, G.; Xu, Q.; Chen, S.C.; Tian, Y.W. Production of native protein by using *Synechocystis* sp. PCC6803 DnaB mini-intein in Escherichia coli. *Protein Expr. Purif.* **2005**, *40*, 340–345.

35. Wu, H.; Hu, Z.; Liu, X.Q. Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9226–9231.

36. Appleby, J.H.; Zhou, K.; Volkmann, G.; Liu, X.Q. Novel split intein for trans-splicing synthetic peptide onto C terminus of protein. *J. Biol. Chem.* **2009**, *284*, 6194–6199.

37. Aranko, A.S.; Zuger, S.; Buchinger, E.; Iwai, H. *In vivo* and *in vitro* protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS One* **2009**, *4*, e5185.

38. Ludwig, C.; Pfeiff, M.; Linne, U.; Mootz, H.D. Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing. *Angew. Chem. Int. Ed. Engl.* **2006**, *45*, 5218–5221.

39. Sun, W.; Yang, J.; Liu, X.Q. Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J. Biol. Chem.* **2004**, *279*, 35281–35286.

40. Lippincott-Schwartz, J.; Patterson, G.H. Development and use of fluorescent protein markers in living cells. *Science* **2003**, *300*, 87–91.

41. Tsien, R.Y. The green fluorescent protein. *Annu. Rev. Biochem.* **1998**, *67*, 509–544.

42. Shaner, N.C.; Steinbach, P.A.; Tsien, R.Y. A guide to choosing fluorescent proteins. *Nat. Methods* **2005**, *2*, 905–909.

43. Tsien, R.Y. Imagining imaging's future. *Nat. Rev. Mol. Cell Biol.* **2003**, *4*, SS16–SS21.

44. Zhang, J.; Campbell, R.E.; Ting, A.Y.; Tsien, R.Y. Creating new fluorescent probes for cell biology. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 906–918.

45. Yeo, D.S.; Srinivasan, R.; Chen, G.Y.; Yao, S.Q. Expanded utility of the reaction. *Chem. Eur. J.* **2004**, *10*, 4664–4672.

46. Chen, I.; Ting, A.Y. Site-specific labeling of proteins with small molecules in live cells. *Curr. Opin. Biotechnol.* **2005**, *16*, 35–40.

47. Hermanson, G.T. *Bioconjugate Techniques*; Academic Press: San Diego, CA, USA, 1996.

48. Smith, M.E.; Schumacher, F.F.; Ryan, C.P.; Tedaldi, L.M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J.R. Protein modification, bioconjugation, and disulfide bridging using bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132*, 1960–1965.

49. Kao, M.W.; Yang, L.L.; Lin, J.C.; Lim, T.S.; Fann, W.; Chen, R.P. Strategy for efficient site-specific FRET-dye labeling of ubiquitin. *Bioconjug. Chem.* **2008**, *19*, 1124–1126.

50. Smith, J.J.; Conrad, D.W.; Cuneo, M.J.; Hellinga, H.W. Orthogonal site-specific protein modification by engineering reversible thiol protection mechanisms. *Protein Sci.* **2005**, *14*, 64–73.

51. Griffin, B.A.; Adams, S.R.; Tsien, R.Y. Specific covalent labeling of recombinant protein molecules inside live cells. *Science* **1998**, *281*, 269–272.

52. Adams, S.R.; Campbell, R.E.; Gross, L.A.; Martin, B.R.; Walkup, G.K.; Yao, Y.; Llopis, J.; Tsien, R.Y. New biarsenical ligands and tetracysteine motifs for protein labeling *in vitro* and *in vivo*: Synthesis and biological applications. *J. Am. Chem. Soc.* **2002**, *124*, 6063–6076.

53. Cao, H.; Chen, B.; Squier, T.C.; Mayer, M.U. CrAsH: A biarsenical multi-use affinity probe with low non-specific fluorescence. *Chem. Commun. (Camb)* **2006**, *2006*, 2601–2603.

54. Halo, T.L.; Appelbaum, J.; Hobert, E.M.; Balkin, D.M.; Schepartz, A. Selective recognition of protein tetraserine motifs with a cell-permeable, pro-fluorescent bis-boronic acid. *J. Am. Chem. Soc.* **2009**, *131*, 438–439.

55. Uchinomiya, S.H.; Nonaka, H.; Fujishima, S.H.; Tsukiji, S.; Ojida, A.; Hamachi, I. Site-specific covalent labeling of His-tag fused proteins with a reactive Ni(II)-NTA probe. *Chem. Commun. (Camb)* **2009**, *2009*, 5880–5882.

56. Strunk, J.J.; Gregor, I.; Becker, Y.; Lamken, P.; Lata, S.; Reichel, A.; Enderlein, J.; Piehler, J. Probing protein conformations by in situ non-covalent fluorescence labeling. *Bioconjug. Chem.* **2009**, *20*, 41–46.

57. Ojida, A.; Honda, K.; Shinmi, D.; Kiyonaka, S.; Mori, Y.; Hamachi, I. Oligo-Asp tag/Zn(II) complex probe as a new pair for labeling and fluorescence imaging of proteins. *J. Am. Chem. Soc.* **2006**, *128*, 10452–10459.

58. Keppler, A.; Gendreizig, S.; Gronemeyer, T.; Pick, H.; Vogel, H.; Johnsson, K. A general method for the covalent labeling of fusion proteins with small molecules *in vivo*. *Nat. Biotechnol.* **2003**, *21*, 86–89.

59. Gautier, A.; Juillerat, A.; Heinis, C.; Correa, I.R., Jr.; Kindermann, M.; Beaufils, F.; Johnsson, K. An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* **2008**, *15*, 128–136.

60. Los, G.V.; Encell, L.P.; McDougall, M.G.; Hartzell, D.D.; Karassina, N.; Zimprich, C.; Wood, M.G.; Learish, R.; Ohana, R.F.; Urh, M.; *et al*. HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **2008**, *3*, 373–382.

61. Antos, J.M.; Chew, G.L.; Guimaraes, C.P.; Yoder, N.C.; Grotenbreg, G.M.; Popp, M.W.; Ploegh, H.L. Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.* **2009**, *131*, 10800–10801.

62. Lin, C.W.; Ting, A.Y. Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins *in vitro* and on the surface of living cells. *J. Am. Chem. Soc.* **2006**, *128*, 4542–4543.

63. Chen, I.; Howarth, M.; Lin, W.; Ting, A.Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2005**, *2*, 99–104.

64. Vivero-Pol, L.; George, N.; Krumm, H.; Johnsson, K.; Johnsson, N. Multicolor imaging of cell surface proteins. *J. Am. Chem. Soc.* **2005**, *127*, 12770–12771.

65. Heal, W.P.; Jovanovic, B.; Bessin, S.; Wright, M.H.; Magee, A.I.; Tate, E.W. Bioorthogonal chemical tagging of protein cholesterylation in living cells. *Chem. Commun. (Camb)* **2011**, *47*, 4081–4083.

66. Cotton, G.J.; Muir, T.W. Peptide ligation and its application to protein engineering. *Chem. Biol.* **1999**, *6*, R247–R256.

67. Cotton, G.J.; Muir, T.W. Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation. *Chem. Biol.* **2000**, *7*, 253–261.

68. Hofmann, R.M.; Cotton, G.J.; Chang, E.J.; Vidal, E.; Veach, D.; Bornmann, W.; Muir, T.W. Fluorescent monitoring of kinase activity in real time: Development of a robust fluorescence-based assay for Abl tyrosine kinase activity. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 3091–3094.

69. Muralidharan, V.; Cho, J.; Trester-Zedlitz, M.; Kowalik, L.; Chait, B T.; Raleigh, D.P.; Muir, T.W. Domain-specific incorporation of noninvasive optical probes into recombinant proteins. *J. Am. Chem. Soc.* **2004**, *126*, 14004–14012.

70. Mukhopadhyay, J.; Kapanidis, A.N.; Mekler, V.; Kortkhonjia, E.; Ebright, Y.W.; Ebright, R.H. Translocation of sigma(70) with RNA polymerase during transcription: Fluorescence resonance energy transfer assay for movement relative to DNA. *Cell* **2001**, *106*, 453–463.

71. Mekler, V.; Kortkhonjia, E.; Mukhopadhyay, J.; Knight, J.; Revyakin, A.; Kapanidis, A.N.; Niu, W.; Ebright, Y.W.; Levy, R.; Ebright, R.H. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **2002**, *108*, 599–614.

72. Xie, N.; Elangwe, E.N.; Asher, S.; Zheng, Y.G. A dual-mode fluorescence strategy for screening HAT modulators. *Bioconjug. Chem.* **2009**, *20*, 360–366.

73. Scheibner, K.A.; Zhang, Z.; Cole, P.A. Merging fluorescence resonance energy transfer and expressed protein ligation to analyze protein-protein interactions. *Anal. Biochem.* **2003**, *317*, 226–232.

74. Becker, C.F.; Seidel, R.; Jahnz, M.; Bacia, K.; Niederhausen, T.; Alexandrov, K.; Schwille, P.; Goody, R.S.; Engelhard, M. C-terminal fluorescence labeling of proteins for interaction studies on the single-molecule level. *Chembiochem* **2006**, *7*, 891–895.

75. Yi, L.; Sun, H.; Itzen, A.; Triola, G.; Waldmann, H.; Goody, R.S.; Wu, Y.W. One-pot dual-labeling of a protein by two chemoselective reactions. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 8287–8290.

76. Yi, L.; Sun, H.; Wu, Y.W.; Triola, G.; Waldmann, H.; Goody, R.S. A highly efficient strategy for modification of proteins at the C terminus. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 9417–9421.

77. Tolbert, T.J.; Wong, C.H. New methods for proteomic research: Preparation of proteins with N-terminal cysteines for labeling and conjugation. *Angew. Chem. Int. Ed. Engl.* **2002**, *41*, 2171–2174.

78. Iakovenko, A.; Rostkova, E.; Merzlyak, E.; Hillebrand, A.M.; Thomä, N.H.; Goody, R.S.; Alexandrov, K. Semi-synthetic Rab proteins as tools for studying intermolecular interactions. *FEBS Lett.* **2000**, *468*, 155–158.

79. Wu, Y.W.; Goody, R.S.; Alexandrov, K. Intein-mediated construction of a library of fluorescent Rab GTPase probes. *ChemBioChem* **2011**, *12*, 2813–2821.

80. De Rosa, L.; Cortajarena, A.L.; Romanelli, A.; Regan, L.; D'Andrea, L.D. Site-specific protein double labeling by expressed protein ligation: Applications to repeat proteins. *Org. Biomol. Chem.* **2012**, *10*, 273–280.

81. Deniz, A.A.; Laurence, T.A.; Beligere, G.S.; Dahan, M.; Martin, A.B.; Chemla, D.S.; Dawson, P.E.; Schultz, P.G.; Weiss, S. Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5179–5184.

82. Appleby-Tagoe, J.H.; Thiel, I.V.; Wang, Y.; Mootz, H.D.; Liu, X.Q. Highly efficient and more general cis- and trans-splicing inteins through sequential directed evolution. *J. Biol. Chem.* **2011**, *286*, 34440–34447.

83. Yang, J.Y.; Yang, W.Y. Site-specific two-color protein labeling for FRET studies using split inteins. *J. Am. Chem. Soc.* **2009**, *131*, 11644–11645.

84. Zettler, J.; Schütz, V.; Mootz, H.D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett.* **2009**, *583*, 909–914.

85. Shah, N.H.; Dann, G.P.; Vila-Perelló, M.; Liu, Z.; Muir, T.W. Ultrafast protein splicing is common among cyanobacterial split inteins: Implications for protein engineering. *J. Am. Chem. Soc.* **2012**, *134*, 11338–11341.

86. Bellmann-Sickert, K.; Baumann, L.; Beck-Sickinger, A.G. Selective labelling of stromal cell-derived factor 1alpha with carboxyfluorescein to study receptor internalisation. *J. Pept. Sci.* **2010**, *16*, 568–574.

87. Pellois, J.P.; Hahn, M.E.; Muir, T.W. Simultaneous triggering of protein activity and fluorescence. *J. Am. Chem. Soc.* **2004**, *126*, 7170–7171.

88. Hahn, M.E.; Pellois, J.P.; Vila-Perello, M.; Muir, T.W. Tunable photoactivation of a post-translationally modified signaling protein and its unmodified counterpart in live cells. *Chembiochem* **2007**, *8*, 2100–2105.

89. Chaisemartin, L.; Chinestra, P.; Favre, G.; Blonski, C.; Faye, J.C. Synthesis and application of a N-1' fluorescent biotinyl derivative inducing the specific carboxy-terminal dual labeling of a novel RhoB-selective scFv. *Bioconjug. Chem.* **2009**, *20*, 847–855.

90. Schuler, B.; Pannell, L.K. Specific labeling of polypeptides at amino-terminal cysteine residues using Cy5-benzyl thioester. *Bioconjug. Chem.* **2002**, *13*, 1039–1043.

91. Yeo, D.S.; Srinivasan, R.; Uttamchandani, M.; Chen, G.Y.; Zhu, Q.; Yao, S.Q. Cell-permeable small molecule probes for site-specific labeling of proteins. *Chem. Commun. (Camb)* **2003**, *2003*, 2870–2871.

92. Gentle, I.E.; de Souza, D.P.; Baca, M. Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug. Chem.* **2004**, *15*, 658–663.

93. Giriat, I.; Muir, T.W. Protein semi-synthesis in living cells. *J. Am. Chem. Soc.* **2003**, *125*, 7180–7181.

94. Dhar, T.; Mootz, H.D. Modification of transmembrane and GPI-anchored proteins on living cells by efficient protein trans-splicing using the Npu DnaE intein. *Chem. Commun. (Camb)* **2011**, *47*, 3063–3065.

95. Xu, R.; Ayers, B.; Cowburn, D.; Muir, T.W. Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 388–393.

96. Xu, R.; Liu, D.; Cowburn, D. Abl kinase constructs expressed in bacteria: Facilitation of structural and functional studies including segmental labeling by expressed protein ligation. *Mol. Biosyst.* **2012**, *8*, 1878–1885.

97. Camarero, J.A.; Shekhtman, A.; Campbell, E.A.; Chlenov, M.; Gruber, T.M.; Bryant, D.A.; Darst, S.A.; Cowburn, D.; Muir, T.W. Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8536–8541.

98. Clerico, E.M.; Zhuravleva, A.; Smock, R.G.; Gierasch, L.M. Segmental isotopic labeling of the Hsp70 molecular chaperone DnaK using expressed protein ligation. *Biopolymers* **2010**, *94*, 742–752.

99. Vitali, F.; Henning, A.; Oberstrass, F.C.; Hargous, Y.; Auweter, S.D.; Erat, M.; Allain, F.H. Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling. *EMBO J.* **2006**, *25*, 150–162.

100. Skrisovska, L.; Allain, F.H. Improved segmental isotope labeling methods for the NMR study of multidomain or large proteins: Application to the RRMs of Npl3p and hnRNP L. *J. Mol. Biol.* **2008**, *375*, 151–164.

101. Chen, J.; Wang, J. A segmental labeling strategy for unambiguous determination of domain-domain interactions of large multi-domain proteins. *J. Biomol. NMR* **2011**, *50*, 403–410.

102. Zhao, W.; Zhang, Y.; Cui, C.; Li, Q.; Wang, J. An efficient on-column expressed protein ligation strategy: Application to segmental triple labeling of human apolipoprotein E3. *Protein Sci.* **2008**, *17*, 736–747.

103. Castaneda, C.A.; Spasser, L.; Bavikar, S.N.; Brik, A.; Fushman, D. Segmental isotopic labeling of ubiquitin chains to unravel monomer-specific molecular behavior. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 11210–11214.

104. Romanelli, A.; Shekhtman, A.; Cowburn, D.; Muir, T.W. Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the *N*-extein-intein junction. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6397–6402.

105. Frutos, S.; Goger, M.; Giovani, B.; Cowburn, D.; Muir, T.W. Branched intermediate formation stimulates peptide bond cleavage in protein splicing. *Nat. Chem. Biol.* **2010**, *6*, 527–533.

106. Anderson, L.L.; Marshall, G.R.; Crocker, E.; Smith, S.O.; Baranski, T.J. Motion of carboxyl terminus of Galpha is restricted upon G protein activation. A solution NMR study using semisynthetic Galpha subunits. *J. Biol. Chem.* **2005**, *280,* 31019–31026.

107. Blaschke, U.K.; Silberstein, J.; Muir, T.W. Protein engineering by expressed protein ligation. *Meth. Enzymol.* **2000**, *328*, 478–496.

108. Karagöz, G.E; Sinnige, T.; Hsieh, O.; Rüdiger, S.G.D. Expressed protein ligation for a large dimeric protein. *Protein Eng. Des. Sel.* **2011**, *24*, 495–501.

109. Moran, S.D.; Woys, A.M.; Buchanan, L.E.; Bixby, E.; Decatur, S.M.; Zanni, M.T. Two-dimensional IR spectroscopy and segmental 13C labeling reveals the domain structure of human γD-crystallin amyloid fibrils. *Proc. Natl. Acad. Sci. USA* **2012**, *128*, 3329–3334.

110. Moran, S.D.; Decatur, S.M.; Zanni, M.T. Structural and Sequence Analysis of the Human γD-Crystallin Amyloid Fibril Core Using 2D IR Spectroscopy, Segmental (13)C Labeling, and Mass Spectrometry. *J. Am. Chem. Soc.* **2012**, *134*, 18410–18416.

111. Busche, A.E.; Aranko, A.S.; Talebzadeh-Farooji, M.; Bernhard, F.; Dotsch, V.; Iwai, H. Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 6128–6131.

112. Minato, Y.; Ueda, T.; Machiyama, A.; Shimada, I.; Iwai, H. Segmental isotopic labeling of a 140 kDa dimeric multi-domain protein CheA from Escherichia coli by expressed protein ligation and protein trans-splicing. *J. Biomol. NMR* **2012**, *53*, 191–207.

113. Yagi, H.; Tsujimoto, T.; Yamazaki, T.; Yoshida, M.; Akutsu, H. Conformational change of H^+-ATPase beta monomer revealed on segmental isotope labeling NMR spectroscopy. *J. Am. Chem. Soc.* **2004**, *126*, 16632–16638.

114. Yagi, H.; Kajiwara, N.; Iwabuchi, T.; Izumi, K.; Yoshida, M.; Akutsu, H. Stepwise propagation of the ATP-induced conformational change of the F1-ATPase beta subunit revealed by NMR. *J. Biol. Chem.* **2009**, *284*, 2374–2382.

115. Muona, M.; Aranko, A.S.; Iwai, H. Segmental isotopic labelling of a multidomain protein by protein ligation by protein trans-splicing. *Chembiochem* **2008**, *9*, 2958–2961.

116. Muona, M.; Aranko, A.S.; Raulinaitis, V.; Iwai, H. Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing *in vivo* and *in vitro*. *Nat. Protoc.* **2010**, *5*, 574–587.

117. Zuger, S.; Iwai, H. Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat. Biotechnol.* **2005**, *23*, 736–740.

118. Kobayashi, H.; Yoshida, T.; Inouye, M. Significant enhanced expression and solubility of human proteins in Escherichia coli by fusion with protein S from Myxococcus xanthus. *Appl. Environ. Microbiol.* **2009**, *75*, 5356–5362.

119. Kobayashi, H.; Swapna, G.V.; Wu, K.P.; Afinogenova, Y.; Conover, K.; Mao, B.; Montelione, G.T.; Inouye, M. Segmental isotope labeling of proteins for NMR structural study using a protein S tag for higher expression and solubility. *J. Biomol. NMR* **2012**, *52*, 303–313.