View Online / Journal Homepage / Table of Contents for this issue

Organic & Biomolecular Chemistry

Dynamic Article Links

Cite this: *Org. Biomol. Chem.*, 2012, **10**, 273

www.rsc.org/obc

PAPER

Site-specific protein double labeling by expressed protein ligation: applications to repeat proteins†

Lucia De Rosa,[a] Aitziber L. Cortajarena,[b,d] Alessandra Romanelli,[c] Lynne Regan[b] and Luca Domenico D'Andrea*[a]

Received 16th August 2011, Accepted 6th October 2011
DOI: 10.1039/c1ob06397a

Published on 06 October 2011 on http://pubs.rsc.org | doi:10.1039/C1OB06397A
Downloaded by Universitaire d'Angers on 08 February 2012

In the last few years, the use of labeled proteins has significantly expanded in the life sciences. Now, labeled proteins are indispensable tools for a wide spectrum of biophysical and chemical biology applications. In particular, the quest for more sophisticated experimental setups requires the development of new synthetic methodology, especially for multiple site-specific labeling. In this paper, we describe a synthetic strategy based on expressed protein ligation to prepare proteins in high purity and homogeneity, in which two different molecular probes are incorporated specifically at any desired position. Proteins are sequentially labeled in solution, with the advantage that a large excess of probes is not required and the labeled fragments are not restricted to peptide synthesis length limitations. This strategy was applied to selectively label a repeat protein with a fluorophores pair in different positions along the protein sequence. The doubly labeled proteins were prepared at high purity and homogeneity, as required for single molecule FRET studies. Remarkably, this approach can be adapted to the introduction of more than two molecular probes.

Introduction

The development of methods for specific protein decoration in multiple sites shows great relevance and impact in many fields of life science. The ability to introduce into proteins new functionalities, such as post-translational modifications, biophysical and biochemical probes, provides a mean by which to characterize and modulate protein function or to endow these macromolecules of new useful properties. So far, different protein engineering approaches, based on the use of both chemistry and/or molecular biology, have been developed. For example, the covalent and selective modification of multiple cysteines in an expressed protein has been achieved, taking advantage of the different solvent accessibility and reactivity of Cys residues.[1–4] These approaches are original, but had only a limited impact because of their partial applicability and the heterogeneity of the samples preparation obtained. An alternative multiple labeling strategy involves the modification of cellular ribosomal machinery for the introduction of unnatural amino acids with functional properties or ability to react selectively with specific probes.[5,6] Enzymatic approaches for the site-specific modification of proteins are also described,[7,8] but they are limited to the labeling of signal peptide or terminal protein regions.

An advantageous strategy for protein modification overcoming most of the above mentioned limitations relies on the use of native chemical ligation (NCL).[9] Such a technique consists on the synthesis of a large polypeptide through the chemoselective ligation of two short fragments bearing a C-terminal α-thioester or an N-terminal free cysteine residue. These fragments can include covalent modifications, affording the preparation of site-specifically modified polypeptides. A related method, named expressed protein ligation (EPL),[10] expands the applicability of NCL to recombinant protein fragments. To achieve C-terminal thioester or N-terminal cysteine containing protein segments, EPL exploits genetically engineered inteins, which are protein elements able to catalyze their self-removing from a flanking polypeptide. EPL and NCL have been used to site-specifically modify proteins with molecular probes, such as fluorophores.[11,12] In all cases peptides were labeled on solid phase and then ligated in solution to a synthetic peptide[11] or an expressed polypeptide.[12] Similarly to EPL, protein *trans* splicing uses a particular class of inteins, called split inteins, to assemble a protein from two fragments and introduce site-specific modifications.[13–16]

Here, we describe a synthetic strategy for double labeling proteins based on the use of EPL. This strategy, which is independent from solid phase labeling, envisages a sequential labeling in solution with the advantage that a large excess of

[a]*Istituto di Biostrutture e Bioimmagini, CNR, via Mezzocannone 16, 80134, Napoli, Italy. E-mail: ldandrea@unina.it; Tel: +390812536679*
[b]*Department of Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, CT-06511, New Haven, U. S. A.*
[c]*Dipartimento delle Scienze Biologiche, Università di Napoli "Federico II", via Mezzocannone 16, 80134, Napoli, Italy*
[d]*IMDEA-Nanociencia and Centro Nacional de Biotecnología (CNB-CSIC), Universidad Autónoma de Madrid, Cantoblanco, Madrid, 28049, Spain*

† Electronic supplementary information (ESI) available: CTPR3 variants sequences, LC-MS analysis and CD spectra of purified double labeled proteins. See DOI: 10.1039/c1ob06397a

probes is not required and the length of the labeled fragments is not subjected to peptide synthesis restrictions. In this way, the two different molecular probes can be incorporated specifically at any desired position. Moreover, this strategy is suitable for the introduction of more than two probes. We applied this strategy to prepare the repeat protein CTPR3 (*Consensus* TetratricoPeptide Repeat)[17,18] selectively labeled with a fluorophores pair in different positions along the protein sequence. These doubly labeled CTPR3 variants were prepared in high purity and homogeneity, as required for single molecule FRET studies.

Experimental Procedures

General materials and methods

pTXB1 and pETM13 plasmids were respectively obtained from New England Biolabs and European Molecular Biology Laboratory. pProEXHTa vector containing the gene construct coding CTPR3 (360 bp) cloned between *BamHI* and *HindIII* sites (pProExHTa-*ctpr3*) has been previously described.[17] All restriction and modification enzymes were purchased from New England Biolabs, except for PfuTurbo DNA Polymerase (Stratagene) and restriction enzyme SacI (Roche). DNA purification kits were purchased from Qiagen. *E. coli* TOP F′10 and DH5α strains were supplied by Invitrogen, while cell strain BL-21(DE3) was purchased from Novagen (Merck). Isopropyl-beta-D-thiogalactopyranoside (IPTG) was from Inalco. Protease inhibitor phenylmethanesulfonyl-fluoride (PMSF) was obtained from Sigma-Aldrich. Fluorophores ATTO 488-maleimide and ATTO 647N-maleimide were purchased from ATTO-TECH. L-cysteine and 2-mercapto-ethane sulfonate sodium (MESNA) were purchased from Sigma-Aldrich, while tris(2-carboxyethyl)phosphine (TCEP) was from STREMM. Fmoc-protected amino acids, coupling reagents and Fmoc–Gly–Wang resin for peptide synthesis, Hydroxybenzotriazole (HOBt) and *O*-benzotriazole-*N*,*N*,*N*′,*N*′-tetramethyl-uronium-hexafluoro-phosphate (HBTU) were provided by NovaBiochem (Merck). *N*,*N*-diisopropylethylamine (DIPEA) was provided by Romil, while piperidine was purchased from Biosolve. Acetic anhydride was supplied by Sigma-Aldrich. Other solvents for peptide synthesis and purification (dimethylformamide (DMF), *N*-methyl-2-pyrrolidone (NMP), dichloromethane (DCM), trifluoroacetic acid (TFA), ethandithiol (EDT), triisopropyl silane (TIS), diethyl ether and acetonitrile were from Sigma-Aldrich or Romil. All reagents and solvents were used without further purification.

Protein purification and analysis methods

Proteins were purified by affinity chromatography using a HisTrap column (GE Healthcare) on a ÄKTA FPLC system (GE HealthCare) and by RP-HPLC on HP 1200 Series instrument (Agilent Technologies). Proteins were purified using Jupiter C18 column (Jupiter, 25 × 1.1 cm, 300 Å, 10 µm; Phenomenex) at a flow rate of 5 mL min^{-1}. The Cys-peptide was purified by RP-HPLC on a Proteo C12 column (25 × 2.2 cm, 90 Å, 10 µm; Phenomenex) at 20 mL min^{-1}. Analytical characterization of peptide and proteins was performed on an LC-MS system (LCQ DECA XP, ThermoFinnigan) equipped with an ESI source, an ion trap mass analyser and a complete Surveyor HPLC system using a C18 column (Jupiter 250 × 2 mm, 5 µm, 300 Å; Phenomenex) at a flow rate of 0.2 mL min^{-1}. HPLC gradients: METHOD A: CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 20% to 70% in 30 min; METHOD B: CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 5% to 70% in 40 min; METHOD C: CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 20% to 25% in 3 min and from 25% to 60% in 40 min.

Protein cloning and expression

Site-specific mutagenesis. pProExHTa-*ctpr3* plasmid was co-digested with restriction enzymes *XhoI* and *SacI* or *SacI* and *HindIII* to extract gene cassettes, including Asn36 or Asn70 codons, respectively, next to their 5′ termini. Excised cassettes (111 bp or 167 bp respectively) were mutagenized by PCR using sense primers including the mutation Asn36Cys or Asn70Cys as well as *XhoI* or *SacI* site (respectively 5′-CCG*CTCGAG*CTGGACCCG**TGC**AACGCA-3′ and 5′-C*GAGCTC*GATCCA**TGT**AATGCAGAG-3′), and antisense primers including *SacI* or *HindIII* site (respectively 5′-G*GAGCTC*AAGCGCCTTTTGATAATA-3′ and 5′-GGG*AAGCTT*CTATCAACCCTGTTTCTG-3′). Mutagenized gene cassettes were then digested with *XhoI* and *SacI* or *SacI* and *HindIII* enzymes and subcloned into the pProExHTa-*ctpr3* vector.

Preparation of CTPR3 variants – Mxe GyrA intein fusion gene. The full *ctpr3* gene containing the mutation Asn36Cys (CTPR3[1_C]) or Asn70Cys (CTPR3[2_C]) was respectively amplified from mutagenic plasmids pProExHTa-*ctpr3*(Asn36Cys) and pProExHTa-*ctpr3*(Asn70Cys) (forward primer 5′-CATG*CCATGG*GTAACTCCGCTGAGGCATGG; reverse primer 5′-TCCCGTGATGCAACCCTGTTTCTGTTTAGCGTT-3′). For CTPR3[N_C] variant, the full *ctpr3* gene was amplified directly from wild-type pProExHTa-*ctpr3* using a sense primer containing an extra Cys codon (Cys -1). To preserve *Nco*I restriction site upstream the gene, an extra Gly residue (position -2) was added in this latter gene (forward primer 5′-CATG*CCATGG*GT**TGC**GGTAACTCCGCTGAGGCA-3′ and reverse primer 5′-TCCCGTGATGCAACCCTGTTTCTGTTTAGCGTT-3′). Instead, for the preparation of CTPR3[1_3], the gene trait corresponding to amino acids 1-104 of CTPR3 was amplified from pProExHTa-*ctpr3*(Asn36Cys) (forward primer 5′-CATG*CCATGG*GTAACTCCGCTGAGGCATGG-3′ and reverse primer 5′-TCCCGTGATGCAGTTCGGGTCCAGTTCCAGCGC-3′). All of the above mentioned genes were amplified using sense primers, including a *NcoI* restriction site and antisense primers designed to generate a 3′ terminal region of 15 bp overlapping the gene of the *Mycobacterium xenopi* DNA gyrase A intein (*Mxe* GyrA intein). The gene coding *Mxe* GyrA intein (594 bp) was amplified from the pTXB1 vector using sense primers, generating a 5′ terminal region of 15 bp overlapping CTPR3 variant coding genes and an antisense primer, including an *EcoRI* restriction site, (forward primer 5′-CAGAAACAGGGTTGCATCACGGGAGATGCACTA-3′ and reverse primer 5′-G*GAATTC*CCAGCGTGGCTGACGAACCCGTT-3′ were used for the amplification of *Mxe* GyrA coding gene overlapping CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C] genes; forward primer 5′-CTGGACCCGAACTGCATCACGGGAGATGCACTA-3′ and reverse primer 5′-G*GAATTC*CCAGCGTGGCTGACGAACCCGTT-3′ were

Published on 06 October 2011 on http://pubs.rsc.org | doi:10.1039/C1OB06397A
Downloaded by Universitaire d'Angers on 08 February 2012

 This journal is © The Royal Society of Chemistry 2012

Published on 06 October 2011 on http://pubs.rsc.org | doi:10.1039/C1OB06397A
Downloaded by Universitaire d'Angers on 08 February 2012

instead used to amplify the *Mxe* GyrA coding gene overlapping the gene coding the fragment 1–104 of CTPR3[1_3]).

Hybrid genes coding the CTPR3 variants fused to *Mxe* GyrA intein were obtained by Overlapping Extension PCR, performed similarly as described in ref. 19. Briefly, the reaction was performed in a total volume of 100 µL using equimolar amounts of *Mxe* GyrA and CTPR3 gene variant, 250 µM dNTPs, 2.5 U (1 µL) of PfuTurbo DNA polymerase, 2.5% formamide and PfuTurbo buffer 1X. Before adding the two primers, 5 cycles of template extension were performed. Then, 30 pmol of each of the two primers annealing at 5′ terminus of CTPR3 variant coding gene and at the 3′ terminus of *Mxe* GyrA were added to the PCR reaction and 25 cycles of amplification were performed. Fusion genes were then digested with *NcoI* and *EcoRI* restriction enzymes and cloned into the expression vector pETM13 upstream an His_6-tag coding sequence.

Protein expression. CTPR3 variants fused to *Mxe* GyrA intein were expressed *E. coli* BL-21(DE3) cells. Cell culture was growth at 37 °C until the $OD_{600\,nm}$ reached 0.7–0.8, then transferred at 22 °C and induced with 0.5 mM IPTG, allowing the expression of the fusion protein to proceed overnight. The cells were then harvested by centrifugation at 7.000 rpm, 10 min at 4 °C and stored at −20 °C.

The experimental procedure is reported in detail for the CTPR3[1_3] protein. The other CTPR variants were prepared following the same synthetic steps with similar yields unless differences are reported.

Preparation of CTPR3[1_3]-thioester protein (fragment 1–104)

Pellet from 1 L culture was resuspended in 50 mL of ice-cold 50 mM phosphate buffer pH 6.5, 300 mM NaCl, 10 mM imidazole (buffer A) containing 1 mM protease inhibitor PMSF and lysed by sonication. The supernatant, containing the soluble His_6-tagged fusion protein, was filtered and loaded onto a 5 mL HisTrap column pre-equilibrated in buffer A. The purification was performed at 4 °C using an ÄKTA FPLC system. Then, fusion protein was eluted from resin with 50 mM phosphate buffer pH 6.5, 300 mM NaCl, 500 mM imidazole (buffer B) (1 mL min^{-1}) and immediately incubated with 50 mM MESNA, at 4 °C and under mild stirring. After ten days of the splicing reaction (six days for CTPR3[N_C], CTPR3[1_C] and CTPR3[2_C]), the protein mixture was dialyzed in 5 mM phosphate buffer pH 6.5 overnight at 4 °C, and then lyophilized. The lyophilized proteins mixture was solubilized in CH_3CN (TFA 0.1%)/water (TFA 0.1%) mixture (2 : 8 v/v), incubated with 20 mM TCEP to reduce Cys residues and purified by reverse-phase HPLC. Collected fractions were analyzed by LC-mass spectrometry using method A. CTPR3[1_3]-thioester: t_R = 19.5 min, MW_{exp} = 12319.9 Da, MW_{calc} = 12319.0 Da; CTPR3[1_C]-thioester: t_R = 17.8 min, MW_{exp} = 13999.5 Da, MW_{calc} = 14000.9 Da; CTPR3[2_C]-thioester: t_R = 18.7 min, MW_{exp} = 13999.4 Da, MW_{calc} = 14000.9 Da; CTPR3[N_C]-thioester: t_R = 17.0 min, MW_{exp} = 14167.4 Da, MW_{calc} = 14172.2 Da.

Synthesis of Cys-Peptide (CTPR3[1_3]-fragment 105–120)

Peptide H–Cys–Ala–Glu–Ala–Lys–Gln–Asn–Leu–Gly–Asn–Ala–Lys–Gln–Lys–Gln–Gly–OH, corresponding to the C-terminal fragment 105–120 of CTPR3[1_3], was synthesized on solid phase using a Fmoc–Gly–Wang resin (0.65 mmol g^{-1}) by standard Fmoc chemistry. For each coupling reaction (1 h) 10 equivalents of the Fmoc-protected amino acid, 9.9 equivalents of HOBt/HBTU and 20 equivalents DIPEA were used. Removal of Fmoc was carried out with a solution of 30% v/v piperidine in DMF (5 min, twice). After each coupling, unreacted N-terminal amino groups were capped with a solution of 2 M acetic anhydride, 0.06 M HOBt, 0.55 M DIPEA in NMP (20 min). Peptide cleavage from the resin and amino acids side chain deprotection were achieved by treatment with TFA, EDT, TIS and water (94 : 2.5 : 1 : 2.5) at room temperature for 3 h. Cold diethyl ether was used to precipitate the peptide. Crude product was collected by centrifugation, washed twice with cold diethyl ether, resupended in water and acetonitrile and lyophilized. The peptide was purified by reverse-phase HPLC. Cys-Peptide was analyzed by LC-MS (method B): t_R = 16.27 min, MW_{exp} = 1687.7 Da, MW_{calc} = 1687.8 Da.

Labeling of CTPR3[1_3]-thioester (fragment 1–104)

Lyophilized thioester protein (7.5 mg, 615 nmol) was resuspended in 4 mL of 20 mM phosphate buffer, pH 7.2 (final protein concentration of 0.15 mM). A two-fold molar excess of ATTO 488-maleimide was dissolved in 20 µL of DMF and added to solution. Reaction mix was incubated for 2 h at 4 °C in the dark, under mild stirring. Exceeding fluorophore was quenched adding MESNA to the reaction mixture. Then, the labeled thioester proteins were dialyzed in 20 mM phosphate buffer pH 7.2 overnight at 4 °C in the dark to afford the labeled protein (7.2 mg, 553 nmol) with a 90% yield. Proteins were analyzed by LC-MS using method A. Mono-labeled CTPR3[1_3]-thioester: t_R = 18.99 min, MW_{exp} = 13026.6 Da, MW_{calcd} = 13031.0 Da; mono-labeled CTPR3[1_C]-thioester: t_R = 17.67 min, MW_{exp} = 14701.5 Da, MW_{calcd} = 14693.0 Da; mono-labeled CTPR3[2_C]-thioester: t_R = 18.97 min, MW_{exp} = 14702.7 Da, MW_{calcd} = 14693.0 Da; mono-CTPR3[N_C]-thioester: t_R = 17.30 min, MW_{exp} = 14874.1 Da, MW_{calcd} = 14864.2 Da.

Synthesis of mono-labeled full CTPR3[1_3] by native chemical ligation

Mono-labeled CTPR3[1_3]- thioester proteins (7.2 mg, 553 nmol) was concentrated and diluted to a final concentration of 0.25 mM with the degassed ligation buffer (100 mM phosphate buffer pH 7.5, 4 M GdnHCl, 1 mM EDTA, 50 mM MESNA, 15 mM TCEP). Then, Cys-peptide (9.2 mg, 5.5 µmol) dissolved in the ligation buffer was added to the protein solution under nitrogen and the reaction was incubated at 4 °C in the dark under mild stirring. After 4 days, the reaction mixture was diluted 1 : 2 with H_2O and dialyzed in 20 mM phosphate buffer pH 7.2 overnight at 4 °C. Mono-labeled full proteins were then lyophilized, resuspended in CH_3CN (TFA 0.1%)/water (TFA 0.1%) mixture (2 : 8 v/v), reduced with 20 mM TCEP and purified by HPLC. 6.7 mg (461 nmol, 83% yield) of pure mono-labeled full CTPR3[1_3] were recovered after lyophilization. The other full length variants were prepared following strategy B using the same experimental conditions. In this case, the mono-labeled thioester proteins were reacted with a 20-fold molar excess of L-Cys overnight.

The mono-labeled full proteins' identity and purity were ascertained by LC-MS (method A). Mono-labeled full CTPR3[1_3]: t_R = 18.15 min, MW_{exp} = 14576.9 Da, MW_{calcd} = 14578.8 Da; mono-labeled full CTPR3[1_C]: t_R = 18.3 min, MW_{exp} = 14690.9 Da,

Published on 06 October 2011 on http://pubs.rsc.org | doi:10.1039/C1OB06397A
Downloaded by Universitaire d'Angers on 08 February 2012

MW_{calcd} = 14693.0 Da; mono-labeled full CTPR3[2_C]: t_R = 18.02 min, MW_{exp} = 14690.6 Da, MW_{calcd} = 14693.0 Da; mono-labeled full CTPR3[N_C]: t_R = 17.0 min, MW_{exp} = 14856.3 Da, MW_{calcd} = 14864.2 Da.

Second labeling reaction on the full CTPR3[1_3]

The mono-labeled full CTPR3[1_3] (6.7 mg, 461 nmol) were dissolved in 5 mL of 20 mM phosphate buffer pH 7.2 and reacted with 2-fold molar excess of the fluorophore ATTO 647N-maleimide (0.79 mg, 912 nmol). This reaction occurred on the Cys residue involved in the NCL reaction (Cys105 for CTPR3[1_3]; C-terminal Cys for CTPR3[1_C], [2_C] and [N_C]) and was carried out in the same experimental conditions described for the first labeling reaction. Finally, doubly labeled CTPR3[1_3] was dialyzed in 20 mM phosphate buffer pH 7.2 overnight at 4 °C to remove exceeding fluorophore and lyophilized. Then, the protein was solubilized in CH_3CN (TFA 0.1%)/water (TFA 0.1%) mixture (2 : 8 v/v) and purified by reverse-phase HPLC. 2.0 mg (130 nmol, 28% yield) of pure doubly labeled full CTPR3[1_3] were recovered after lyophilization. Lyophilizated proteins were diluted in the appropriate buffer before spectroscopic analysis. Proteins were analyzed by LC-MS using method C. Doubly labeled CTPR3[1_3]: t_R = 25.5–27.0 min, MW_{exp} = 15346.4 Da, MW_{calcd} = 15346.8 Da; doubly labeled CTPR3[1_C]: t_R = 28.65/29.33 min, MW_{exp} = 15455.7 Da, MW_{calcd} = 15460.9 Da; doubly labeled CTPR3[2_C]: t_R = 28.24/28.90 min, MW_{exp} = 15461.6 Da, MW_{calcd} = 15460.9 Da; doubly labeled CTPR3[N_C]: t_R = 27.97/28.65 min, MW_{exp} = 15631.3 Da, MW_{calcd} = 15632.2 Da.

UV-Vis absorption spectroscopy

Protein concentration was estimated by UV-Vis absorption spectroscopy using a Jasco V-550 (Jasco Corporation, Japan) and a 1-cm pathlength quartz cuvette (Hellma, Italy). Spectra were acquired in the range 220–700 nm with a scan speed of 100 nm min^{-1} and 1 nm data pitch. Protein concentration was estimated by absorbance (A) at 280 nm after correction for fluorophores contribution. Corrected absorbance, A_{280}(corr), for fluorophore contribution at 280 nm was evaluated by the formula: $A_{280}(corr) = A_{280} - [A_{\lambda max}\ (dye) \times CF_{280}\ (dye)]$; where A_{280} is absorbance of the sample at 280 nm, $A_{\lambda max}$ (dye) is the absorbance at the absorption maximum (λ_{max}) of the dye (ATTO488: 501 nm; ATTO647N: 644 nm), CF_{280} (dye) is the correction factor for the dye at 280 nm and it is defined as the ratio $A_{280}/A_{\lambda max}$ of the dye. The following CF values and molar extinction coefficients were used: CF (ATTO 488, 280 nm) = 0.1; CF (ATTO 647 N, 280 nm) = 0.05; ε_{280} (CTPR3) = 43320 M^{-1} cm^{-1}(calculated from www.expasy.ch using the ProtParam tool); ε_{501} (ATTO 488) = 9×10^4 M^{-1} cm^{-1}; ε_{644} (ATTO 647 N) = 1.5×10^5 M^{-1} cm^{-1}.

The coupling efficiency (CE) for a dye was calculated according to following formula: CE (Dye) = [ε_{280} (CTPR3) × $A_{\lambda max}$ (Dye)]/[A_{280}(corr) × $\varepsilon_{\lambda max}$ (Dye)], where A_{280} was corrected for both dyes' contribution. Measurements were carried out in 6 M guanidinium to avoid dye aggregation effects.

Circular dichroism spectroscopy

Circular dichroism analyses were performed on AVIV Model 215 CD spectophotometer (AVIV Instruments, Lakewood, NJ). CD experiments were performed in 50 mM phosphate buffer, 150 mM NaCl pH 6.5 and acquired in the far UV range 190–260 nm. The CD spectra were recorded at a protein concentration of 6 μM, at 25 °C, with a band width of 1 nm, a step of 1 nm and an average time of 10 s, in a 0.1 cm path length cuvette.

Fluorescence spectroscopy

Fluorescence spectra were recorded using a PTI Quantamaster C-61 two channel fluorescence spectrophotometer (Photon Technology International). All fluorescence spectra were acquired at 25 °C in a PTI Quantamaster C-61 two channel fluorescence spectrophotometer in blackened quartz cuvettes with 0.3 cm pathlength. All protein samples were dissolved in filtered and degassed 50 mM phosphate, 150 mM NaCl pH 7.0 buffer at the concentration of 3 μM. The sample was excited at 503 nm with a 2 nm slit and the fluorescence emission was recorded between 505 nm and 700 nm through a 2 nm slit. FRET efficiency (E) was calculated using the formula: $E = F_A/(F_A + \gamma F_D)$ where F_A and F_D correspond respectively to the acceptor and donor fluorescence intensities at 669 nm and 523 nm, while γ is a correction factor for the quantum yield and the detector efficiency. In the system used for the experiments, γ was assumed 1. The equation $E = (R_0^6)/(R_0^6 + R^6)$ was used to fit the experimental E values *vs.* the X-ray calculated dye distance, where R_0 = 49 Å is a constant for the specific dye pair, corresponding to the inter-dye distance at which $E = 0.5$.

Results and Discussion

The strategy developed for the preparation of doubly functionalized proteins is schematically illustrated in Fig. 1. This approach is based on the use of the expressed protein ligation and requires the splitting of the protein of interest in two fragments, each bearing a single cysteine. The N-terminal fragment is expressed in bacterial host as a fusion protein with the *Mycobacterium xenopi* gyrase A intein (*Mxe* GyrA intein). The expressed chimeric protein is also endowed of a C-terminal His_6-tag, allowing its purification by affinity chromatography. After purification, the fusion protein is incubated with a thiol that induces the splicing of *Mxe* GyrA intein and the release of the protein fragment as a C-terminal thioester. The thioester protein is then isolated from the intein and selectively labeled on its single Cys residue upon reaction with a probe containing a maleimide group as sulfhydryl-reactive moiety. The C-terminal fragment, carrying an N-terminal Cys residue, can be chemically synthesized or expressed in bacteria depending on the fragment size. The native chemical ligation reaction between the mono-labeled thioester fragment and the C-terminal fragment affords the full-length mono-labeled protein. The Cys residue involved in NCL reaction is then exploited to introduce the second probe into the protein (Fig. 1, strategy A). A simpler approach was used when the second probe is located at C-terminus of the protein of interest. In this case, the full-length protein containing a single Cys at the required position, is expressed as an intein fusion protein. After purification and first labeling, the full-length thioester protein is reacted with L-Cys in a NCL reaction. The newly introduced Cys residue is then labeled with the second probe (Fig. 1, strategy B). A final re-folding step is required before their use. A slightly modified "strategy A" can allow the preparation of a triple labeled protein. In fact, the fragment chemically synthesized

 This journal is © The Royal Society of Chemistry 2012

Published on 06 October 2011 on http://pubs.rsc.org | doi:10.1039/C1OB06397A
Downloaded by Universitaire d'Angers on 08 February 2012

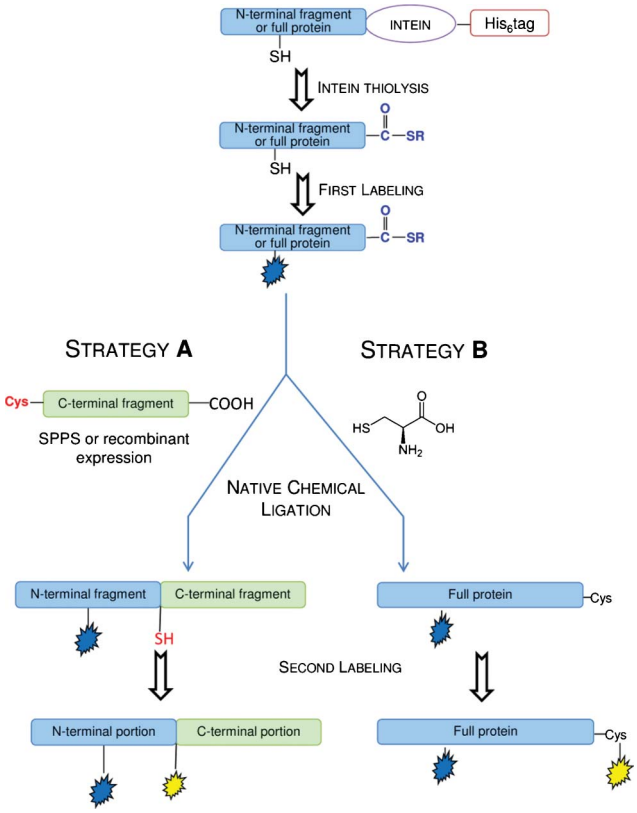

Fig. 1 A schematic representation of the semi-synthetic strategy adopted for the preparation of doubly labeled proteins. The N-terminal portion of the protein was prepared as a C-terminal thioester through expression with the *Mxe* GyrA intein as fusion partner. After splicing, the N-terminal fragment of the protein, carrying a C-terminal thioester group, was labeled with the first probe and then reacted by native chemical ligation with the C-terminal remaining portion of the protein containing an N-terminal Cys residue (strategy A) or a L-Cys (strategy B). Finally, the full protein is labeled with the second probe on the Cys reactive in the native chemical ligation reaction.

and containing the N-terminal Cys could be selectively labeled on solid phase in any position. After the NCL reaction between thioester protein and the synthetic peptide, both carrying a probe, the Cys involved in NCL can be exploited for the introduction of the third probe.

The described semi-synthetic approach was adopted for the preparation of four variants of the protein CTPR3[17] labeled with two different fluorophores. The variants are schematically represented in Fig. 2 (primary sequences are reported in Fig. S1†) and are named CTPR3[N_C], CTPR3[1_C], CTPR3[2_C] and CTPR3[1_3]. The nomenclature adopted indicates the protein regions in which the fluorescent probes (acceptor and donor) are located. The variants differ for the distance between the two fluorophores and for the localization in internal *vs.* peripheral protein regions, useful for the folding studies by FRET. The same synthetic strategy was applied to prepare the same variants with the dye positions swapped. The preparation of doubly labeled CTPR3[1_3] (Fig. 2, strategy A) is described in detail as an example of the synthetic strategy as all of the other variants were prepared following similar steps.

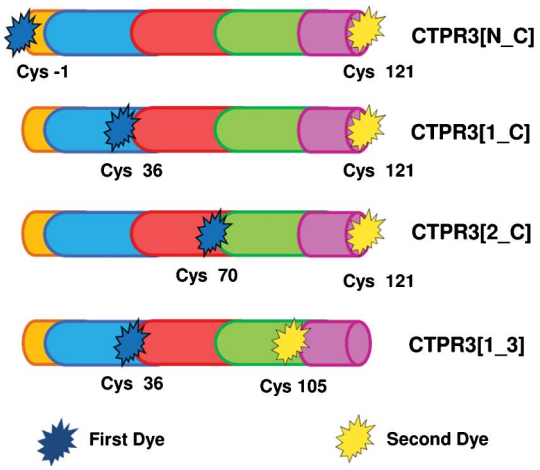

Fig. 2 A schematic representation of the doubly labeled CTPR3 variants. Protein nomenclature and Cys point mutations with respect to wild-type CTPR3[17] are indicated. Dyes were covalently linked to Cys side chains. TPR repeats are reported in blue, red and green. The N-terminal capping region and the extra helix are shown in orange and pink, respectively.

Cloning, expression and purification of the thioester protein

The gene coding for the fragment 1–104 of CTPR3[1_3] variant was fused by Overlapping Extension PCR reaction to the gene coding the *Mxe* GyrA intein. Using this cloning strategy, no additional amino acids were placed between the two components of the fusion protein so that, after intein splicing, the CTPR3 fragment did not inherit any residual amino acids at its C-terminus. The resulting chimeric gene was cloned into the pETM13 vector, upstream of a His_6-tag coding sequence. In this way, the His_6-tag was placed at C-terminus of the intein, which is a self-removing element. Accordingly, no protease-mediated tag removal step was required. Expression performed overnight at 22 °C yielded the protein in the soluble fraction. The same was observed for all the other CTPR3 variants. Intein splicing *in vivo* was not observed during the expression, as expected for a protein sequence possessing an Asn residue (or Gly for the other variants) located at the protein–intein junction.[20] Fusion protein purification was accomplished by Ni-NTA affinity purification. The affinity chromatography was performed at pH 6.5 in order to prevent thioester hydrolysis and the use of Tris buffer was avoided, as it has been demonstrated to react with the thioester protein forming a stable adduct and reducing the thioester protein final yield.[21] The C-terminal thioester derivative of the CTPR3[1_3] variant was obtained by thiolysis, immediately incubating the intein-tagged protein with a large excess of the splicing-inducing thiol mercaptoethanesulfonic acid (MESNA). The C-terminal thioester protein fragment was purified by reverse-phase HPLC from the spliced intein, unspliced fusion protein and side-products, such as intramolecular cyclic derivatives (Fig. 3A). The identity of the thioester proteins was assessed by ESI mass spectrometry and found to be in agreement with expected values. The final yield of the thioester protein was 15 mg per liter of culture. The preparation of the other thioester CTPR3 variants proceeded with similar yields.

Protein labeling reaction: first probe

The thioester fragment 1–104 of CTPR3[1_3] was site-specifically labeled on the single Cys residue (position 36) with the maleimide

Published on 06 October 2011 on http://pubs.rsc.org | doi:10.1039/C1OB06397A
Downloaded by Universitaire d'Angers on 08 February 2012

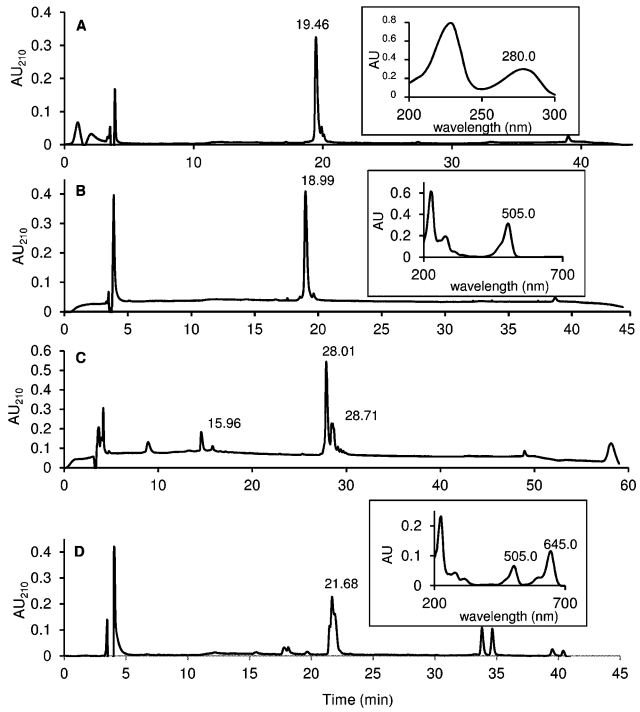

Fig. 3 The preparation of doubly labeled CTPR3[1_3]. HPLC profiles at 210 nm of the synthetic steps and the absorbance spectrum (inset) are reported. A) The chromatographic profile of CTPR3[1_3]-thioester. B) The chromatographic profile of the first labeling reaction of CTPR3[1_3]-thioester. C) The chromatographic profile of the native chemical ligation reaction between the mono-labeled CTPR3[1_3]-thioester with Cys-peptide. Peak identification: Cys-peptide (15.96 min); mono-labeled full CTPR3[1_3] (28.01 min); CTPR3[1_3]-thioester and its carboxylic acid derivative (28.71 min). D) The chromatographic profile of the second labeling reaction of mono-labeled full CTPR3[1_3]. The double peak around 35 min corresponds to the unreacted dye.

derivative of ATTO-488 fluorophore (or ATTO-647N). The labeling reaction was performed at pH 7.2 in an aqueous buffer using a two-fold molar excess of fluorophore. The labeling reaction occurred almost completely within 2 h (Fig. 3B). No reducing agent was used, even if many protocols for Cys side-chain labeling with maleimide derivative compounds recommend the use of the non-thiolic reducing agent TCEP. We observed that TCEP reacted with the maleimide moiety, preventing Cys labeling. This observation was also supported by previous works reported in literature.[22,23] Excess fluorophore was quenched with the thiol MESNA and then the reaction mixture was extensively dialyzed in phosphate buffer for the next chemical ligation reaction. LC-MS analysis showed a protein mass increment of 712 Da, with respect to the thioester protein mass, which is consistent with the addition of a ATTO 488-maleimide molecule. The absorption spectrum of the protein showed the peculiar absorption features of the ATTO 488 fluorophore (Fig. 3B inset).

Native Chemical Ligation with Cys-peptide

The mono-labeled thioester protein was reacted with a synthetic peptide (Cys-peptide) corresponding to the remaining portion of CTPR3[1_3] construct and bearing a Cys residue as the N-terminal amino acid (protein fragment 105–120). The reaction was carried out for four days and afforded the full-length protein (Fig. 3C) with a yield of 85% (based on HPLC profile). A small amount of side-products was observed, such as the mono-labeled CTPR3[1_3]-COOH and the unreacted reagent (Fig. 3C). The NCL reaction on the other CTPR3 variants with L-Cys proceeded even with higher yield. After NCL with the Cys-peptide, the mono-labeled full CTPR3[1_3] protein was purified to homogeneity by reverse-phase HPLC and lyophilized. The identity of the final product was ascertained by mass spectrometry.

Protein labeling reaction: second probe

Mono-labeled CTPR3[1_3] protein variant was subjected to a second labeling reaction with the acceptor dye ATTO 647N-maleimide (or ATTO 488-maleimide). The labeling reaction was carried out in the same experimental conditions adopted for the first labeling and afforded the doubly labeled protein in a similar yield (Fig. 3D). Finally, doubly labeled protein was purified by reverse-phase HPLC and lyophilized. LC-MS characterization of all purified doubly labeled CTPR3 variants are reported in the supplementary material (Figs S2–S5†). All of the purified doubly labeled CTPR3 variants presented a double peak in the HPLC profile (Figs S2A–S5A†), as the ATTO-647N fluorophore is supplied as a mixture of two diastereoisomers (see dye technical report). The two peaks showed an identical mass spectrum, consistent with the molecular weight of the double labeled full protein (Figs S2B–S5B†). In our HPLC conditions, the doubly labeled CTPR3[1_3] protein is eluted as a broad, instead of a double, peak, with a unique mass value corresponding to the dual labeled full protein (Fig. S2A†). Likely, the two ATTO 647 N diasteroisomers in this specific location have a less defined hydrophobic character, which widens the chromatographic peak of the two diasteromers. All of the doubly labeled CTPR3 variants showed the correct stoichiometry (1 : 1 : 1 protein : acceptor : donor), as ascertained by the presence of a single species in the ESI-mass spectrum (Figs S2B–S5B†). Moreover, we calculated the coupling efficiency (CE) with respect to each dye for all of the doubly labeled proteins. All of the calculated CE values are in the range 0.95–1.1, which is consistent with the expected stoichiometry.

Finally, the absorption spectrum (Figs S2C–S5C†) show the characteristic features of the ATTO-647N dye, *i.e.* the presence of a band with a maximum around 640 nm and a shoulder at 600 nm. The protein final yield was 4 mg per liter of culture.

Spectroscopic characterization

A preliminary spectroscopic characterization of the CTPR3 variants was performed to assess the integrity of the proteins and dyes. All of the prepared proteins showed a CD spectrum characteristic of an all helical protein (Fig. S6†), which is very similar to the wild-type protein, indicating that the overall protein fold was not modified. Doubly labeled protein samples were excited at the donor maximum absorbance wavelength (503 nm) and the fluorescence emission spectra were recorded in the wavelength range 505–700 nm, including both donor and acceptor emission maxima (523 nm and 669 nm respectively). The fluorescence spectra recorded for the four doubly labeled variants of CTPR3

 This journal is © The Royal Society of Chemistry 2012

Published on 06 October 2011 on http://pubs.rsc.org | doi:10.1039/C1OB06397A
Downloaded by Universitaire d'Angers on 08 February 2012

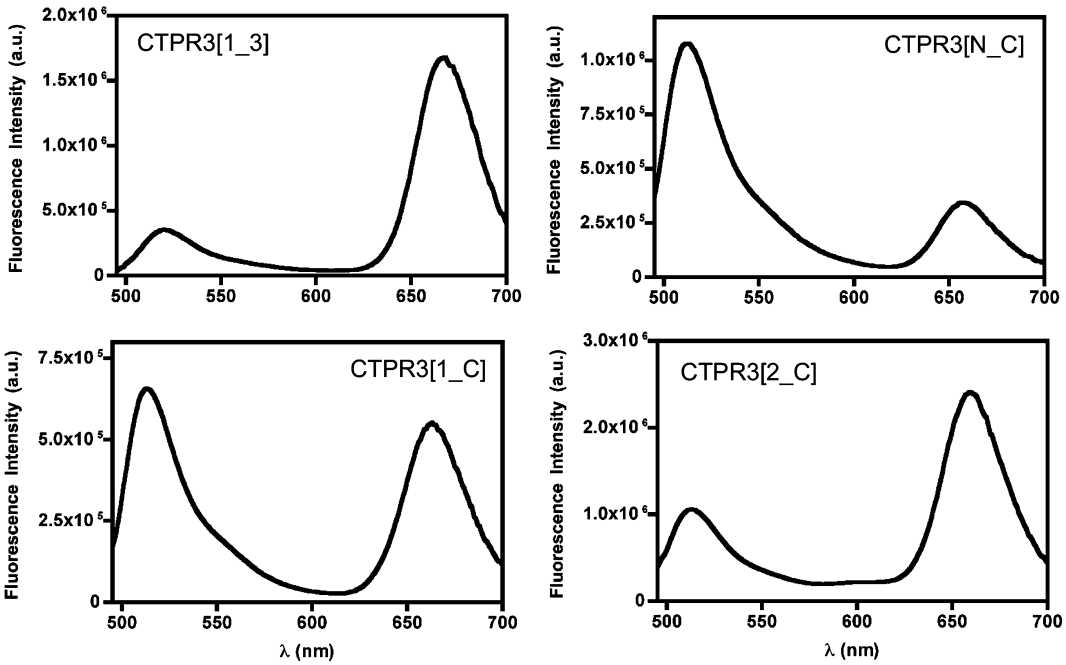

Fig. 4 Fluorescence emission spectra of CTPR3 protein variants doubly labeled with ATTO 488 and ATTO 647 N dyes in 50 mM phosphate buffer pH 7.0, 150 mM NaCl excited at 503 nm.

Table 1 FRET efficiency of CTPR3 variants labeled with ATTO 488 as the donor and ATTO 647 N as the acceptor (R_0 = 49 Å)

Protein	ATTO-647N[a]	ATTO-488[a]	Distance dyes (Å)[b]	*E*
CTPR3[N_C]	N-terminus	C-terminus	52.0	0.21
CTPR3[1_C]	36	C-terminus	40.3	0.40
CTPR3[2_C]	70	C-terminus	31.3	0.67
CTPR3[1_3]	36	105	28.4	0.82

[a] Dye position along the protein sequence. [b] Approximate distance measured from the X-ray structure of CTPR3 (1NA0.pdb) considering the C^{α} of label amino acids.

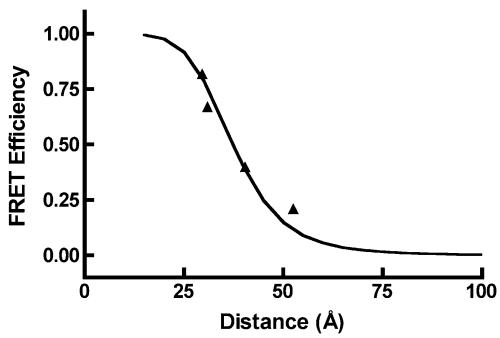

Fig. 5 FRET efficiency (*E*) *vs.* donor–acceptor distance (*R*) for the CTPR3 variants. The experimental *E* values for the four CTPR3 variants are plotted against the approximate distance between the donor and acceptor dyes (filled triangles) calculated from the X-ray crystal structure of the CTPR3 protein (PDB ID: 1na0). The best fit of the data to the equation that relates *E* and distance between the two probes ($E = (R_0^6)/(R_0^6 + R^6)$) is shown as a solid line.

displayed an energy transfer between the dyes, as expected for integral dyes (Fig. 4). In particular, FRET efficiencies (*E*), calculated for the CTPR3 variants (Table 1), correlate with the dyes distances (Fig. 5).

Conclusion

This paper describes a method to prepare proteins labeled with two different probes in high purity and homogeneity. The synthetic approach is based on the use of EPL and its effectiveness has been assessed through the preparation of a series of repeat protein variants, in which two fluorophores were incorporated at different, specific positions. The doubly labeled proteins were preliminary characterized by circular dichroism and fluorescence spectroscopy and were correctly folded and suitable for single molecule FRET studies. The reported synthetic procedure is based on the sequential labeling of a convenient amino acid (in our case a cysteine) in solution. This allows for the use of only a mild excess of probe, probes with a chemical structure that is not compatible with solid phase synthesis, and it takes advantage of a variety of commercial probes developed for protein bioconjugation. Furthermore, the labeling reactions are performed in solution on unprotected polypeptides and are, therefore, suitable for the multiple labeling of proteins, implying that solid phase synthesis is not required for the preparation of the fragments. As a consequence, even a large fragment can be prepared by molecular biology techniques, expanding the applicability of the described protocol. This latter characteristic is not trivial, as it could allow for the preparation of a protein labeled in any desired position and,

indeed, the preparation of proteins with probes shifting along the amino acid sequence. Finally, this approach could be adapted to the introduction of more than two molecular probes. To conclude, this method is versatile and allows for the preparation of labeled proteins with high purity and homogeneity. For these reasons it is useful for a wide range of biophysical and chemical biology applications.

Acknowledgements

This work was supported by Human Frontier Science Program Organization (Research Grant 2007, # RGP44/2007) and NIH RO1 GM080515 (LR).

References

1 V. Ratner, E. Kahana, M. Eichler and E. Haas, A general strategy for site-specific double labeling of globular proteins for kinetic FRET studies, *Bioconjugate Chem.*, 2002, **13**, 1163–1170.
2 M. W. Kao, L. L. Yang, J. C. Lin, T. S. Lim, W. Fann and R. P. Chen, Strategy for efficient site-specific FRET-dye labeling of ubiquitin, *Bioconjugate Chem.*, 2008, **19**, 1124–1126.
3 J. J. Smith, D. W. Conrad, M. J. Cuneo and H. W. Hellinga, Orthogonal site-specific protein modification by engineering reversible thiol protection mechanisms, *Protein Sci.*, 2005, **14**, 64–73.
4 M. Jäger, X. Michalet and S. Weiss, Protein–protein interactions as a tool for site-specific labeling of proteins, *Protein Sci.*, 2005, **14**, 2059–2068.
5 E. M. Brustad, E. A. Lemke, P. G. Schultz and A. A. Deniz, A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer, *J. Am. Chem. Soc.*, 2008, **130**, 17664–17665.
6 S. J. Miyake-Stoner, A. M. Miller, J. T. Hammill, J. C. Peeler, K. R. Hess, R. A. Mehl and S. H. Brewer, Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan, *Biochemistry*, 2009, **48**, 5953–5962.
7 J. M. Antos, G. L. Chew, C. P. Guimaraes, N. C. Yoder, GM Grotenbreg, MW Popp and HL Ploegh, Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity., *J. Am. Chem. Soc.*, 2009, **131**, 10800–10801.
8 A. P. Crochet, M. M. Kabir, M. B. Francis and C. D. Paavola, Site-selective dual modification of periplasmic binding proteins for sensing applications, *Biosens. Bioelectron.*, 2010, **26**, 55–61.
9 P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. Kent, Synthesis of proteins by native chemical ligation, *Science*, 1994, **266**, 776–779.
10 T. W. Muir, D. Sondhi and P. A. Cole, Expressed protein ligation: a general method for protein engineering, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6705–6710.
11 A. A. Deniz, T. A. Laurence, G. S. Beligere, M. Dahan, A. B. Martin, D. S. Chemla, P. E. Dawson, P. G. Schultz and S. Weiss, Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 5179–5184.
12 G. J. Cotton and T. W. Muir, Generation of a dual-labeled fluorescence biosensor for Crk-II phosphorylation using solid-phase expressed protein ligation, *Chem. Biol.*, 2000, **7**, 253–261.
13 J. Yang and W. Y. Yang, Site-specific two-color protein labeling for FRET studies using split inteins, *J. Am. Chem. Soc.*, 2009, **131**, 11644–11645.
14 T. Kurpiers and H. D. Mootz, Regioselective cysteine bioconjugation by appending a labeled cystein tag to a protein by using protein splicing in trans, *Angew. Chem., Int. Ed.*, 2007, **46**, 5234–5237.
15 G. Volkmann and X. Liu, Protein C-terminal labeling and biotinylation using synthetic peptide and split-intein, *PLoS ONE 4*, 2009, 12.
16 C. Ludwig, M. Pfeiff, U. Linne and H. D. Mootz, Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing, *Angew. Chem., Int. Ed.*, 2006, **45**, 5218–5221.
17 E. R. Main, Y. Xiong, M. J. Cocco, L. D'Andrea and L. Regan, Design of stable alpha-helical arrays from an idealized TPR motif, *Structure*, 2003, **11**, 497–508.
18 L. D. D'Andrea and L. Regan, TPR proteins: the versatile helix, *Trends Biochem. Sci.*, 2003, **28**, 655–662.
19 A. N. Warrens, M. D. Jones and R. I. Lechler, Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest, *Gene*, 1997, **186**, 29–35.
20 M. W. Southworth, K. Amaya, T. C. Evans, M. Q. Xu and F. B. Perler, Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein, *Biotechniques*, 1999, **27**, 110–120.
21 E. A. Peroza and E. Freisinger, Tris is a non-innocent buffer during intein-mediated protein cleavage, *Protein Expression Purif.*, 2008, **57**, 217–225.
22 D. E. Shafer, J. K. Inman and A. Lees, Reaction of Tris(2-carboxyethyl)phosphine (TCEP) with maleimide and alpha-haloacyl groups: anomalous elution of TCEP by gel filtration, *Anal. Biochem.*, 2000, **282**, 161–164.
23 K. Tyagarajan, E. Pretzer and J. E. Wiktorowicz, Thiol-reactive dyes for fluorescence labeling of proteomic samples, *Electrophoresis*, 2003, **24**, 2348–2358.

Published on 06 October 2011 on http://pubs.rsc.org | doi:10.1039/C1OB06397A
Downloaded by Universitaire d'Angers on 08 February 2012

 This journal is © The Royal Society of Chemistry 2012