CHEMICAL SYNTHESIS OF PROTEINS: A TOOL FOR PROTEIN LABELING

Lucia De Rosa

Dottorato in Scienze Biotecnologiche – XXIII ciclo
Indirizzo Biotecnologie Industriali e Molecolari
Università di Napoli Federico II

Dottorato in Scienze Biotecnologiche – XXIII ciclo
Indirizzo Biotecnologie Industriali e Molecolari
Università di Napoli Federico II

CHEMICAL SYNTHESIS OF PROTEINS: A TOOL FOR PROTEIN LABELING

Lucia De Rosa

Dottoranda:	Lucia De Rosa
Relatore:	Prof. Ettore Benedetti
Correlatore:	Dr. Luca D. D'Andrea
Coordinatore:	Prof. Giovanni Sannia

Ai miei genitori,
il punto di riferimento della mia vita.

INDEX

ABBREVIATIONS

Acm	acetamidomethyl
AW	azatryptophan
BAL	backbone amide linker
Boc	*t*-butoxycarbonyl
BSA	bovine serum albumin
c-Abl	Abelson nonreceptor protein tyrosine kinase
CD	circular dicroism
ctpr3	*Consensus* Tetratrico Peptide Repeat protein 3 gene
CTPR3	*Consensus* Tetratrico Peptide Repeat protein 3
Dab	dabcyl
Dansyl	5-(dimethylamino)-naphtalene-sulfonamide
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
Dbz	diaminobenzoic acid
DCM	dichloromethane
DIPEA	N,N-diisopropylethylamine
DMAP	4-Dimethylaminopyridine
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxy-nucleotide tri-phosphate
ε	extinction coefficient
E	FRET efficiency
E. coli	*Escherichia coli*
EDT	ethandithiol
EDTA	ethylene-diamino-tetraacetic acid
EPL	Expressed Protein Ligation
ESI	electron spray ionization source
FCS	Fluorescence Correlation Spectroscopy
Fmoc	fluorenylmethyloxycarbonyl chloride
FP	Fluorescence Polarization
FPLC	Fast Protein Liquid Chromatography

FRET	Förster Resonance Energy Transfer
GdnHCl	guanidinium hydrochloride
GFP	Green Fluorescent Protein
Gla	*γ*-carboxy-glutamic acid
GPI	glycosylphosphatidylinositol
HAT	histone acetyltransferase
HOBt	Hydroxybenzotriazole
HBTU	O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate
HF	fluorhydric acid
HPLC	High Performance Liquid Chromatography
Hsp	heat shock protein
H2B	Histone H2B
H3	Histone H3
Int^N	N-terminal intein domain
Int^C	C-terminal intein domain
IPTG	isopropylbeta-D-thiogalactopyranoside
KCL	Kinetically Controlled Ligation
LB	Luria-Bertani Broth
LC-MS	Liquid Chromatography Mass Spectrometry
Mca	Methoxycoumarin
MetAPs	methionyl aminopeptidases
MESNA	2-mercapto-ethane sulfonate sodium
MPAA	p-mercaptophenylacetic acid
***Mth* RIR1**	*Methanobacterium thermo-autotrophicum* ribonucleoside diphosphate reductase
Mw	Molecular Weight
***Mxe* GyrA**	*Mycobacterium xenopi* DNA gyrase A
NCL	Native Chemical Ligation
NMP	N-methyl-2-pyrrolidone
NMR	Nuclear Magnetic Resonance
***Npu* DnaE**	*Nostoc punctiforme* DnaE
OD	Optical Density
PCR	Polymerase Chain Reaction

PDA	photo diode array
pI	isoelectric point
PMSF	protease inhibitor phenylmethanesulfonyl-fluoride
PPII	polyproline II
PTMs	post-translational modifications
PTS	Protein *Trans*-Splicing
PVDF	polyvinylidene difluoride
R	distance between fluorophores
R_0	Förster radius
RabGGTase	Rab geranylgeranyl transferase
REP	RabGGTase escort protein
RNA	ribonucleic acid
***Sce* VMA**	*Saccharomyces cerevisiae* vacuolar ATPase subunit
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SH	Src homology domain
SPPL	Solid-Phase Expressed Protein Ligation
SPPS	Solid-Phase Peptide Synthesis
***Ssp* DnaB**	*Synechocystis* sp. PCC6803 DnaB helicase
STAT	signal transducer and activator of transcription
SUMO	Small Ubiquitin-like Modifier
TAE	Tris acetate EDTA
T*β*R-I	transforming growth factor *β* receptor
t-BuSH	*tert*-butyl mercaptan
TCEP	Tris(2-carboxyethyl)phosphine
TEV	protease from tobacco etch virus
TFA	trifluoroacetic acid
TGF-*β*	transforming growth factor *β*
Thz	thiazolidine
Tm	melting temperature
TPR	Tetratrico Peptide Repeat
Tris	Tris (hydroxy methyl) amino methane
U	units
Ub	ubiquitin
UV	ultra violet

VA-044	2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride
Vis	visible

The three letters code is used for amino acids.

SUMMARY

An ingenious stratagem useful to understand and modulate the structural and functional features of the proteins refers to the modification of their chemical structure. In this regard, the chemical synthesis of proteins appears a key tool, as it allows the unlimited modification of a polypeptide chain with any kind and number of labels. The last decade has seen the introduction of several techniques of chemical protein synthesis that allow the manipulation of proteins structure [Hahn, M.E. & Muir, T.W. 2005; Kent, S.B.H. 2009]. Among these approaches, the Native Chemical Ligation (NCL) appear as the most useful and advantageous strategy [Dawson, P.E. *et al.* 1994]. The method leads to the formation of a protein molecule through the chemoselective ligation of two peptide segments containing the former a C-terminal α-thioester group and the latter an N-terminal Cys residue. Each peptide segment is prepared by solid-phase peptide synthesis (SPPS) and thus can contain any useful chemical modifications, such as unnatural amino acids, post-translational modifications, isotopes and fluorophores. A semisynthetic version of NCL is the Expressed Protein Ligation (EPL), in which one or both of the peptide building blocks are made by recombinant DNA expression, while the actual ligation step is still a chemical reaction [Muir, T.W. *et al.* 1998; Evans T.C., *et al.* 1998]. This PhD thesis represents a practical guide to protein chemical synthesis and modification, as it presents and discusses a general protocol useful for the preparation of proteins site-specifically functionalized with two molecular probes. The designed synthetic approach is based on the use of EPL and thus requires the splitting of the protein of interest in two fragments. The N-terminal protein fragment is expressed by recombinant means and is endowed of a C-terminal thioester group. Such fragment can be selectively labeled on its unique Cys residue with a probe containing a thiol-reactive moiety. After first labeling, the thioester fragment is ligated to the C-terminal remaining protein portion, which is synthesized by SPPS and carries an N-terminal Cys residue, affording the full-length mono-labeled protein. The Cys residue involved in the ligation reaction is finally exploited to introduce the second probe into the protein, giving the doubly-functionalized full protein molecule. The effectiveness of the proposed semisynthetic protocol has been assessed through the preparation of a series of CTPR3 (*Consensus* Tetratrico Peptide Repeat protein 3) protein variants in which two fluorophores are incorporated at different, specific positions. The protocol enabled the preparation of four homogeneously doubly-labeled variants, defined CTPR3[N_C], CTPR3[1_C], CTPR3[2_C] and CTPR3[1_3], whose name suggest the positions in which the two fluorophores are located. The obtained fluorescent variants are useful for the conduction of folding studies by FRET (Förster resonance energy transfer). In this thesis are presented the results of a preliminary biophysical investigation performed on the obtained fluorescent CTPR3 variants by circular discroism and FRET. Such analysis opens the way to further folding studies on the fluorescent variants of CTPR3 by the use of modern biophysical spectroscopic methods, such as the single molecule approaches.

RIASSUNTO

INTRODUZIONE

In questo lavoro di tesi è proposto un efficiente protocollo di sintesi chimica utile per la produzione di proteine sito-specificamente funzionalizzate con una coppia di sonde molecolari.

La sintesi chimica di proteine rappresenta una vantaggiosa alternativa ai metodi di espressione in organismi geneticamente ingegnerizzati, poiché offre la possibilità di modificare la struttura chimica di una proteina in maniera controllata e sito-specifica, introducendo in essa un'ampia gamma di modifiche come isotopi stabili, amminoacidi non naturali, modifiche post-traduzionali e sonde fluorescenti, aprendo la strada a svariati studi strutturali e funzionali [Hahn, M.E. & Muir, T.W. 2005; Kent, S.B.H. 2009]. Attualmente, le strategie di sintesi chimica di proteine sono per lo più basate sull'impiego di approcci di *Chemical Ligation* [Schnölzer, M. & Kent, S.B.H. 1992], i quali permettono di sintetizzare una catena polipeptidica mediante l'unione in sequenza dei segmenti peptidici che la compongono, ciascuno dei quali è prodotto mediante sintesi in fase solida (SPPS). In origine, gli approcci di CL comportavano la formazione di legami non peptidici in corrispondenza del sito di unione tra due peptidi, introducendo nella proteina alterazioni indesiderate dello scheletro polipeptidico. Successivamente, tale svantaggio è stato superato con l'introduzione della *Native Chemical Ligation* (NCL) che al momento è la strategia di CL più ampiamente adoperata per la sintesi chimica di proteine [Dawson, P.E. *et al.* 1994]. Nelle reazioni di NCL, un peptide contenente un gruppo tioestere all'estremità C-terminale è fatto reagire con un secondo segmento peptidico contenente un residuo di Cys a monte, portando ad un unico prodotto polipeptidico tramite la formazione inizialmente di un legame tioestere, il quale è poi rapidamente e irreversibilmente convertito in un legame peptidico. Un'ingegnosa estensione della NCL è la strategia di semisintesi proteica nota come *Expressed Protein Ligation* (EPL) [Muir, T.W. *et al.* 1998; Evans T.C. *et al.* 1998]. L'EPL è una variante delle strategie di CL in cui uno o più dei segmenti peptidici da unire in sequenza non è di origine sintetica bensì è espresso in forma ricombinante. Infatti, è possibile preparare i segmenti polipeptidici contenenti un gruppo tioestere al C-terminale o una cisteina all'estremità N-terminale per *via* ricombinante mediante la loro espressione sotto forma di proteine di fusione con particolari proteine, le inteine. Queste ultime sono in grado di promuovere un processo noto come *splicing* proteico, attraverso cui il polipeptide fuso ad esse è rilasciato con un gruppo tioestere al C-terminale o con una cisteina all'N-terminale, a seconda che tale polipeptide sia stato posto rispettivamente a monte o a valle dell'inteina.

Le strategie di sintesi chimica di proteine hanno trovato ampia applicazione in svariati campi scientifici, come quello biotecnologico, biofisico e biochimico [Muralidharan, V. & Muir, T.W. 2006]. In questo scenario si inquadra il presente lavoro di tesi.

RISULTATI

Strategia semisintetica

La strategia semisintetica ideata, schematicamente rappresentata in Fig. 1, prevede la preparazione della proteina di interesse in due frammenti. Il frammento N-terminale della proteina, contenente un unico residuo di cisteina, è espresso in *E. coli* sotto forma di proteina di fusione con l'inteina *Mxe* GyrA. Dopo l'espressione, la proteina chimerica è isolata dall'estratto citosolico dell'ospite batterico mediante cromatografia di affinità, sfruttando la presenza di una coda di poli-istidine al C-

terminale del costrutto di fusione. La proteina di fusione è poi incubata con un eccesso di tiolo, il quale è capace di indurre lo *splicing* dell'inteina con il conseguente rilascio del polipeptide fuso ad essa sotto forma di tioestere C-terminale. Successivamente, tale polipeptide tioestere è isolato dall'inteina mediante HPLC a fase inversa e marcato con il primo *probe* della coppia in corrispondenza dell'unico residuo di cisteina, sfruttando la reattività verso i tioli del gruppo maleimmide associato alla sonda. La rimanente porzione C-terminale della proteina, invece, è sintetizzata chimicamente mediante SPPS e espone un residuo di cisteina come amminoacido N-terminale. Tale residuo media una reazione di NCL con il polipeptide tioestere singolarmente marcato, portando alla formazione della proteina intera. Il residuo di cisteina coinvolto nella reazione di NCL può infine essere sfruttato per introdurre il secondo *probe* (Fig. 1 strategia A).

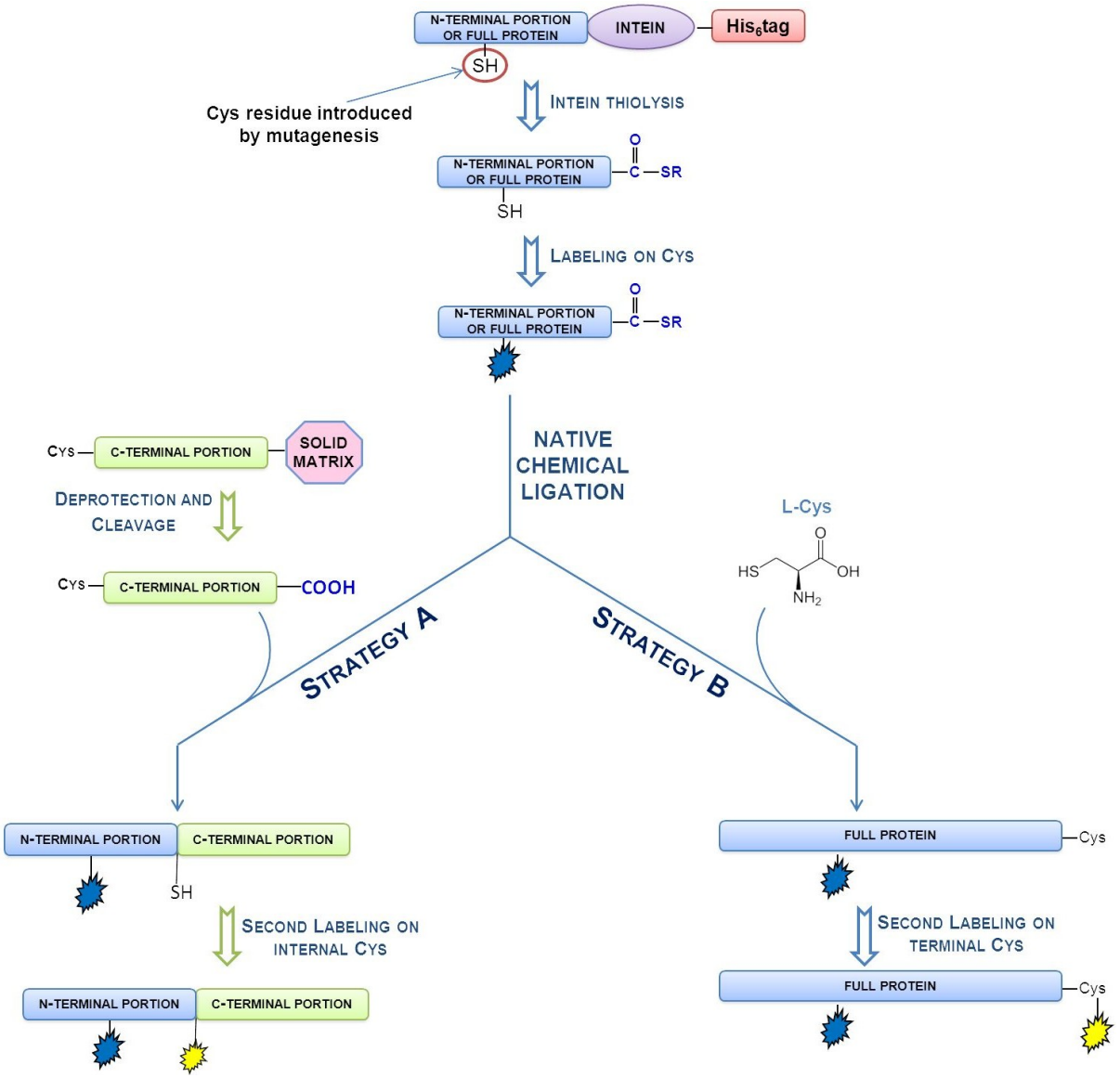

Figura 1 Rappresentazione schematica della strategia semisintetica progettata per la preparazione di proteine funzionalizzate con una coppia di sonde molecolari. **Strategia A)** Approccio adottato qualora entrambe le sonde debbano essere introdotte in regioni interne della proteina. **Strategia B)** Approccio sintetico adoperato per la preparazione di varianti proteiche che espongono una delle due sonde in corrispondenza dell'estremità C-terminale.

Una variante più semplice della strategia semisintetica presentata è invece necessaria per la preparazione di varianti della proteina di interesse doppiamente marcate recanti il secondo *probe* all'estremità C-terminale. In questo caso, l'intera proteina è espressa sotto forma di costrutto di fusione con l'inteina e così dotata di un gruppo funzionale tioestere al C-terminale. Dopo il suo isolamento e la marcatura con il primo *probe*, la proteina è fatta reagire mediante NCL con L-Cisteina introducendo così il sito per la marcatura con la seconda sonda (Fig. 1 strategia B).
L'efficacia della strategia semisintetica ideata per la preparazione di proteine funzionalizzate con una coppia di sonde molecolari è stata dimostrata attraverso la preparazione di quattro varianti di una *Tetratrico Peptide Repeat protein* (*TPR protein*) marcate con sonde fluorescenti, utili per la conduzione di studi di *folding* mediante *Förster Resonance Energy Transfert* (FRET).

Applicazione del protocollo di semisintesi: preparazione di varianti fluorescenti di CTPR3

Le *TPR proteins* sono caratterizzate da ripetizioni in sequenza di un motivo strutturale elicoidale di 34 amminoacidi, le quali complessivamente si organizzano a formare strutture estese e regolari. Le proteine appartenenti a questa classe sono coinvolte in svariati processi biologici, quali la regolazione del ciclo cellulare, il controllo trascrizionale, il trasporto a livello mitocondriale e perossisomale, la neurogenesi ed il *folding* proteico. Il loro compito consiste generalmente nel mediare interazioni proteina-proteina, esponendo ampie superfici che fungono da *scaffold* per l'assemblaggio di complessi macchinari biologici [Main, E.R.G. *et al.* 2003]. Attraverso un'analisi statistica delle sequenze amminoacidiche di domini TPR contenuti in proteine naturali, il gruppo di ricerca diretto dalla Prof.ssa L. Regan (*Department of Molecular Biophysics & Biochemistry*, *Yale University*) ha identificato una sequenza consenso del motivo TPR. Il *design* di un motivo TPR "ideale", il quale è stato adoperato come unità di base per costruire nuove proteine TPR contenenti un numero variabile di ripetizioni (definite con la sigla CTPRn, *Consensus* TPR numero di ripetizioni), costituisce un valido strumento per lo studio di questa particolare classe di *repeat proteins* [Kajander, T. *et al.* 2006]. Numerose analisi condotte mediante CD ed NMR sulle proteine CTPR suggeriscono che il processo di *unfolding* di tali proteine ripetute non segue il classico modello a due stati, caratterizzato da conformazioni completamente destrutturate o completamente strutturate, ma si adatta piuttosto al cosiddetto modello di Ising [Kajander *et al.* 2005; Cortajarena, A.L. *et al.* 2008; Javadi, Y. & Main, E.R.G. 2009]. Tale modello prescrive un processo di *unfolding* gerarchico, in cui le eliche più esterne delle proteine CTPR si destrutturano prima di quelle più interne, cosicché globalmente il processo di denaturazione proceda dalle estremità N- e C-terminale verso la regione centrale della proteina, attraversando quindi un'ampia popolazione di stati parzialmente strutturati.
Al fine di ottenere ulteriori informazioni sul meccanismo di *folding* di questa particolare classe di proteine, la strategia di semisintesi messa a punto per la preparazione di proteine doppiamente funzionalizzate con sonde molecolari discussa in questo lavoro di tesi è stata adoperata per la sintesi di quattro varianti della proteina CTPR3 (Fig. 2) sito-specificamente funzionalizzate con una coppia di fluorofori, utili per la conduzione di esperimenti mediante FRET.
Le quattro varianti di CTPR3 prodotte attraverso l'approccio semisintetico ideato sono state definite CTPR3[N_C], CTPR3[1_C], CTPR3[2_C] e CTPR3[1_3] e si distinguono per la posizione dei due fluorofori lungo la sequenza. La variante CTPR3[1_3] presenta entrambe i fluorofori in corrispondenza di siti interni della

proteina, precisamente in corrispondenza della prima e della terza ripetizione del motivo CTPR. Pertanto, per la sua preparazione è stata adottata la strategia semisintetica A di Fig. 1. Le rimanenti tre varianti CTPR3[N_C], CTPR3[1_C] e CTPR3[2_C], le quali presentano il secondo fluoroforo in corrispondenza dell'estremità C-terminale della proteina, sono invece state preparate attraverso la strategia B di Fig. 1. I siti di *labeling* nelle quattro varianti sono stati selezionati in modo tale che in ciascun costrutto i fluorofori fossero posizionati a differente distanza tra loro e in regioni progressivamente più interne della proteina.

Figura 2 Rappresentazione grafica della proteina CTPR3.

Caratterizzazione spettroscopica delle varianti fluorescenti di CTPR3

Le quattro varianti fluorescenti di CTPR3 sono state sottoposte ad una caratterizzazione strutturale preliminare mediante CD, da cui è risultato che le modifiche introdotte (mutazioni amminoacidiche e fluorofori) non compromettono significativamente la struttura nativa della proteina, sebbene ne riducano la stabilità termica. Studi di anisotropia di fluorescenza hanno inoltre dimostrato che i fluorofori non mediano fenomeni di aggregazione o interazione nelle proteine marcate. In seguito, sono stati registrati gli spettri di fluorescenza di ciascun costrutto nella forma nativa e chimicamente denaturata. Tale analisi ha mostrato una sensibile variazione dell'efficienza di FRET tra i due fluorofori della coppia passando dalla forma nativa a quella denaturata di ciascuna variante di CTPR3, dimostrando che in tutti i costrutti i fluorofori scelti per la marcatura (ATTO 488 e ATTO 647N) sono sensibili alla variazione di distanza che si verifica in seguito ad *unfolding*.

Considerando che il meccanismo di denaturazione delle proteine TPR sia di tipo sequenziale, procedendo dalle estremità terminali verso la regione centrale della proteina [Kajander *et al.* 2005; Main, E.R.G. *et al.* 2005; Cortajarena, A.L. *et al.* 2008; Hagay, T. & Levy, Y. 2008; Javadi, Y. & Main, E.R.G. 2009], abbiamo ipotizzato che, monitorando il processo di *unfolding* in ciascuna variante di CTPR3 mediante CD e FRET, si sarebbero ottenuti risultati differenti per ciascun costrutto doppiamente marcato. Infatti, mentre la spettroscopia CD permette di seguire il processo di *unfolding* di una proteina nel suo complesso, perché osserva variazioni del contenuto totale di struttura secondaria, al contrario la FRET è in grado di seguire il processo di *unfolding* della regione compresa tra i due fluorofori ed è dunque sensibile ad eventi di destrutturazione locale. Per verificare tale ipotesi, le quattro varianti doppiamente marcate sono state sottoposte ad esperimenti di denaturazione chimica condotti monitorando il processo *unfolding* sia mediante CD che FRET. Per i costrutti

CTPR3[1_3] e CTPR3[2_C] sono stati ottenuti risultati in accordo con il meccanismo di *unfolding* predetto per tali proteine. Invece, per le varianti CTPR3[N_C] e CTPR3[1_C] gli studi hanno portato ad osservazioni non in accordo con un meccanismo di *unfolding* sequenziale.

CONCLUSIONI

In conclusione, la strategia di semisintesi proteica progettata fornisce un protocollo di carattere generale per la funzionalizzazione multipla e sito-specifica di proteine. In questo lavoro di tesi la strategia ideata è stata impiegata per applicazioni biofisiche, attraverso la preparazione di quattro varianti della proteina CTPR3 marcate con due fluorofori utili per la conduzione di studi di *folding* tramite FRET e CD. Il risultato intrigante di tali studi è indicativo di una deviazione da un semplice scenario in cui le transizioni FRET sono funzione esclusivamente della perdita di struttura elicoidale. Tali studi inoltre pongono le basi per la conduzione di più approfondite analisi mediante tecniche all'avanguardia, come la *single molecule*-FRET, volte a chiarire i dettagli ancora insoluti delle dinamiche di *folding* delle proteine TPR ed in generale di tutte le *repeat proteins*.

INTRODUCTION

The need for protein modification strategies

Providing a complete understanding of proteins structure and function has emerged as one of the great challenges in biochemistry in the post-genomics era. In the analysis of structure-activity relationship, the ability to manipulate the chemical structure of proteins appears as an ingenious stratagem to understand the basis of their function and structure. For the past thirty years, DNA recombinant techniques have been widely employed to systematically vary the amino acid sequence of a polypeptide in order to correlate the observed changes in the properties of the mutant protein with the change in the amino acid sequence. This approach, named “protein engineering”, has provided much useful insight into how the structure of a protein gives rise to its functional properties [Wells, J.A. & Estell, D.A. 1998]. The power of this method includes its applicability both *in vitro* and *in vivo*. However, from a chemist’s point of view the molecular biology approach is subject to severe limitations: only the twenty genetically encoded amino acids can be introduced or substituted into a protein molecule and site-specific incorporation of post-translational modifications or molecular probes are both technically difficult and impossible to control precisely. Great efforts have been made to overcome these limitations, for instance by the use of cell-free translational system to expand the repertoire of ribosomal synthesis to include non-canonical amino acids as building blocks [Mendel, D. *et al.* 1995; Xie, J. & Schultz, P.G. 2005], and by the use of group-modifying reagents which facilitate the selective tagging of proteins isolated from living cells. These sophisticated approaches are original tools, but so far had only limited impact because relative subtle modifications are often not possible, such as replacement of single atom in a protein with an isotope thereof, or another atom. In addition, it is generally not possible to incorporate multiple, different molecular probes into discrete sites or regions of a protein using standard labeling procedures. These and other restrictions continued to fuel the development of new approaches for protein modifications [Kent, S.B.H. 2009].

Chemical synthesis of protein

Great innovations in the field of proteins modification were expected to come from the chemical synthesis. The use of synthetic chemistry promises the unlimited variation of the covalent structure of a polypeptide chain, enabling to effect, at-will, any desired change in the protein molecule. Chemistry ensures the ability to systematically tune the properties of a protein molecule in a completely general fashion. Chemical synthesis also allows the labeling of a protein without limitations with respect to the type and number of labels, yet with the atom-by-atom precision that is essential, for instance, to the modern biophysical methods. For example, chemical synthesis is critical for realizing the full potential of single molecule and fluorescence studies [Kent, S.B.H., 2009]. The last decade, in particular, has seen the introduction of several techniques of chemical protein synthesis that allow the covalent structure of proteins to be manipulated with an unprecedented level of control [Hahn, M.E. & Muir, T.W., 2005].

Classical organic synthesis in solution

The quest to make enzymes and other protein molecules by chemical synthesis was one of the major challenges for organic chemistry in the twentieth century. The desire to achieve this goal led to the development of a vast array of synthetic peptide chemistry methods which culminated in the synthesis of a number of proteins, such as human insulin protein (51 amino acids) and a series of chemical insulin analogues [Sieber, P. *et al.* 1978], the enzyme ribonuclease A (124 amino acids) [Yajima, H. & Fujii, N. 1981] and a consensus lysozyme enzyme molecule (129 amino acids) [Kenner, G. W. 1977]. The protein syntheses mentioned above were carried out in solution and involved the preparation of short fully-protected peptide segments and their subsequent condensation in organic solvents (Fig. 1).

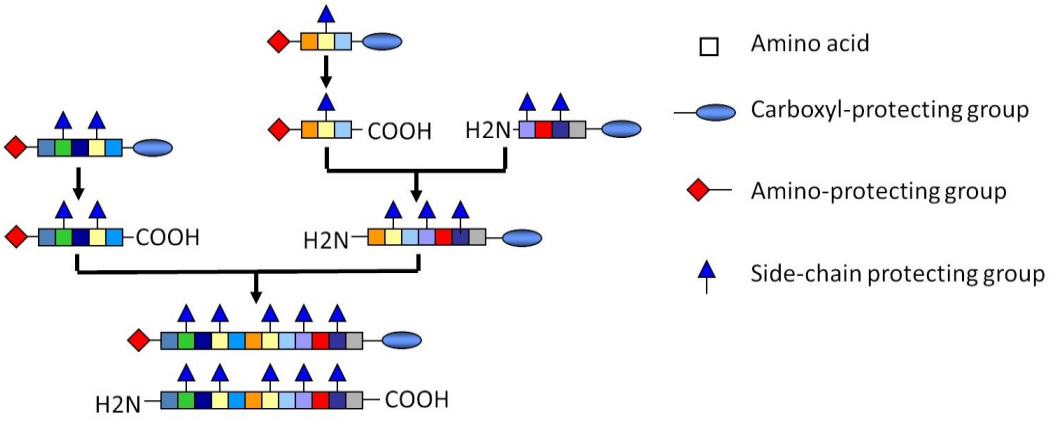

Figure 1 Protein synthesis in solution phase. Peptide segment building blocks are reacted in a fully convergent strategy; all non-reacting functional groups in both peptide segments are masked by reversible protecting groups. In the final step, all the protecting groups are removed to give the full-length polypeptide product [Kent, S.B.H. *et al.* 2009].

This classical solution approach to the chemical synthesis of polypeptide chains suffered from a number of shortcomings, including laborious and technically demanding preparation of the protected segments, lack of chiral homogeneity in peptide bond formation, inability to purify and characterize fully-protected peptides due to their low solubility [Kent, S.B.H., 1988]. Moreover, the sophisticated chemistry used required an exceptionally high level of skill because syntheses were arduous, requiring teams of expert chemists to carry out them effectively. Although a number of proteins were successfully synthesized by classical methods in solution, known as convergent synthesis, the limitations encountered using maximally protected peptide segments were never overcome, so that this sophisticated synthesis was abandoned.

Solid phase peptide synthesis

In 1963 Merrifield introduced a novel synthetic technique that greatly facilitated the chemical synthesis of peptides, the stepwise solid-phase peptide synthesis (SPPS). In SPPS, the C-terminal amino acid residue of a target peptide sequence, with side-chain properly protected, is covalently attached to an insoluble polymeric support; for the introduction of the subsequent amino acid residue, the N(α)-protecting group (a *t*-butoxycarbonyl –Boc- or 9-fluorenylmethoxycarbonyl –Fmoc- group) of the resin-bound first amino acid is removed and the resin is purified by filtration and washing; then, the next amino acid with the N(α)- and side chain group protected is added in a carboxyl-activated form; after formation of the new peptide bond, excess of activated

amino acid and reaction by-products are removed by filtration and washing. These steps are sequentially repeated until the target resin-bound protected peptide chain is completely assembled. In the final step, all side-chain protecting groups are removed and simultaneously the covalent link to the resin is cleaved to release the fully-unprotected crude peptide product (Fig. 2).

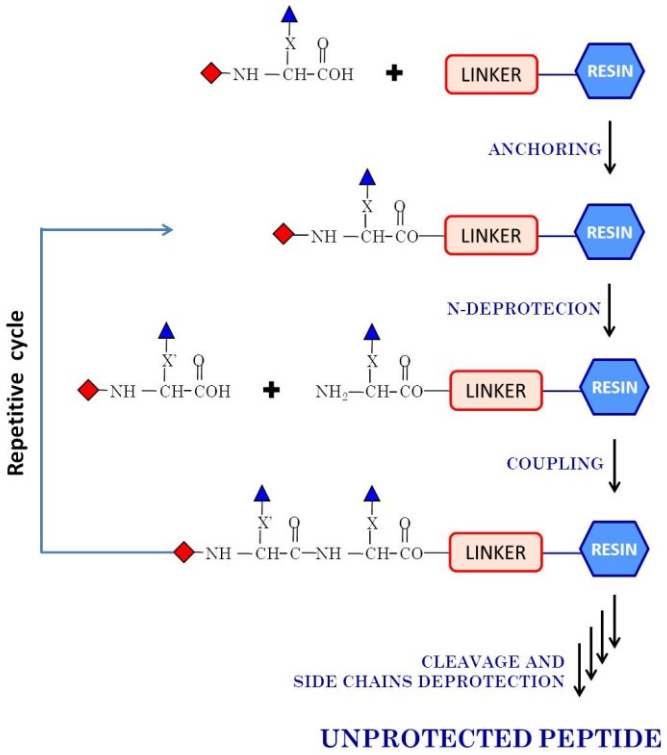

Figure 2 Solid phase peptide synthesis. The fully-protected peptide chain is built up in stepwise fashion from the C-terminal amino acid, which is covalently attached to an insoluble polymer support. In the final step, all the protecting groups are removed and the covalent link to the polymer support is cleaved to give the full-length polypeptide product in solution [Kent, S.B.H. *et al.* 2009].

Facile purification by filtration and washing at each step of the synthesis enables the use of large excess of activated amino acids for each peptide bond forming step. Consequently, reactions are rapid and near-quantitative. Handling losses are minimized because the growing peptide chain remains bound to the insoluble support at all stages of the synthesis. As a result, with correctly designed and well-executed chemistry, crude products containing a high proportion of the desired peptide are obtained in high yield. The solid-phase method revolutionized the chemical synthesis of peptides because attachment of the growing protected peptide chain to the resin support overcomes the problems of the solution approaches. Despite the extraordinary power of SPPS in respect to the classical synthesis in solution, SPPS had no impact on the maximum size of polypeptides that could be made by chemical means because the lack of full-quantitative reactions and the presence of impurities in the solvents and reagents leads to the formation of resin-bound by-products at each step of reaction. This accumulation of co-products limits the ultimate size of high-purity polypeptides that can be effectively prepared by this way to chains of ~50 amino acids in length. This corresponds to only the very smallest proteins and protein domains. Clearly, a more powerful approach based on novel principles was needed.
A number of attempts were made to take advantage of the gained ability to make, characterize and handle unprotected peptides, such as the preparation of large

polypeptides from synthetic segments through transpeptidation reactions catalyzed by enzymes specifically engineered by molecular biology for this purpose [Chang, T.K., *et al.*, 1994; Braisted A.C., *et al.,* 1997]. The principal obstacle to general utility of enzymatic ligation has proven to be the limited solubility of even the unprotected peptide segments under the physiological conditions compatible with the enzymes used. So, despite enzymes such as subtiligase and sortase still remain in use today for notable applications in proteomics [Wildes D. & Wells J.A., 2010; Yoshihara H.A., 2008] and biotechnology [Antos J. M. *et al.*, 2009; Proft, T., 2009], such methods have not found widespread use.

Chemical ligation reaction

Chemical ligation of unprotected peptide segments

A truly useful approach to chemical synthesis of proteins would take advantage of the ability to easily make unprotected peptide shorter than 50 amino acids in length and would consist of a practical way to stich such synthetic segments together to give polypeptide of any desired length, and hence the corresponding folded protein molecules. Based on this premise, in the early 1990s was introduced a novel concept that revolutionized the field of chemical synthesis of proteins, the chemoselective condensation of unprotected peptides [Schnölzer, M. & Kent, S.B.H., 1992]. In essence, in a chemoselective condensation reaction between unprotected peptides, two unique, mutually reactive groups are employed, one on each of the reacting unprotected peptides; these two reactive functionalities are designed to react with one another, but to not react with any of the other functional groups present in peptides, thus giving a single reaction polypeptide product (Fig. 3).

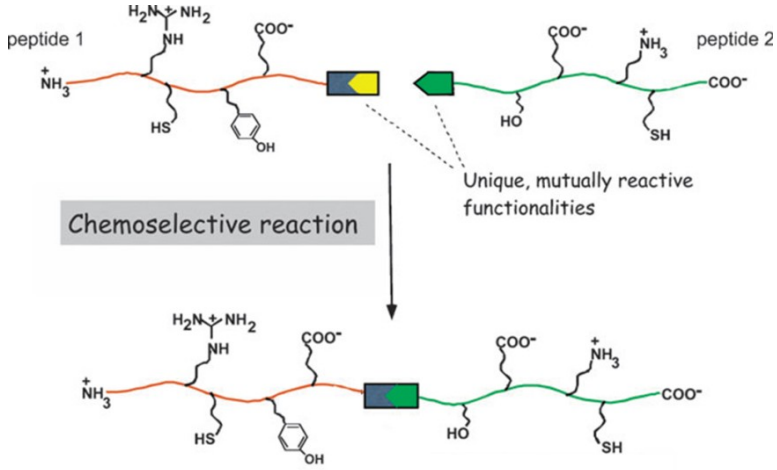

Figure 3 Principles of chemical ligation. Two unprotected peptide segments are covalently joined by the chemoselective reaction of unique, mutually reactive functional groups, one on each reacting segments [Kent, S.B.H. 2009].

This novel approach, named “chemical ligation”, has proven to be simple to implement, highly effective and generally applicable [Muir, T.W. & Kent, S.B.H., 1993]. Examples of the first proteins prepared by this way include HIV-1 protease [Schnolzer, M. & Kent, S.B.H., 1992] and the mirror image enzyme D-HIV-1 protease [deLisle-Milton, R. *et al.*, 1993], the fully functional covalent heterodimer of b/HLH/Z transcription factor cMyc-Max [Canne, L. E. *et al.*, 1995]. These and lots of other protein syntheses performed by the chemical ligation methods demonstrated that

proteins could be made in high yield and good purity from unprotected peptide building blocks and that unnatural analogs, containing for example post-translational modifications, molecular probes or unnatural amino acids, could be readily prepared to investigate new aspects of protein structure and function.

Chemical ligation reactions yielding a non-native link

A variety of chemistries has been used for the chemical ligation of unprotected peptides but they usually involved the formation of an non-peptide bond at the site of covalent linking between two reacting segments. However, these unnatural structures are often well tolerated within the context of a folded protein and numerous examples exist of fully active proteins prepared by chemical ligation strategies yielding a non-native bond.

From the mid-1980, hydrazone and oxime ligations found widespread use in synthesis of proteins [Gaertner, H.F. *et al.* 1992]. The reaction between a carbonyl group and an hydrazine or an aminooxy group is highly chemoselective and both reacting groups can be straightforwardly introduced in peptide by means of chemical organic synthesis. Slow reaction rate limited the application of this ligation methods, but the ability of anilines to act as a nucleophilic catalyst for hydrazone and oxime reaction, increasing reaction rate of three orders of magnitude, suggests a promising future for these kinds of imine ligations [Dirksen, A. & Dawson, P.E. 2008].

The Cu(I)-catalyzed azide-alkine [3+2] cycloaddition, better known as click-chemistry [Kolb, H.C. *et al.* 2001], is applied in peptide ligation, even if not as often as in the labeling of biological macromolecules [Dirksen, A. & Dawson, P.E., 2008]. A notable example of head-to-tail cyclodimerization of a protected peptide on a resin was described by Finn and co-workers [Punna, S. *et al.* 2005]. The addition of the Cu(I) catalyst was essential for the reaction to occur.

Another non-native ligation chemistry employs the simple nucleophilic reaction of a peptide-thiocarboxylate with a bromoacetyl-peptide in aqueous solution at low pH, to give a thioester-linked polypeptide product. The thioester moiety is a reasonable facsimile of a peptide bond and the product thioester containing polypeptide is stable in lyophilized form [Schnolzer, M. & Kent, S.B.H., 1992].

Developed by Saxon and Bertozzi in 2000, the Staudinger ligation reaction proceeds through the nucleophilic attack of a phosphine on an azide to give an azaylide intermediate, which is subsequently trapped in an intramolecular fashion by a methoxycarbonyl group on one of the aryl rings of the phosphine to give an amide-linked phosphine oxide after hydrolys. The reaction proceeds more rapidly in protic, polar solvents, such as water [Lin, F.L. *et al.* 2005] and has been successfully applied in cellular labeling and in a FRET-based assay for live-cell imaging [Hangauer, M.J. & Bertozzi, C.R. 2008].

Several other ligation chemistries using chemoselective reaction to form a non-native covalent link between two unprotected peptides have been introduced and used for the chemical synthesis of proteins [Dawson, P.E. & Kent, S.B.H. 2000]. One notable achievement has been the exploration of the Diels-Alder reaction for the ligation unprotected peptides. The chemoselective reaction involves a C-terminal hexadienyl ester and an N-terminal maleimide group and proceeds in high yield under mild aqueous conditions [Dantas de Araújo, A. *et al.* 2006]. Since both functionality can be readily introduced on the solid-phase, the method may find broad utility with peptides not containing free cysteines, due to the reactivity of maleimide toward thiol groups.

Another promising strategy is the application of [Ru]-catalyzed ring-opening cross-metathesis to the ligation of peptides. In some examples, the conjugation of

hydrophobic peptides in organic solvent yielded a linkage that could be used to replace proline or to introduce β–turn stabilizing units [Michaelis, S. & Blechert, S. 2007].

Native Chemical Ligation

In 1994 Dawson *et al.* introduced an ingenious extension of the chemistries used for the chemoselective ligation of unprotected peptides, named Native Chemical Ligation (NCL) [Dawson, P.E. *et al.* 1994]. The mechanism of NCL is shown in Fig. 4.

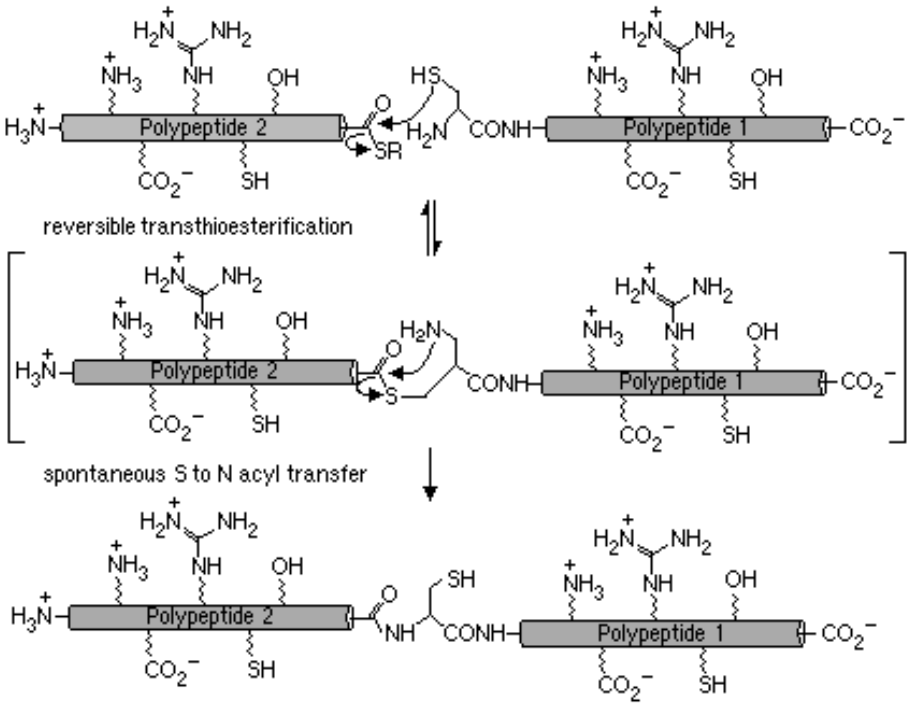

Figure 4 Principle of native chemical ligation. Both peptide reactants are fully unprotected and the reaction proceeds in water at or around neutral pH. Polypeptide building blocks contain the required reactive groups, namely a α-Cys or α-thioester [Dawson, P.E. & Kent, S.B.H. 2000].

The method leads to the formation of a single polypeptide chain bearing a native peptide bond at the ligation site simply mixing together two peptide segments one containing a carboxy-terminal α-thioester group and the other an amino-terminal Cys residue. The reaction is performed in aqueous solution at neutral pH under denaturing condition and proceeds through a trans-thioesterification reaction to the transient formation of a thioester intermediate, as in the strategy giving thioester-linked polypeptide product mentioned before in this section. In NCL, however, this thioester-linked adduct is not stable in the conditions used for the reaction and undergoes a spontaneous rearrangement *via* intramolecular S-to-N acyl shift, giving the desired amide-linked product. The target full-length polypeptide chain is obtained without any further manipulation. In the presence of exogenous thiol, added as catalyst, the formation of the thioester intermediated is freely reversible under the neutral aqueous condition used for the reaction, instead intramolecular nucleophilic attack forming the amide bond at the ligation site is irreversible under the same conditions, so that, the overall reaction is shifted towards the formation of the native polypeptide product [Dawson, P.E. & Kent, S.B.H. 2000]. The yield and rate of native

chemical ligation is strictly dependent on the nature of the thioester moiety involved into the reaction. Peptide α-thioalkyl-esters are commonly used because of their ease of preparation. These thioalky-ester are rather unreactive, so the ligation reaction is generally catalyzed by *in situ* transthioesterification with thiols additives, usually thiophenol/benzyl mercaptan, sodium 2-mercaptoethanesulfonate (MESNA) or p-mercaptophenylacetic acid (MPAA) [Johnson, E.C.B. & Kent, S.B.H. 2006]. Furthermore, the addition of a thiol excess not only increases the reactivity by forming new thioesters through transthioesterification but also keeps the thiol-functions reduced by preventing oxidation of the N-terminal Cys. The addition of solubilizing agents such as urea or guanidinium hydrochloride (GdnHCl) can be used to increase the concentration of peptide segments, resulting in higher yields. Thus, after ligation, a refolding step may be required to obtain a functional protein. Similarly, an oxidation step is needed if the protein contains native disulfides. All the 20 proteinogenic amino acids can be placed at the C-terminus of the thioester peptide, in the -X-Cys- region. However, they show different compatibility and efficiency in a NCL reaction. In fact, Thr, Val, Ile and Pro are reported to react slowly [Hackeng, T.M. *et al.* 1999]; Asp and Glu are less favorable because of the formation of side products [Villain, M. *et al*. 2003]. An interesting feature of NCL is that ligation occurs at a unique N-terminal Cys residue, even if many additional internal Cys are present in either segment [Dawson, P.E. *et al.* 1994; Hackeng, T.M. *et al*., 1997]. No protecting group are necessary for any of the side-chain functional groups normally found in natural proteins and near-quantitative yields of the ligation product are obtained. Another remarkable feature of NCL is the absence of racemization in the coupling reaction. Detailed studies have been carried out and no racemization was detected in the ligation product to a limit of <1% D-amino acid content [Lu, W.Y. *et al.* 1996]. All these features have meant that NCL, together with the correlated Expressed Protein Ligation (EPL) (see below), is nowadays the most widely used method for chemical protein synthesis.

Expressed Protein Ligation

EPL is a semisynthetic version of NCL in which one or both of the peptide building blocks are made by recombinant DNA expression, while the actual ligation step is still a chemical reaction between a thioester moiety and a Cys [Muir, T.W. *et al.* 1998; Evans T.C., *et al.* 1998]. EPL overcomes most of the shortcomings of precedent protein synthesis and modification approaches due to the truly interdisciplinarity of its execution which combines the strength of organic synthesis and recombinant DNA technology. In EPL, a recombinant C-terminal thioester segment reacts with a chemically synthetized or expressed Cys-segment (or *vice versa*) under the same condition of NCL. As in NCL, even if of recombinant origin, each fragments can include useful modifications for biotechnological or biophysical applications, such as synthetic labels. The advantage of this approach respect to the classical NCL is that the expressed part is not limited in size, expanding the access of semisynthetic approaches also to very large proteins.

Performing Native Chemical Ligation and Expressed Protein Ligation

The great challenge of NCL and EPL refers to the ability of preparing C-terminal thioester or N-terminal Cys-containing polypeptides by chemical or recombinant means. Robust approaches are available for introducing these reactive handles into peptides and polypeptides.

Production of thioester peptides

Currently, the most effective chemical protocols for thioester peptides production are performed primarily by Boc-SPPS, using thioester linkers. This kind of stepwise solid-phase peptide synthesis chemistry necessitates of trifluoroacetic acid (TFA) treatments for N^{α}-Boc deprotection and treatment with anhydrous fluorhydric acid (HF) for final side-chains deprotection and cleavage from solid support. Many laboratories do not have the facilities needed for performing HF cleavage, due both to the cost of the apparatus and the hazards inherent in its operation. However, the extremely strong acid conditions used in Boc-SPPS are not compatible with many peptide modifications introduced during synthesis, such as post-translational modifications like glycosylation. On the other hand, the preparation of peptide thioesters by the alternative SPPS approach, the Fmoc-chemistry, is not directly accessible due to the nucleophilicity of the secondary amine, the piperidine, required for Fmoc removal that is incompatible with the use of the thioester linkers. The development of methods for the production of thioester peptides by Fmoc-SPPS has therefore been a major challenge over the past decade. Several approaches have been developed to address this problem, for example the use of an optimized Fmoc deprotection cocktail containing 25% 1-methylpyrrolidine, 2% hexamethyleneimine, 2% hydroxybenzotriazole (HOBt) (v/v/w) in 50:50 N-methylpyrrolidone (NMP) and dimethylsulfoxide (DMSO) [Li, X.Q. *et al.* 1998] or the use of 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) as non-nucleophilic base to remove Fmoc group [Clippingdale, C.J. *et al.* 2000]. A further possibility is the synthesis of the peptide on an acid-labile resin, like Cl-trityl-resin, and the cleavage of the fully-protected peptide chain. The thioester can be obtained by the treatment of the protected peptide with activating reagents and thiols. After, the peptide is deprotected with strong acid and purified [von Eggelkraut-Gottanka, R. *et al*. 2003]. The Pessi group proposed the safety-catch system [Ingenito R. *et al.*, 1999] which uses the thiol labile Kenner's sulfonamide linker for Fmoc-based synthesis of C-terminal peptide thioesters (Fig. 5A). A novel backbone amide linker (BAL) strategy has been developed by Barany and extended for Fmoc-SPPS of peptide thioesters [Alsina, J. *et al.* 1999]. This strategy comprises (i) start of peptide synthesis by anchoring the penultimate residue, with its carboxyl group orthogonally protected by a removable allyl ester, through the backbone nitrogen, (ii) continuation with standard protocols for peptide chain elongation in the *C* → *N* direction, (iii) selective orthogonal removal of the carboxyl allyl ester protecting group, (iv) solid-phase activation of the pendant carboxyl and coupling with the α-amino group of the desired thioester C-terminal residue, and (v), finally, cleavage/deprotection to release the free peptide product into solution (Fig. 5B). Another post-SPPS thioesterification reaction use C-terminal *N*-alkyl cysteine as N to S migratory device [Hojo, H. *et al.* 2007] (Fig. 5C). A useful approach for peptide thioesters production compatible with Fmoc-chemistry was presented by Gaertner and co-workers [Botti, P. *et al.* 2004]. In this strategy, Fmoc-Cys(tButhio)-OH is coupled to amino-Rink-PEGA resin and after Fmoc removal with piperidine, the N-terminal amino group is reacted with KNO_2 to generate a 3-tert-

butyldisulfanyl-2-hydroxypropionamide template as a racemic mixture. The first amino acid of the sequence is then attached to the template *via* HBTU-mediated coupling with DMAP as catalyst. The peptide synthesis is completed by standard Fmoc- chemistry and the peptide C^{α}carboxyester of the 3-tert-butyldisulfanyl-2-hydroxypropionamide is obtained by acid cleavage from resin (Fig. 5D 1). This species could be converted into a thioester peptide through *in situ* O to S acyl shift prior or during ligation (Fig. 5D 2). Recently, a novel method for thioester peptides preparation was reported based on the formation of a backbone pyroglutamyl imide, which, after displacement by a thiol, provides the peptide thioester [Tofteng, A.P. *et al.* 2009] (Fig. 5E). Notably, Blanco-Canosa and Dawson introduced a really ingenious novel strategy that promises to effectively enable the use of standard Fmoc-chemistry solid phase synthesis for the routine production of peptide thioesters [Blanco-Canosa, J.B. & and Dawson, P.E. 2008]. This alternative approach is based on the formation, following chain elongation, of a C-terminal N-acylurea functionality through the use of the more stable intermediates o-aminoanilides (Fig. 5F). Starting with Rink-PEG-PS resin, 3-Fmoc-4-diaminobenzoic acid (Fmoc-Dbz) is coupled and then deprotected with piperidine. Subsequent amino acids are coupled by standard Fmoc-SPPS. Following chain assembly completion, with the last amino acid introduced as N^{α}-Boc-protected, the C-terminus is activated with p-nitrophenylchloroformate followed by the addition of a base to promote intramolecular attack of the anilide to form the resin bound N-acyl-urea peptide. The conversion to acyl-urea is quantitative and the reaction leads to the formation of two isomers in a ratio of 4:1. However, subsequent treatment of the peptide with MPAA rapidly yields a single thioester product. The peptide is deprotected and cleaved from resin with TFA. Importantly, this mechanism leads to less than 2% of epimerization of the C-terminal amino acid.

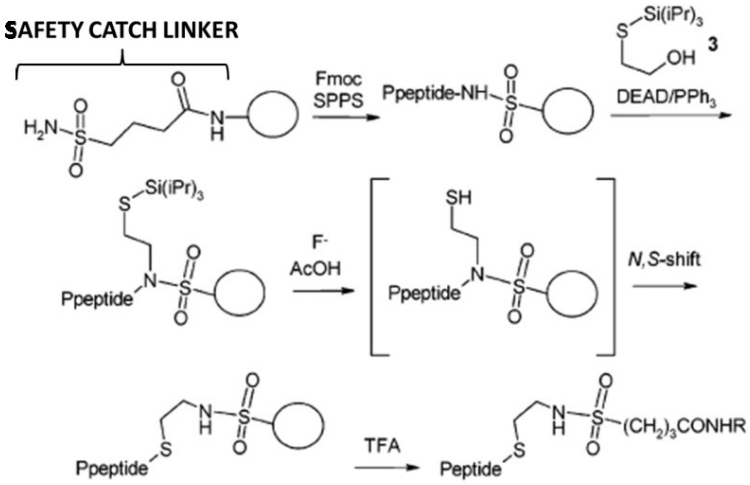

Figure 5 A) Synthesis of thioester peptides through the use of Kenner safety-catch linker and Fmoc SPPS.

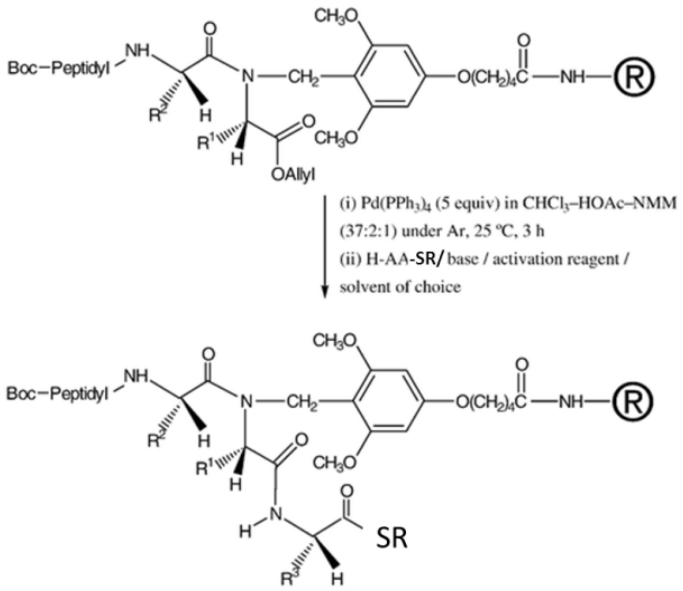

Figure 5 B) Preparation of unprotected peptide thioesters using Backbone Amide Linker (BAL) approach for Fmoc SPPS [Alsina, J. *et al.* 1999].

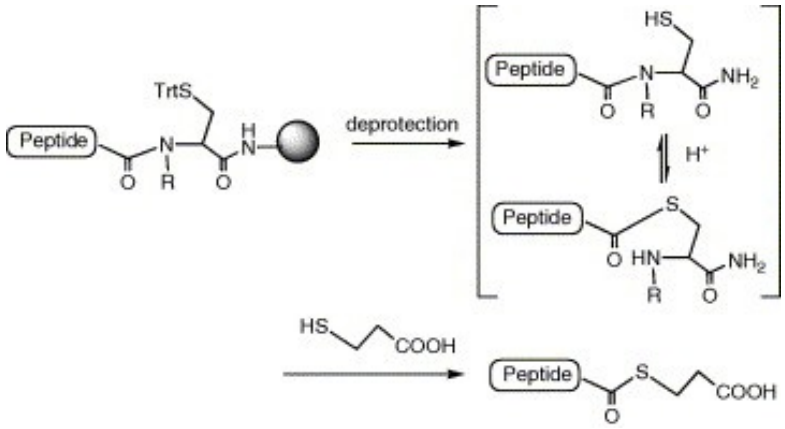

Figure 5 C) Post Fmoc-SPPS thioesterification using *N*-alkyl cysteine as *N*- to *S*-acyl transfer device [Hojo, H. *et al.* 2007].

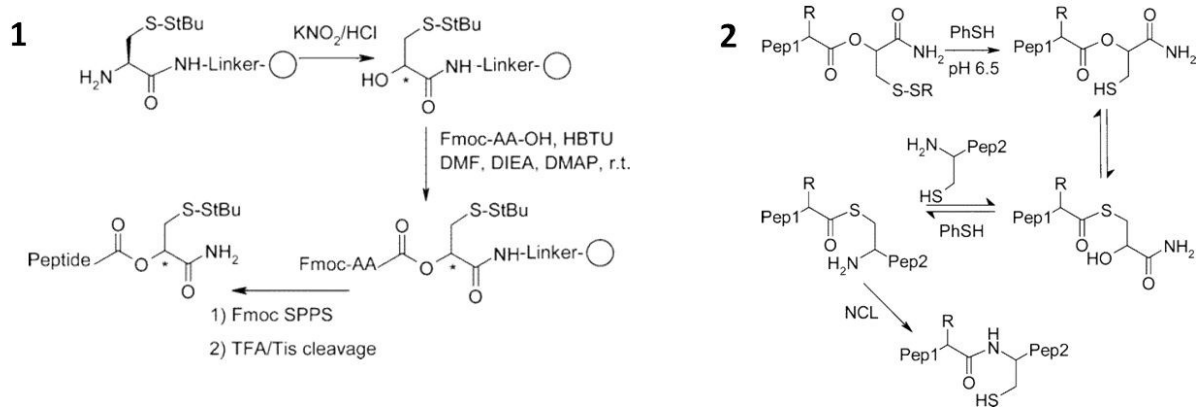

Figure 5 D, 1) Preparation of peptide-C$^{\alpha}$oxy-(2-tert-butyldisulfanyl-1-1carboxyamide) ethyl ester starting from Rink Amide PEGA Resin; **2)** Native Chemical Ligation through *in situ* O to S acyl shift [Botti, P. *et al.* 2004].

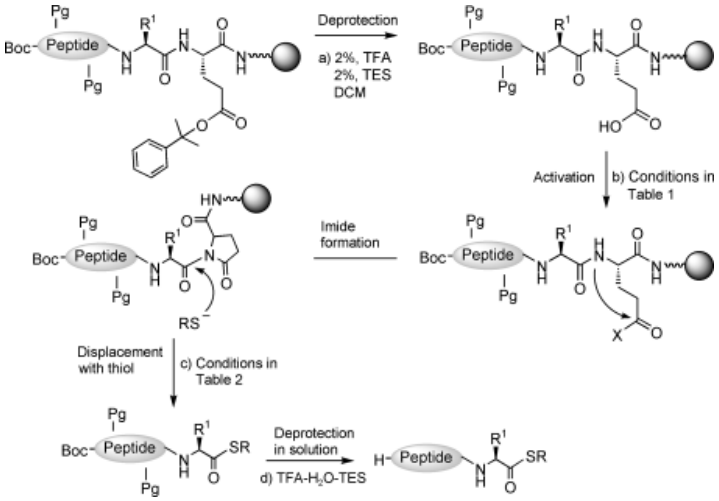

Figure 5 E) Synthesis of peptide thioesters through a backbone amide activation [Tofteng, A.P. *et al.* 2009].

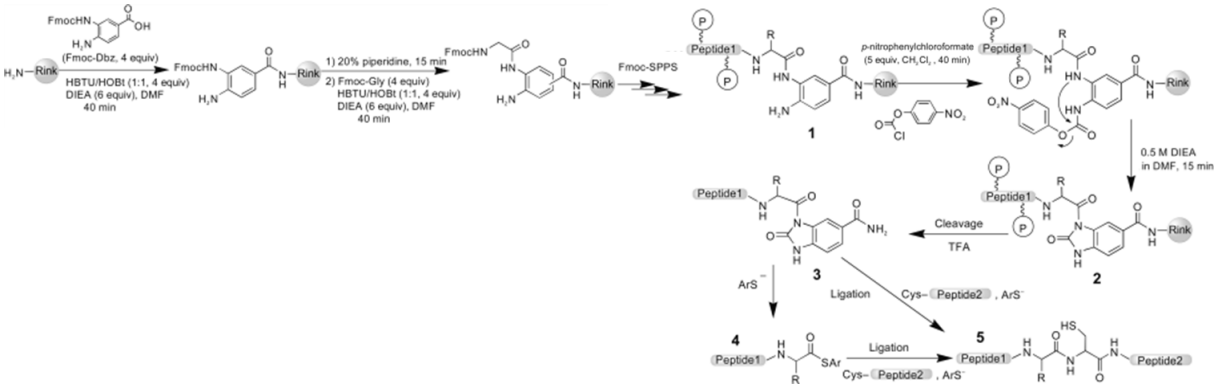

Figure 5 F) Synthetic strategy for Fmoc-SPPS of thioester peptides through a C-terminal N-acylurea functionality [Blanco-Canosa, J.B. & and Dawson, P.E. 2008].

The question of how to prepare recombinant polypeptides carrying a C-terminal thioester functionality needed for EPL reactions finds a solution into a natural occurring process known as protein splicing. Protein splicing is a post-translational process in which a precursor protein undergoes self-catalyzed intramolecular rearrangements that result in the precise excision of an internal segment, referred to as an intein, and ligation of the two flanking portions, termed exteins [Muir, T.W. 2003; Perler, F.B. 2005]. Thus, inteins can be described as analogues of self-splicing RNA introns. Inteins occur in organism from all three kingdoms of life as well as in viral proteins and are predominantly found in enzymes involved in DNA replication and repair [Perler, F.B. 2000]. Inteins are promiscuous with respect to the sequence of the two flanking exteins, so there are no specific amino acids required in either of the two exteins. In contrast, inteins are characterized by several conserved key residue required for the basic chemical steps in splicing mechanism. As shown in Fig. 6, the first step in the protein splicing involves an N → S (or N → O) acyl shift in which the N-extein unit is transferred to the side chain -SH or -OH of a Cys/Ser residue, located at the immediate N-terminus of the intein. This rearrangement is thermodynamically unfavorable; however, the intein structure catalyzes this step by twisting the scissile amide-bond into a higher energy conformation, thereby helping to push the equilibrium to the (thio)ester side [Romanelli, A. *et al.* 2004]. In the next

step, the entire N-extein unit is transferred to a second Cys/Ser/Thr residue at the intein-C-extein boundary (or +1 position) through a trans(thio)esterification reaction. The resulting branched intermediate is then resolved through cyclization reaction involving a conserved asparagine residue at the C-terminus of the intein and assisted by the penultimate His residue. The intein is thus excised as a C-terminal succinimide derivative. In the final and concerted step, an amide bond is formed between the two exteins as a result of an S → N (or O → N) acyl shift [Muir, T.W. 2003].

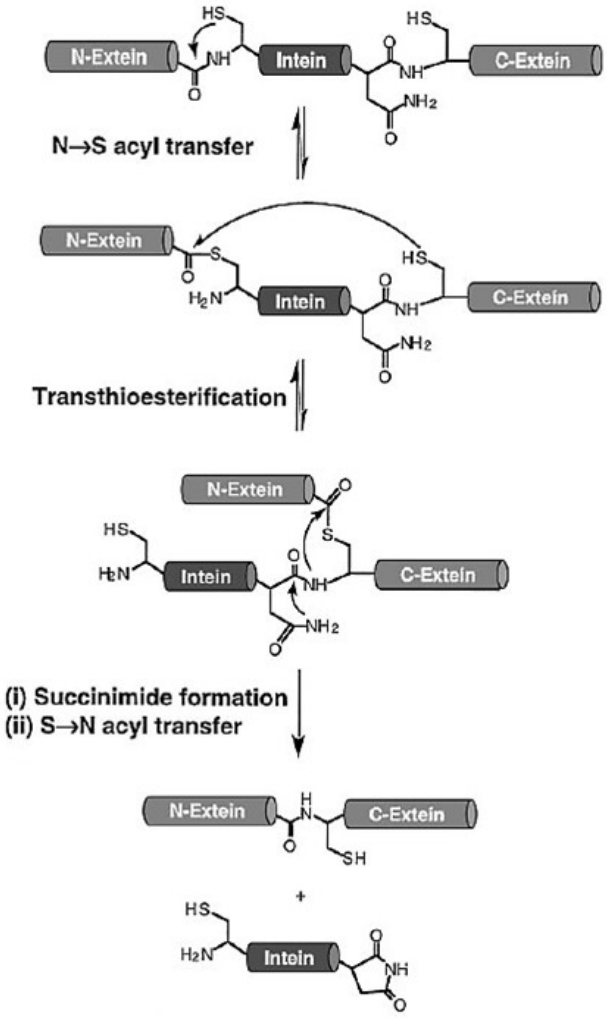

Figure 6 Mechanism of protein splicing. The process involves a series of acyl rearrangements and internal reactions catalyzed by the central intein protein domain. This process results in linkage of the two flanking polypeptides, the N- and C- exteins, *via* a normal peptide bond [Muir, T. W. 2003].

Although the biological role of protein splicing remains an open question, the process has been extensively exploited in the area of biotechnology and biochemistry for its useful application in protein semisynthesis by EPL. In fact, a number of mutant mini-inteins containing a C-terminal Asn → Ala mutations have been designed that can promote only the first step of protein splicing [Xu, M.Q. *et al.* 1996]. Protein expressed as in frame N-terminal fusion partner with one of these engineered inteins can be cleaved by thiols in an intramolecular transthioesterification reaction. Importantly, C-terminal thioester proteins can be prepared by this route and used as building blocks in EPL (Fig. 7).

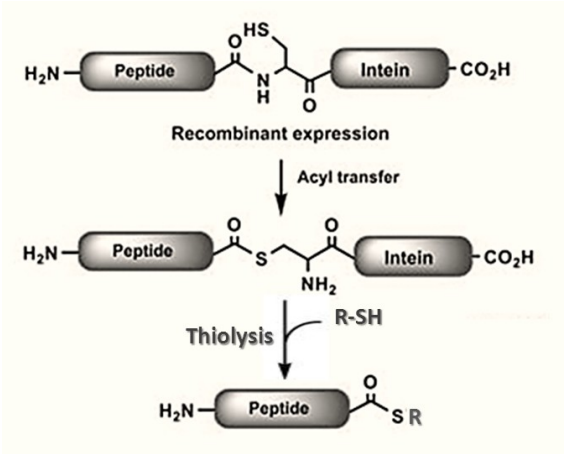

Figure 7 Recombinant C-terminal thioester proteins production. Mutation of the C-terminal Asn residue within the intein to an Ala residue blocks the final step in protein splicing. Proteins expressed as in frame N-terminal fusions to much mutant inteins can be cleaved by thiols to give the corresponding thioester protein derivatives [Muir, T. W. 2003].

E. coli expression vectors are commercially available that allow the production of fusion constructs with engineered inteins, such as *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 [Telenti, A. *et al.*, 1997; Southworth, M.W. *et al.* 1999]. *Mxe* GyrA intein is the most widely used for this kind of application due to its modest size (198 amino acids), high expression level in bacterial hosts and efficiency to refold from inclusion bodies, its ability to cleave efficiently with a variety of thiols to give the corresponding thioester protein even in presence of moderate concentration of denaturants, detergents or organic solvents [Valiyaveetil, F.I. *et al.* 2002]. For the thiolysis of an intein fusion protein, the choice of a certain thiol depends on the accessibility of the catalytic pocket of the extein/intein splicing domain and on the properties of the target protein. In general, the thiol should be enough small to enter the catalytic pocket and attack the thioester bond connecting the extein and the intein. Besides, the newly-formed thioester should be at the same time stable to hydrolysis, in order to be isolated, and reactive in EPL. As yet stated, thioalkyl-esters are quite stable but not enough reactive. Contrariwise, thioaryl-esters are much more reactive but instable. Commonly, a good compromise is reached using the thiol MESNA, which also has the advantage of being a non-malodourous thiol, facilitating thioester proteins handling and purification. Similarly as in NCL, an important factor to consider in choosing the protein fragment to produce as thioester for EPL reaction is the identity of the amino acid at the ligation site, avoiding -Ile-Cys-, -Thr-Cys-, -Val-Cys-, -Glu-Cys-, -Asp-Cys- and -Pro-Cys- junctions. A further related issue here is the effect of the -1 residue on the efficiency of the protein-intein thiolysis step. Each intein has its set of preferences at this position: for example, certain residues are associated with increased levels of premature cleavage *in vivo* (e.g. Asp), while Pro is associated with no cleavage at all [Muir, T.W. *et al.* 2003; Southworth, M.W. *et al.* 1999].

Thanks to their self-removing nature, the use of inteins tagged with an affinity appendage as fusion partner also suggests a protease-free purification system that can be exploited for isolating recombinant proteins without need to remove any sort of tag by the use of specific protease. The application of the *Mxe* GyrA intein for thioester proteins production using a protease-free purification strategy will be widely

mentioned in the Experimental Section of this PhD thesis due to its involvement into the semisynthetic protocol conceived and applied to the production of doubly functionalized proteins.

Production of N-terminal Cys peptide

A critic to NCL and EPL is that they require a Cys residue at the ligation site. It is often stated that this is a severe limitation of these approaches, because Cys is not so commonly found in proteins. This criticism is misleading; while Cys may be the least common amino acid found in proteins, it is equally true that there are many thousands of Cys-rich disulfide containing proteins in nature, and that the most common structural motif in the human genome is the Cys-rich zinc finger domain. All of these Cys containing proteins are accessible targets for chemical synthesis by NCL or semisynthesis by EPL. For the other proteins, lots of chemistries have been developed in order to extend these applications to non-Cys sites [Kent, S.B.H. 2009]. The most useful of these was the use of catalytic desulfurization of Cys in ligation product, enabling the chemical ligation to be performed also at X-Ala sites. Global desulfurization of the Cys residue in the final polypeptide product can be achieved by reduction with Raney Nickel method [Pentelute, B.L. & Kent, S.B. 2007]. Alternatively, a selective desulfurization can be achieved by the use of the water-soluble free radical initiator VA-044 (2,2'-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride), TCEP (Tris(2-carboxyethyl) phosphine) and t-BuSH (*tert*-butyl mercaptan) [Wan, Q. & Danishefsky, S.J. 2007]. Importantly, this latter method is compatible with the presence of Met residues, a wide variety of functional groups, such as post-translational modifications and biotin, and protecting groups as acetamidomethyl (Acm) and thiazolidine (Thz). The desulfurization approach has been generalized also to other β- and γ-thiol amino acids analogs, expanding the number of amino acids that could be obtained following desulfurization also to Val and Thr [Chen, J. *et al.* 2008; Chen, J. *et al.* 2010]. Also ligation at X-Met site is possible exploiting the latent thiol moiety of Met by the use of homocysteine; after ligation of the N-terminal homocysteine peptide with the C-terminal thioester moiety, S-methylation with excess p-nitrobenezensulfonate yields the native Met at the ligation site [Tam, J.P. & Yu, Q. 1998]. Indeed, recently the X-Phe ligation site became accessible for native chemical ligation by introducing β-mercaptophenylalanine at the ligation site [Crich, D. & Banerjee, A. 2007]. A limit of this strategy is that desulfurization acts uniformly throughout all the protein, also reducing other Cys residues eventually present into the protein. In this regard, Dawson group very recently proposed the use of selenocysteine mediated chemical ligation in combination with a protocol for the conversion of this amino acid to alanine which preserve native Cys residues present in the sequence without the need for side chain protection [Metanis, N. *et al.* 2010]. A completely different kind of approach to ligation at non-Cys sites rely on the use of thiol-containing N(α) auxiliaries that mimic the presence of a Cys residue at the N-terminal of the peptide and which can be selectively removed from the newly-formed amide bond after the ligation reaction. The use of thiol containing auxiliaries is highly sensitive to the nature of the amino acid at the ligation site and is currently much less effective than chemical ligation at Cys. An elegant example proposed was the photolabile auxiliary based on the o-nitrobenzyl group [Marinzi, C. *et al.* 2004]. The use of auxiliaries as well as the use of amino acids analogs is applicable only to chemically synthesized Cys-peptide, due to the impossibility of introducing such modifications during recombinant expression. NCL and EPL have also been used for the chemical synthesis or semisynthesis of a wide range of fully functional proteins by

the simple expedient of introducing a Cys residue as needed at the ligation site. Cys scanning mutagenesis experiments in many proteins have shown that Cys residue can be introduced in the majority of positions of a protein molecule without affecting folding or function [Kent, S.B.H. 2009].
While the chemical production of N-terminal Cys peptide is achievable simply by ending SPPS with Cys, methods for generating recombinant proteins possessing amino-terminal Cys residue all rely on the cleavage of an appropriate precursor protein (Fig. 8). In the simplest scenario, a Cys is introduced next to the initiating methionine in the protein sequence. Endogenous methionyl aminopeptidases (MetAPs) remove the initial Met, thereby generating amino-terminal Cys protein. However, this *in vivo* processing event is often incomplete. Most approaches involve the *in vitro* use of exogenous proteases. In these approaches, a factor Xa recognition sequence is appended immediately in front of the cryptic N-terminal Cys in the polypeptide of interest [Erlanson, D.A. *et al.* 1996]. Besides, Tolbert & Wong showed that the protease from tobacco etch virus (TEV) can also be used to release amino-terminal Cys proteins from suitable precursors demonstrating that TEV protease could accept Cys in the P1' of the recognition site rather than the usual Ser or Gly [Tolbert, T.J. & Wong, C.H. 2002]. Moreover, a N-terminal Cys protein could be obtained from fusion construct with Small Ubiquitin-like Modifier (SUMO) by digestion of the chimeric protein with a specific protease that recognizes and selectively hydrolyzes the fusion partner SUMO, releasing the fused protein without any additional amino acid, allowing to unmask an N-terminal Cys-residue [Weeks, S.D. *et al*. 2007]. Protein splicing can also be exploited to prepare amino-terminal Cys proteins. Inteins have been mutated so that cleavage at the C-terminal splice junction, between the intein and the C-extein, can be induced in a pH and temperature-dependent fashion. Splicing of such an intein allows the release of an amino-terminal Cys protein without any further treatment. The *Mxe* GyrA intein was also adapted to this other kind of application [Southworth, M.W. *et al.* 1999], as well as the *Synechocystis* sp. PCC6803 DnaB helicase (*Ssp* DnaB) intein [Mathys, S. *et al.* 1999]. A drawback of this intein-based approach is that spontaneous cleavage can occur, resulting in premature loss of intein and, more importantly, the loss of the associated affinity tag.

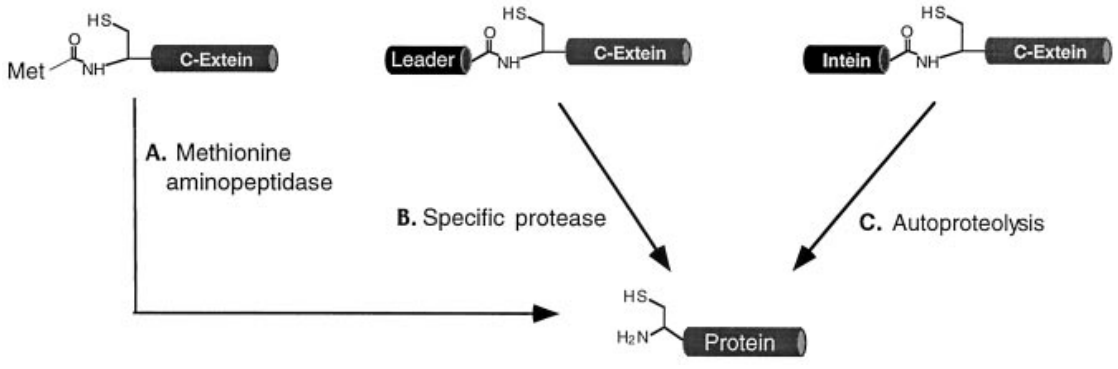

Figure 8 Recombinant N-terminal Cys proteins production. A Cys-protein can be produced by specific proteolysis of recombinant protein containing a cryptic Cys residue in the C-terminal position of a protease cleavage site. In one strategy, the leader can be simply a Met, which is removed *in vivo* by specific amino-peptidase (**A**), or it can be a short peptide recognition sequence removed *in vitro* by cleavage with heterologous protease (**B**). In alternative strategy, the leader sequence consists of a mutated intein that undergoes specific cleavage at the intein C-extein junction in response to changes in pH or temperature (**C**) [Muir, T.W. 2003].

Ligation of multiple peptide fragments

When three or more peptide fragments are to be ligated sequentially for the synthesis or semisynthesis of larger protein, the purification of each intermediate ligation product by HPLC often leads to large product losses, resulting in a low overall yield. To solve the problem, Kent and co-workers invented the kinetically controlled ligation (KCL), that relies on the large reactivity difference between a peptide-α-thioarylester and a peptide-α-thioalkylester [Bang, D. *et al.* 2006]. They first applied the KCL approach to the synthesis of crambrin (46 amino acid). The major challenge of the work was to control the intrinsic dual reactivity of a bifunctional Cys-peptide-thioester, impeding its self-ligation. This was achieved by protecting the N-terminal Cys residue of thioester peptide with a Thz group. Once the first ligation reaction reached the completion, the N-terminal thiazolidine was converted to cysteine with methoxylamine at pH 4. After that, the pH was readjusted around neutrality and the next thioester peptide was introduced for the subsequent ligation reaction. The thiazolidine is not the only one way to mask a Cys. An alternative enzymatically approach involves the use of proteases, such as Factor Xa or TEV, to give a new N-terminal reactive Cys. The proteolytic cleavage of an intermediate polypeptide containing a Cys-residue covered by a proteolysis site allows the appearance of a new N-terminal Cys that can be involved in a new ligation step.

Protein semisynthesis by *trans*-splicing

Several inteins can be split into two pieces that individually have no activity but, when mixed together, they reassociate *non*-covalently and fold into an active self-splicing unit. This process is called Protein *Trans*-Splicing (PTS) and proceeds in the same way as regular protein splicing [Southworth, M.W. *et al.* 1998]. The two fragments of the intein can reconstitute even if each portion is fused to extein (Fig. 9).

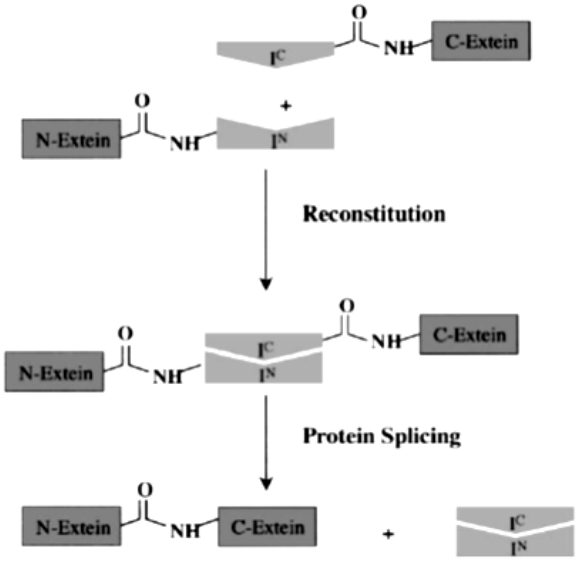

Figure 9 Protein *trans*-splicing by split inteins. Inteins can be cut into N- and C-terminal pieces, that individually have no activity but that when combined, associate non-covalently to give a functional splicing element [Muir, T.W. 2003].

Thus, the result of a PTS reaction is similar to a NCL or EPL: two polypeptide chains linked through a native peptide bond. Protein production through PTS bears several potential advantages over NCL or EPL. For example, no synthesis and isolation of a thioester or treatments with protease are required and the reaction does not need thiol excess. Furthermore, the reaction proceeds efficiently also at lower concentrations of the two split intein fragments. However, solubility of fusion proteins with many split intein fragments could be observed due to the exposure of hydrophobic patches leading to misfolding and aggregation and requiring renaturing procedure [Mootz, H.D. 2009]. Split inteins are found in nature (i.e. *Synechocystis* DnaE intein) and can also be engineered from regular (or cis) inteins. Natural or artificially split inteins, such as *Mxe* GyrA and *Ssp* DnaB inteins, have been examined in laboratory to identify different possible sites of split. Interestingly, investigations on split sites in the *Ssp* DnaB intein showed that an Int^N fragment of only 11 amino acids and its corresponding 143 amino acids Int^C counterpart were capable of protein *trans*-splicing both *in vivo* and *in vitro* [Sun, W. *et al.* 2004; Ludwig, C. *et al.* 2006]. Similarly, Iwai and co-workers demonstrated that the Int^C fragment from *Nostoc punctiforme* DnaE intein (*Npu* DnaE intein) could be shortened to 15 residues without abolishing protein *trans*-splicing activity [Aranko, A.S. *et al.* 2009]. In a screen of other inteins, Liu and co-workers identified a split *Ssp* GyrB intein with an Int^C fragment of only 6 amino acids able to reconstitute with the 150 amino acids remaining N-terminal region [Appleby, J.H. *et al.* 2009]. The ability to obtain inteins with N-terminal or C-terminal fragments so short allows the preparation of this segments and their exteins by SPPS, that appears as an attractive and suitable way for the incorporation into proteins of chemical modified segments.

Applications of NCL and EPL

NCL and EPL have found widespread application in biochemical, biotechnological and biophysical areas, allowing the study and the characterization of many proteins and leading to a great number of discoveries. The following paragraphs provide an overview of the major applications of NCL and EPL in protein engineering, with a particular attention to the use of these synthetic techniques for the production of labeled proteins useful for biophysical studies, which is the aim of this PhD thesis.

Introduction of fluorescent probes

Because of their extremely high sensitivity, fluorescence spectroscopy approaches, such as Förster resonance energy transfer (FRET), fluorescence correlation spectroscopy (FCS) and fluorescence polarization (FP), are powerful biophysical methods for studying protein folding, function, conformational fluctuations and protein-protein interactions. All biophysical techniques make use of molecular probes that can either be naturally present within the protein or introduced through labeling procedures. Nature provides us with two fluorescent amino acids, tyrosine and tryptophan, but their utility as fluorescent probes decreases as the size of the protein increases. Moreover, most large proteins contain several Trp and Tyr residues, thereby reducing the resolution of the analysis. For many applications it is desirable to employ fluorescent probes placed in specific locations inside protein and possessing spectroscopic properties quite different from those of the two intrinsic fluorophores. Conventional methods for protein derivatization with fluorophores use the reactivity of native or engineered Cys or Lys side chains. In fact, a large number

of useful fluorescent probes are available as maleimide *or N*-hydroxy-succinimide derivates that react selectively with Cys thiol or Lys ε-amino group. However, this approach is limited by the occurrence of multiple Cys or Lys in the same protein, resulting in heterogeneous and incomplete labeling. Chemical ligation approaches appear as the most appropriate and general methods for proteins site-specific labeling with fluorescent probes. In EPL and NCL approaches the traditional bioconjugation reactions can be performed on each recombinant or synthetic building block which are then ligated to obtain a site-specifically labeled protein. A large number of fluorescent proteins have been prepared by chemical ligation [for reviews see Muir, T.W. 2003; David, R. *et al.* 2004; Schwarzer, D. & Cole, P.A. 2005; Muralidharan, V. & Muir, T.W. 2006]. In the majority of the examples reported the probe is attached to the side-chain of an amino acid within the synthetic peptide building block, taking advantage of the several orthogonal protection strategies available that allow fluorophores to be introduced site-specifically into peptide during SPPS. The simplest application of protein fluorescent labeling involves the monitoring of the intrinsic fluorescence of a single probe during the biological process under investigation. Indeed, quantum yields and/or Strokes shifts of several fuorophores are known to be sensitive to changes in the dielectric constant of their immediate environment. Thus, the fluorescence of such a probe changes if it is located in a region that undergoes a structural change in the protein, for example following the interaction with a ligand or the activation through post-translational modifications. In one of the first applications, a sequential EPL procedure was used to incorporate the environmentally sensitive fluorophore 5-(dimethylamino)-naphtalene-sulfonamide (dansyl) between the Src homology 3 (SH3) and Src homology 2 (SH2) domains of the Abelson nonreceptor protein tyrosine kinase (c-Abl) [Cotton, G.J. *et al.* 1999]. The fluorescence properties of semisynthetic c-Abl were found to be sensitive to changes in domain orientation, allowing this construct to act as a biosensor for ligand interaction. In another example, Iakovenko and co-workers incorporated a dansyl probe into a semisynthetic version of Rab7, a small GTPase [Iakovenko, A. *et al.* 2000]. The fluorophore was incorporated close to a region at the C-terminus of the protein that is known to be post-translation prenylated by Rab geranylgeranyl transferase (RabGGTase). They used this fluorescent semisynthetic Rab7 protein to evaluate the affinity of Rab7 for RabGGTase and for the escort protein REP-1. In a similar approach, Muir group used *in vivo* amino acid replacement to incorporate the tryptophan analog 7-azatryptophan (7-AW) into SH3 domain of Crk protein. Ligation *via* EPL of the 7AW-labeled SH3 domain to the SH2 domain of the same protein generated a multidomain protein with a single domain specifically labeled. Studies of this labeled protein show that the biochemical and thermodynamic properties of the SH3 domain do not change within the context of the larger multidomain protein [Muralidharan, V. *et al.* 2004]. A different example of biophysical application of chemical ligation was presented by Bertozzi group, that synthesized a series of glycosylphosphatidylinositol (GPI)-protein analogues differing for the glycan component at their C-terminus. The GPI is a posttranslational modification occurring at the C-terminus of many eukaryotic proteins that anchors the modified proteins in the outer leaflet of the plasma membrane. A combination of EPL and NCL was used to attach these (GPI)-protein analogues to the green fluorescent protein (GFP), used as fluorescence probe. These modified GFPs were incorporated in lipid bilayers, and their mobility was analyzed using FCS. The data from these experiments suggest that the GPI anchor is more than a simple membrane-anchoring device; it also may

prevent transient interactions between the attached protein and the underlying lipid bilayer, thereby permitting rapid diffusion in the bilayer [Paulick, M.G. *et al.* 2007].
A useful and widely explored application of fluorescently labeled proteins refers to their use in FRET studies. FRET is a physical phenomenon that occurs in systems having two fluorophores, one serving as donor and the other as acceptor. In a FRET pair the emission spectra of the donor should overlap the excitation wavelength of the acceptor so that, upon excitation of the donor, energy can be transferred from the donor to the acceptor, resulting in emission at the acceptor's emission wavelength. The energy transfer occurs only if donor and acceptor are spatially displaced within a discrete distance. Furthermore, the efficiency of energy transfer (E) is function of the distance (R) between donor and acceptor according to the equation $E = [1 + (R/R_0)^6]^{-1}$, where R_0 is the Förster Radius, a constant depending on the kind of fluorophores employed. FRET is a high sensitive method that can be used to monitor dynamic phenomena that bring the two fluorophores spatially closer or farther, such as folding/unfolding processes and protein-protein interactions. Deniz *et al.* reported a NCL based methodology for the chemical synthesis of the chymotrypsin inhibitor 2 and its functionalization with a couple of fluorophores. Protein was split into two fragments, both synthesized by SPPS using Boc-chemistry. N-terminal fragment was labeled on its amino group with tetramethylrhodamine succinimide ester before cleavage from resin. The peptide was then cleaved and produced as thioester. Labeled thioester peptide was reacted with the second fragment containing an N-terminal Cys. A thiopyridil-activated disulfide was used to label the Cys involved in NCL. Doubly labeled protein was used in single-molecule FRET assays [Deniz, A.A. *et al.* 2000]. Cotton and Muir described a solid-phase expressed protein ligation (SPPL) strategy that enables large semisynthetic proteins to be assembled on a solid support by the controlled sequential ligation of a series of recombinant and synthetic polypeptide building blocks. This modular approach was used to label the amino and carboxyl termini of the Crk-II adapter protein with the FRET pair fluorescein and tetramethylrhodamine. Phosphorylation by the nonreceptor protein tyrosine kinase c-Abl of the dual labeled Crk-II construct resulted in a fluorescence change of the FRET pair [Cotton, G.J. & Muir, T.W. 2000]. In subsequent studies, they designed and synthesized an extremely sensitive doubly labeled truncated version of Crk-II that enabled real-time monitoring of c-Abl kinase activity and provided a rapid approach to screening potential c-Abl kinase inhibitors [Hofmann, R.M. *et al.* 2001]. At my knowledge, this is the only one example of protocol based on EPL described for the double labeling of a protein with fluorescence probes. The Ebright group used EPL-labeling and cysteine-labeling to introduce the fluorophores of a FRET pair into different subunits of the *E. coli* RNA polymerase holoenzyme. The researchers studied the mechanism of transcription initiation and RNA polymerase translocation by using doubly labeled complex containing promoter DNA, σ^{70} factor and RNA polymerase core enzyme, this latter prepared and labeled by EPL [Mukhopadhyay, J. *et al.* 2001]. The group carried out numerous FRET experiments on this complex and elaborated a structural model for the RNA polymerase holoenzyme [Mekler, V. *et al.* 2002]. Xie *et al.* described an approach for the identification and characterization of histone acetyltransferases (HATs) inhibitors using FRET and fluorescence polarization. EPL was used to label the HATs PCAF and p300 with Dabcyl (Dab) as FRET acceptor. Methoxycoumarin (Mca) was conjugated to HAT substrate analogues through SPPS to function as fluorescent donor. When a ligand-protein interaction occurs, the fluorescent intensity of the donor fluorophore Mca decreases due to FRET quenching by the acceptor

Dab. Meanwhile, the formation of ligand-protein complexes causes reduction of the molecular mobility of the donor fluorophore, resulting in increased fluorescence anisotropy. Thus, dual modes of fluorescence measurement, FRET and anisotropy, are integrated in the same assay system [Xie, N. *et al.* 2009]. A FRET based approach to observe protein oligomerization was described by Scheibner *et al.* [Scheibner, K.A. *et al.* 2003]. A synthetic dipeptide Cys-Lys(ε-fluorescein) (donor) or Cys-Lys(ε-rhodamine) (acceptor) was attached via EPL to the C-terminus of a recombinant protein and a mixture of the two singly-labeled proteins analogs with the donor or the acceptor fluorophore was analyzed for FRET potential. Three proteins were analyzed by this mean, the known protein dimer glutathione S-transferase, that showed evidence of FRET, the known protein monomer SH2 domain phosphatase-1, that did not display FRET and finally the previously uncharacterized serotonin *N*-acetyltransferase, that displayed significant FRET, indicating its likely propensity for dimerization or more complex oligomerization. These results establish the potential of the union of EPL and FRET in the analysis of protein-protein interactions.
Also protein *trans*-splicing have been widely used for site-specific protein labeling with fluorophores, as presented in estimable works of Mootz group [Kurpiers, T.& Mootz, H.D. 2008 and 2007; Ludwig, C. *et al.* 2006] in which they show strategies to append a labeled tag to N-terminal or C-terminal of a protein of interest using split inteins. Protein *trans*-splicing was also used for the introduction of a FRET pair into a single protein [Yang, J.Y. & Yang, W.Y. 2009]. The general applicability of this approach is however limited by the instability of the split protein fragments which can aggregate in inclusion bodies.

Introduction of isotopic probes

EPL, PTS and NCL are of great utility for proteins characterization by nuclear magnetic resonance (NMR) spectroscopy because they have allowed to extend the use of such biophysical technique also to large, multidomain or repeat proteins. In fact, a major drawback of NMR applied to macromolecules are its size limitation, caused by a progressive loss of resolution as the number of amino acids increases in a protein, and the difficulties in analyzing repeat proteins in which most of the chemical shifts are overlapped. The production of proteins by ligation methods resolves this issue by allowing selected portions of a protein to be specifically labeled with isotopes for NMR studies. The great advantage of this synthetic approach, named segmental isotopic labeling, refers to the ability to specifically label with isotopes only a single segment, domain or repeat of a protein, and analyze it by NMR spectroscopy in its native multidomain context. In a general segmental isotopic labeling strategy, a uniformly labeled protein fragment, usually of recombinant origin, is ligated to an unlabeled one. As an example, in pioneering studies Muir group used EPL to introduce an ^{15}N-labeled domain within the Src-homology domain derived from c-Abl protein tyrosine kinase [Xu, R. *et al*. 1999]. In this EPL based strategy, a part of the protein of interest was bacterially expressed in a growth medium enriched with the ^{15}N isotope. Subsequent ligation of this labeled protein portion with the unlabeled remaining part yielded the desired segmentally labeled full-product. Recently, Gierasch group used EPL to create an Hsp70 molecular chaperone where only one of the two constituent domains was labeled with NMR active isotopes, allowing visualization of the single domain in the context of the two domains full protein. The NMR spectrum of the segmentally labeled protein was considerably simplified, enabling unequivocal signals assignment and enhanced analysis of dynamics, as a prelude to exploring the energy landscape for allostery in the Hsp70

family [Clerico, E.M. *et al*. 2010]. An elegant approach was proposed by Zhao *et al.*, that reported an efficient on-column EPL strategy for the production of human apolipoprotein E (apoE) triply labeled with ^{2}H, ^{15}N and ^{13}C. They prepared by recombinant route a deuterated N-terminal domain of the protein as C-terminal thioester by fusion to the *Mxe* GyrA intein; this segment was then ligated to $^{15}N/^{13}C$ isotopically labeled remaining portion of the apoE, similarly produced in bacterial host, and containing an appositively introduced N-terminal Cys [Zhao, W. *et al.* 2008]. EPL, PTS and NCL are useful approaches even for the site-specific introduction of one or few stable isotopes within a protein, allowing straightforward assignment of resonance from the labeled amino acids. A comprehensive example, was reported by Romanelli *et al.* Using EPL, the scissile peptide bond at the N-extein-intein junction of *Mxe* GyrA intein was uniquely labeled with ^{13}C and ^{15}N nuclei. This allowed the amide $^{1}J_{NC'}$ coupling constant to be measured on the active intein. Based on this analysis, this amide appears to be unusually polarized, perhaps because of nonplanarity. These studies supported the "ground-state destabilization" reaction mechanism for the first cleavage step in protein splicing [Romanelli, A. *et al.* 2004]. In another example, a series of ^{13}C-labeled amino acids were incorporated into the C-terminus of the α-subunit of a heterotrimeric G protein [Anderson, L.L. *et al.* 2005]. 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. Segmental labeling was also carried-out by trans-splicing, as reported by Iwai group which described a strategy to label an internal region of a protein of interest using split inteins [Busche, A.E. *et al.* 2009].

Introduction of post-translational modifications: phosphorylation, glycosylation, lipidation, ubiquitylation

The functional state of many proteins is regulated through post-translational modifications (PTMs). A full molecular understanding of how a modification affects protein function and structure can be achieved only through studies on the homogeneously modified protein. Recombinant expression of post-translational modified full-proteins is not trivial and often results in heterogeneous preparations. Similarly, methods for preparing proteins modified with PTMs by enzymatic treatments not always proceed with high yields, and the position of the modification cannot be controlled. EPL and NCL provide a solution to the problem allowing the preparation of appropriately modified peptides that can be ligated to give the full functional protein that can be biophysically and biochemically characterized. One of the first applications refers to the incorporation of phosphorylated residues into a protein. The availability of semisynthetic phosphorylated proteins allowed their high-resolution structure to be determined by biophysical technique such as crystallography. Phosphorylation is a critical modification for the transmission of information through cellular signaling pathway. The utility of synthetic approaches for shedding light on the role of phosphorylation is well illustrated by the work of Muir group on TGF-*β* signaling [Muir, T.W. 2008]. A tetra-phosphorylated derivative of the cytoplasmic domain of type 1 transforming growth factor *β* receptor (T*β*R-I) was prepared by ligation of a tetra-phosphopeptide thioester to a recombinantly expressed fragment [Huse, M. *et al.* 2000]. Access to homogeneous preparation of hyperphosphorylated T*β*R-I allowed the study of molecular receptor activation. In

related experiments, EPL was used to study the effect of phosphorylation on protein Smad2 transcription factor, the natural substrate of T*β*R-I. Phosphorylated Smad2 was prepared through the ligation of a synthetic bis-phosphopeptide or mono-phosphopeptide to a recombinant C-terminal thioester protein [Wu, J.W. *et al.* 2001; Ottesen, J.J. *et al.* 2004]. An intriguing focus of Muir group in TGF-*β* signaling has therefore been the development of chemistries to control the activity of a post-translational modified protein inside living system. They developed two strategies that allow the control of Smad2 activation with light. In one approach, EPL was used to chemically ligate a recombinant protein fragment of Smad2 to a synthetic segment bearing two phosphate moieties masked with 2-nitrophenylethyl protecting groups, which could be removed using UV light [Hahn, M.A. & Muir, T.W. 2004]. Biochemical and biological assays demonstrated that the resulting caged semisynthetic protein was activated upon irradiation with UV light. Subsequently, a bis-phoshorylated Smad2 was modified with a fluorophore and a photocleavable linker that acted as both a caging group and a fluorescence quenching group. UV irradiation led to the removal of the photolinker, with release of fluorescence, and activation of the protein [Hanh, M.E. *et al.* 2007]. This technology allows the direct imaging of a protein-activation event and its titration. A further target of protein phosphorylation by chemical synthesis is the tyrosine kinase Csk, which has been the first protein semisynthesized by EPL [Muir, T.W. *et al.* 1998]. A more recent example was reported by Becker group [Lahiri, S. 2010]. They described an approach to selectively analyze the impact of phosphorylation of the signal transducer and activator of transcription STAT6 on STAT dimerization, DNA binding and nuclear import by accessing six types of semisynthetic STAT6 proteins via EPL.

Understanding how glycosylation regulates protein function was also hampered by difficulties associated with the production of homogeneous samples. Indeed, the complex structure of oligosaccharide modifications make the production of chemically defined glycosylation a considerably more daunting challenge than the synthesis of phosphoproteins. Nonetheless, rapid progresses continue to be made on the chemical and enzymatic synthesis of large carbohydrates and glycopeptides [Grogan, M.J. *et al.* 2002] that can be successfully combined to EPL and NCL for the production of glycosylated proteins. Bertozzi group pioneered the field, being the first group to apply NCL and EPL to the preparation of O-linked glycoproteins. For example, they used EPL for the synthesis of three semisynthetic variants of the physiologically relevant glycoprotein GlyCAM-1 containing saccharides at predetermined sites [Macmillan, D. & Bertozzi, C.R. 2004]. Imperiali group have developed NCL based approach for the production of a N-linked chitobiose glycoprotein analogue of immunity protein Im7 with an Ala29Cys mutation. The semisynthetic approach relies on the solid-phase peptide synthesis (SPPS) of N-terminal thioesters, in glycosylated or unglycosylated form, in combination with the expression of the C-terminal fragment of Im7 [Hackenberger, C.P. *et al.* 2005]. Yamamoto *et al.* presented the synthesis of a single glycoform of monocyte chemotactic protein-3 (MCP-3), a CC-chemokine that consists of 76 amino acids and one N-glycosylation site. A three-segments native chemical ligation strategy was employed using unprotected peptides and glycopeptides [Yamamoto, N. *et al.* 2008].

Lipidation represents another important class of protein modification whose role in targeting proteins to cellular membranes is well documented. As with phosphorylation and glycosylation, the preparation of homogeneously lipidated protein is problematic due to problems of solubility and stability of the resulting conjugate in aqueous

solution. Once again, EPL and NCL may provide a practical solution. A noteworthy application of synthetic approaches to the assembly of lipid-modified proteins for biophysical studies refers to the co-crystal structure of a prenylated version of the GTPase Ypt1 in complex with Rab GDP dissociation inhibitor [Rak, A. *et al.* 2003].

Another prominent PTM is ubiquitylation. Ubiquitin (Ub) is a higly conserved 76-residue-long eukaryotic protein that plays essential role in cellular processes such as proteosomal degradation, endocytosis of membrane receptors, regulation of gene transcription. In the cell, Ub is enzymatically activated as C-terminal thioester and reacts with a Lys ε-NH_2 of the target protein, generating an isopeptide bond. Ubiquitylation is even more challenging to perform than other PTMs due to the size of the modification itself (8 kDa) [Muir, T.W. 2008]. EPL and NCL are useful to the resolution of the problem. A recent key strategy involved the use of a photocleavable ligation auxiliary, which is coupled to the side-chain of a target lysine residue in a synthetic peptide. The auxiliary functions as a N-terminal Cys surrogate and reacts in an EPL reaction with the recombinant Ub thioester. Following the ligation, the auxiliary is removed by photolysis [Chatterjee, C. *et al.* 2007]. This approach has been used for the production of ubiquitylated histone H2B. The target protein was assembled from three pieces, two recombinant C-terminal thioester and one synthetic peptide, and employed two orthogonal EPL reaction. The first ligation reaction involved the photolytic ligation auxiliary described above, whereas the second ligation was carried out standardly and involved a desulfurization step to convert the Cys residue at the ligation site back into the native Ala of H2B [McGinty, R.K. *et al.* 2008; Muir, T.W 2008]. Because histones are proteins heavily post-translationally modified with a wide variety of possible modifications regulating structure and function of chromatin in transcription, DNA replication, repair and condensation, they have been a common target of chemical synthesis and semisynthesis. In a pioneering study, Peterson group prepared by NCL an H3 variant modified with a phosphoserine at position 10 [Shogren-Knaak, M.A. *et al.* 2003]. By similar methods, a site-specifically Lys16-acetylated H4 derivative was prepared [Shogren-Knaak, M. *et al.* 2006].

Introduction of unnatural amino acids

NCL and EPL allow the introduction into a protein of an almost unlimited range of unnatural amino acids at any specific site. Multiple substitutions can also be incorporated into a protein molecule, enabling combination of number, type and position of noncoded amino acids. Noncoded amino acids are frequently found in native proteins *in vivo*. These arise from specific post-translational enzymatic modification of coded amino acid residues. One common modification of this type is γ-carboxy-glutamic acid (Gla), found for example in the eponymous Gla domains of plasma proteins. A polypeptide construct containing the first three domains of human plasma protein S, including a Gla rich module with 11 Gla residues, has been synthetized from three segments by NCL [Hackeng, T.M. *et al.* 2000]. Cole group used EPL to incorporate a tyrosine phosphonate, a nonhydrolyzable stable analog of phosphotyrosine, at the C-terminal of SHP-2 phosphatase, allowing to study the effect of tyrosine phosphorylation on structure and function of the protein [Lu, W. *et al.* 2001]. In other examples from the same group, a series of unnatural tyrosine analogues, such as fluorinated tyrosines, aminophenylalanine, methyltyrosines and ATP-linked phenylalanine were incorporated through EPL into the protein tyrosine kinase Src in place of the natural tail tyrosine residue [Wang, D. & Cole, P.A. 2001;

Shen, K. & Cole P.A., 2003]. In further experiment on Src kinase, Virdee *et al.* performed the semisynthesis of Src homology 2 (SH2) ligand binding domain from three fragments by NCL. Using this strategy, they replaced a key lysine residue with lysyl derivatives possessing progressively shorter aliphatic side chains. Biophysical characterization of these SH2 domain analogs allowed a systematic dissection of the side chain length contribution from a lysine residue to ligand binding and showed that the specificity of the SH2 domain of the Src kinase can be altered by incorporation of such lysyl derivatives [Virdee, S. *et al.* 2010]. Unnatural amino acids can also be introduced at the ligation site within semisynthetic proteins by changing the chemical apparatus used in NCL and EPL. In this regard, it has been demonstrated that N-terminal homocysteine and selenocysteine peptide can be used in ligation reactions [Tam, J.P. & Yu, Q. 1998; Hondal, R.J. *et al.* 2001], with useful applications respectively in peptide synthesis and protein structural studies by crystallography. The introduction of unnatural amino acids can also be used to modify the backbone structure of a protein. For example, Raines group reported the use of EPL to alter the ribonuclease A (RNase A) overall geometry, replacing an entire element of secondary structure, a *β*-turn, with the reverse-turn mimetic *R*-nipecotic acid-*S*-nipecotic acid [Arnold, U. *et al.* 2002]. The RNase analog showed the same catalytic activity of the wild-type protein and an increased thermal stability.
The study of the structure and function of ion channels has greatly benefited from protein synthesis by NCL and EPL. Synthetic protocols have been developed to generate folded channels and channel analogs containing unnatural amino acids. Notably, Muir and co-workers developed an EPL-based protocol to generate folded tetrameric potassium channel KcsA. This semisynthetic method was adopted to produce a KcsA analog with Gly77 replaced by D-Ala that resulted in channel functional loss, confirming the contribution of Gly77 in the selectivity filter [Valiyaveetil, F.I. *et al.* 2004]. Recently, Valiyaveetil described the semisynthesis by EPL of NaK, a bacterial *non*-selective cation channel. In the semisynthesis, the NaK polypeptide was assembled from a recombinant thioester peptide and a chemically synthesized peptide. Semisynthesis was also used to substitute Asp66 in the selectivity filter region of the NaK channel with the unnatural amino acids homoserine and cysteine sulfonic acid. Functional analysis of these mutants suggested that the presence of a negatively charged residue in the proximity of the ion binding sites is necessary for optimal flux of ions through the NaK channel [Linn, K.M. *et al.* 2010].

Production of insoluble or cytotoxic proteins

The recombinant expression of some proteins, like membrane proteins, have proven to be very hard due to the high propensity of the overexpressed protein to aggregate as inclusion bodies in the cytosol. As yet affirmed, integral membrane proteins can be successfully synthesized in large amounts by chemical ligation approaches [Valiyaveetil, F.I. *et al.* 2002; Linn, K.M. *et al.* 2010]. Another problem is that, in some cases, the expression of a desired protein in bacteria can cause cytotoxic side-effects because the target protein competes with cellular components of the host. By using EPL or NCL techniques this can be avoided through modular synthesis of the target protein. After ligation and refolding, the native conformation and biological functionality of a cytotoxic protein can be recovered [David, R. *et al.* 2004]. The potential cytotoxic RNase A was expressed by this method [Evans, T.C. Jr. *et al.* 1998]. One part of the protein was expressed as intein fusion construct and obtained as C-terminal thioester after thiolysis. The C-terminal truncated RNase A was then joined with a fragment synthesized by SPPS that contained a naturally occurring Cys

residue at N-terminus. Ligation of both enzymatic inactive fragments led to the full-length protein, which reconstituted its activity after several renaturation steps. Intein based *trans*-splicing system also seem to be adequate workhorse for applications in cytotoxic proteins production.

Backbone cyclization

Backbone cyclization improves the stability and activity of peptides and proteins and reduces their conformational flexibility. Many pharmaceutically important natural products, including several antibiotics and immunosuppressants, are based on cyclic peptides. Thus, the ability to synthesize backbone cyclic peptides has important implications for drug development. Head-to-tail cyclized protein can be prepared using an intramolecular version of NCL or EPL, through the formation of a new peptide bond between N- and C-termini. To this purpose, the precursor polypeptide have to contain both an N-terminal Cys and a C-terminal thioester. Such a polypeptide can be prepared by recombinant means, flanking the protein of interest with two inteins. The N-terminal modified intein (for example *Ssp* DnaB intein) can be cleaved by temperature or pH, releasing the α-Cys, whereas the C-terminal intein (for example *Mxe* GyrA intein) is cleaved by thiols, leading to the formation of the complementary reactive group α-thioester. This “two intein system” allows the preparation of N-terminal Cys and C-terminal thioester containing polypeptides that can react to give cyclic peptides or multimers. In another approach, circularization occurs after unmasking of the N-terminal Cys of a thioester polypeptide, for example by proteolysis. The resulting Cys attacks the thioester resulting in the backbone circularization through a ligation reaction. Several approaches use intramolecular *trans*-splicing for the generation of cyclic backbone *in vivo* and *in vitro*. In this case, the split intein is not coupled to a cleaved protein or two proteins which should knotted, but the two split intein parts flank the N-terminal and C-terminal region of the same polypeptide chain in a rearranged order (Int^C–target protein–Int^N). Intein reconstitution and self-cleavage result in the releasing of a cyclic protein. In a recent application of protein cyclization, a mutant of green fluorescent protein (GFP) was cyclized *in vitro* and *in vivo* by the use of a naturally split intein *Ssp* DnaE. The circular mutant GFP was more thermostable than the linear form, significantly more resistant to proteolysis of exopeptidase and also displayed increased relative fluorescence intensity [Zhao, Z. *et al.* 2010].

Tetratrico Peptide Repeat (TPR) proteins

The Tetratrico Peptide Repeat (TPR) is a structural motif widely diffused in proteins from all kingdoms of life that mediates protein–protein interactions in a multitude of biological processes, such as cell cycle regulation, transcriptional control, mitochondrial and peroxisomal protein transport, neurogenesis, protein folding and cancer. As denoted by its name, the basic TPR motif consists of 34 amino acids which assemble into a characteristic helix-turn-helix structure. TPR domains contain multiple copies of this α-helical motif arrayed in tandem to form an elongated structure in which each motif stacks on the top of the next one, thereby forming a regular, right-handed superhelical arrangement that can be imagined as a spiral staircase [D'Andrea, L.D. & Regan, L. 2003]. In theory, there should be no limit to the number of repeats possible within a TPR domain. However, modules with three tandem TPR repeats are by far the most common in nature, and presumably represent the minimal functional binding unit [Main, E.R.G. *et al.* 2003]. As a consequence of their elongated and repetitive architecture, TPR proteins, as well as other repeat proteins, have shape, structure and properties very different from those of typical globular proteins [Hagay, T. & Levy, Y. 2008]. This class of proteins is dominated by short-range and regularized interactions between amino acids relatively close together in the sequence, whereas globular proteins exhibit complex topologies that frequently result in numerous long-range interactions [Main, E.R.G. *et al.* 2005]. The innovative modular structure, the high stability and the ability to bind to different ligands using a common scaffold make TPR proteins an intriguing and interesting research target.
Through a statistical analysis of TPR-containing protein sequences, Regan and co-workers designed a TPR *consensus* sequence by identifying the amino acid with the highest propensity for each of the 34 positions within the motif [Main, E.R.G. *et al.* 2003; D'Andrea, L.D. & Regan, L. 2003]. To this purpose, a non-redundant proteins database was scanned, all TPR sequences were retrieved and the amino acid distribution for each of the 34 TPR positions was determined (Fig. 10 and 11). Although the TPR motif shows a highly degenerate sequence, with no positions invariant, there is a consistent pattern of small and large hydrophobic residues that are essential for the structural integrity of the TPR fold. These residues, which are boxed in yellow in Fig. 10, present the highest global propensity values, and can be thought of as the TPR motif's "signature residues" because they form many of the fundamental inter- and intra-repeat packing interactions. As shown in Fig. 11, the sequence Gly-Asn-Ser and an additional "solvating" helix were respectively added at the N-and C-terminus of the TPR motif to stabilize the sequence [Main, E.R.G. *et al.* 2003]. *Consensus* TPR modules can be arrayed in tandem to create new optimized TPR proteins containing a different number of repeats (CTPRn) that can be used as a general tool to study repeat proteins, allowing a better understanding of the features of this class of proteins (Fig. 12).

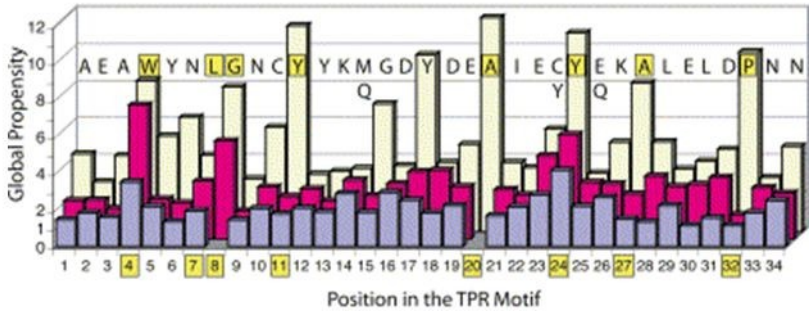

Figure 10 Histogram showing the global propensity of the three most prevalent amino acids at each of the 34 positions of the TPR motif. Values were obtained from a statistical analysis with the PFAM database. The column corresponding to the amino acid with highest global propensity is shaded pale yellow; the second highest, pink; the third highest, blue. The sequence corresponding to the amino acids with the highest global propensity at each position is shown along the top of the figure. The highly conserved "TPR signature residues" and their numbers are boxed in yellow [Main, E.R.G. *et al.* 2003].

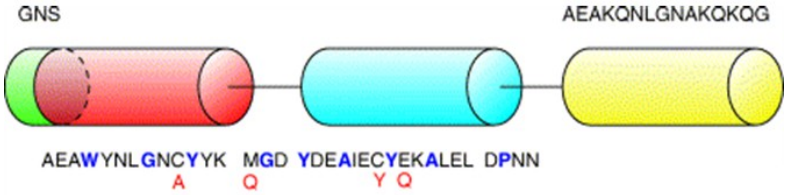

Figure 11 Schematic illustration of the designed CTPR protein. Shown are the N-cap (green), TPR motif (helix A, red; helix B, blue), link regions (black), and solvating helix sequence (yellow). The sequence of the N-cap region and the solvating helix are shown above the figure. The solvating helix sequence is derived from helix A, in which residues exposed to the solvent were mutated to increase solubility (W4K, Y5Q, Y11K and Y12Q, where the number refers to TPR position). The sequence corresponding to the amino acids with the highest global propensity at each position is shown below the figure. The highly conserved 'TPR signature residues' are shown in blue. Mutations reported to be inserted in the TPR prototype sequence are shown in red. The final sequence of the designed consensus variant was GNS (AEAWYNLGNAYYKQGDYDEAIEYYQKALELDPNN)n AEAKQNLGNAKQKQG [Main, E.R.G. *et al.* 2003].

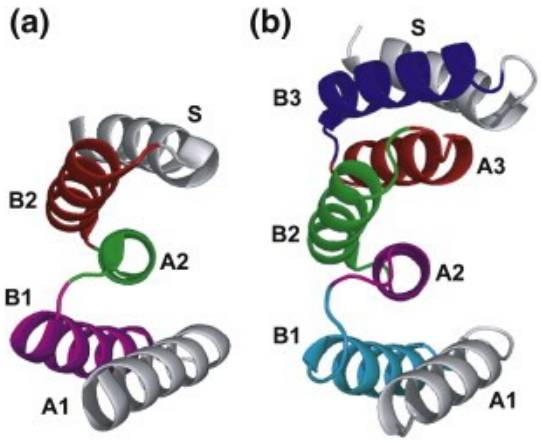

Figure 12 Ribbon representation of **A)** CTPR2, PDB ID 1NA3 and **B)** CTPR3, PDB ID 1NA0 crystal structures. The two helices of each TPR motif are named as helix A and B and identified with the number of the repeat. In the picture the helices of CTPR2 and CTPR3 are represented in different colors [Main, E.R.G. *et al.* 2003].

Synthetic genes encoding the novel CTPR proteins with an increasing number of repeats were created, and the CTPR proteins overexpressed and purified from *E.coli*. All the variants produced adopt the typical TPR fold and were shown to be stable, monomeric and highly soluble [Main, E.R.G. *et al.* 2003]. The design of idealized proteins that can adopt a TPR typical fold allowed to better understand the features of this class of proteins. Using these ideal TPR proteins, Regan group studied the folding mechanism of repeat proteins and demonstrated [Kajander, T. *et al.* 2005; Cortajarena, A.L. *et al.* 2008] that this class of proteins does not conform the typical two-state folding mechanism, generally assumed for globular proteins and consisting of only fully folded or unfolded species. Rather, TPR folding can be quantitatively described by the classical one-dimensional Ising model [Zimm, B.H. & Bragg, J.K. 1953], which prescribes the existence of a wide population of partially folded configurations with significant statistical weight. The Ising model predicts a hierarchical unfolding process in which each helix behaves as a single folding unit and the outermost helices show a significantly higher probability of being unfolded than the inner helices. Kajander and coworkers [Kajander, T. *et al.* 2005] observed that, as the number of the repeats increased, the folding/unfolding transition midpoint occurred at increased temperature or GdnHCl concentrations and, equally, also the slope of the unfolding curve increased (Fig. 13). These observations are expected for proteins that fold following the Ising model rather than a two-state mechanism. Hydrogen/deuterium exchange (HX) studies lend support to this picture. Regan and coworkers measured the native state hydrogen exchange for the *consensus* proteins CTPR2 and 3, and observed that the distribution of protection factors along the polypeptide backbone was consistent with the “end frying” predicted by the Ising model. These HX data demonstrated that helices closer to the end unfold first and, globally, the protein unfolds from both N and C termini toward the center [Main, E.R.G. *et al.* 2005; Cortajarena, A.L. *et al.* 2008]. Javadi and Main also supported this theory exploring the energy landscape of a series of designed CTPR proteins. They investigated the kinetic folding cooperativity and folding energy landscapes on increasing repeat protein length by comparing the folding kinetics of a series of designed CTPR proteins. The results showed that as a repeat number increases the landscape became more complex, with intermediates increasingly populated [Javadi, Y. & Main, E.R.G. 2009]. A very interesting feature of CTPR proteins was described in Cortajarena *et al.* 2008. They used FCS to analyze the hydrodynamic radius (R_h) of a set of CTPR proteins in the denatured state. The results revealed that, in presence of GdnHCl, R_h values observed for repeat proteins containing 2-20 tandem copies of TPR motif strongly deviated from those predicted for random coil structures. R_h values observed indicated a significantly more compact structure of the unfolded state of CTPR proteins. The origin of this unexpected behavior can be ascribed to the presence of an extensive nonnative polyproline II (PPII) helical structure in the denatured proteins, that was detected by CD spectroscopy. In fact, it is established that PPII helix have a distinctive CD spectrum, with a maximum at 229 nm [Woody, R.W. 1992], as observed in the spectra of CTPR proteins at different concentrations of GdnHCl (Fig. 14). In the same work, Cortajarena and coworkers identified the segment of the repeat proteins which assumes a PPII helical conformation. The segment overlaps the last helix of a motif (helix B) and the first helix of the next motif (helix A). Moreover, to better understand the effect of PPII structure on the dimensions of the denatured CTPR proteins, they performed simulation using simple lattice model of polymers. They showed that chains with a regularly-ordered sequence of short and long links (the latter representing PPII segments) may attain a

more compact structure than that predicted for standard random structure. The incorporation of interactions between PPII segments in the model leads further compaction of the chains, with size scaling similar to that observed in the experiment. Concluding, PPII structure induces significant compaction of denatured proteins, with high effects on the dimension of the denatured protein state. The formation of nonnative structure in denatured state may have implications for the folding mechanism of CTPR proteins, which will be explored in this PhD thesis.

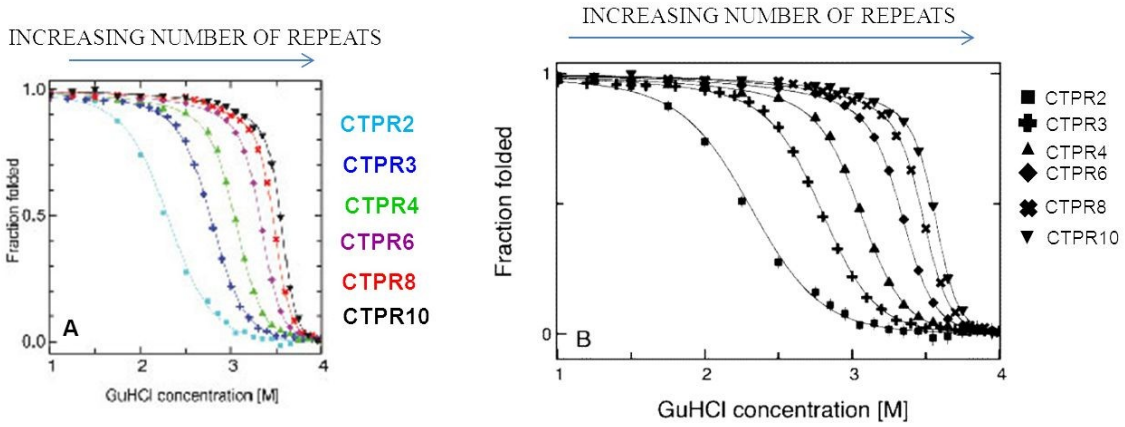

Figure 13 Fraction folded *vs* T (**A**) or [GdnHCl] (**B**) for CTPR constructs containing 2-10 repeats [Kajander, T. *et al*. 2005].

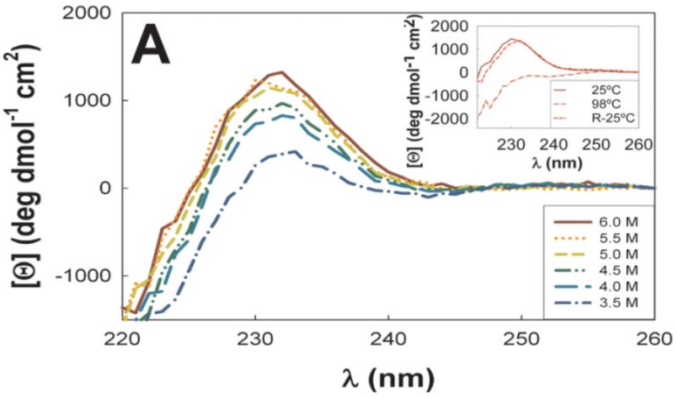

Figure 14 CD spectra of CTPR8 at increasing GuHCl concentrations showing PPII helical structure. The inset shows CD spectra of CTPR8 at 6 M GuHCl at 25°C (solid line), at 98°C (dashed and dotted line) and at 25°C after the sample was heated and cooled down (dashed line).

Aim of the work

Chemical modification of proteins with biophysical probes appears an extremely valuable tool for investigating and characterizing the molecular mechanisms controlling biological phenomena, such as protein folding. For long time the difficulties associated with the site-specific introduction of more than one probe into a protein have hindered the implementation of such approach. The introduction of Chemical Ligation methods represents a key event in this field, allowing the site-specific labeling of proteins with a broad range of molecular probes. Chemical ligation strategies in combination with standard biophysical techniques have greatly improved our ability to understand protein function, structure and folding mechanism [Muralidharan, V. & Muir, T.W. 2006]. This PhD thesis describe a useful and general protocol based on the use of EPL for the preparation of proteins site-specifically functionalized at multiple sites. The power of the semisynthetic protocol developed was demonstrated through the production of a series of CTPR3 protein variants in which two fluorophores were incorporated at different and specific positions, to allow the conduction of FRET experiments. The four doubly labeled variants prepared were named CTPR3[1_C], CTPR3[2_C], CTPR3[N_C] and CTPR3[1_3] depending on the location of the two fluorophores (for instance, the construct CTPR3[1_C] harbors the two fluorofores of the pair respectively in the first CTPR motif and at C-terminus). Besides, this PhD thesis reports the results of preliminary biophysical studies performed by CD and ensemble-FRET on the obtained labeled constructs. The semi-synthetic strategy proposed is generally applicable to the preparation of proteins in which a combination of two different labels is incorporated at any desired position within the protein. The approach described, moreover, could be adapted to the introduction of more than two probes and could be used for a wide range of applications not only in biophysics but even in biotechnology.

EXPERIMENTAL SECTION

MATERIALS AND INSTRUMENTS

pTXB1 (Fig. 15A) and pETM13 (Fig. 15B) plasmids were respectively obtained from New England Biolabs and European Molecular Biology Laboratory. pProEXHTa vector (Fig. 15C) containing the gene construct coding CTPR3 (360 bp) cloned between *Bam*HI and *Hind*III sites (pProExHTa-*ctpr3*) was provided by *Prof*. Lynne Regan from the Department of Molecular Biophysics and Biochemistry of Yale University (New Haven, CT-USA). All restriction and modification enzymes were purchased from New England Biolabs, except for *PfuTurbo* DNA Polymerase (Stratagene) and restriction enzyme *Sac*I (Roche). Molecular weight markers for nucleic acids and loading dye for the preparation of DNA samples were supplied by NEB and Roche. DNA purification kits were purchased from Qiagen. The synthesis of the oligonucleotides was committed at Sigma-Genosys, while plasmids sequencing was committed at Primm. *E. coli* TOP F'10 and DH5α strains, used for cloning, were supplied by Invitrogen, while cell strain BL-21(DE3), used for proteins expression, was purchased from Novagen (Merck). Isopropylbeta-D-thiogalactopyranoside (IPTG) was from Inalco. Protease inhibitor phenylmethanesulfonyl-fluoride (PMSF) was obtained from Sigma-Aldrich. Protein molecular weight markers, reagents for the preparation of bacterial growth media, buffers, polyacrylamide and agarose gel electrophoresis were provided from Sigma-Aldrich and Applichem. Fluorophores ATTO 647N-maleimide and ATTO 488-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) from STREMM. Fmoc-protected amino acids and coupling reagents for peptide synthesis were purchased from InBios and AnaSpec. Fmoc-Gly-Wang resin was provided by NovaBiochem (Merck). N,N-diisopropylethylamine (DIPEA) was provided from Romil, while piperidine was purchased from Biosolve. Acetic anhydride was supplied by Sigma-Aldrich. Other solvents for peptide synthesis and purification (dimethylformamide, N-methyl-2-pyrrolidone, dichloromethane, trifluoroacetic acid, ethandithiol, triisopropyl silane, diethyl ether, acetonitrile) were from Sigma-Aldrich or Romil. All reagents and solvents were used without further purification.
PCR reactions were performed on an Eppendorf Mastercycler personal apparatus. Electroporation was carried out in M-Medical cuvettes with an electrode gap of 0.2 cm and using a Bio-Rad Gene Pulser Xcell™ electroporator. Cell lysis was carried out using a Misonix Sonicator 3000. 0.22 μm polyvinylidene difluoride (PVDF) filters were acquired from Millipore. Dialysis membrane was from Spectra/Por. HisTrap affinity chromatography columns and ÄKTA FPLC system were from GE HealthCare. Polyacrylamide and agarose gel were visualized using a Chemi doc XRS apparatus (Bio-Rad). Analytical characterization of peptide and proteins was performed on an LC-MS system comprising an LCQ DECA XP ion trap mass spectrometer (ThermoElectron) equipped with an ESI source and a complete Surveyor HPLC system (including MS pump, autosampler and photo diode array [PDA]). The LC-MS analysis of peptide and proteins was carried out on a Jupiter C18 column, 5 μm, 300 Å, 250 x 2.0 mm (Phenomenex). Purifications by HPLC were performed on an HP 1200 Series (Agilent Technologies) using a Jupiter C18 column, 10 μm, 250 x 22 mm (Phenomenex), Jupiter C18 column, 10 μm, 250 x 10 mm (Phenomenex) and a Proteo C12 column, 10 μm, 250 x 22 mm (Phenomenex). Methods used for the analyses and purifications are specified in the following paragraphs. UV-VIS spectra

were recorded on a Jasco V-550 UV-Vis spectrophotometer and quartz cuvettes with a path length of 1 cm. Circular Dicroism (CD) analyses were performed on AVIV Model 215 CD spectophotometer (AVIV Instruments) and an Applied Photophysics Chirascan CD spectrometer (Applied Photophysics) equipped with Peltier temperature controller. Fluorescence spectra were recorded using a PTI Quantamaster C-61 two channel fluorescence spectrophotometer (Photon Technology International). CD and fluorescence analyses were performed using quartz cuvette with path length of 0.3 cm.

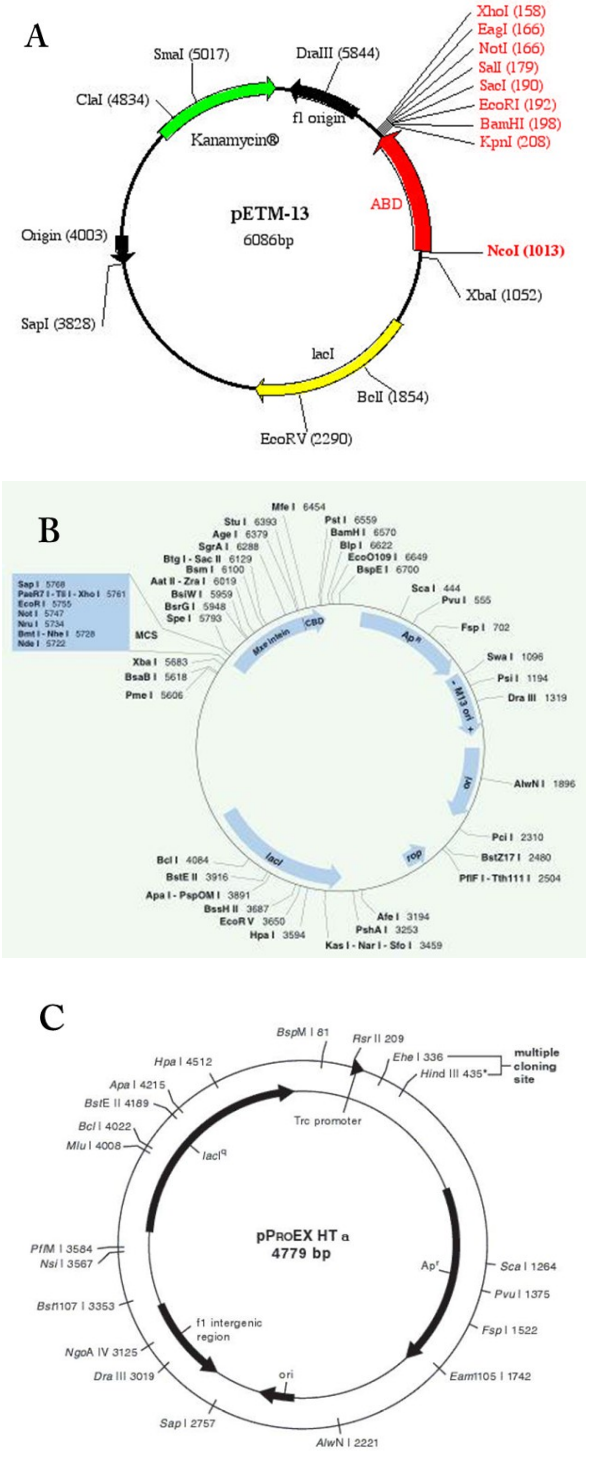

Figure 15 Structural organization of pETM13 (A), pTXB1 (B) and pProEXHTa (C) plasmids.

METHODS

Antibiotics

Ampicillin and kanamycin (commercially available as ampicillin sodium salt and kanamycin sulphate) were solubilized in deionized water at a concentration of 1000X, filter sterilized and stored at –20°C until use. Ampicillin was used at a concentration of 100 µg/mL while kanamycin was used at a concentration of 50 µg/mL in both solid and liquid media.

Solid and liquid media for bacterial strains

LB broth was prepared as follows: 10 g/L tripton, 5 g/L yeast extract, 10 g/L NaCl. Solid medium was prepared dissolving complete LB-agar powder in deionized water at a concentration of 40 g/L. Culture media were supplemented with antibiotic.

Preparation of *E. coli* TOP F'10 competent cells and transformation by electroporation

2.5 mL of an overnight culture of *E. coli* TOP F'10 cells were inoculated into 250 mL of LB medium. The cells were grown up to mid-log phase (0.6 OD_{600nm}) at 37°C, stored on ice for 20-30 min and then harvested by centrifugation (6000 rpm, 10 min, 4°C). The pellet was washed in 250 mL of sterile water. After the second centrifugation the cells were washed in 125 mL of sterile water. The third washing was performed in 5 mL of 10% glycerol; the cells were then harvested by centrifugation and the pellet was resuspended in 750 µL of 10% glycerol. Aliquots of 30 µL were mixed with 1.5 µL of ligase reaction, incubated for 1 min on ice and transferred into chilled plastic cuvettes. High voltage electroporation (25 µF) was performed at a field strength of 2.5 kV/cm and 200 Ohm. A shock pulse was applied to competent cells producing pulse length of ~5.0-5.5 ms. Immediately after electroporation cell mixtures were diluted to 1 mL with LB medium and incubated for ~1 h at 37°C under shaking. The cells were then plated onto selective solid medium supplemented with 100 µg/mL ampicillin or 50 µg/mL kanamycin to isolate the recombinant clones. Single clones were inoculated in 10 mL LB medium with the same antibiotics and grown overnight at 37°C under shaking. Finally, cells were harvested by centrifugation (7,000 rpm, 10 min, 4°C) and processed to extract plasmidic DNA by using the QIAprep Spin Miniprep Kit or the QIAGEN plasmid Maxi Kit.

Preparation of *E. coli* competent cells and transformation by heat shock

Single clone of BL-21(DE3) or DH5α *E. coli* strain, grown at 37°C in LB-agar, was inoculated into 2.5 mL of LB medium and incubated overnight at 37°C on a shaker. The cells were inoculated into 250 mL of LB medium and the culture was grown up to mid-log phase (0.6 OD_{600}) at 37°C, stored on ice for 20-30 min and then harvested by centrifugation (6000 rpm, 10 min, 4°C). The pellet was washed in 125 mL of cold 50 mM $CaCl_2$ and stored on ice for 30 min. Successively, cells were harvested by centrifugation and the pellet resuspended in 16 mL of cold 50 mM $CaCl_2$. Aliquots of 200 µL of competent cells were mixed with 50 ng of plasmidic DNA or 10 µL of ligase reaction mix and stored on ice for ~30 min. The cells were transferred at 42°C for 45 sec, on ice for 2 min (heat shock) and then diluted to 1 mL with LB medium. An incubation for ~1 h at 37°C under shaking was performed before plating the cells onto solid medium supplemented with the opportune antibiotic.

Electrophoretic analysis of DNA
DNA samples were prepared in 1X loading dye and then loaded on 1% or 2% agarose gel prepared in TAE buffer 1X (1 mM EDTA, 40 mM Tris base. Add acetic acid until pH 7.8) and containing 5 µg/mL ethidium bromide. Electrophoresis was performed in the same buffer TAE 1X for 30 min at 100V.

Electrophoretic analysis of proteins (SDS-PAGE)
The electrophoresis on polyacrylamide gel in the presence of sodium dodecyl sulfate (SDS-PAGE) was performed according to Laemmli's protocol [Laemmli, U. 1970]. The samples were denatured at 100°C for 10 min in loading buffer (1% SDS, 0.001% bromophenol blue and 10% glycerol) without β-mercaptoethanol. The samples were then loaded on a 15% polyacrylamide gel and electrophoresed in running buffer 1X (0.025 M Tris-HCl, 0.2 M glycine pH 8.3 and 0.1% SDS). The proteins were then revealed by Coomassie Brilliant-Blue staining; the gel was first submerged in the prestaining solution (20% ethanol, 10% acetic acid) and then in the staining solution (0.01% Coomassie Brilliant-Blue R250, 5% ethanol, and 7.5% acetic acid) with gentle agitation. The gel was finally washed in H_2O to remove the excess of Coomassie.

Determination of the protein concentration
The concentration of proteins in solution was determined according to the Bradford's method [Bradford, M. 1976]. The Coomassie Brilliant (Bio-Rad) reagent was added to the samples and the absorbance at 595 nm was recorded. A solution of 1.5 µg/µL of bovine serum albumin (BSA) was used as standard. Protein concentration was also measured by UV spectroscopy, reading the tryptophan absorbance at 280 nm. Labeled proteins concentration was estimated by UV-Vis spectroscopy, detracting from the absorbance at 280 nm the contribution of the fluorophore/s, using the following equation:

$$Abs_{280\ nm}^{corrected} = Abs_{280\ nm} - \left[\left(Abs_{\lambda\ max}^{I\ Fluorophore} \times CF^{I\ fluorophore}\right) + \left(Abs_{\lambda\ max}^{II\ Fluorophore} \times CF^{II\ fluorophore}\right)\right]$$

where CF are correction factors suggested by the fluorophores manufacturer (ATTO 647N-maleimide CF 0.05; ATTO 488-maleimide CF 0.1).

Bioinformatic tools
Protein sequences were processed using *ProtParam* tool available on the ExPASy Server (http://www.expasy.ch), in *proteomics tools* section. The tool was used to compute pI (isoelectric point), Mw (molecular weight) and ε (extinction coefficient) of proteins.

Cloning procedure
Introduction of the mutations Asn36Cys or Asn70Cys into ctpr3 gene
pProExHTa-*ctpr3* plasmid was mutagenized by cassette-mutagenesis to introduce mutations Asn36Cys or Asn70Cys into *ctpr3* gene. To this purpose, pProExHTa-*ctpr3* plasmid was co-digested with restriction enzymes *Xho*I and *Sac*I or *Sac*I and *Hind*III to extract gene cassettes including respectively Asn36 or Asn70 codon next to their 5' termini. Each co-digestion was performed on 30 µg of the recombinant plasmid pProExHTa-*ctpr3* in 33 mM Tris acetate, 10 mM magnesium acetate, 66 mM potassium acetate, 0.5 mM dithiothreitol pH 7.9, for 2.5 h at 37°C using 4 $U_{enzyme}/µg_{DNA}$. Excised cassettes (111 bp or 167 bp respectively) and digested vectors were extracted from agarose gel using the QIAquick Gel Extraction Kit. Gene

cassettes were then mutagenized by PCR using sense primers including the mutation Asn36Cys or Asn70Cys as well as *Xho*I or *Sac*I site, and antisense primers including *Sac*I or *Hind*III site (Table I). The mutagenic PCR reactions were performed in a total volume of 50 µL containing 50 ng of gene-cassette preparation as template, 30 pmol of each primer, 250 µM dNTP, 2.5 U (1 µL) of *PfuTurbo* DNA polymerase in the supplied buffer 1X. PCR reactions were performed following the procedure indicated below:

Initial denaturation step 3 min at 95°C

Denaturation (step 2) 1 min at 95°C
Annealing (step 3) 1 min at the convenient temperature for each primers pair
Elongation 1 min at 72°C
from step 2 for 30 cycles

Final elongation (step 4) 15 min at 72°C

PCR products were purified by using the QIAquick PCR Purification Kit, digested with *Xho*I and *Sac*I or *Sac*I and *Hind*III in the conditions described before in this paragraph and subcloned into pProExHTa-*ctpr3* vector between the same restrictions sites. Ligation reactions were carried out in a total volume of 10 µL, using 100 ng of the digested plasmid, a three and five fold molar excess of the insert, 1 µL of 400 U/µL T4 DNA ligase in the supplied buffer 1X. Ligation mixtures were incubated 3 h at room temperature and then 1.5 µL of each mixture was transformed by electroporation in competent *E. coli* TOP F'10 cells, following the transformation protocol described before in this section. After transformation, cells were plated onto LB-agar plates supplemented with ampicillin. The day after, clones obtained on plates were screened by colony-PCR. Clones were resuspended in 20 µL of sterile water and 5 µL of each cell suspension was used as template in PCR reactions. PCR was carried out in the same conditions and with the same primers adopted for mutagenic PCR before described. The remaining 15 µL of cell suspension of each positive clone were inoculated in 10 mL of LB supplemented with ampicillin and incubated at 37°C overnight. Recombinant plasmids were isolated from cells using QIAprep Spin Miniprep Kit and verified by sequencing.

Primer	Sequence 5'-3'	Application
Forward	CCGCTCGAGCTGGACCCGTGCAACGCA *Xho*I site Cys	Introduction of mutation Asn36Cys into *ctpr3* gene cassette
Reverse	GGAGCTCAAGCGCCTTTTGATAATA *Sac*I site	
Forward	CGAGCTCGATCCATGTAATGCAGAG *Sac*I site Cys	Introduction of mutation Asn70Cys into *ctpr3* gene cassette
Reverse	GGGAAGCTTCTATCAACCCTGTTTCTG *Hind*III site	

Table I PCR primers used for the introduction of mutations Asn36Cys and Asn70Cys into CTPR3 coding gene.

Preparation of the genes coding CTPR3 variants and Mxe GyrA intein

For the variant CTPR3[1_3], the gene trait corresponding to amino acids 1-104 of CTPR3 (312 bp) was amplified from pProExHTa-*ctpr3*(Asn36Cys). Instead, for CTPR3[1_C] and CTPR3[2_C] variants, the full *ctpr3* gene (360 bp, 120 amino acids) containing the mutation Asn36Cys or Asn70Cys was respectively amplified from mutagenic plasmids pProExHTa-*ctpr3*(Asn36Cys) and pProExHTa-*ctpr3*(Asn70Cys). For CTPR3[N_C] variant preparation, 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 site upstream the gene an extra Gly residue in position -2 was also added. All the above mentioned genes were amplified using sense primers including *Nco*I 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) (Table II).

Primer	Sequence 5'-3'	Application
Forward	CATGCCATGGGTAACTCCGCTGAGGCATGG *Nco*I site	Amplification of CTPR3[1_3]-fragment 1-104 coding gene from pProExHTa-*ctpr3(Asn36Cys)*
Reverse	***TCCCGTGATGCA***GTTCGGGTCCAGTTCCAGCGC *Mxe GyrA overlapping region*	
Forward	CATGCCATGGGTAACTCCGCTGAGGCATGG *Nco*I site	Amplification of CTRP3[1_C] and CTPR3[2_C] coding genes from ProExHTa-*ctpr3(Asn36Cys) or (Asn70Cys)*
Reverse	***TCCCGTGATGCA***ACCCTGTTTCTGTTTAGCGTT *Mxe* GyrA overlapping region	
Forward	CATGCCATGGGT**TGC**GGTAACTCCGCTGAGGCA *Nco*I site Cys	Amplification of CTPR3[N_C] coding gene from wild-type pProExHTa-*ctpr3*
Reverse	***TCCCGTGATGCA***ACCCTGTTTCTGTTTAGCGTT *Mxe* GyrA overlapping region	

Table II PCR primers used for the amplification of CTPR3 variants gene.

The gene coding *Mxe* GyrA intein (594 bp) was amplified from pTXB1 vector using sense primers generating a 5' terminal region of 15 bp overlapping CTPR3 variant coding genes and an antisense primer including *Eco*RI restriction site (Table III).

Primer	Sequence 5'-3'	Application
Forward	***CAGAAACAGGGT***TGCATCACGGGAGATGCACTA overlapping region	Amplification of *Mxe* GyrA coding gene from pTXB1 and addition of a 5'terminal region of overlap with CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C] genes
Reverse	GGAATTCCCAGCGTGGCTGACGAACCCGTT *Eco*RI site	
Forward	***CTGGACCCGAAC***TGCATCACGGGAGATGCACTA overlapping region	Amplification of *Mxe* GyrA coding gene from pTXB1 and addition of a 5' terminal region of overlap with the fragment 1-104 of CTPR3[1_3]
Reverse	GGAATTCCCAGCGTGGCTGACGAACCCGTT *Eco*RI site	

Table III PCR primers used for the amplification of *Mxe* GyrA intein gene.

All PCR reactions were performed as described in the previous paragraph, using the polymerase *PfuTurbo* which is endowed of 3'-5' exonuclease activity. After amplification, all the genes were purified using QIAquick PCR Purification Kit.

Fusion genes construction by Overlapping Extension PCR

Hybrid genes coding the CTPR3 variants fused to *Mxe* GyrA intein were obtained by Overlapping Extension PCR. Each reaction was performed in a total volume of 100 µL using 50 ng of *Mxe* GyrA gene and an equimolar amount of the CTPR3 gene variant, 250 µM dNTPs, 2.5 U (1 µL) of *PfuTurbo* DNA polymerase, 2.5% formamide, *PfuTurbo* buffer 1X. Before adding the two primers, 5 cycles of template extension were performed as follows:

Initial denaturation step	3 min at 95°C
Denaturation (step 2)	1 min at 95°C
Annealing (step 3)	1 min at 55°C
Elongation	1 min at 72°C
for 5 cycles from step 2	
Final elongation (step 4)	15 min at 72°C

Then, 30 pmol (3 µL of a 10 µM dilution) of each of the two primers annealing 5' terminus of CTPR3 variant coding gene and 3' terminus of *Mxe* GyrA were added to the PCR reaction and 25 cycles of amplification were performed as follows:

Initial denaturation step	3 min at 95°C
Denaturation (step 2)	1 min at 95°C
Annealing (step 3)	1 min at the convenient temperature for each primers pair
Elongation	1 min at 72°C
for 25 cycles from step 2	
Final elongation (step 4)	15 min at 72°C

PCR mixtures were then separated on a 2% agarose gel and the fragments corresponding to the fusion genes were extracted from agarose gel using QIAquick Gel Extraction Kit. Fusion genes were then digested with *Nco*I and *Eco*RI restriction enzymes. Digestions were performed in 20 mM Tris acetate, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM dithiothreitol pH 7.9, for 2.5 h at 37°C using 4 $U_{enzyme}/\mu g_{DNA}$. Similarly, also the expression vector pETM13 was digested with the restriction enzymes *Nco*I and *Eco*RI and treated with calf intestine phosphatase (CIP, 10U) for 1 h at 37°C in the supplied buffer (50 mM Tris-HCl, 100 mM NaCl, 10 mM $MgCl_2$, 1 mM dithiothreitol pH 7.9) to remove 5' phosphoryl groups. Digested fusion genes and vector were purified and, finally, fusion genes were cloned into pETM13 upstream an His_6-tag coding sequence. Ligation mixtures were prepared as before described in this text and transformed in chemically competent *E. coli* DH5α cells. Cells were spread on LB-agar supplemented with kanamycin. Positive clones were screened by colony-PCR as yet described and sequenced.

Expression procedure
Recombinant pETM13 plasmid harboring chimeric DNA coding CTPR3 variant fused to *Mxe* GyrA intein was transformed in chemically competent *E. coli* BL-21(DE3) cells. Cells were plated on LB-agar supplemented with kanamycin and incubated overnight at 37°C. A single colony was inoculated in 25 mL of LB media with kanamycin and incubated overnight a 37°C while shaking. The small scale culture was used to inoculate 1 L of LB media at 0.06 OD_{600nm}; the culture was incubated at 37°C until the OD_{600nm} 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.

Thioester proteins purification procedure
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 proteases inhibitor PMSF. The cells were lysated by sonication and the crude extract was centrifuged at 16,000 rpm for 30 min at 4°C. The supernatant, containing the soluble His_6-tagged fusion protein, was filtered using a 0.22 µm PVDF filter and loaded onto a 5 mL HisTrap column pre-equilibrated in buffer A. The purification was performed using an ÄKTA FPLC system at 4°C. After soluble fraction loading, the column was extensively washed with buffer A (1 mL/min) until reaching UV baseline. 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) and immediately incubated at 4°C, under mild stirring, with 50 mM MESNA. After splicing, the proteins mixture was extensively dialyzed in 5 mM phosphate buffer pH 6.5, using a membrane of molecular cut-off of 6-8 kDa, and then lyophilized. The lyophilized proteins mixture was resuspended in 20% CH_3CN, 0.1% TFA in water and incubated with 20 mM TCEP to reduce Cys residues. C-terminal thioester derivative of the CTPR3 variants were isolated by reverse-phase HPLC using a Jupiter C18 column, 10 µm, 250 x 22 mm performed at a flow rate of 20 mL/min. Thioester protein CTPR3[1_3]-fragment 1-104 was purified using a step-gradient of CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 20% to 30% in 3 min and from 30% to 40% in 20 min; thioester proteins CTPR3[1_C] and [N_C] were purified using a step-gradient of CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 20% to 29% in 5 min and from 29% to 33% in 30 min; thioester protein CTPR3[2_C] was purified using a step-gradient of CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 20% to 29% in 3 min and from 29% to 36% in 20 min. Collected fractions were analyzed by LC-mass spectrometry using a Jupiter C18 column, 5 µm, 300 Å, 250 x 2.0 mm using the same methods adopted for proteins purification before described and a flow rate of 0.2 mL/min. Fractions containing pure CTPR3 variants as MESNA C-terminal thioester were pooled and lyophilized.

Cys-Peptide (CTPR3[1_3]-fragment 105-120) synthesis
Peptide NH_2-Cys-Ala-Glu-Ala-Lys-Gln-Asn-Leu-Gly-Asn-Ala-Lys-Gln-Lys-Gln-Gly–COOH, corresponding to the C-terminal fragment 105-120 of CTPR3[1_3] was synthesized on solid phase using a Fmoc-Gly-Wang resin by standard Fmoc chemistry. Peptide was synthesized on a 0.1 mmol scale. All the incubation steps were performed under stirring. For each coupling reactions were used 10 equivalents of the Fmoc-protected amino acid, 9.9 equivalents of Hydroxybenzotriazole (HOBt)/ O-Benzotriazole-N,N,N',N'-tetramethyl-uronium-hexafluoro-phosphate (HBTU) and 20 equivalents of the base DIPEA (1 h each). 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). Between each reaction step, 5 wash of 1 min each in DMF were performed to remove exceeding reagent and reaction side-products. Peptide cleavage from the resin and amino acids side-chain deprotection were achieved by treatment with TFA, EDT, triisopropyl silane 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 on a Proteo C12 column, 10 µm, 250 x 22 mm using a linear gradient of CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 5% to 40% in 50 min at a flow rate of 20 mL/min. Fractions were analyzed by LC-MS, pooled and lyophilized.

First labeling reaction
Lyophilized thioester proteins were resuspended in 20 mM phosphate buffer, pH 7.2 at a concentration of 0.25 mM. A two-fold molar excess of the first fluorophore (ATTO 647N-maleimide) was dissolved in 20 µL of DMF and added to protein. Reaction mix was incubated for 2 h at 4°C on the dark, under mild stirring. Then, exceeding fluorophore was quenched adding to the reaction mixture a number of µmoles of MESNA equal to those of fluorophore added before. Then, the labeled thioester proteins were dialyzed in 20 mM phosphate buffer pH 7.2 at 4°C on the dark.

Native Chemical Ligation with L-Cys
For the preparation of the variants CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C], the mono-labeled thioester proteins were incubated overnight at 4°C, on the dark, with a 20 fold molar excess of L-Cys. After reaction occurred, mono-labeled proteins were extensively dialyzed in 20 mM phosphate buffer pH 7.2 to remove the L-Cys excess. Mono-labeled proteins were then lyophilized, resuspended in CH_3CN 20%, TFA 0.1% in water and incubated at 4°C in 20 mM TCEP to reduce C-terminal extra Cys residue. After, proteins were purified by reverse-phase HPLC performed using a Jupiter C18 column, 10 µm, 250 x 10 mm and a linear gradient of CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 20% to 70% in 40 min at a flow rate of 5 mL/min. Fractions were analyzed by LC-MS, pooled and lyophilized.

Native Chemical Ligation with Cys-peptide
For the preparation of the variant CTPR3[1_3], mono-labeled thioester protein corresponding to the first 104 amino acids of the construct was incubated with a tenfold molar excess of the Cys-peptide. The composition of the NCL reaction mixture was: 0.25 mM mono-labeled thioester CTPR3[1_3]-fragment 1-104, 2.5 mM Cys peptide, 100 mM phosphate buffer pH 7.5, 4 M GdnHCl, 1 mM EDTA, 50 mM MESNA, 15 mM TCEP. Reaction vessel was closed under N_2 and incubated at 4°C on the dark with mild stirring. The reaction was monitored by LC-MS and allowed to proceed until the disappearance of the thioester protein peak. After, the NCL reaction mixture was diluted 1:2 with H_2O and extensively dialyzed in 20 mM phosphate buffer pH 7.2. Mono-labeled full protein was then lyophilized, resuspended in 20% CH_3CN, 0.1% TFA in water, reduced with 10 mM TCEP and purified by HPLC as before described for the other constructs on a Jupiter C18 column, 10 µm, 250 x 10 mm at a flow rate of 5 mL/min. CTPR3[1_3] labeled with 647N on Cys36 was purified using a

gradient of CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 10% to 40% in 5 min and from 40% to 50% in 40 min.

Second labeling reaction

Single labeled constructs were dissolved in 20 mM phosphate buffer pH 7.2 and reacted with the second fluorophore of the FRET pair. The second labeling reaction (performed with fluorophore ATTO 488-maleimide) occurred on the Cys residue involved in the NCL reaction (C-terminal Cys for CTPR3[1_C], [2_C] and [N_C]; Cys105 for CTPR3[1_3]). Second labeling was carried out in the same experimental conditions described for the first labeling reaction. After reaction, exceeding label was quenched with MESNA as described for the first labeling. Finally, doubly labeled proteins were newly dialyzed in 20 mM phosphate buffer pH 7.2 and lyophilized. Doubly labeled proteins were then resuspended in 20% CH_3CN, 0.1% TFA in water and all purified by reverse-phase HPLC using a Jupiter C18 column, 10 µm, 250 x 10 mm performed at a flow rate of 5 mL/min. Doubly labeled variant CTPR3[1_C] and CTPR3[2_C] were purified using a linear gradient of CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 20% to 70% in 30 min. Doubly labeled CTPR3[N_C] was purified using a gradient of CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 20% to 70% in 25 min. Doubly labeled CTPR3[1_3] was purified using a gradient of CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from 20% to 70% in 50 min. Fractions were analyzed by LC-MS and those containing the doubly labeled protein were pooled and lyophilized.

Control constructs preparation

Control constructs harboring only the donor or the acceptor fluorophore or none of them were also prepared using the same approach adopted for the preparation of the doubly-labeled species. Unlabeled proteins containing the two Cys residues in the same position as the doubly labeled proteins were obtained carrying out NCL reactions between unlabeled thioester moieties and L-Cys or Cys-peptide as described for the preparation of the labeled analogs. For the preparation of unlabeled CTPR3[1_C] and CTPR3[2_C], after NCL, proteins were reduced with 10 mM TCEP, extensively dialyzed in 5 mM phosphate buffer pH 7.2 and lyophilized. A further purification step was not necessary, due to the homogeneity of the sample obtained. CTPR3[N_C] unlabeled protein was not prepared because this variant does not contain internal mutations but only insertions at N- and C-termini and we assume it to retain the same stability of the wild-type protein. For the preparation of unlabeled CTPR3[1_3], after NCL with the Cys-peptide, the protein was dialyzed in 5 mM phosphate buffer pH 7.2 and lyophilized. Protein was resuspend in 20% CH_3CN, 0.1% TFA in water, reduced with 10 mM TCEP and purified to homogeneity by reverse-phase HPLC using a Jupiter C18 column, 10 µm, 250 x 10 mm, performed at 5 mL/min and using a step gradient from 20% to 30% in 3 min and from 30% to 55% in 30 min. Mono-labeled proteins, containing only the first fluorophore of FRET pair ATTO 647N in position N-terminal, 36 or 70 and the second Cys free, were prepared concomitantly with the preparation of the doubly-labeled variants. The control construct containing only the second fluorophore of the pair, ATTO 488, at C-terminal position was prepared expressing the wild-type CTPR3 protein as fusion construct with *Mxe* GyrA intein; thioester CTPR3, which does not contain Cys residues, was reacted in NCL reaction with L-Cys, adding a single Cys residue at the C-terminus. Cys was labeled with ATTO 488 similarly as before described and purified by reverse phase HPLC using a linear gradient of CH_3CN (TFA 0.1%) in H_2O (TFA 0.1%) from

20% to 45% in 45 min on a Jupiter C18 column, 10 µm, 250 x 10 mm, performed at 5 mL/min. Pure fractions were pooled and lyophilized.

Spectroscopic characterization of CTPR3 variants

CD analyses

All CD experiments were performed in 50 mM phosphate buffer, 150 mM NaCl pH 7.0 and acquired in the far UV range 190-260 nm. The CD spectra were recorded at a protein concentration of 3 µM, at 25°C, with a band width of 1 nm, a step of 1 nm and a time per point of 3 sec, in a 0.3 cm path length cuvette. Each CD spectrum was corrected for buffer blank obtained under identical conditions.
Thermal denaturation was monitored by following the ellipticity signal of the helical structure at 222 nm from 10°C to 94°C and in the reverse direction from 94°C to 10°C. The temperature ramp was performed in 1°C steps with an equilibration time at each temperature of 1 min. The estimation of protein unfolded fraction content at each temperature was evaluated giving 0% value to mdegree registered in the native state (10°C) and 100% value to mdegree estimated in the unfolded state (94°C).
Chemical denaturation experiments were performed at 25°C, using urea or GdnHCl as denaturant. 150 µL of sample at protein concentration of 10 µM was titrated with 50 mM phosphate buffer, 150 mM NaCl, containing 9 M urea or 7 M GdnHCl. At each titration point the concentration of urea or GdnHCl was increased by 0.25 M and the sample was equilibrated for 2 min before acquiring the CD spectrum. Ellipticity at 222 nm was corrected considering the dilution factor. The estimation of protein unfolded fraction content at each titration point was evaluated giving 0% value to mdegree registered in the native state (0 M [urea] or [GdnHCl]) and 100% value to mdegree estimated in the unfolded state (8.25 M [urea] or 5 M [GdnHCl]).

FRET analyses

All fluorescence spectra were acquired at 25°C. 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. To minimize sample variability effects, CD and FRET chemical denaturation experiments were performed on the same sample and in the same cuvette with path length of 0.3 cm, recording before the fluorescence spectrum and immediately after CD spectrum at each GdnHCl or urea concentration. Chemical denaturation experiments were performed at 25°C, using urea or GdnHCl as denaturant. 150 µL of sample at protein concentration of 10 µM was titrated with 50 mM phosphate buffer, 150 mM NaCl, containing 9 M urea or 7 M GdnHCl. At each titration point the concentration of urea or GdnHCl was increased by 0.25 M and the sample was equilibrated for 2 min before acquiring the fluorescence spectrum. At each titration point, fluorescence intensities of donor and acceptor fluorophores at their emission maximum wavelengths were recorded and their ratio was used to estimate unfolded protein content at each denaturant concentration values. Acceptor fluorophores ATTO 647N emission maximum was observed at 669 nm; donor fluorophores ATTO 488 emission maximum was recorded at 523. Fluorescence intensities ratio at 0M [urea] or [GdnHCl]) was correlated to 0% unfolded fraction, while the same ratio at 5 M GdnHCl or 8.25 M urea was associated to 100% unfolded fraction. Titration with the denaturant caused a slight deviation of emission maximum wavelengths from these values.
For the analysis of FRET data the following equations were considered. The efficiency of the nonradiative FRET process (*E*) is given by:

$$E = \frac{R_0^6}{R^6 + R_0^6} = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \qquad [1]$$

where R is the donor–acceptor distance and R_0 is the Förster radius, defined to be the distance at which E is 50%. To relate fluorescence intensities of the donor and acceptor to FRET efficiency, the following expression was used:

$$E = \frac{F_A}{F_A + \gamma F_D} \qquad [2]$$

where F_A and F_D correspond to the acceptor and donor fluorescence intensities, respectively, when the donor is excited, and γ is correction factor for the quantum yield and the detector efficiency described as:

$$\gamma = \frac{\eta_A \Phi_A}{\eta_D \Phi_D} \qquad [3]$$

where η_A and η_D corresponds to the detector corrections for acceptor and donor while Φ_a and Φ_d are the fluorescence quantum yields for the donor and acceptor of the aceptor and donor. In the system used for the experiments, γ values for ATTO labels was assumed 1. By equating expressions 1 and 2, the donor–acceptor distance R can be measured.

Fluorescence anisotropy assays

Fluorescence anisotropy studies were performed in 50 mM phosphate buffer, 150 mM NaCl pH 7.0 at 3 µM protein concentration in a 0.3 cm path-length cuvette at 25ºC. To perform fluorescence anisotropy experiments the fluorescence spectrophotometer (reported in Material and Instrument paragraph of this section) was equipped with excitation and emission polarizers. ATTO 647N excitation was accomplished with a 4 nm slit-width at 647 nm and the emission recorded at 669 nm with slit widths of 4 nm. ATTO 488 excitation was accomplished with a 4 nm slit-width at 495 nm and the emission recorded at 521 nm with slit widths of 4 nm.
G-factor corrections to account the difference in transmission efficiencies of the two emission channels were applied to calculate anisotropy using the equation:

$$G - factor = \frac{I_{HV} - I_{B,HV}}{I_{HH} - I_{B,HH}}$$

where I_{HV} is the vertical emission (0º) of a standard solution of free fluorophore with excitation in horizontal orientation (90º), I_{HH} is the horizontal emission of a standard solution with excitation in horizontal orientation, $I_{B,HV}$ is the vertical emission of a blank solution with excitation in horizontal orientation and $I_{B,HH}$ is the horizontal emission of a blank solution with excitation in horizontal orientation. For excitation at vertical orientation (0º) the expression for anisotropy (r) containing the G-factor is:

$$r = \frac{(I_{VV} - I_{B,VV}) - G(I_{VH} - I_{B,VH})}{(I_{VV} - I_{B,VV}) + G(I_{VH} - I_{B,VH})}$$

where G is the G-factor, I_{VV} and I_{VH} are the vertical and horizontal emission of the sample, respectively, and $I_{B,VV}$ and $I_{B,VH}$ the intensity of emission of the blank with emission polarizer at vertical and horizontal orientation, respectively.

RESULTS AND DISCUSSION

Semisynthetic approach

The strategy developed for the preparation of doubly-functionalized proteins is schematically illustrated in Fig. 16.

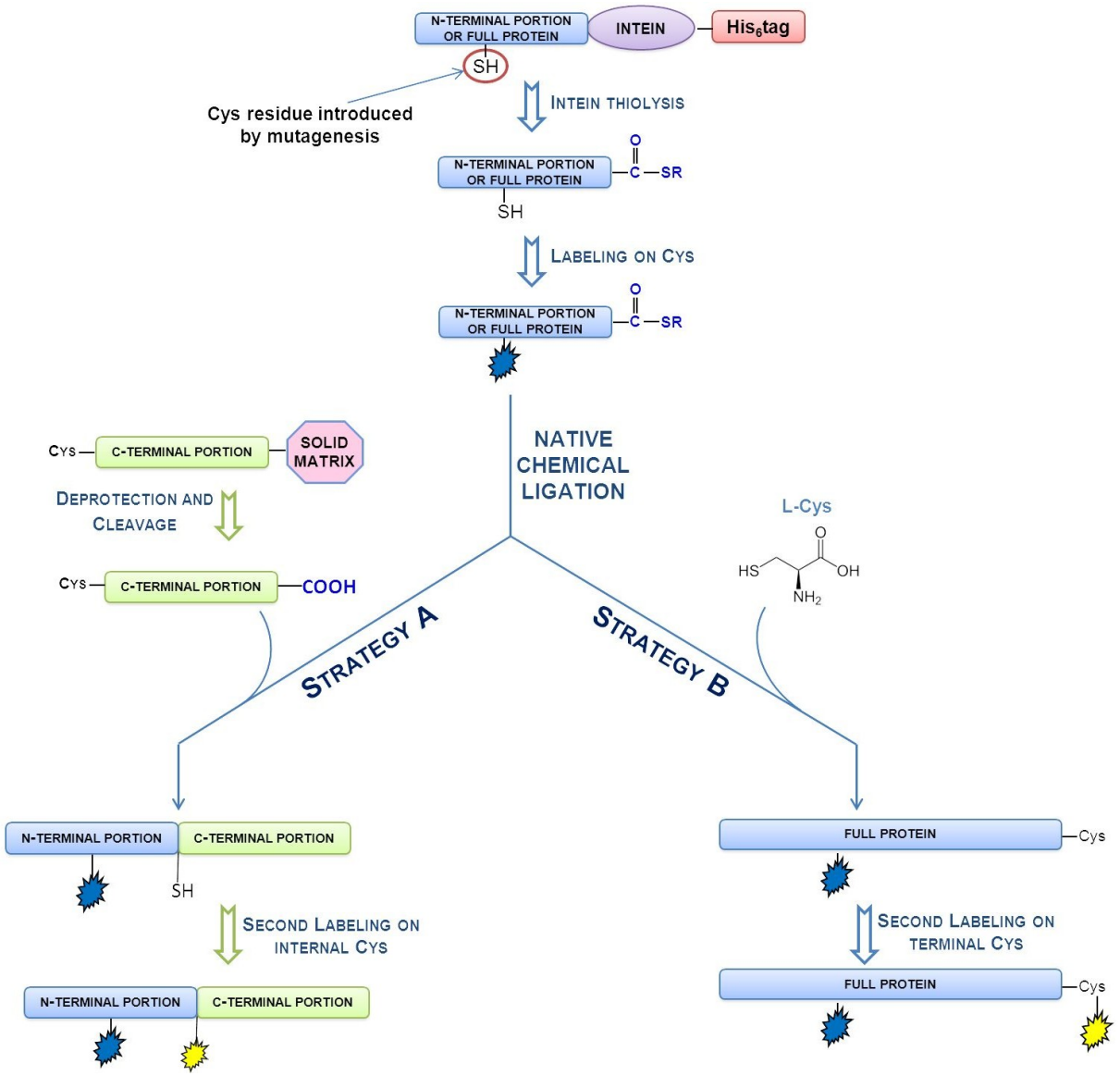

Figure 16 Schematic representation of the semisynthetic strategy adopted for the production of doubly-labeled proteins. **Strategy A)** Approach adopted for the preparation of a protein carrying both probes inside the sequence. The N-terminal portion of the protein was produced as C-terminal thioester through the expression with the *Mxe* GyrA intein as fusion partner. After purification, the N-terminal fragment of the protein of interest, carrying a C-terminal thioester group, was labeled with the first probe and then reacted by NCL with a chemically synthesized peptide corresponding to the C-terminal remaining portion of the protein and containing an N-terminal Cys residue. Cys involved in NCL was then used for the site-specific labeling with the second probe. **Strategy B)** Approach adopted for the preparation of proteins carrying the second probe at C-terminus. The full protein was expressed as fusion construct with the intein and after thioester purification and labeling with the first molecular probe, the protein was reacted by NCL with L-Cys, so that an extra-sequence Cys residue was inserted at C-terminus. C-terminal Cys was finally used for the second labeling.

The approach is based on the use of the Expressed Protein Ligation (EPL) and requires the splitting of the protein of interest in two fragments. The N-terminal fragment, containing a single Cys residue, is expressed in bacterial host as 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 splicing of *Mxe* GyrA intein and the release of the protein fragment as C-terminal thioester. Thioester protein is then isolated from intein and selectively labeled on its single Cys residue with a probe containing a maleimide group as sulfhydryl-reactive moiety. Instead, the C-terminal remaining fragment, carrying an N-terminal Cys residue, is chemically synthesized. The Native Chemical Ligation (NCL) reaction between the mono-labeled thioester fragment and the synthetic peptide 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. 16 strategy A). A simpler approach could be 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 a required position is expressed as 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. 16 strategy B).
A slightly modified protocol of strategy A could allow the preparation of a triple labeled protein. In fact, the fragment chemically synthesized and containing the N-terminal Cys, could be selectively labeled on solid phase in any position. After NCL reaction between thioester protein and the synthetic peptide, both carrying a probe, the Cys involved in NCL could be used for the introduction of the third probe.

Doubly-labeled CTPR3 protein variants

The semisynthetic approach proposed was adopted for the preparation of four variants of the protein CTPR3 (Fig. 17) labeled with a fluorophores pair. The variants are schematically represented in Fig. 17C and are named CTPR3[1_3], CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C]. The nomenclature adopted indicates the protein regions in which the fluorescent probes (an acceptor and a 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. In CTPR3[1_3] variant (Fig. 17C, #1), the two fluorophores were respectively placed in the first and third CTPR motif (amino acid sequence positions 36 and 105). Among the variants prepared, CTPR3[1_3] exhibits the shortest spatial distance between the fluorophores (28.4 Å, from CTPR3 x-ray structure; pdb code:1NA0) and is the only construct harboring both fluorophores on internal regions of the protein. The other three variants exhibit at least one of the probes on terminal protein regions. In fact, CTPR3[1_C] (Fig. 17C, #2) and CTPR3[2_C] (Fig. 17C, #3) constructs bear only the first probe on internal protein region, respectively in the first and second CTPR motif (position 36 and 70 respectively), while the second probe is placed at C-terminus (position 121). The distance between fluorophores is 40.3 Å in CTPR3[1_C] and 31.3 Å in CTPR3[2_C] variant. CTPR3[N_C] (Fig. 17C, #4) presents both fluorophores in the terminal protein regions, the first probe at N-terminus while the second at C-terminus (position -1 and 121). CTPR3[N_C] shows the higher spatial distance between the fluorophores (52.0 Å).

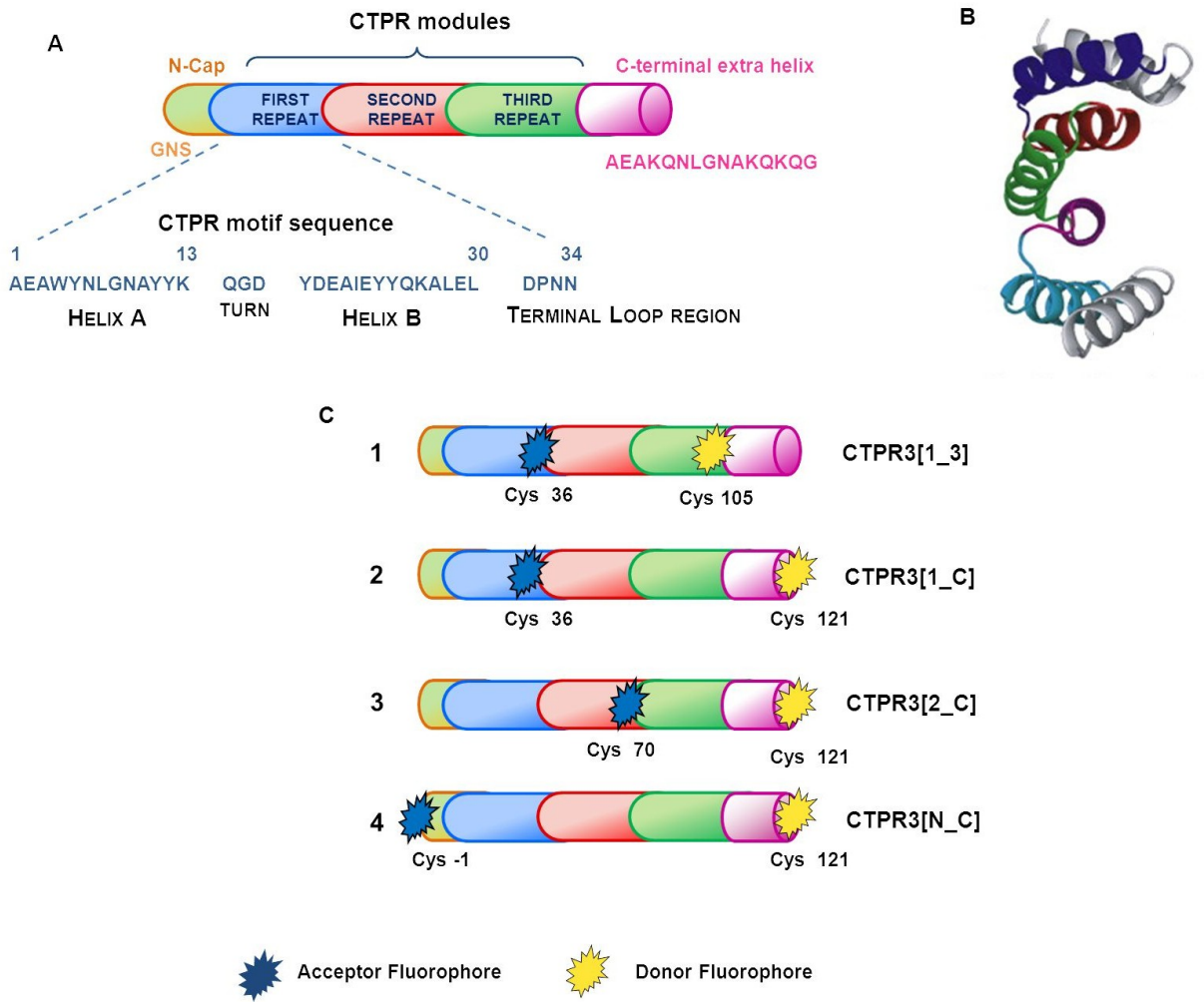

Figure 17 A) Schematic representation of CTPR3 wild-type. The protein contains three identical repeats of the 34 amino acids *consensus* TPR module flanked by a N-terminal cap sequence and a C-terminal solvating extra-helix. **B)** Ribbon representation of CTPR3, PDB ID 1NA0 crystal structure. **C)** Schematic representation of the doubly-labeled CTPR3 variants prepared by the semisynthetic approach. The doubly-labeled variants differ for fluorophores location, which could be placed inside the protein or in the terminal regions.

CTPR3[1_3] variant preparation

The strategy A illustrated in Fig. 16 was adopted for the semisynthesis of CTPR3[1_3] protein variant (Fig. 17C, #1).

Design of the variant: choice of the sites of labeling

The sites of labeling were selected in order to assure a minimal perturbation of the native protein structure. Therefore, we decided to introduce the two labels in the loop regions connecting a repeat with the next one, so that the probes were oriented towards the outside of the protein, limiting their effect on protein folding (Fig. 17A and B). Among the four amino acids which constitute this loop region (Asp-Pro-Asn-Asn), the first Asn residue is associated with the lowest global propensity value for that

position [Main, E.R.G. *et al.* 2003] and thus, the mutation of such amino acid should not affect significantly native protein structure. Therefore, we selected this residue as site of labeling. Accordingly, the first labeling site in the variant CTPR3[1_3] was located in position 36, corresponding to the less conserved Asn residue within the terminal loop of the first CTPR motif. However, the positioning of also the second fluorophore on the correspondent Asn residue in the loop of the third repeat (Asn in position 104), in terms of semisynthetic strategy, implies the fusion of the protein fragment 1-103, exhibiting a Pro as last amino acid, to the *Mxe* GyrA intein. Pro residue as amino acid directly flanking the *Mxe* GyrA intein is known to prevent intein splicing [Muir, T.W. *et al.* 2003; Southworth, M.W. *et al.* 1999]. Thus, in order to guarantee the success of the semisynthetic protocol designed, we moved the second labeling site to the subsequent amino acid of the loop (Asn in position 105 instead than Asn 104). In this way, in fact, the fragment 1-104 had to be fused to the intein, and thus an Asn is placed in the splice junction instead than Pro, which assures the feasibility of the splicing reaction. Even if the Asn 105 presents a higher global propensity value than the precedent Asn 104, it is not included between the "TPR signature residues", allowing the mutation of the residue without the loss of the CTPR fold.

Cloning procedure

The cloning strategy adopted is schematized in Fig. 18. We performed cassette-mutagenesis on the recombinant pProEXHTa plasmid harboring wild-type CTPR3 coding gene (pProEXHTa-*ctpr3* plasmid), in order to replace with a Cys residue the native Asn in position 36, located inside the terminal loop of the first repeat. To this purpose, pProEXHTa-*ctpr3* plasmid was digested with *Xho*I and *Sac*I restriction enzymes to extract a gene cassette including the Asn 36 codon (Fig. 18A). The excised gene cassette was recovered by extraction from agarose gel and then mutagenized by PCR, performing the reaction with a forward primer carrying the desired mutation. After PCR, the cassette was newly digested with *Xho*I and *Sac*I and subcloned into pProEXHTa-*ctpr3* plasmid between the corresponding restriction sites. The resulting mutagenic plasmid (pProEXHTa-*ctpr3*Asn36Cys) was verified by sequencing and then used as template in a further PCR reaction by which the gene trait corresponding to the first 104 amino acids of CTPR3[1_3] was obtained. The PCR was performed employing a forward primer which contained *Nco*I restriction site and a chimeric reverse primer, which added at 3' end of the gene trait a region of overlap with the gene coding *Mxe* GyrA intein. Similarly, *Mxe* GyrA intein gene was amplified from pTXB1 plasmid using a forward primer adding at the 5' end of intein gene a region of overlap with the gene trait coding the N-terminal fragment of CTPR3[1_3], and a reverse primer including *Eco*RI restriction site (Fig. 18B).

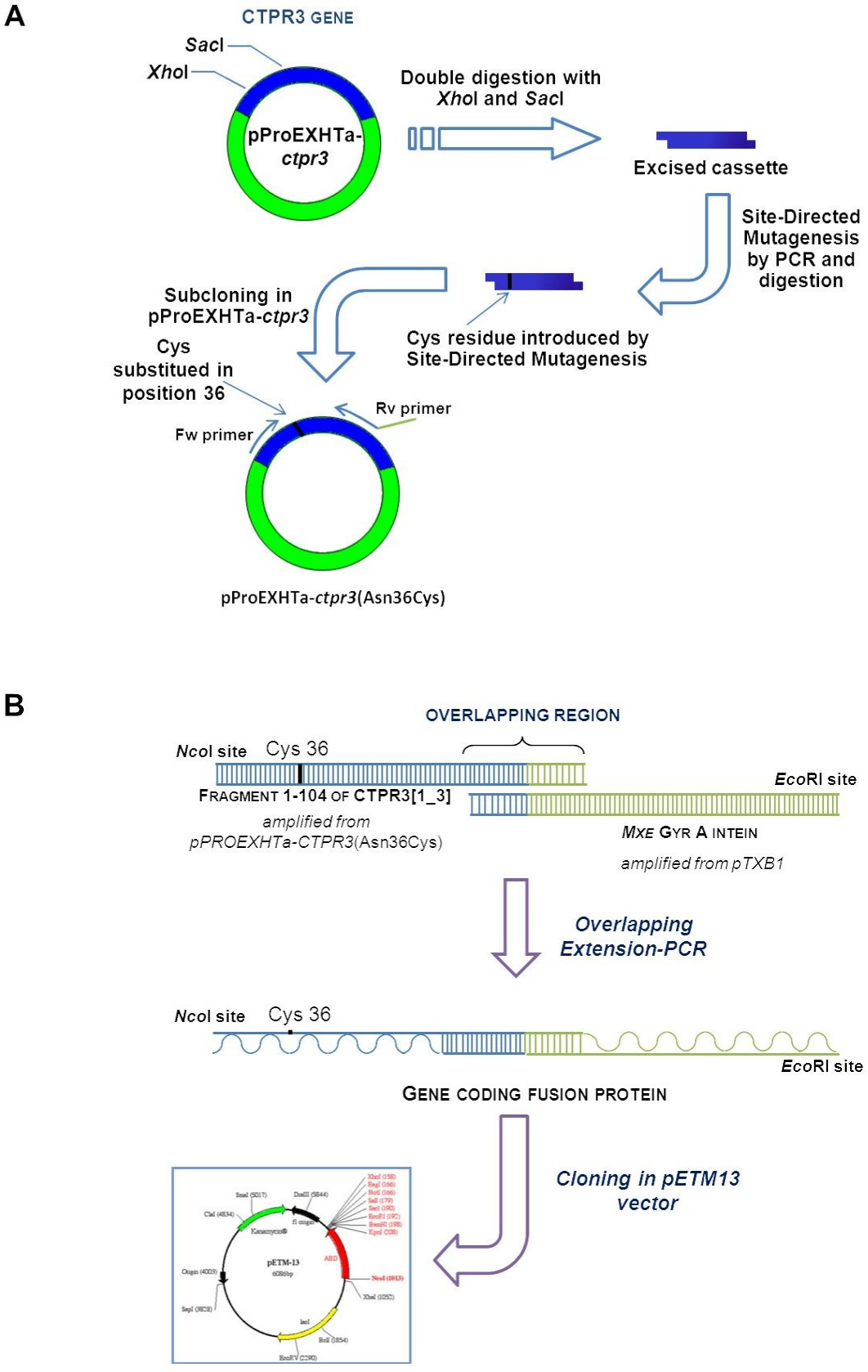

Figure 18 Schematic representation of the cloning strategy adopted for CTPTR3[1_3] variant synthesis. **A)** Preparation of the mutagenic plasmid pProEXHTa-*ctpr3*(Asn36Cys) by cassette mutagenesis and its use as template for the amplification of gene coding the N-terminal fragment of CTPR3[1_3] (amino acids 1-104). **B)** Construction of the chimeric gene coding the fragment 1-104 of CTPR3[1_3] fused to *Mxe* GyrA intein by Overlapping Extension PCR and cloning in pETM13.

After amplification and purification (Fig. 19A), the two genes were mixed in equimolar amounts in an Overlapping Extension PCR reaction [Ho, S.N. *et al.* 1989], in which their complementary regions (30 bp) annealed and served as primers for the extension of the fusion gene (Fig. 19B).

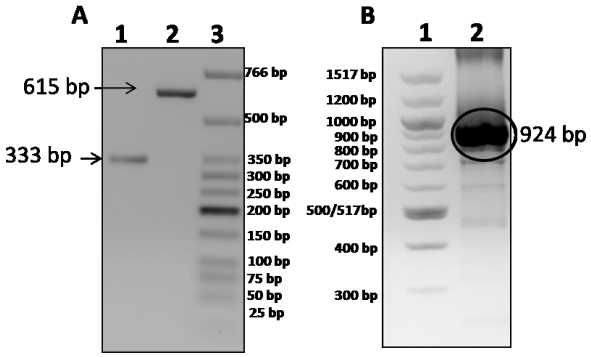

Figure 19 Analysis by agarose gel electrophoresis of: **A)** gene coding the fragment 1-104 of CTPR3[1_3] amplified by PCR on pProEXHTa-*ctpr3*(Asn36Cys) (*lane 1*); *Mxe* GyrA intein gene obtained from PCR performed on pTXB1 vector (*lane 2*); ladder 25 bp (*lane 3*); **B)** ladder 100 bp (*lane 1*); Overlapping Extension PCR reaction between overlapping genes coding the fragment 1-104 of CTPR3[1_3] and *Mxe* GyrA intein (*lane 2*). The circle highlights fusion gene band.

The resulting chimeric gene was recovered by agarose gel extraction, digested with *Nco*I and *Eco*RI restriction enzymes and cloned into pETM13 vector, upstream an His_6 tag coding sequence. Recombinant plasmid (pETM13-*ctpr3[1_3](fragment 1-104)- mxe gyra intein*) was verified by sequencing and used for fusion protein expression.
By the use of 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. Moreover, His_6 tag was placed at C-terminus of the fusion partner which is a self-removing element. Accordingly, no tag removal mediated by protease was required for the implementation of the discussed semisynthetic protocol.

Expression

The recombinant pETM13 plasmid coding the fragment 1-104 of CTPR3[1_3] variant fused to *Mxe* GyrA intein was utilized to transform bacterial cells, in order to express the corresponding fusion protein. The host cell chosen for fusion protein expression was the BL21-(DE3) strain of *E. coli.* The optimal expression conditions were selected by comparing different temperatures (22°C and 37°C), IPTG concentrations (0.3, 0.5 and 0.8 mM) and time of induction (5 h and overnight). Fusion protein accumulated in the soluble fraction of bacterial cells during the expression at 22°C, in all the conditions tested, while expression at 37°C caused the partial accumulation of the recombinant protein in inclusion bodies. The highest expression level was observed using 0.5 mM IPTG and performing the expression overnight at 22°C (Fig. 20). No significant splicing *in vivo* of the intein was observed during expression, according to the data reported in literature assessing that, when an Asn residue is located in the protein-intein junction, no *in vivo* premature intein cleavage is observed [Southworth, M.W. *et al.* 1999].

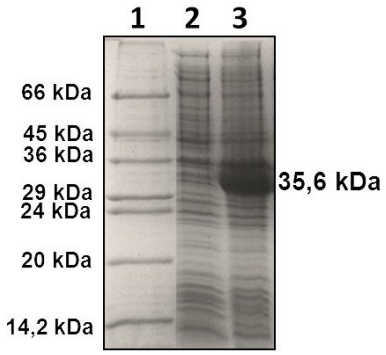

Figure 20 SDS-PAGE analysis of the fusion protein (CTPR3[1_3](fragment 1-104)-*Mxe* GyrA intein) (35.6 kDa) expressed in *E. coli* BL21-(DE3). The 15% polyacrylamide gel shows a molecular weight marker (*lane 1*), uninduced (*lane 2*) and induced (*lane 3*) cells samples. The SDS-PAGE analysis was performed using loading buffer without β-mercaptoethanol, in order to avoid splicing of the fusion protein.

Purification procedure

Fusion protein was purified from soluble fraction of bacterial cells by nickel affinity chromatography, exploiting the presence of the histidines tail situated at the C-terminus of intein. All purification steps were performed at low temperature (4°C), in order to minimize intein splicing before the addition of thiols, which led to the release of the CTPR3 variant as a C-terminal carboxylic-acid, unable to react in NCL reaction. For the same reason, purification by nickel affinity chromatography was performed at lower pH value (pH 6.5) than that usually adopted for this kind of chromatography (pH 8.0). In fact, as described in the introductive chapter, intein splicing can be induced by temperature and pH as well as by thiols. Furthermore, we avoided the use of the tris(hydroxymethyl)aminomethane (Tris) buffer, which is usually preferred for this kind of affinity purification, because it has been demonstrated to react with the thioester protein forming a stable adduct and causing not negligible consequence on thioester protein final yields [Peroza, E.A & Freisinger, E. 2008]. Bacterial pellet from 1 L of culture was resuspended with ice-cold 50 mM phosphate buffer pH 6.5, 300 mM NaCl, 10 mM imidazole (buffer A), containing the proteases inhibitor PMSF to avoid fusion protein degradation. Cells suspension was sonicated and the lysate was centrifuged to recover the soluble fraction (supernatant). Cytosolic extract was then loaded onto a Ni^{2+}-NTA agarose resin column pre-equilibrated in buffer A. Fusion protein was eluted using a buffer at high concentration of imidazole (50 mM phosphate buffer pH 6.5, 300 mM NaCl, 500 mM imidazole - buffer B) (Fig. 21). Elution of fusion protein was not performed applying a gradient of imidazole to the column, but in one step, to reduce the purification time and minimize the spontaneous cleavage of the intein. We obtained a yield of 100 mg of fusion protein from 1 L of culture. Once eluted, fusion protein was immediately incubated with a large excess of the splicing-inducing thiol MESNA (50 mM) at 4°C, under mild stirring, in order to obtain the C-terminal thioester derivatives of the N-terminal fragment of CTPR3[1_3] variant.

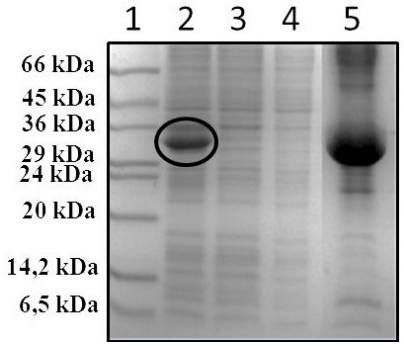

Figure 21 SDS-PAGE analysis of the fusion protein CTPR3[1_3](fragment 1-104)-*Mxe* GyrA intein (35.6 kDa) after purification by affinity chromatography. The 15% polyacrylamide gel shows: molecular weight marker (*lane 1*); soluble fraction of induced *E. coli* BL21-(DE3) cells (*lane 2*), in the circle is indicated the fusion protein band; unbound proteins eluted from column (*lane 3*); column wash with buffer A (*lane 4*); fusion protein eluted with buffer B (*lane 5*). SDS-PAGE analysis was performed using a loading buffer without β-mercaptoethanol, in order to avoid splicing of the fusion protein.

Intein splicing reaction was followed by time-course SDS-PAGE analysis (Fig. 22). After 6 days of incubation with thiols, ~50% of fusion protein underwent splicing. The observed rate of splicing is not in agreement with the data reported in literature [Southworth, M.W. *et al.* 1999], in which is stated that in presence of an Asn residue at the protein-intein splicing junction, an almost complete cleavage of the fusion protein is observed after overnight incubation with a thiol at 23°C. Probably, the lower temperature at which we performed the splicing (4°C) contributes to slowing down the splicing reaction rate. After 15 days of incubation with thiols, the percentage of cleaved protein is similar to that observed after 6 days. This observation suggests that, probably, the remaining part of the fusion protein failed to cleave because of misfolding or aggregation, justifying the slowing-down of the splicing reaction rate.

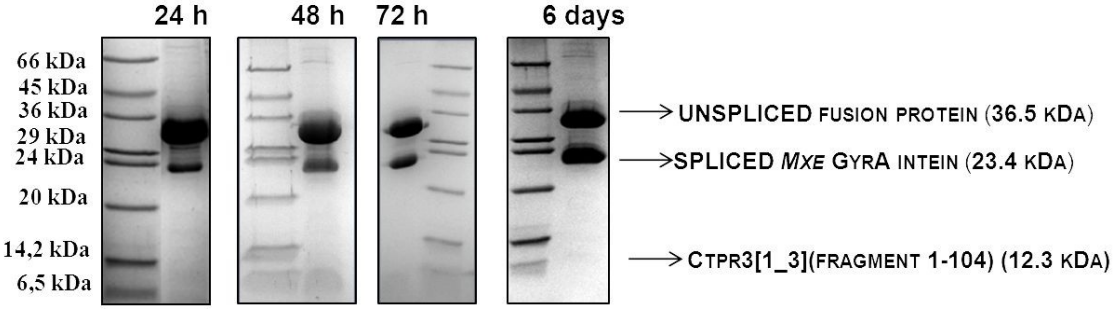

Figure 22 Time-course analysis by SDS-PAGE of the splicing reaction. CTPR3[1_3](fragment 1-104)-*Mxe* GyrA intein after 24 h, 48 h, 72 h and 6 days of incubation with 50 mM MESNA.

Thioester group was stable in the experimental conditions of splicing reaction. In fact, the hydrolysis of thioester group was controlled adopting a slightly acidic pH value (6.5) and an excess of thiol reagent. After splicing, thioester protein was extensively dialyzed against 5 mM phosphate buffer pH 6.5 to remove salt and thiol excess, and then lyophilized. Then, C-terminal thioester protein fragment was reduced with TCEP and purified from spliced intein and unspliced fusion protein by reverse-phase HPLC. The thioester protein was obtained in a good purity (Fig. 23A). The identity of the thioester protein was verified by ESI mass spectrometry and found in agreement with expected value (Fig. 23B). The final yield of thioester protein was of about 15 mg from 1 L of culture.

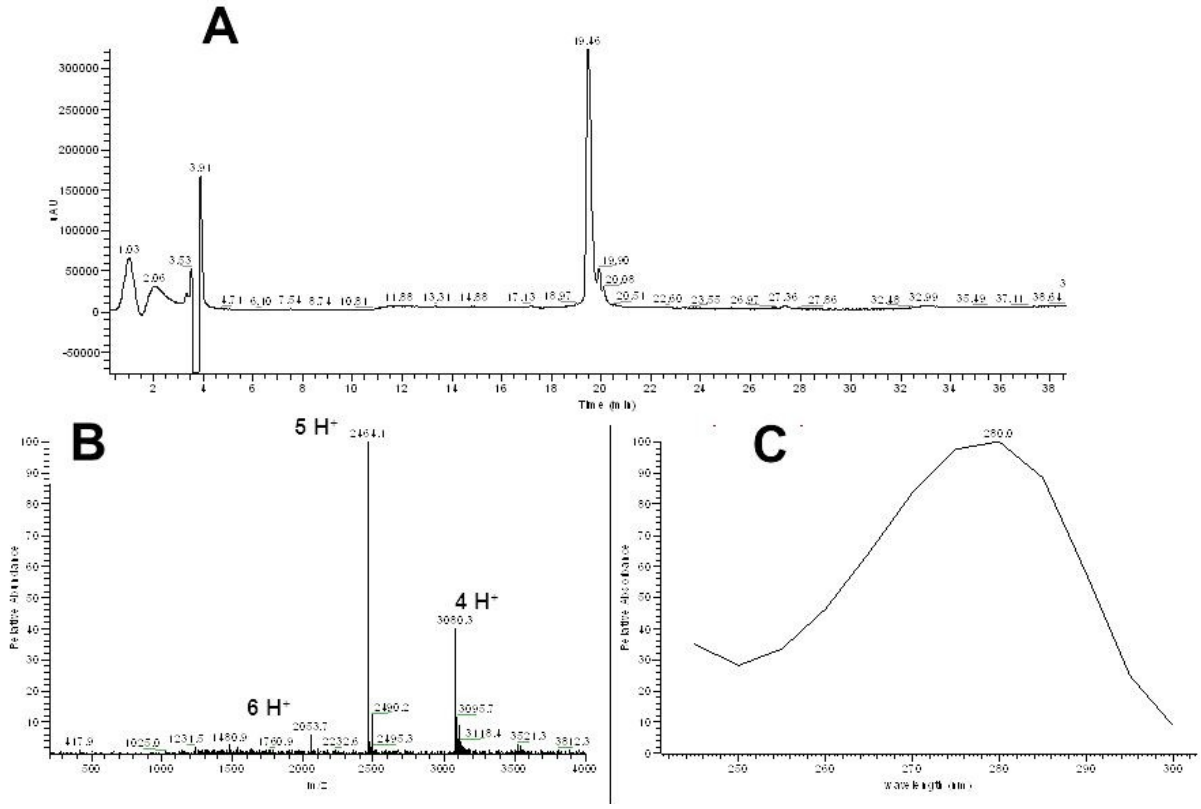

Figure 23 LC-MS analysis of the thioester protein corresponding to the N-terminal fragment of CTPR3[1_3] after HPLC purification. **A)** Chromatographic profile revealed at 210 nm. **B)** Mass spectrum of the chromatographic peak at 19.46 min. Experimental mass (12319.8 Da) is in agreement with the theoretical one (12319.0 Da). **C)** Absorption spectrum of the chromatographic peak at 19.46 min.

First labeling reaction

The thioester fragment 1-104 of CTPR3[1_3] was site-specifically labeled on the single Cys residue (position 36) with the maleimide derivative of the first molecular probe (acceptor fluorophore ATTO 647N). As ATTO 647N-maleimide is reported to react at pH value comprised between 7.0 and 7.5, we performed the labeling reaction at pH 7.2 in aqueous buffer (20 mM phosphate buffer), using a mild excess of fluorophore (two-fold molar excess). Reaction was carried out on the dark and at 4°C, as we found ATTO 647N fluorophore significantly thermo unstable, losing its spectroscopic properties at room temperature. No reducing agent was used as, 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 [Shafer, D. E. *et al.* 2000; Tyagarajan, K. *et al.* 2003]. The labeling reaction was checked by LC-MS analysis and occurred with yield >90% within 2 h. As shown in the HPLC profile reported in Fig. 24, labeled thioester protein elutes as double peak, because ATTO 647N fluorophore is supplied as mixture of two isomers with identical absorption and emission properties. Mass value obtained from the deconvolution of the m/z signals obtained for the labeled protein showed an increment of 768 Da (Fig. 24B) respect to the thioester protein mass, consistent with the addition of a molecule of ATTO 647N-

maleimide. The absorption spectrum of the protein peak shows the peculiar absorption features of the fluorophore ATTO 647N (Fig. 24C).

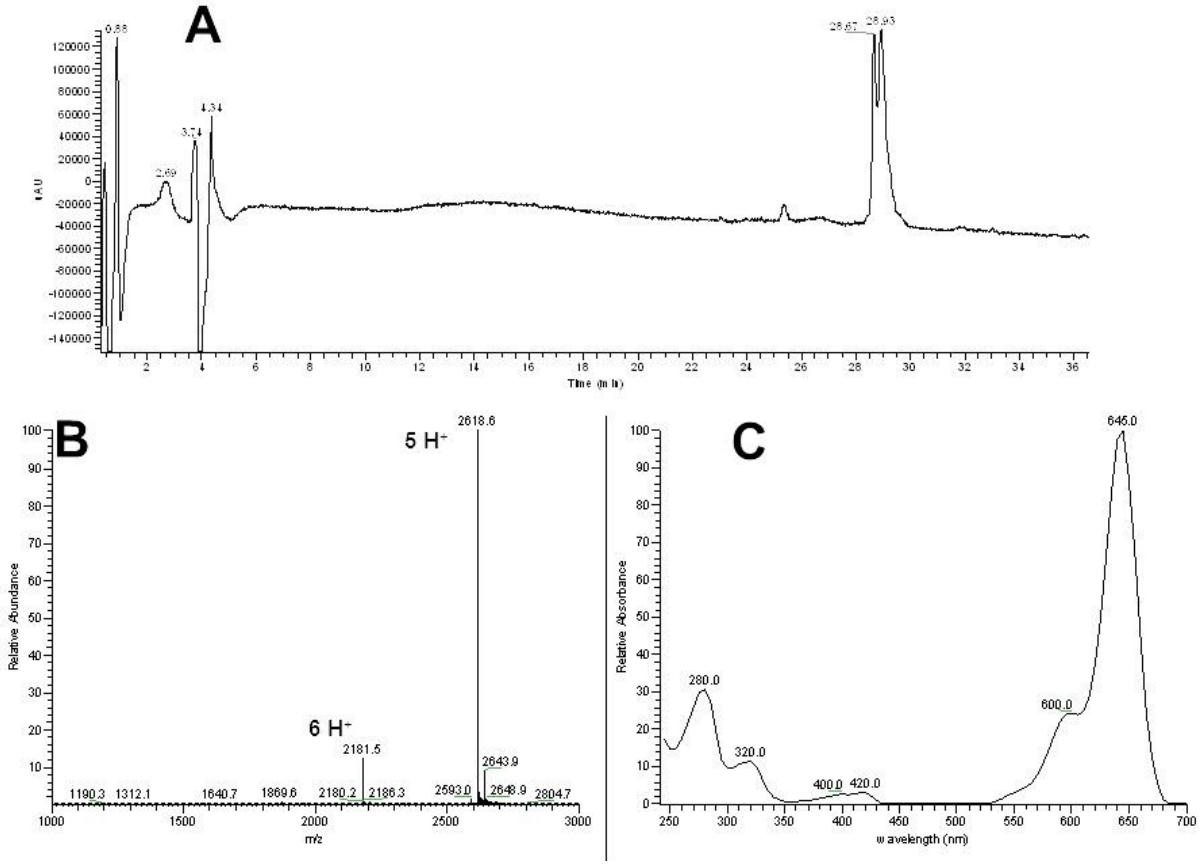

Figure 24 LC-MS analysis of the labeling reaction of the thioester protein with ATTO 647N-maleimide. **A)** Chromatographic profile revealed at 210 nm. **B)** Mass spectrum of the chromatographic peaks at 28.67 min and 28.93 min. Experimental mass (13090.0 Da) is in agreement with the theoretical one (13087.0 Da). **C)** Absorption spectrum of the chromatographic peaks at 28.67 min and 28.93 min.

To prevent unspecific labeling on multiple sites, exceeding fluorophore was quenched by the addition of the thiol MESNA and then removed by dialysis against 20 mM phosphate buffer pH 7.2. Dialysis was a pivotal step for the success of the subsequent chemical ligation reaction, because contributed to restore a pH value around the neutrality at which NCL occurs.

Native Chemical Ligation with Cys-peptide

The labeled thioester protein, corresponding to the fragment 1-104 of CTPR3[1_3] variant, was reacted with a synthetic peptide corresponding to the remaining portion of CTPR3[1_3] construct (protein fragment 105-120) (Fig. 25). The peptide, containing a Cys residue as N-terminal amino acid, was synthesized by Fmoc chemistry (see Experimental Section).

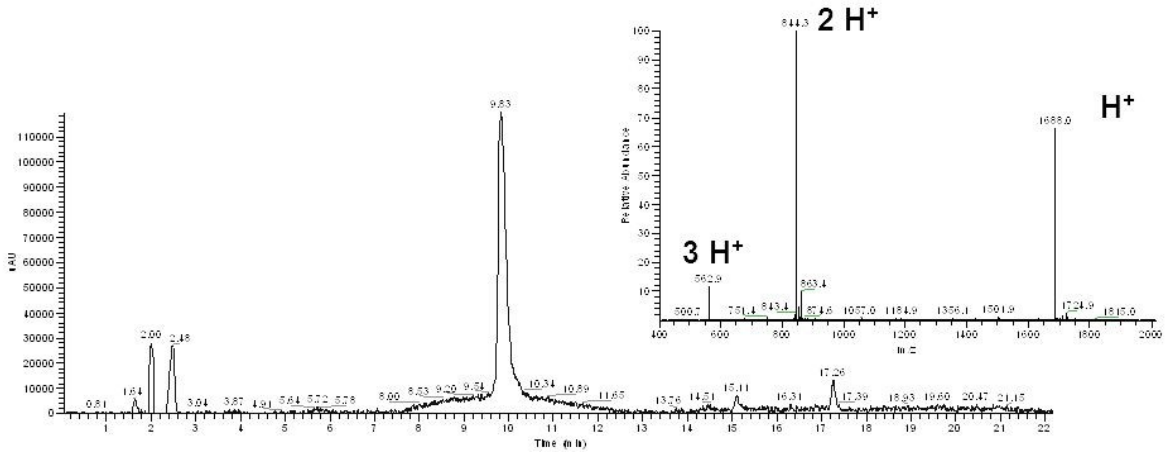

Figure 25 LC-MS analysis of the synthetic peptide corresponding to the C-terminal fragment of CTPR3[1_3] protein variant (amino acids 105-120). The Figure shows the chromatographic profile revealed at 210 nm. The insert shows mass spectrum of the chromatographic peak at 9.83 min. Experimental mass (1687.6 Da) is in agreement with the theoretical one (1687.8 Da).

The experimental conditions for the NCL reaction were screened by comparing the yield of full-length protein obtained using different peptide molar excess (2, 5 and 10 fold), thioester protein concentration (0.25 mM, 0.5 mM and 1 mM), pH values (7.2, 7.5 and 7.8) and temperature (room temperature and 4°C), always in denaturing conditions. Moreover, in order to speed up the reaction, we also evaluated the effects of adding a thiol (MESNA) and a reducing agent (TCEP) to the reaction mixture. The best yield of full-length protein was obtained performing NCL at pH 7.5 in 100 mM phosphate buffer, 4 M GdnHCl,1 mM EDTA, 50 mM MESNA and 15 mM TCEP, and employing a ten-fold molar excess of peptide. NCL reaction was performed at 4°C to avoid fluorophore damage. The ligation reaction was checked by LC-MS and allowed to proceed until the disappearance of the thioester protein peak (~4 days at 4°C). In these conditions, ~70% of the thioester protein reacted with the peptide affording the full-length protein (Fig. 26). The remaining portion of thioester protein underwent hydrolysis, becoming unable to react with the peptide. After NCL with the Cys-peptide, reaction mixture was dialyzed and lyophilized. Protein was then incubated with TCEP to reduce Cys residue in position 105, purified to homogeneity by reverse-phase HPLC and lyophilized.

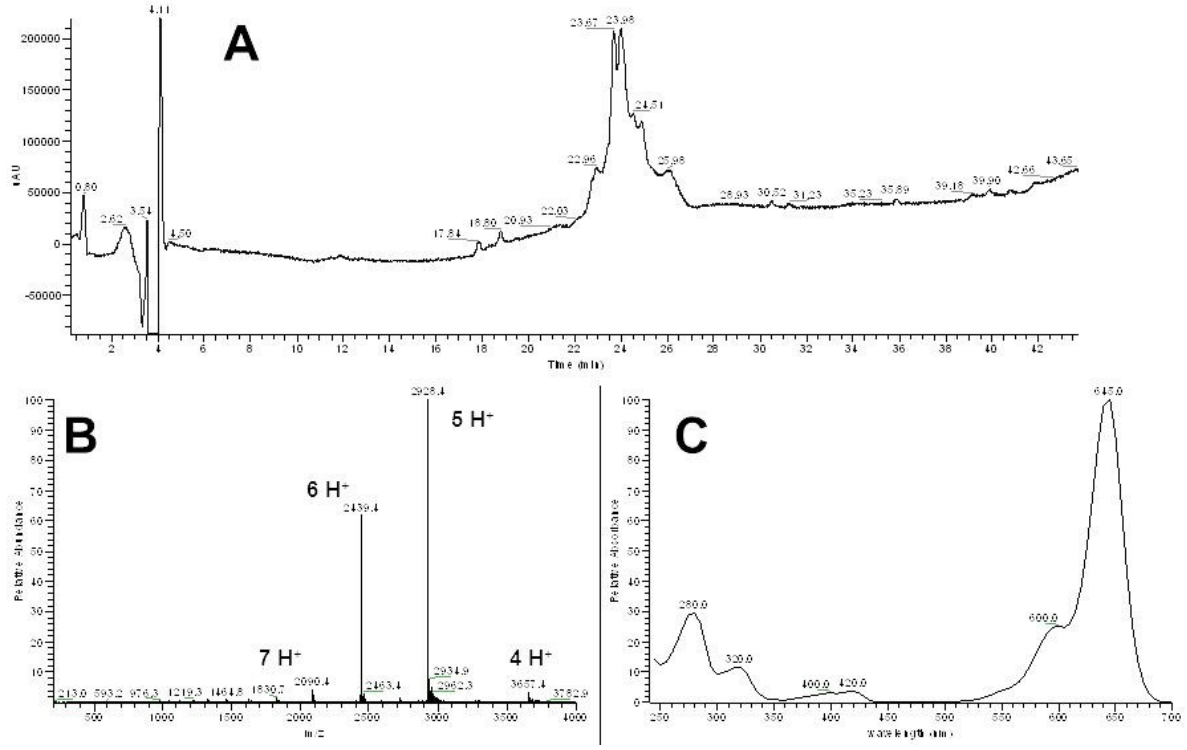

Figure 26 LC-MS analysis of the acceptor labeled CTPR3[1_3] protein obtained from NCL reaction. **A)** Chromatographic profile revealed at 210 nm. **B)** Mass spectrum of the chromatographic peaks at 23.67 min and 23.98 min corresponding to the full-length CTPR3[1_3] protein labeled with ATTO 647N-maleimide. Experimental mass (14637.0 Da) is in agreement with the theoretical one (14634.8 Da). **C)** Absorption spectrum of the chromatographic peaks eluted at 23.67 min and 23.98 min.

Second labeling reaction

Mono-labeled CTPR3[1_3] protein variant was resuspended in phosphate buffer 20 mM pH 7.2 and subjected to the second labeling reaction, performed with the donor dye ATTO 488-maleimide. Labeling reaction was carried out in the same experimental conditions adopted for the first labeling and, as verified by LC-MS analysis, it was completed within 2 h at 4°C. Deconvolution of the ESI spectrum gave a mass increment of 712 Da (Fig. 27B), as expected from the addition of a molecule of ATTO 488-maleimide. Absorption spectrum (Fig. 27C) revealed the appearance of an additional maximum, according to the ATTO 488 spectroscopic properties. After reaction occurred, to avoid multiple and unspecific labeling, exceeding fluorophore ATTO 488 was quenched with MESNA, the protein was dialyzed and lyophilized. Finally, doubly-labeled protein was reduced with TCEP, purified by reverse-phase HPLC and lyophilized (Fig. 27).

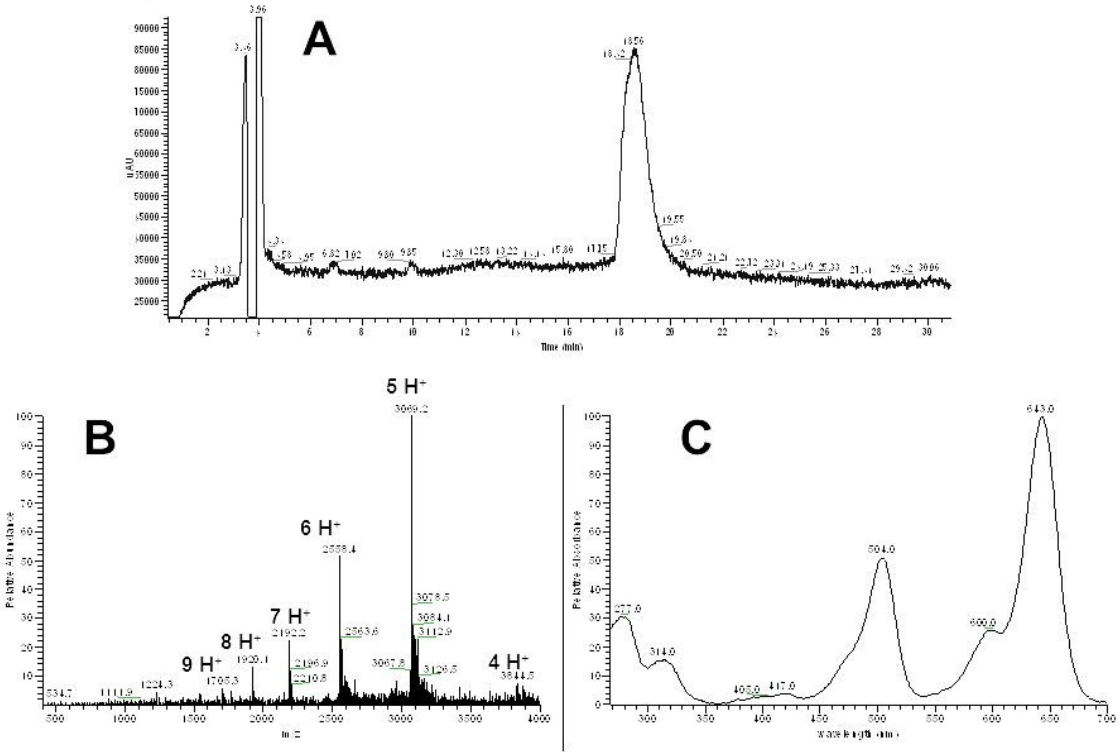

Figure 27 LC-MS analysis of the doubly-labeled CTPR3[1_3] variant. **A)** Chromatographic profile revealed by registering absorbance at 210 nm. **B)** Mass spectrum of the chromatographic peak at 18.56 min. Experimental mass of the species (15344.7 Da) is in agreement with the theoretical one (15346.8 Da). **C)** Absorption spectrum of the chromatographic peak eluted at 18.56 min.

CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C] variants preparation

Doubly-labeled CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C] constructs (schematically shown in Fig. 17C, #2, #3 and #4) were synthesized through the approach denoted as strategy B in Fig. 16. Due to the similarity of the procedures employed, the preparation of these three constructs will be discussed simultaneously in the following paragraphs.

CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C] variants design

CTPR3[N_C], CTPR3[1_C] and CTPR3[2_C] variants share the position of the second probe, located at C-terminus of the protein, while the first probe is located in progressively more internal protein regions, going from the N-terminus to the central repeat. The sites of labeling in these three variants were selected basing on the same guidelines adopted for the CTPR3[1_3] variant design. Therefore, even in these constructs labeling on internal regions is performed on the Asn residue of the loops showing the less global propensity value for that position, avoiding protein structure perturbation. Labeling at N- or C- terminal regions is performed, instead, on residues inserted as extra-sequence amino acids, avoiding the introduction of mutations into the N-Cap and the C-terminal solvating extra-helix, which are known to be key elements of the CTPR protein stability [Main, E.R.G. *et al.* 2003]. Thus, in CTPR3[1_C] variant (Fig. 17C, #2) the first probe is located in position 36, on the more convenient Asn residue within the loop of the first repeat, while the second probe is placed at C-terminus (position 121). CTPR3[2_C] construct (Fig. 17C, #3) harbors the first probe in position 70, within the loop region of the second repeat, while second probe is placed at C-terminus, in position 121. CTPR3[N_C] (Fig. 17C, #4) bears the first fluorophores at N- terminus (position -1), while the second at the C-terminus (position 121).

Cloning procedure

The cloning strategy adopted for the preparation of CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C] constructs is schematized in Fig. 28.
For CTPR3[1_C] variant preparation, the gene coding the full CTPR3 protein (amino acids 1-120) and containing the mutation Asn36Cys (site of first labeling) was amplified from pProEXHTa-*ctpr3*(Asn36Cys) plasmid, whose production by cassette-mutagenesis has been previously described in the section dedicated to CTPR3[1_3]. Instead, for CTPR3[2_C] variant preparation, in which the first label should be placed in position 70, the plasmid pProEXHTa-*ctpr3* was newly mutagenized in order to introduce the mutation Asn70Cys, which is located in the second repeat of the protein. The same mutagenic approach used for the introduction of the mutation Asn36Cys was adopted, performing cassette-mutagenesis on the pProEXHTa-*ctpr3* plasmid. In this case, the plasmid was digested with *Sac*I and *Hind*III restriction enzymes. The excised cassette was mutagenized by PCR, newly digested with *Sac*I and *Hind*III and subcloned into pProEXHTa-*ctpr3.* The mutagenic plasmid obtained (pProEXHTa-*ctpr3*(Asn70Cys)) was used as template in a PCR reaction to amplify the gene coding CTPR3 including the mutation Asn70Cys (amino acids 1-120). Finally, for the preparation of the variant CTPR3[N_C], the gene coding full CTPR3 protein with an N-terminal extra Cys was obtained through a PCR reaction performed

on the wild-type pProEXHTa-*ctpr3* plasmid, using a mutagenic forward primer containing an extra Cys codon.
As well as previously described for CTPR3[1_3], the genes coding CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C] variants were amplified from their respective templates using forward primers adding *Nco*I restriction site and reverse primers generating a region of overlap with *Mxe* GyrA intein coding gene. *Mxe* GyrA intein gene was amplified from pTXB1 plasmid using forward primers generating a region of complementarity with the 3' end of the genes coding the CTPR3 variants. Then, the genes were fused by Overlapping Extension PCR reaction (Fig. 29). Each chimeric gene was digested with *Nco*I and *Eco*RI restriction enzymes and cloned into the pETM13 vector, which allowed the expression of C-terminal His_6 tagged fusion proteins. All recombinant plasmids were verified by sequencing and used to transform bacterial cells for the expression of the fusion proteins.

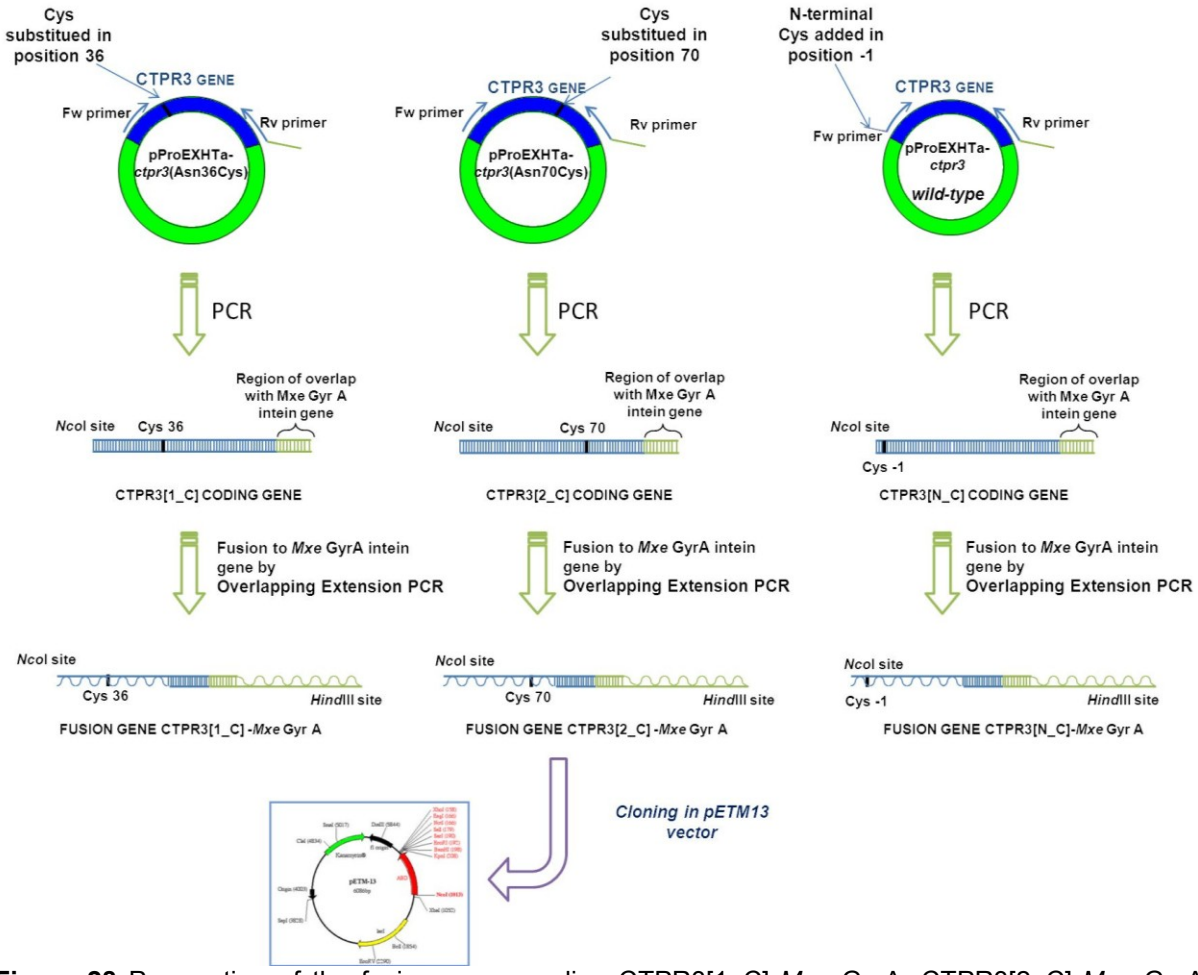

Figure 28 Preparation of the fusion genes coding CTPR3[1_C]-*Mxe* GyrA, CTPR3[2_C]-*Mxe* GyrA and CTPR3[N_C]-*Mxe* GyrA and cloning into pETM13. The gene coding CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C] were obtained through PCR reaction carried out respectively on pProEXHTa-*ctpr3*(Asn36Cys), pProEXHTa-*ctpr3*(Asn70Cys) and wild-type pProEXHTa-*ctpr3*. Construction of the chimeric genes was then performed by Overlapping Extension PCR, exploiting the complementary region introduced into the two components of the fusion gene. Genes were finally cloned in pETM13 vector between *Nco*I and *Hind*III restriction sites.

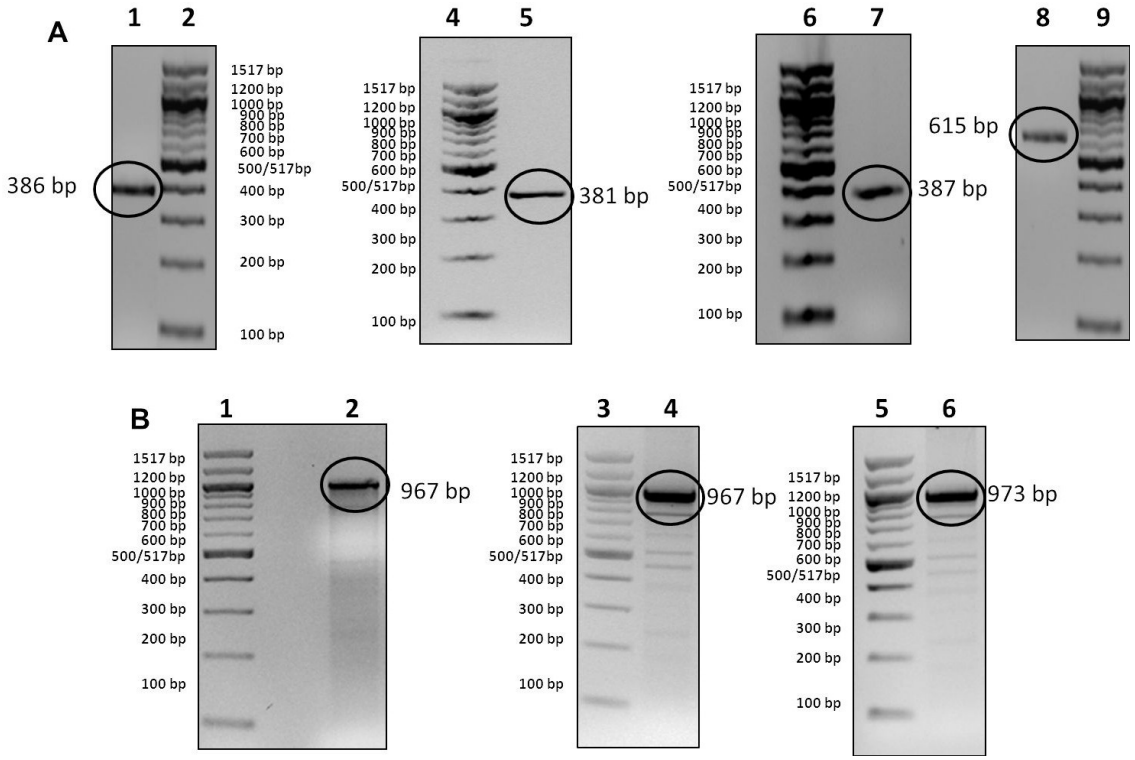

Figure 29 Analysis by agarose gel electrophoresis of: **A)** genes coding CTPR3[1_C] (*lane 1),* CTPR3[2_C] *(lane 5)*, CTPR3[N_C] *(lane 7)* and *Mxe* GyrA intein *(lane 8)* obtained from PCR reactions performed respectively on pProEXHTa-*ctpr3*(Asn36Cys), pProEXHTa-*ctpr3*(Asn70Cys), pProEXHTa-*ctpr3* wild-type and pTXB1; in each gel is also loaded a molecular weight DNA marker (ladder 100 bp) (*lane 1, 4, 6, 9*). **B)** Overlapping Extension PCR reactions between CTPR3[1_C] and *Mxe* GyrA intein *(lane 2)*, CTPR3[2_C] and *Mxe* GyrA intein *(lane 4)*, CTPR3[N_C] and *Mxe* GyrA intein *(lane 6)*. The black circles highlight bands corresponding to the fusion genes. Ladder 100 bp is also loaded in each agarose gel (*lane 1, 3, 6*).

Expression

Expression of the fusion proteins CTPR3[1_C]-*Mxe* GyrA intein, CTPR3[2_C]-*Mxe* GyrA intein and CTPR3[N_C]-*Mxe* GyrA intein was performed in the same experimental conditions selected for the expression of CTPR3[1_3](fragment 1-104)-*Mxe* GyrA intein (expression was inducted with 0.5 mM IPTG and performed at 22°C overnight). All fusion proteins accumulate in the soluble fraction of *E. coli* BL21-(DE3) bacterial cells and exhibit comparable expression levels (Fig. 30). Even in these cases, no splicing *in vivo* of the intein was observed during the expression, as a Gly residue is located at the splicing junction in all the three variants and such amino acid is not included among the amino acids causing *in vivo* premature cleavage of the intein [Southworth, M.W. *et al.* 1999].

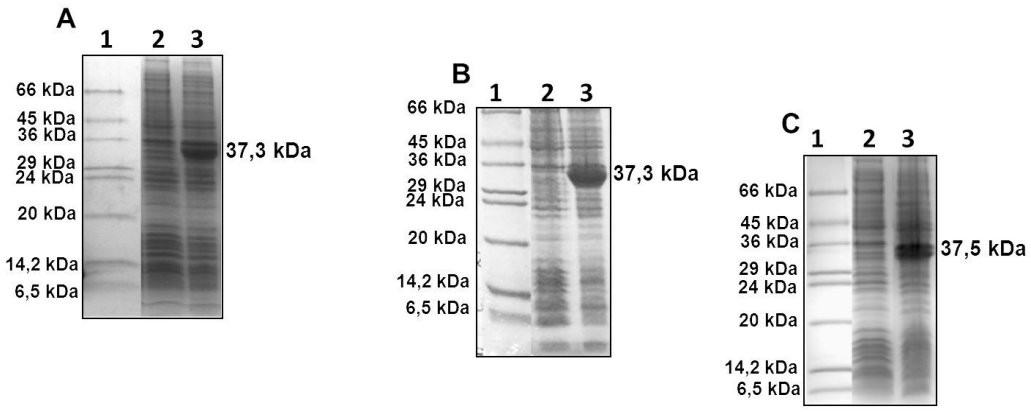

Figure 30 SDS-PAGE analysis of fusion proteins **A)** CTPR3[1_C]-*Mxe* GyrA intein (37.3 kDa), **B)** CTPR3[2_C]-*Mxe* GyrA intein (37.3 kDa) and **C)** CTPR3[N_C]-*Mxe* GyrA intein (37.5 kDa) expressed in *E. coli* BL21-(DE3). Each 15% polyacrylamide gel shows uninduced (*lane 2*) and induced (*lane 3*) cells samples. In each gel is also loaded a molecular weight marker (*lane 1*). All the SDS-PAGE analyses were performed using loading buffer without β-mercaptoethanol, to avoid splicing of the fusion proteins.

Purification

Fusion proteins purification was accomplished as described for CTPR3[1_3] variant. Therefore, all the His_6-tagged fusion proteins were isolated by affinity chromatography, following the same procedure previously discussed (Fig. 31).

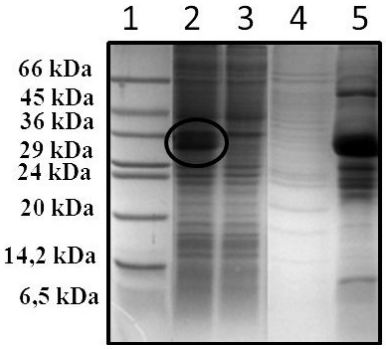

Figure 31 SDS-PAGE analysis of fusion protein CTPR3[N_C]-*Mxe* GyrA intein (37.5 kDa) after purification by affinity chromatography. Molecular weight marker (*lane 1*); soluble fraction of induced *E. coli* BL21-(DE3) cells (*lane 2*), in the black circle is highlighted the fusion protein band; unbound proteins eluted from column (*lane 3*); column wash with buffer A (*lane 4*); fusion protein eluted with buffer B (*lane 5*). Similar purification profiles were obtained for CTPR3[1_C]-*Mxe* GyrA intein and CTPR3[2_C]-*Mxe* GyrA intein fusion proteins purification (data not shown). SDS-PAGE analysis was performed using a loading buffer without β-mercaptoethanol, in order to avoid splicing of the fusion protein.

After isolation of the fusion proteins, C-terminal thioester derivative of each CTPR3 variant was obtained through intein thiolysis, reacting the intein-tagged proteins eluted from the resin with the thiol MESNA. Intein splicing was accomplished in the same experimental conditions adopted for CTPR3[1_3] variant. However, in this case an almost complete cleavage of the intein was observed after 6 days of incubation with 50 mM MESNA at 4°C, as observed following intein splicing reaction for the fusion proteins CTPR3[N_C]-*Mxe* GyrA intein by time-course SDS-PAGE analysis (Fig. 32). Even if Gly residue in the splicing junction is reported to promote splicing

reaction with similar efficiency than Asn, in this case the splicing proceeded faster than in CTPR3[1_3]. In fact, in the same conditions and after the same incubation time, only 50% of fusion protein CTPR3[1_3](fragment 1-104)-*Mxe* Gyr A intein was observed undergoing splicing, because of the aggregation or misfolding of part of the fusion protein (compare Fig. 22 with Fig. 31). In this case, instead, all the preparation of fusion protein was well folded and underwent splicing, so that, globally, splicing reaction appeared to proceed with higher rate.

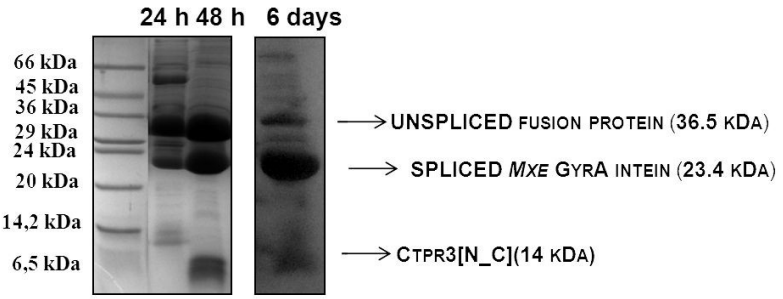

Figure 32 Time-course analysis by SDS-PAGE of the splicing reaction. CTPR3[N_C]-*Mxe* GyrA intein after 24 h, 48 h and 6 days of incubation with MESNA 50 mM. Similar results were observed for the other two fusion proteins, CTPR3[1_C]- *Mxe* GyrA intein and CTPR3[2_C]-*Mxe* GyrA intein, harboring a Gly residue at the splicing junction.

After splicing reaction occurred, reaction mixture was dialyzed against 5 mM phosphate buffer pH 6.5, lyophilized and purified by reverse-phase HPLC to isolate thioester proteins from intein and residual unspliced fusion protein. The thioester proteins were obtained in good purity (Fig. 33, 34 and 35). The identity of the thioester proteins was verified by ESI mass spectrometry and found in agreement with theoretical values (Fig. 33B, 34B and 35B). The final yield of thioester proteins was, on average, of 20 mg from 1 L of culture.

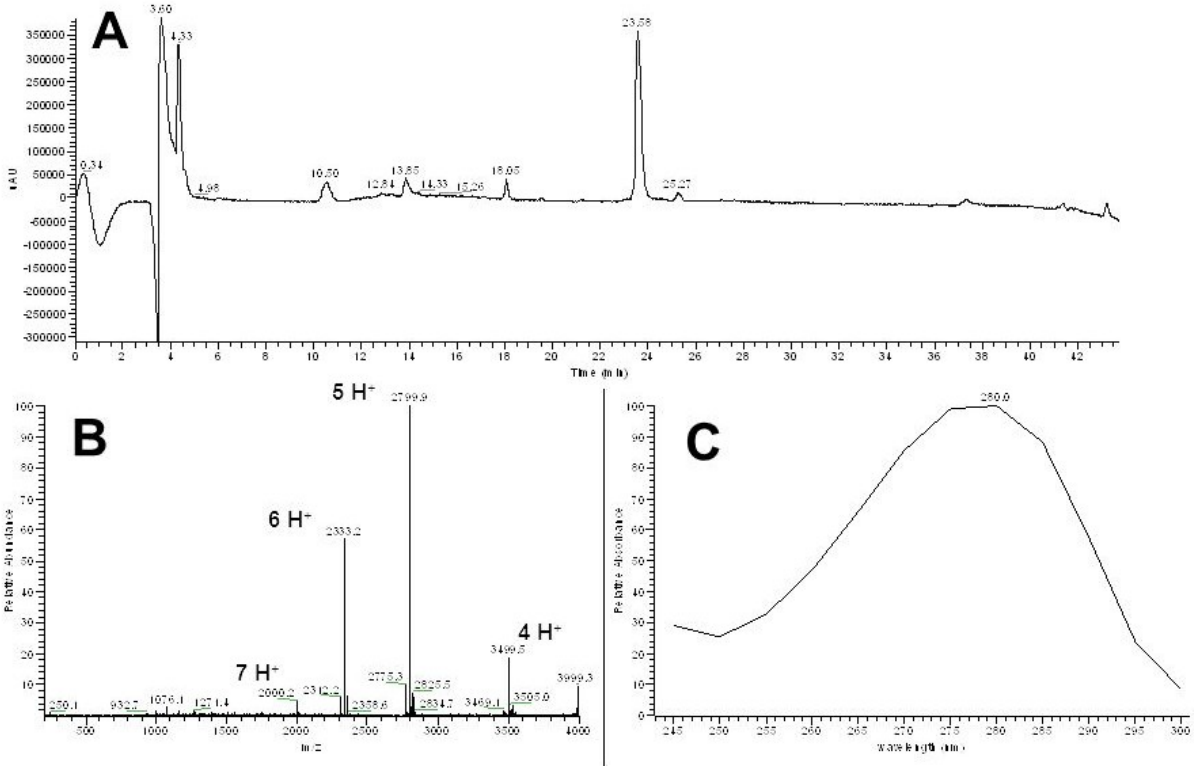

Figure 33 LC-MS analysis of the thioester protein CTPR3[1_C] after HPLC purification. **A)** chromatographic profile revealed at 210 nm. **B)** Mass spectrum of the chromatographic peak revealed at 23.58 min. Experimental mass (13999.5 Da) is in agreement with the theoretical one (14000.9 Da). **C)** Absorption spectrum of the of the chromatographic peak at 23.58 min.

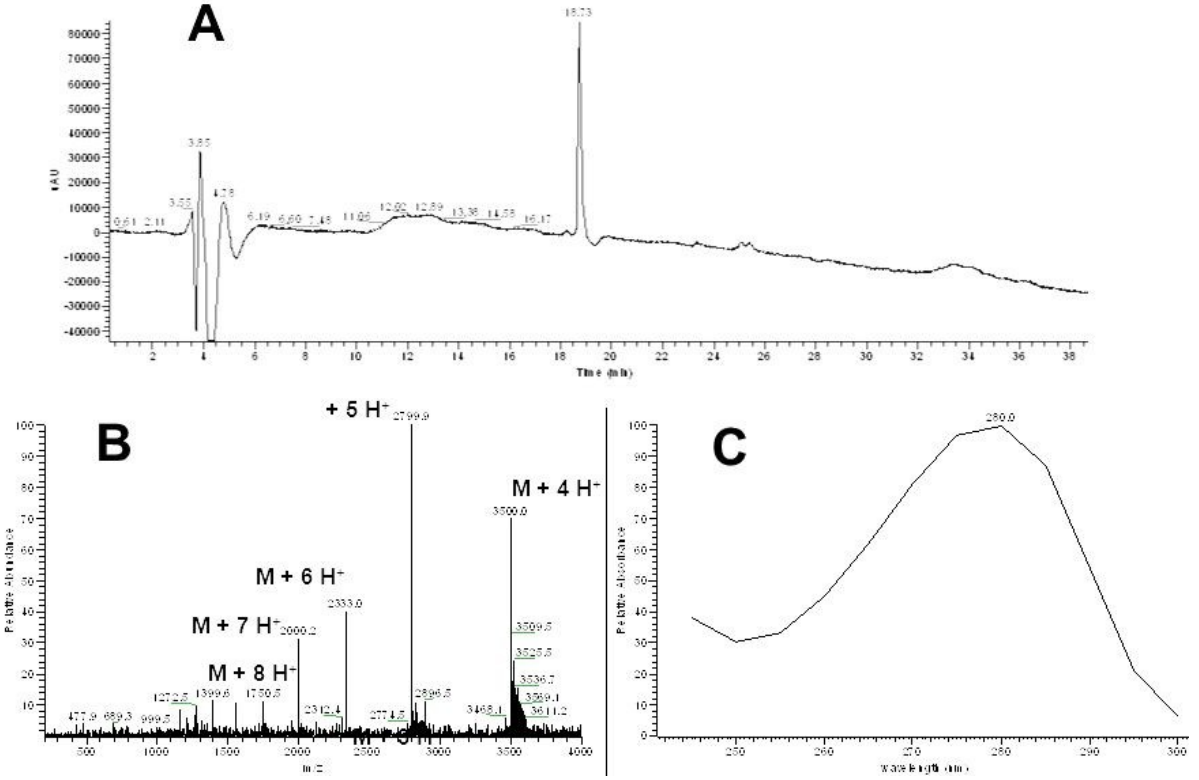

Figure 34 LC-MS analysis of the thioester protein CTPR3[2_C] after HPLC purification. **A)** chromatographic profile revealed at 210 nm. **B)** Mass spectrum of the chromatographic peak revealed at 18.73 min. Experimental mass (13999.4 Da) is in agreement with the theoretical one (14000.9 Da). **C)** Absorption spectrum of the of the chromatographic peak at 18.73 min.

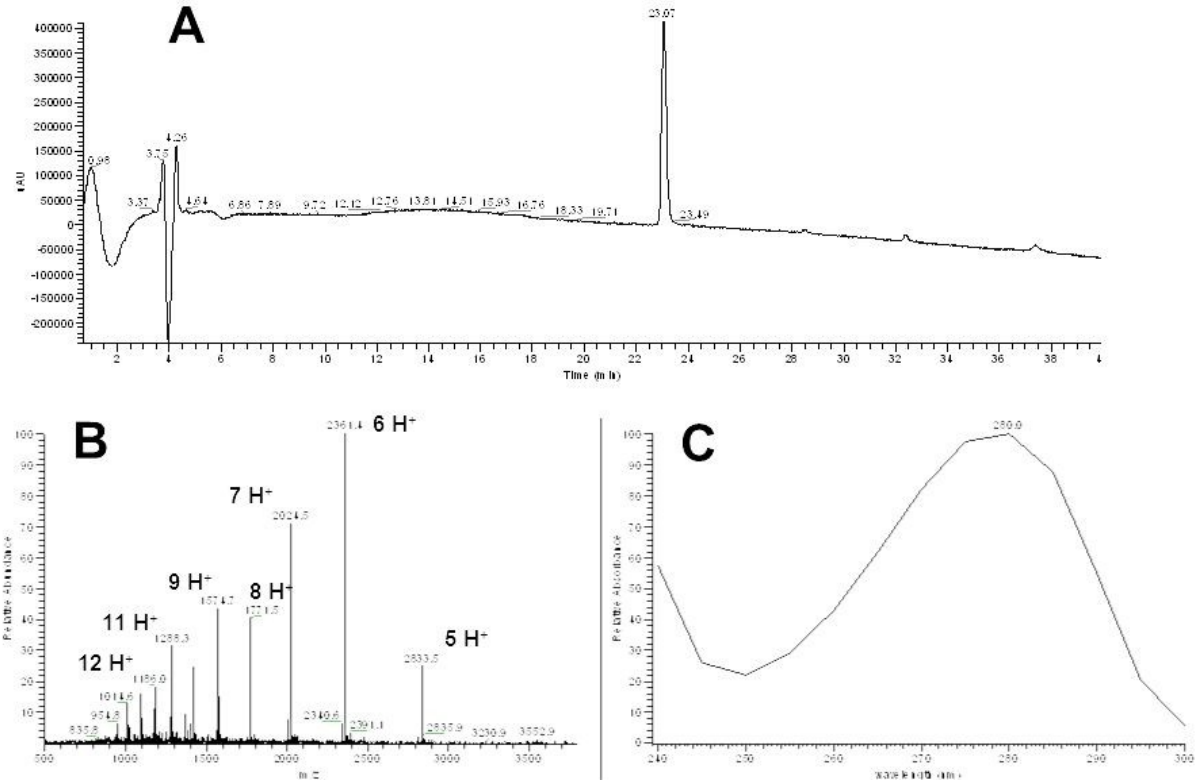

Figure 35 LC-MS analysis of the thioester protein CTPR3[N_C] after HPLC purification. **A)** chromatographic profile revealed at 210 nm. **B)** Mass spectrum of the chromatographic peak revealed at 23.07 min. Experimental mass (14166.8 Da) is in agreement with the theoretical one (14172.18 Da). **C)** Absorption spectrum of the of the chromatographic peak at 23.07 min.

First labeling reaction

Thioester derivatives of CTPR3[1_C], CTPR3[2_C] and CTPR3[N_C] proteins were site-specifically labeled on their Cys residue (respectively located in position 36, 70 and N-terminal) with the acceptor fluorophore ATTO 647N-maleimide. The reaction was performed as previously described for the labeling of the fragment 1-104 of CTPR3[1_3], in 20 mM phosphate buffer pH 7.2, at 4°C on the dark, using a two-fold molar excess of ATTO 647N-maleimide, and occurred with a similar yield (above 90% within 2 h) (Fig. 36). Exceeding fluorophore was quenched with the thiol MESNA and then reaction mixture was extensively dialyzed against 20 mM phosphate buffer pH 7.2. This step of dialysis ensured restoring of a pH value of 7.2, at which subsequent NCL with L-Cysteine was performed.

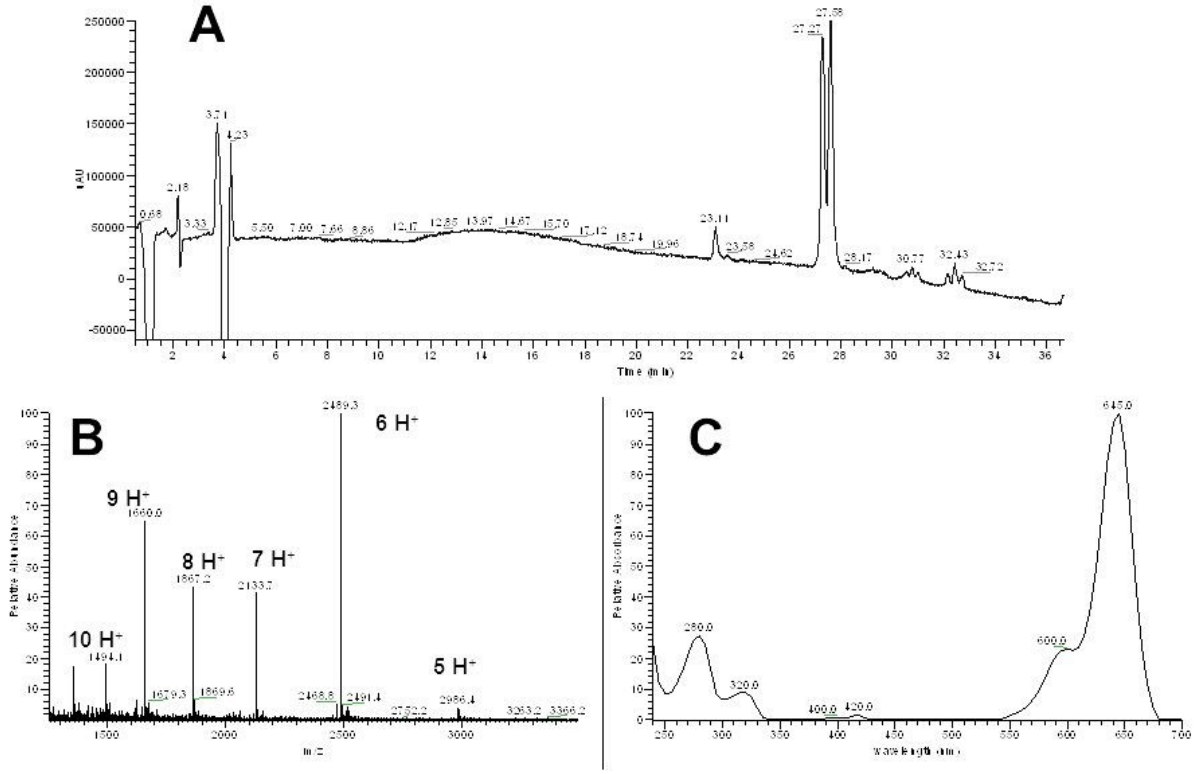

Figure 36 LC-MS analysis of the labeling reaction of the thioester protein CTPR3[N_C] with ATTO 647N-maleimide. **A)** Chromatographic profile revealed at 210 nm. **B)** Mass spectrum of the chromatographic peaks at 27.27 min and 27.58 min. Experimental mass (14934.8 Da) is in agreement with the theoretical one (14940.2 Da). **C)** Absorption spectrum of the thioester protein peaks at 27.27 and 27.58 min. Similar data were obtained for thioester protein CTPR3[1_C] and CTPR3[2_C] (data not shown).

Native Chemical Ligation with L-Cys

Mono-labeled thioester proteins were reacted with L-Cys through NCL, in order to introduce an additional Cys residue at C-terminal position. The best NCL conditions were selected by comparing reaction yield at different pH value (in the range 7.0-8.0), temperature (room temperature and 4°C), L-Cys excess (in the range 10-100 fold) and protein concentration (in the range 0.1-1.0 mM) in presence or absence of a denaturing agent (GdnHCl). NCL with L-Cys occurred with high yield (above 90%) in all the conditions tested. We decided to perform the NCL reaction in native conditions, using an aqueous buffer at pH 7.2 and a 20 fold molar excess of L-Cys. Reaction went to completion after an incubation overnight at 4°C, under stirring (Fig. 37, 38, 39). Ligation reaction was performed at 4°C to preserve ATTO 647N integrity. After NCL occurred, reaction mixtures were lyophilized and purified by reverse-phase HPLC. Before performing the chromatography, samples were incubated with TCEP to reduce C-terminal Cys residue.

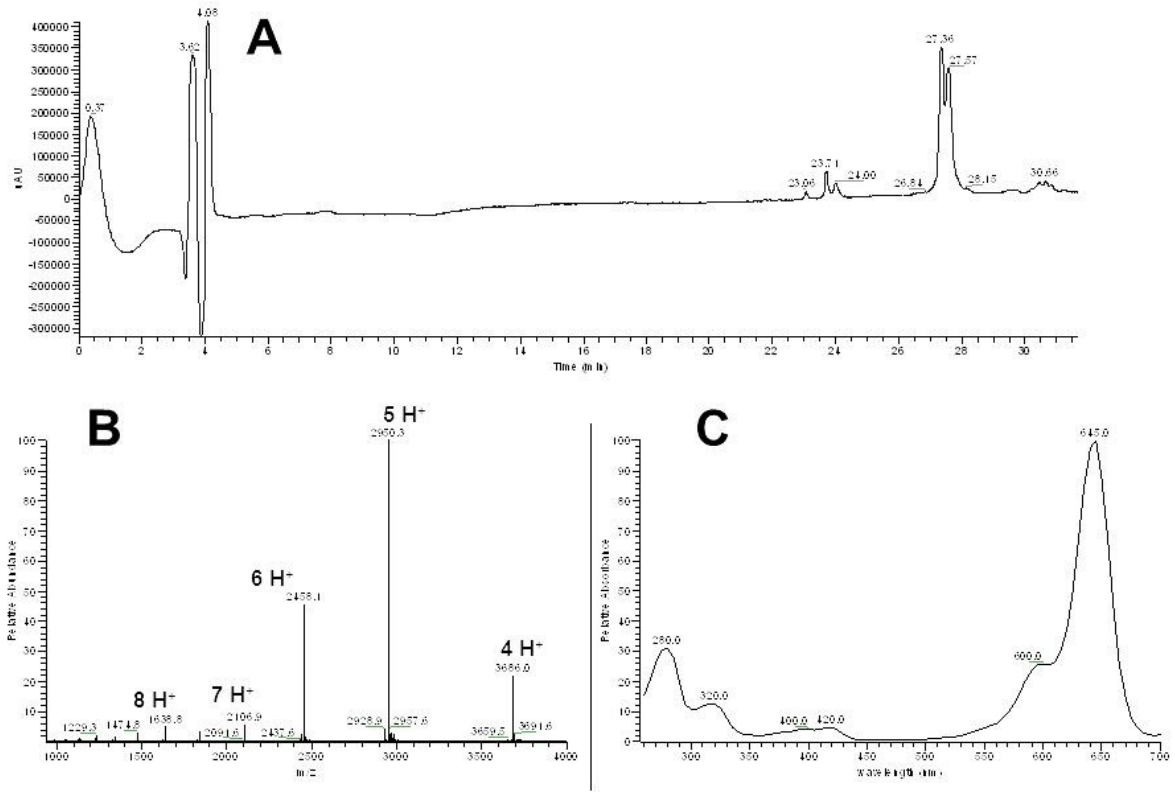

Figure 37 LC-MS analysis of CTPR3[1_C] variant obtained from NCL with L-Cys. **A)** Chromatographic profile revealed by at 210 nm. **B)** Mass spectrum of the chromatographic peaks at 27.36 min and 27.57 min. Experimental mass (14749.4 Da) is in agreement with the theoretical one (14748.9 Da). **C)** Absorption spectrum of the protein peaks at 23.36 min and 27.57 min. Sample was incubated with TCEP before LC-MS analysis to reduce C-terminal Cys.

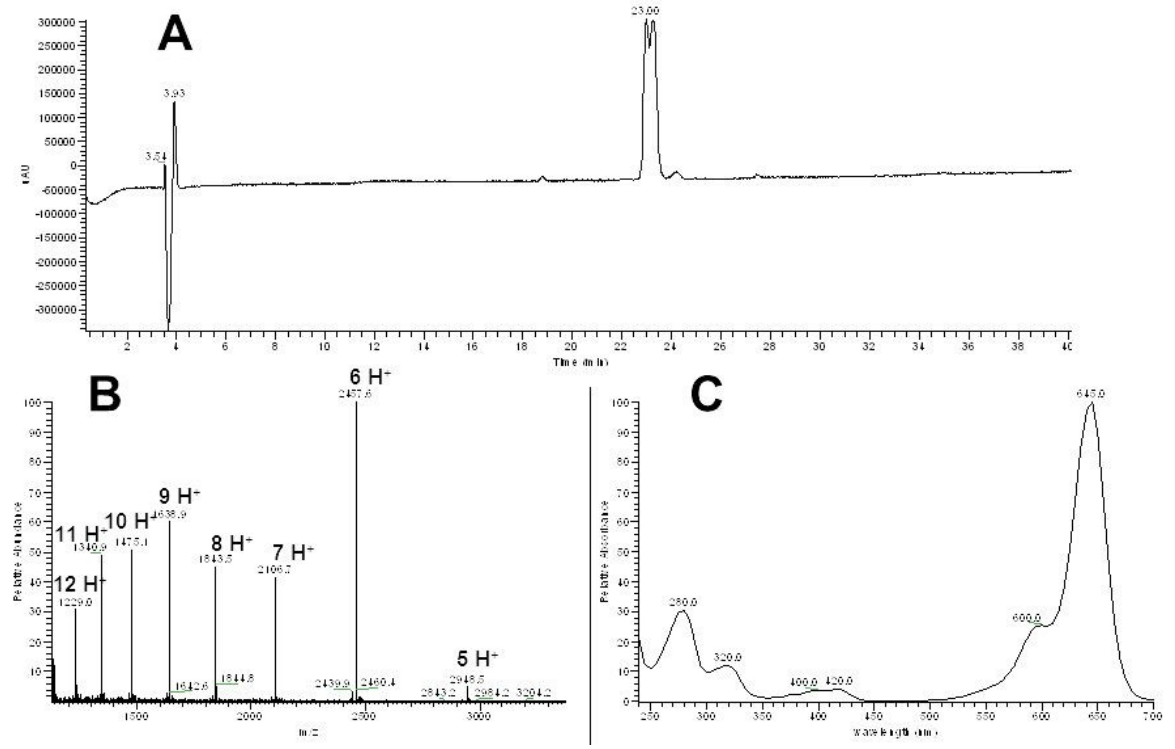

Figure 38 LC-MS analysis of CTPR3[2_C] variant obtained from NCL with L-Cys. **A)** Chromatographic profile revealed by at 210 nm. **B)** Mass spectrum of the chromatographic peaks at 27.36 min and 27.57 min. Experimental mass (14741.9 Da) is in agreement with the theoretical one (14748.9 Da). **C)** Absorption spectrum of the protein peaks at 23.0 min. Sample was incubated with TCEP before LC-MS analysis to reduce C-terminal Cys.

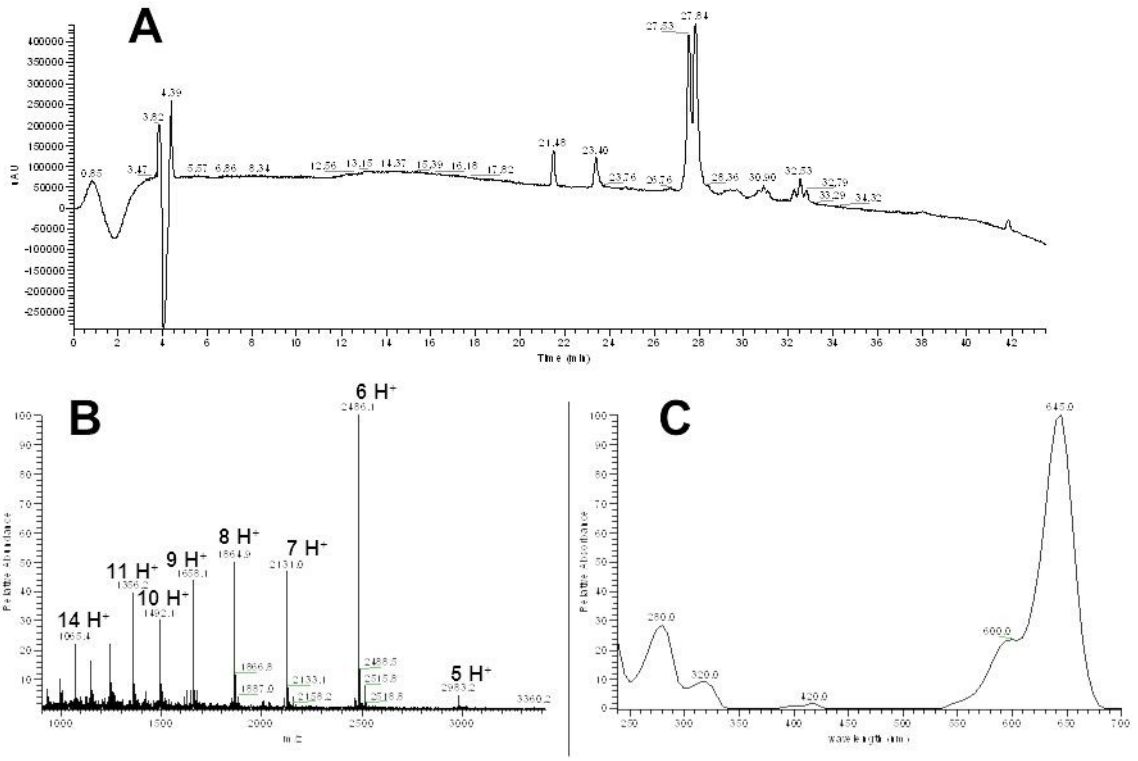

Figure 39 LC-MS analysis of CTPR3[N_C] variant obtained from NCL with L-Cys. **A)** Chromatographic profile revealed by at 210 nm. **B)** Mass spectrum of the chromatographic peaks at 27.53 min and 27.84 min. Experimental mass (14916.2 Da) is in agreement with the theoretical one (14920.2 Da). **C)** Absorption spectrum of the protein peaks at 23.0 min. Sample was incubated with TCEP before LC-MS analysis to reduce C-terminal Cys.

Second labeling reaction

All the obtained mono-labeled variants were resuspended in 20 mM phosphate buffer pH 7.2 and subjected to the second labeling reaction, performed with the donor dye ATTO 488-maleimide, as also described for CTPR3[1_3]. Doubly-labeled proteins were purified by reverse-phase HPLC and lyophilized (Fig. 40, 41 and 42). The minor peaks in the chromatographic profiles reported in Fig. 40, 41 and 42 (19.03 min, 18.68 min, 17 min respectively) show the m/z signal of the doubly-labeled species but their absorption spectra do not show the absorption maximum at 644 nm of ATTO 647N but only the one of ATTO 488 at 501 nm. This observation indicates that in an aliquot of sample ATTO 647N fluorophore is damaged.

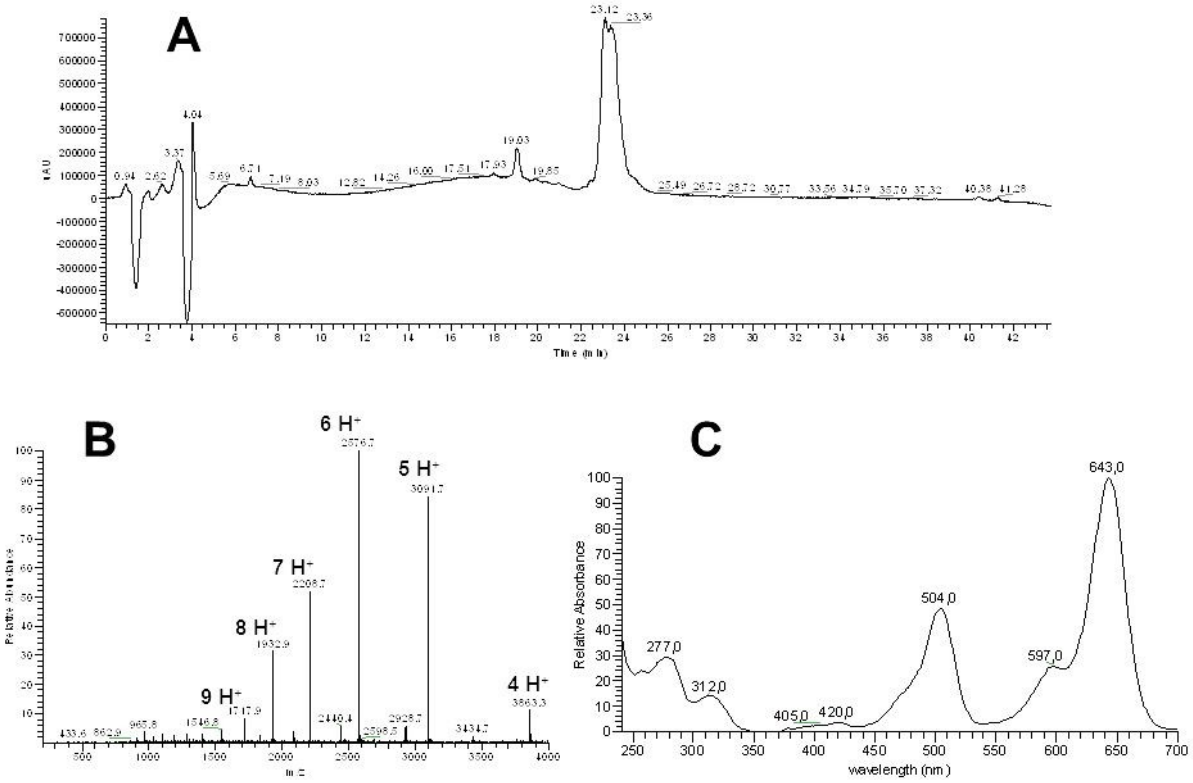

Figure 40 LC-MS analysis of the doubly-labeled CTPR3[1_C]. **A)** Chromatographic profile revealed at 210 nm. **B)** Mass spectrum of the protein peaks at 23.12 min and 23.36 min. Experimental mass (15455.9 Da) is in agreement with the theoretical one (15460.9 Da). **C)** Absorption spectrum of the chromatographic peaks at 23.12 min and 23.36 min.

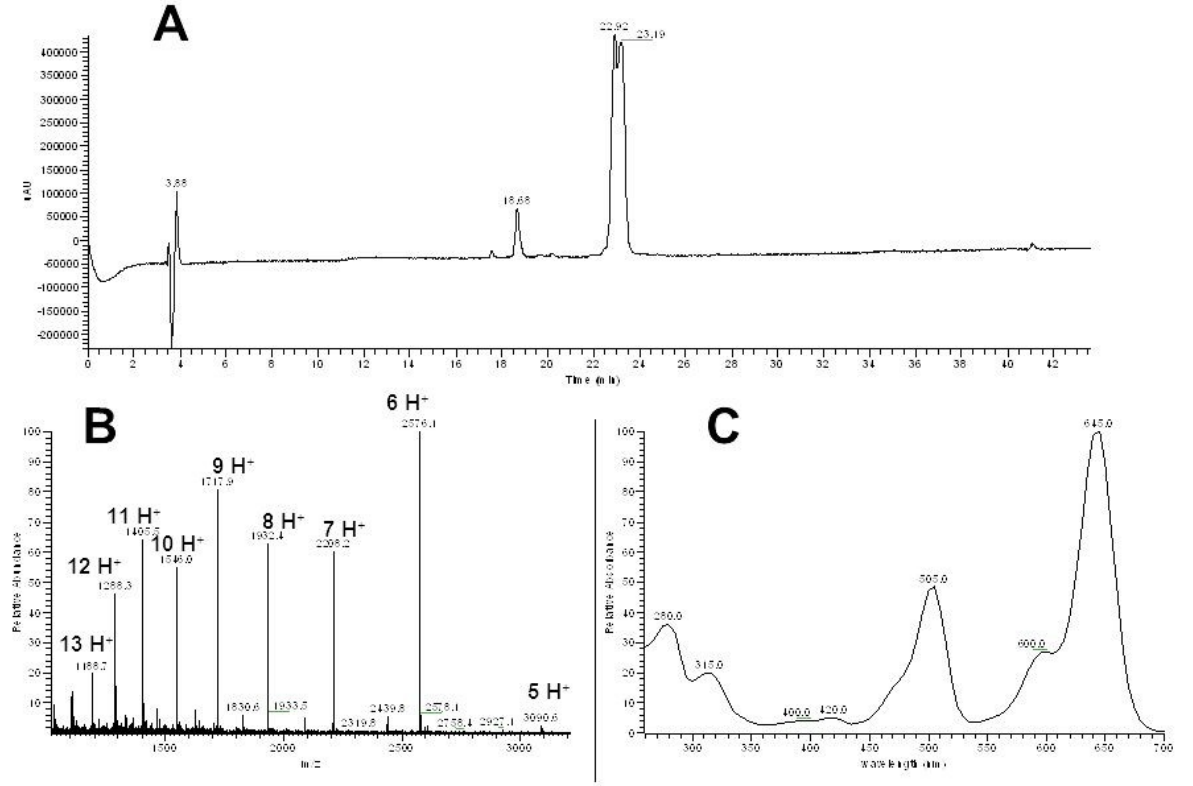

Figure 41 LC-MS analysis of the doubly-labeled CTPR3[2_C]. **A)** Chromatographic profile revealed at 210 nm. **B)** Mass spectrum of the protein peaks at 22.92 min and 23.19 min. Experimental mass (15455.8 Da) is in agreement with the theoretical one (15460.9 Da). **C)** Absorption spectrum of the chromatographic peaks at eluted at 23.12 min and 23.36 min.

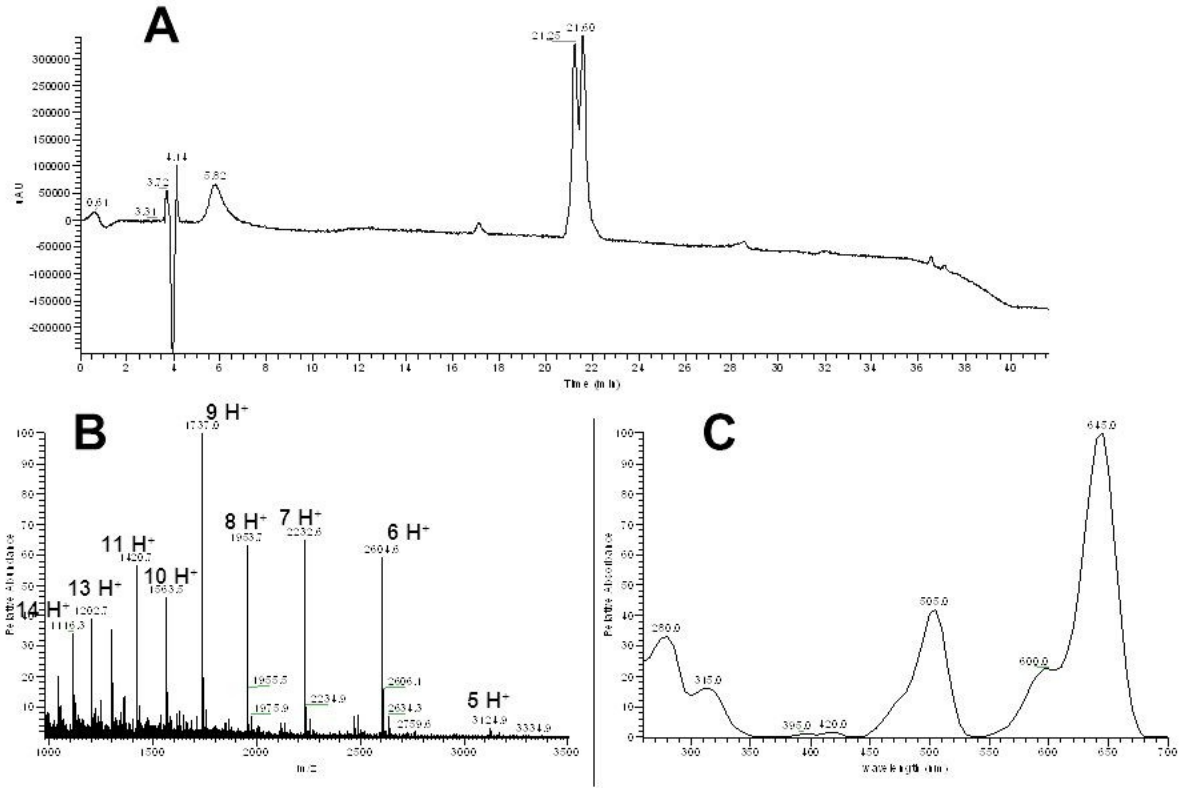

Figure 42 LC-MS analysis of the doubly-labeled CTPR3[N_C]. **A)** Chromatographic profile revealed at 210 nm. **B)** Mass spectrum of the protein peaks at 21025 min and 21.50 min. Experimental mass (15623.2 Da) is in agreement with the theoretical one (15632.2 Da Da). **C)** Absorption spectrum of the chromatographic peaks at eluted at 23.12 min and 23.36 min.

Mono-labeled and unlabeled control constructs preparation

Control constructs harboring only the donor or the acceptor fluorophore or none of them were also prepared using the same approach adopted for the semisynthesis of the doubly-labeled species. Such control constructs were needed to carry out a complete spectroscopic characterization of the doubly-labeled species, in order to assess the effects of the introduction of each fluorophore and of the mutations on the stability of CTPR3 and to perform fluorescence anisotropy studies. All the doubly-labeled, mono-labeled and unlabeled constructs prepared are listed in Table IV. Unlabeled proteins containing the two Cys residues in the same position as the doubly-labeled proteins (Table IV #2-4) were obtained carrying out NCL reactions between unlabeled thioester moieties and L-Cys or Cys-peptide, as described for the preparation of the labeled analogs. Unlabeled CTPR3[N_C] protein was not prepared because this variant does not contain internal mutations but only insertions at N- and C-termini, and we assume it to retain the same stability of the wild-type protein (Table IV #1). Mono-labeled proteins, containing only the first fluorophore of FRET pair ATTO 647N in position N-terminal, 36 or 70 and the second Cys free (Table IV, #6-9), were prepared concomitantly with the preparation of the doubly-labeled variants. The control construct containing only the second fluorophore of the pair, ATTO 488, at C-terminal position (Table IV #5) was prepared expressing the wild-type CTPR3 protein as fusion construct with *Mxe* GyrA intein; thioester CTPR3, which does not contain Cys residues, was reacted in NCL reaction with L-Cys, adding a single Cys residue at the C-terminus. Cys was labeled with ATTO 488 similarly as before described.

Spectroscopic characterization of CTPR3 protein variants

It was showed [Kajander *et al.* 2005; Main, E.R.G. *et al.* 2005; Hagay, T. & Levy, Y. 2008; Javadi, Y. & Main, E.R.G. 2009] that CTPR3 unfolding follows the Ising model and proceeds from the external helices towards the protein interior [Cortajarena, A.L. *et al.* 2008]. This behavior could be verified using the doubly-labeled CTPR3 variants prepared monitoring the unfolding process by CD and FRET. These studies were performed in collaboration with Prof. Lynne Regan and Dr. Aitziber Lopez Cortajarena from the Department of Molecular Biophysics & Biochemistry of Yale University (New Haven, CT-USA).

CD characterization

In order to verify if the semisynthetic CTPR3 variants retained the typical CTPR3 protein folding despite the presence of mutations and fluorophores, all the prepared constructs (doubly-labeled proteins as well as the mono-labeled and unlabeled control constructs, Table IV) were analyzed by far-UV CD spectroscopy. CD spectra, recorded in the 190-260 nm region, clearly demonstrate that all the variants exhibit an overall helical structure, indicating that the introduction of mutations and fluorophores into CTPR3 does not lead to a substantial disruption of the native protein fold (Fig. 43).

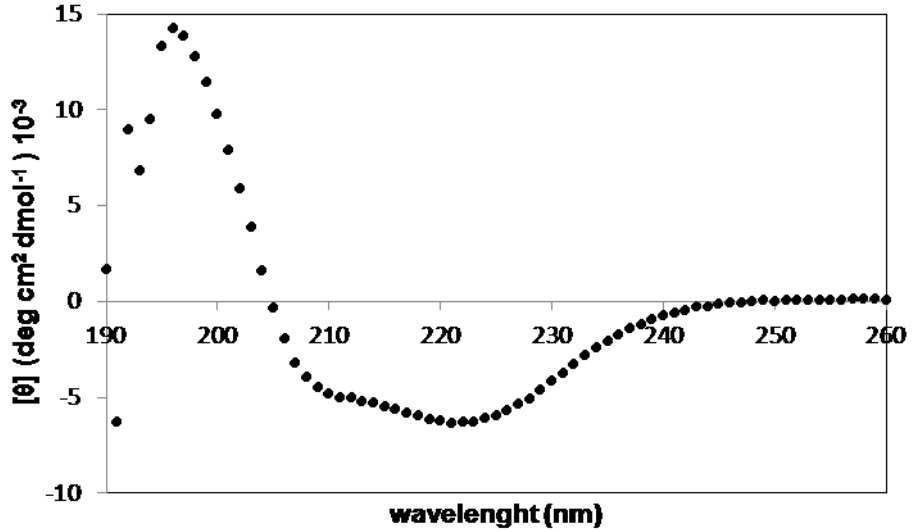

Figure 43 Far-UV CD spectrum of doubly-labeled CTPR3[1_C]. Similar spectra were observed for all the other variants.

Next, we verified if the mutations and the labels introduced into CTPR3 affect protein global stability. This control is needed to compare the results obtained for each construct from the folding studies. To this aim, we carried out heat-denaturation experiments on all the doubly-labeled, mono-labeled and unlabeled CTPR3 variants. The thermal unfolding process of the CTPR3 variants was followed by CD, monitoring ellipticity at 222 nm, a signature of the helical structure, in the temperature range 10-94°C. Protein unfolded fraction at each temperature was estimated assigning 0% and 100% values to the ellipticity recorded respectively at 10°C and 94°C. Table IV shows the Tm values (melting temperature, the temperature at which the protein unfolded fraction is 50%) estimated for the CTPR3 variants.

Entry number	CTPR variant name	Construct type	Mutations	ATTO label position	Tm (°C)
1	CTPR3 WT	wild-type	----	-----	80.0
2	CTPR3[1_C] UL	unlabeled	- Asn36Cys - Cys added at C-terminus	-----	78.5
3	CTPR3[2_C] UL	unlabeled	- Asn70Cys - Cys added at C-terminus	----	78.5
4	CTPR3[1_3] UL	unlabeled	- Asn36Cys - Asn105Cys	----	78.5
5	CTPR3-D	mono-labeled with the DONOR	- Cys added at C-terminus	ATTO 488 at C-terminus	81.1
6	CTPR3[N_C]-A	mono-labeled with the ACCEPTOR	- Cys added at N-terminus - Cys added at C-terminus	ATTO 647N at N-terminus	77.0
7	CTPR3[1_C]-A	mono-labeled with the ACCEPTOR	- Asn36Cys - Cys added at C-terminus	ATTO 647N at position 36	71.5
8	CTPR3[2_C]-A	mono-labeled with the ACCEPTOR	- Asn70Cys - Cys added at C-terminus	ATTO 647N at position 70	75.1
9	CTPR3[1_3]-A	mono-labeled with the ACCEPTOR	- Asn36Cys - Asn105Cys	ATTO 647N at position 36	69.5
10	CTPR3[N_C]-AD	doubly-labeled	- Cys added at N-terminus - Cys added at C-terminus	ATTO 647N at N-terminus ATTO 488 at C-terminus	76.0
11	CTPR3[1_C]-AD	doubly-labeled	- Asn36Cys - Cys added at C-terminus	ATTO 647N at position 36 ATTO 488 at C-terminus	72.5
12	CTPR3[2_C]-AD	doubly-labeled	- Asn70Cys - Cys added at C-terminus	ATTO 647N at position 70 ATTO 488 at C-terminus	76.8
13	CTPR3[1_3]-AD	doubly-labeled	- Asn36Cys - Asn105Cys	ATTO 647N at position 36 ATTO 488 at position 105	65.5

Table IV List of all the CTPR3 variants analyzed by CD. All the Unlabeled constructs are denoted with UL. The mono-labeled constructs carrying only the acceptor ATTO 647N or the donor ATTO 488 are marked with A (=Acceptor) or D (=Donor) respectively. Doubly-labeled proteins are indicated with AD (Acceptor and Donor). The values of the melting temperatures (Tm) are reported in last column.

Fig. 44 shows the denaturation curves of the unlabeled species compared with that observed for CTPR3 wild-type. Fig. 45, 46, 47 and 48 show the denaturation curves respectively obtained for doubly-labeled protein CTPR3[N_C], CTPR3[1_C], CTPR3[2_C] and CTPR3[1_3] in comparison with their mono-labeled and unlabeled counterparts.

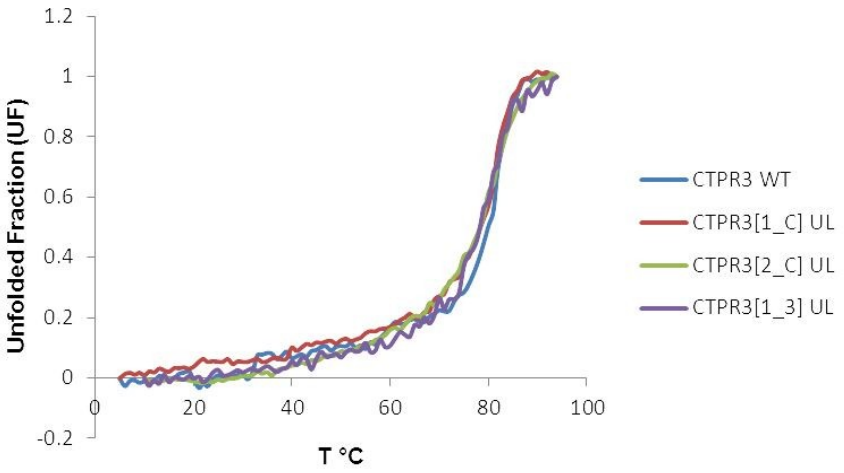

Figure 44 Thermal denaturation curves of CTPR3 unlabeled control constructs (CTPR3 UL) and CTPR3 wild-type (CTPR3 WT).

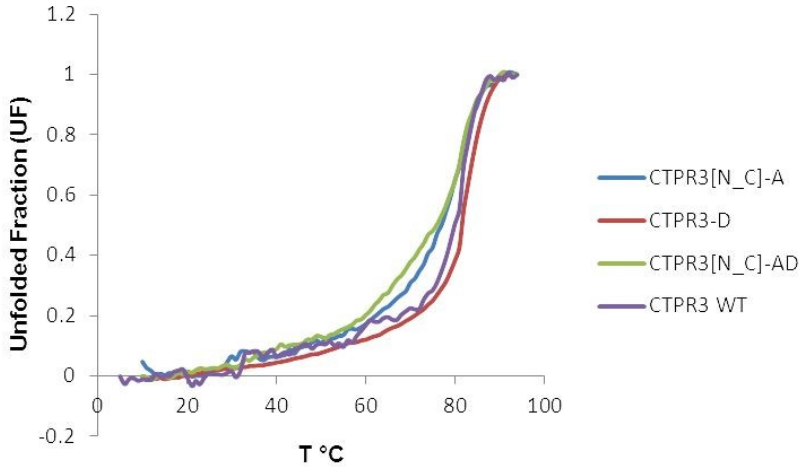

Figure 45 Thermal denaturation curves of doubly-labeled CTPR3[N_C] (CTPR3[N_C]-AD), mono-labeled constructs (CTPR3[N_C]-A and CTPR3-D) and CTPR3 wild-type (CTPR3 WT).

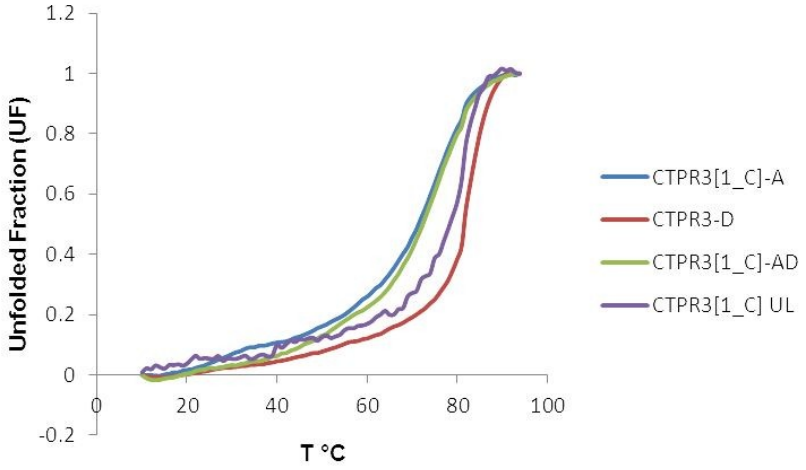

Figure 46 Thermal denaturation curves of doubly-labeled CTPR3[1_C] (CTPR3[1_C]-AD), mono-labeled constructs (CTPR3[1_C]-A and CTPR3-D) and unlabeled variant (CTPR3[1_C] UL).

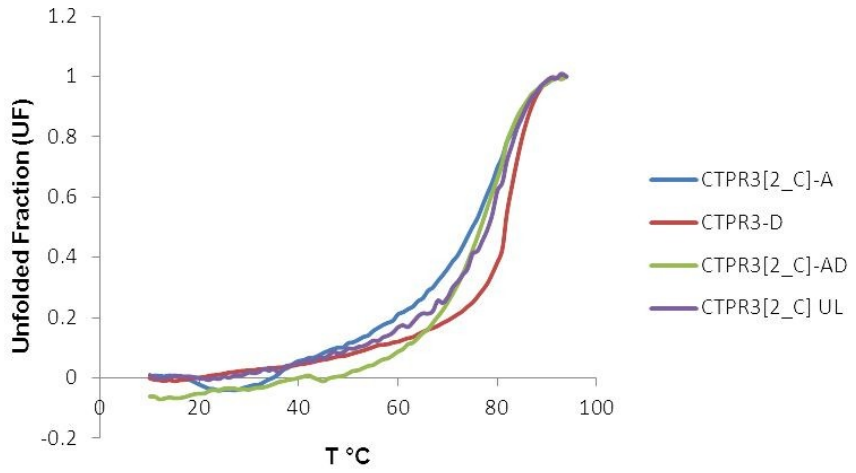

Figure 47 Thermal denaturation curves of doubly-labeled CTPR3[2_C] (CTPR3[2_C]-AD), mono-labeled constructs (CTPR3[2_C]-A and CTPR3-D) and unlabeled variant (CTPR3[2_C] UL).

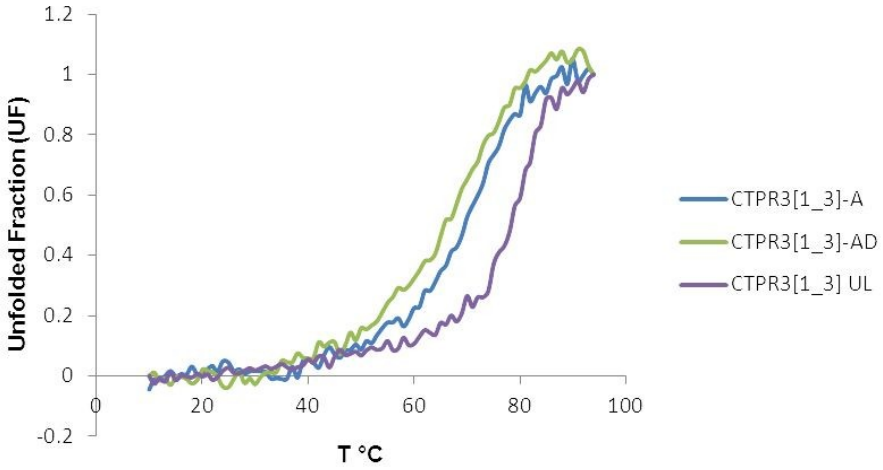

Figure 48 Thermal denaturation curves of doubly-labeled CTPR3[1_3] (CTPR3[1_3]-AD), acceptor mono-labeled construct (CTPR3[1_3]-A) and unlabeled variant (CTPR3[1_3] UL).

All the unlabeled CTPR3 variants (Table IV #2, 3 and 4, Fig. 44), which harbor the two mutations but no fluorophores, exhibit a thermal stability (Tm=78.5°C) comparable with CTPR3 wild-type (Tm=80°C), demonstrating that the mutations introduced into CTPR3 sequence do not affect the protein stability. Similarly, CTPR3[N_C]-A (Table IV #5), carrying only the acceptor label at N-terminus, and CTPR3-D (Table IV #6), harboring only the donor fluorophore at C-terminal position, exhibit Tm values of 77°C and 81°C, which are comparable to those estimated for the unlabeled and wild-type species. These data suggest that labeling on terminal regions does not affect the protein structural stability. On the contrary, CTPR3[1_3]-A (Table IV #7) and CTPR3[1_C]-A (Table IV #9), labeled only with the acceptor fluorophore in position 36, show Tm values slightly lower with respect to those estimated for the unlabeled mutant and the wild-type protein (69.5 and 71.5°C). These data demonstrate that, when placed inside the protein sequence instead than

on the terminal regions, fluorophores slightly destabilize the protein. However, CTPR3[2_C]-A (Table IV #8), which carries the fluorophore ATTO 647N in position 70, shows a Tm value (75.1°C) higher than those observed for CTPR3[1_C]-A and CTPR3[1_3]-A. This observation could be justified considering that in CTPR3[2_C]-A the fluorophore is placed in the protein loop connecting the second and the third repeat, which is a more internal and hence higher stable region respect to the first loop. Thermal stability data obtained for the doubly-labeled species are in agreement with the results achieved for the mono-labeled or unlabeled control constructs. Doubly-labeled CTPR3[N_C]-AD (Table IV #10) exhibits a Tm of 76.0°C, showing only a slightly destabilization according to the presence of the two labels on the terminal regions. Similar results are obtained for CTPR3[2_C]-AD (Table IV #12) (Tm 76.8°C). Contrariwise, Tm values estimated for doubly-labeled CTPR3[1_C]-AD (Table IV #11) and CTPR3[1_3]-AD (Table IV #13) are respectively 72.5°C and 65.5°C, according to the presence of the acceptor dye within the first loop that we assessed to be a destabilizing factor. CTPR3[1_3] shows the lowest stability that could be attributed to the presence of both fluorophores inside the protein sequence.
Globally, CD data suggest that neither the mutations nor the labels introduced into protein sequence affect considerably CTPR3 native fold, which is an essential conditions for the conduction of folding studies on the CTPR3 variants. However, mutations and labels slightly destabilize proteins.

Fluorescence anisotropy measurements

To assess whether the fluorophores are freely rotating when attached to the protein or, conversely, interactions occur between protein and dyes, we carried out fluorescence anisotropy experiments on the mono-labeled CTPR3 protein variants. The anisotropy of these constructs was compared to that exhibited by dyes in their free form. All the CTPR3 variants present similar anisotropy values, independently on the position of dyes, and moreover the values obtained suggest that the dyes are freely rotating when attached to the protein, ensuring that labels do not mediate protein aggregation (Table V).

Protein variant or dye	Anisotropy value
Free acceptor dye (A)	0.137
Free donor dye (D)	0.017
CTPR3[N_C]-A	0.162
CTPR3[1_C]-A	0.167
CTPR3[2_C]-A	0.158
CTPR3[1_3]-A	0.169
CTPR3-D	0.073

Table V Fluorescence anisotropy measurements on mono-labeled control constructs.

Ensemble-FRET studies on Doubly-labeled CTPR3 protein variants

ATTO 488 and ATTO 647N were respectively selected as donor and acceptor dyes because they were predicted to be well suited for probing by FRET the transition from the native to the unfolded state in the doubly-labeled CTPR3 protein variants. In fact, the distances between fluorophores in the native state of the four doubly-labeled constructs are predicted to be 52 Å in CTPR3[N_C], 40.3 Å in CTPR3[1_C], 31.3 Å in CTPR3[2_C] and 28.4 Å in CTPR3[1_3] from CTPR3 x-ray structure. Such distance values are similar to the the R_0 of ATTO 647N/488 pair (49 Å). R_0 of a fluorophores pair is defined as the distance (R) at which FRET efficiency (E) is 0.5 and corresponds to the inflection point of the plot in Fig.49, which reports E for ATTO 647N/488 in function of the distance (R) between the two dyes. Therefore, the increase of the distance between fluorophores occurring upon protein unfolding should result in appreciable variation of FRET efficiency (E) between the two dyes.

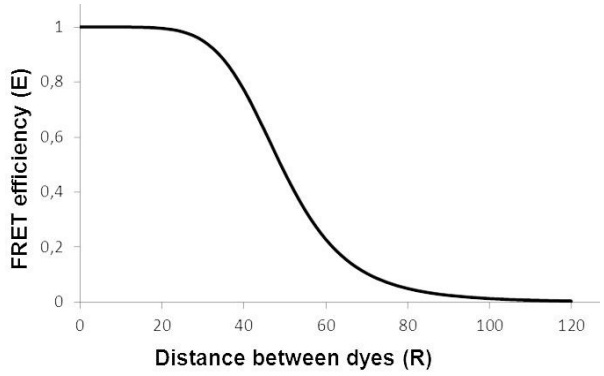

Figure 49 FRET efficiency (E) dependence from distance (R) for ATTO 647N/488 fluorophores pair (R_0 = 49 Å).

We verified the ability of the selected pair, ATTO 647N and ATTO 488, to reveal global folding-unfolding transitions by FRET experiments. To this purpose, fluorescence spectra of the doubly-labeled CTPR3 variants were registered in the native (0 M GdnHCl) and chemically-unfolded state (4 M GdnHCl) (Fig. 50, 51, 52 and 53).

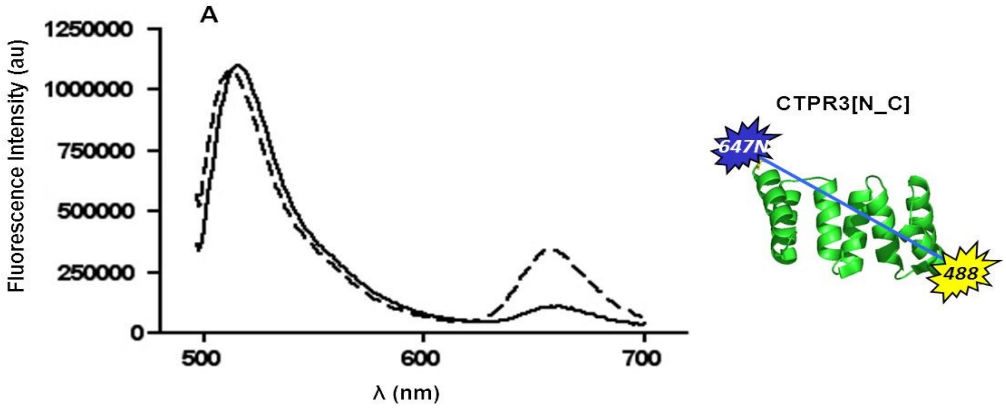

Figure 50 Emission spectra of doubly-labeled CTPR3[N_C] protein in native ([GdnHCl] = 0 M) (dashed line) and denaturing conditions ([GdnHCl] = 4 M) (solid line). The doubly-labeled construct under investigation is schematically represented on the right of the graph.

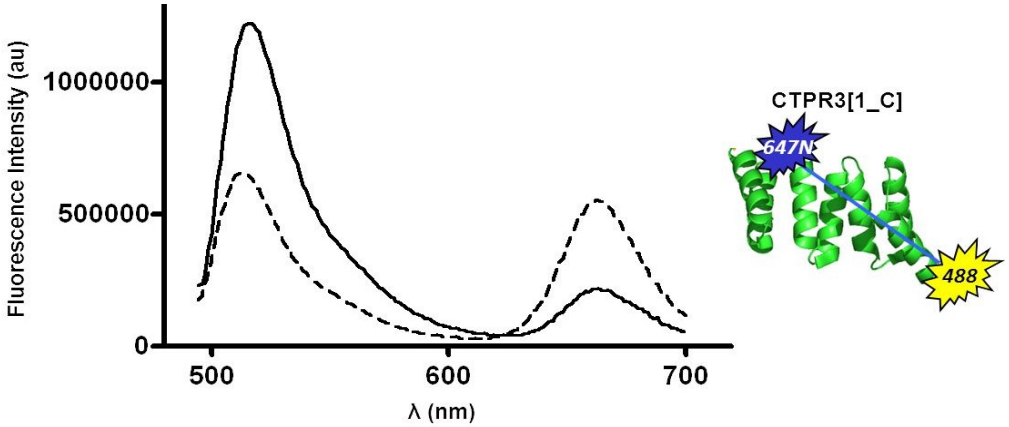

Figure 51 Emission spectra of doubly-labeled CTPR3[1_C] protein in native ([GdnHCl] = 0 M) (dashed line) and denaturing conditions ([GdnHCl] = 4 M) (solid line). The doubly-labeled construct under investigation is schematically represented on the right of the graph.

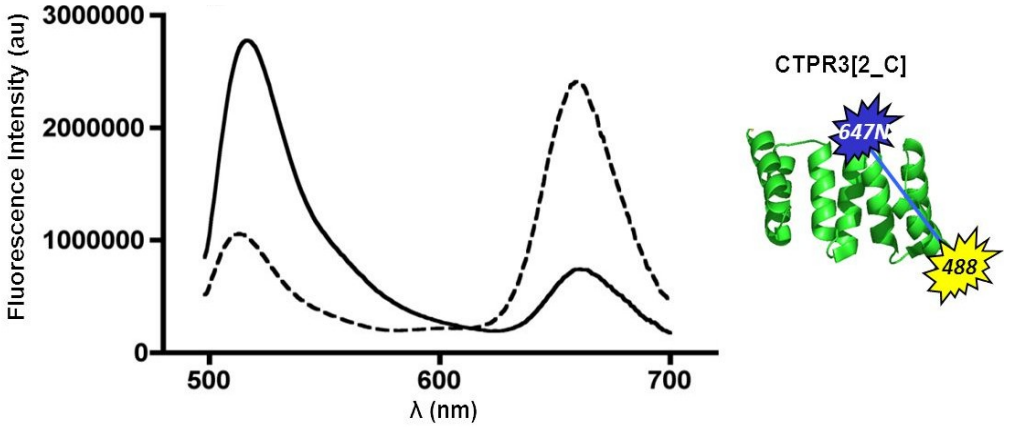

Figure 52 Emission spectra of doubly-labeled CTPR3[2_C] protein in native ([GdnHCl] = 0 M) (dashed line) and denaturing conditions ([GdnHCl] = 4 M) (solid line). The doubly-labeled construct under investigation is schematically represented on the right of the graph.

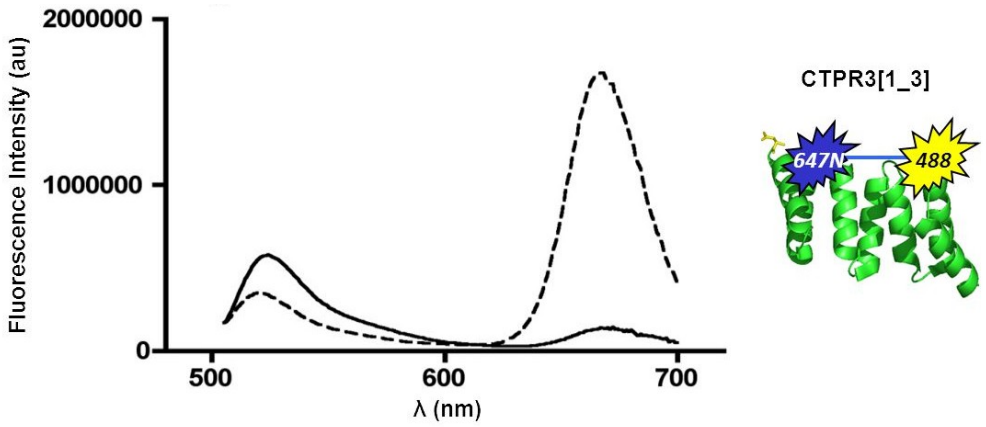

Figure 53 Emission spectra of doubly-labeled CTPR3[1_3] protein in native ([GdnHCl] = 0 M) (dashed line) and denaturing conditions ([GdnHCl] = 4 M) (solid line). The doubly-labeled construct under investigation is schematically represented on the right of the graph.

Unfolding was performed by chemical means because of the thermal instability of ATTO 647N dye. 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 nm-700 nm, including both donor and acceptor emission maximum (523 nm and 669 nm respectively). Fluorescence spectra recorded for the four doubly-labeled variants of CTPR3 in native conditions display evidence of FRET, showing E values of 0.21 in the construct CTPR3[N_C] (Fig. 50), 0.40 in CTPR3[1_C] (Fig. 51), 0.67 in CTPR3[2_C] (Fig. 52) and 0.82 CTPR3[1_3] (Fig. 53) (Table VI). As expected, in the native state the four doubly-labeled proteins exhibit progressively higher values of FRET efficiency going from CTPR3[N_C] to CTPR3[1_C], CTPR3[2_C] and CTPR3[1_3] constructs, according to a progressive decrease of the distance between dyes. In particular, the donor fluorescence is quenched almost completely in CTPR3[1_3], indicating that the distance between donor and acceptor is well within the Förster distance of the FRET pair. The discrepancy between the distances calculated from the E values obtained for the proteins in the native state and the theoretical values (Table VI) could be explained considering that these latter values are estimated from CTPR3 x-ray structure considering the C^{α}-C^{α} distance of the labeled amino acids. The distance increment due to the fluorophores was not taken into consideration. As expected, addition of a denaturant (GdnHCl) causes the donor fluorescence to increase, while acceptor fluorescence decreases, demonstrating that in all constructs FRET process is sensitive to changes in distance between the fluorophores going from the native to the unfolded state. A different result was observed in the fluorescence spectra of the denatured state of CTPR3[N_C] (Fig. 50). In fact, upon unfolding, we registered a decrease of the acceptor emission intensity fluorophore without a correlated increase of donor dye emission intensity. This unusual behavior could be due to the quenching of the donor dye mediated by the side-chain of neighboring residue, which is instead impeded by its restricted orientation in the native helical structure. When the native structure unwinds upon chemical denaturation, stereochemical restriction is relaxed and the quenching effect is more efficient. A similar observation was also reported by Ratner and coworkers (Ratner, V. *et al.* 2002). E values after denaturation fall down to 0.06 for CTPR3[N_C] (Fig. 50), 0.1 in CTPR3[1_C] (Fig. 51), 0.14 in CTPR3[2_C] (Fig. 52) and 0.19 CTPR3[1_3] (Fig. 53), according to the progressively shorter distance between dyes in the polypeptide chain (Table VI).

CTPR3 variant	R predicted from x-ray structure	E_{folded} (0 M GdnHCl)	R calculated from E_{folded} values	$E_{unfolded}$ (4 M GdnHCl)	R calculated from $E_{unfolded}$ values
CTPR3[N_C]-AD	52.0	0.21	61.0	0.06	77.5
CTPR3[1_C]-AD	40.3	0.40	53.4	0.10	70.6
CTPR3[2_C]-AD	31.3	0.67	43.4	0.14	66.3
CTPR3[1_3]-AD	28.4	0.82	38.0	0.19	62.3

Table VI FRET efficiency (E) and distance between fluorophores (R) predicted from CTPR3 x-ray structure and calculated from FRET experiments data.

These data demonstrate the efficacy of the FRET pair employed in yielding information about changes in the distance between the donor and the acceptor dyes, opening the way to further folding studies by fluorescence spectroscopy approaches.

Chemical denaturation studies by CD and FRET

According to the end-frying unfolding model followed by TPR proteins [Kajander *et al.* 2005; Main, E.R.G. *et al.* 2005; Cortajarena, A.L. *et al.* 2008; Hagay, T. & Levy, Y. 2008; Javadi, Y. & Main, E.R.G. 2009] the unfolding process starts with the unwinding of terminal regions and proceeds towards the middle of the protein. Therefore, we speculated that, monitoring unfolding process of the doubly-labeled constructs by CD and FRET, we should obtain different results for each variant. In fact, while CD monitors the total protein helical content revealing the unfolding process in its globality, FRET is sensible to local denaturation events. Consequently, we expected, comparing unfolding experiments by CD and FRET, to reveal differences in the unfolding transitions depending on the location of the two probes. For example, the unfolding transition by FRET for CTPR3[N_C], which carries the two fluorophores in the terminal regions, should precede that obtained by CD. On the contrary, the unfolding transition for CTPR3[1_3], which probes are placed in the central region, should appear before by CD and then by FRET. In fact, the distance between the two fluorophores is expected to change only at the end of the unfolding process, when the content of secondary structure is very low. Accordingly, CTPR3[1_C] and CTPR3[2_C] should exhibit an intermediate behavior.
To verify our hypothesis, we carried out chemical denaturation experiments on the doubly-labeled variants monitoring the unfolding process simultaneously by CD and FRET and comparing the obtained plots (Fig. 54, 55, 56 and 57).
Chemical denaturation was carried out by titration with GdnHCl within the concentration range 0 M – 5 M. CD and FRET spectra of the doubly-labeled protein samples were acquired at each denaturant concentration, respectively in the range 190-260 nm and 505-700 nm. Protein unfolded fraction content at each chemical denaturant concentration was evaluated assigning 0% and 100% value to the ellipticity and to the donor/acceptor fluorescence intensities ratio recorded respectively in the native ([GdnHCl] 0 M) and unfolded state ([GdnHCl] 5 M).

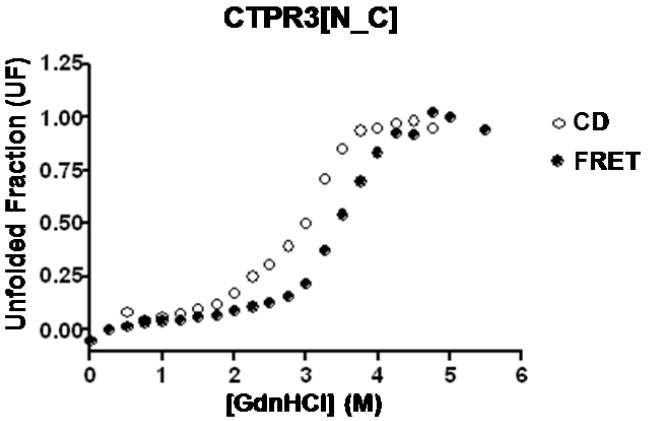

Figure 54 Chemical denaturation curves obtained for the doubly-labeled protein CTPR3[N_C] monitoring the unfolding process by CD and FRET.

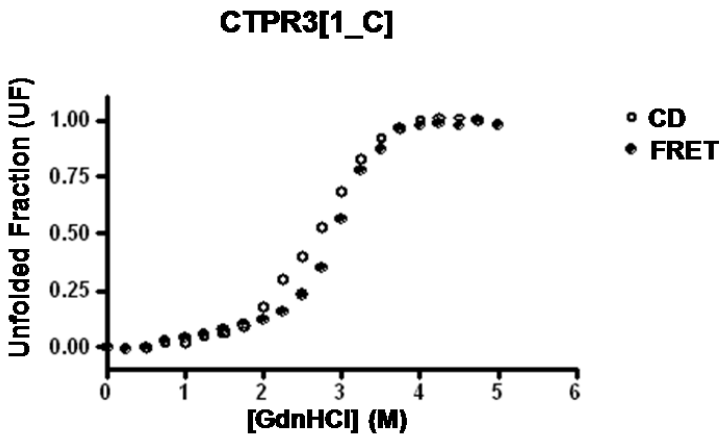

Figure 55 Chemical denaturation curves obtained for the doubly-labeled protein CTPR3[1_C] monitoring the unfolding process by CD and FRET.

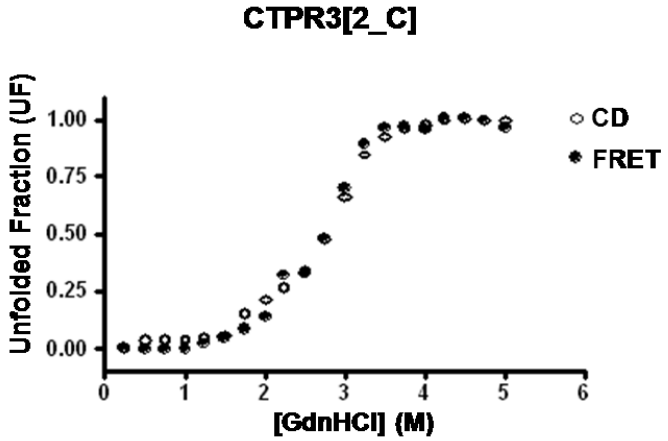

Figure 56 Chemical denaturation curves obtained for the doubly-labeled protein CTPR3[2_C] monitoring the unfolding process by CD and FRET.

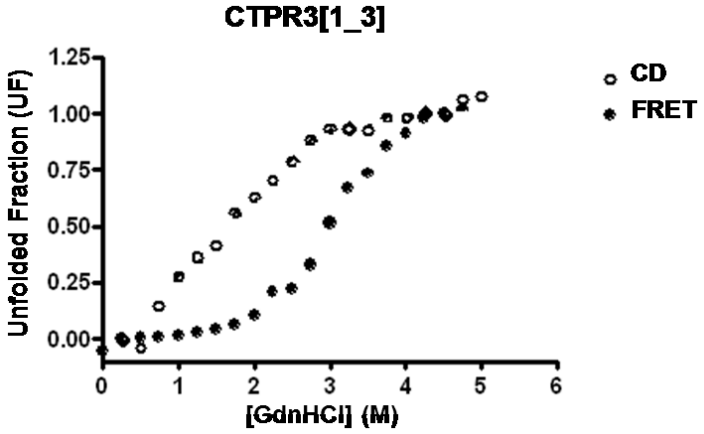

Figure 57 Chemical denaturation curves obtained for the doubly-labeled protein CTPR3[1_3] monitoring the unfolding process by CD and FRET.

FRET unfolding curves obtained for the doubly-labeled CTPR3 variants overlap CD plots, except for CTPR3[1_3] whose CD plot does not seem to depict a cooperative process. The CD plot observed for CTPR3[1_3] could be ascribed to its instability, as previously demonstrated through thermal denaturation studies, so that the protein unfolds at lower GdnHCl concentration than the other variants. To obtain a better resolution of the CD and FRET unfolding plots, we decided to perform the same kind of experiments using a weaker denaturant, such as the urea. Besides allowing a better resolution of the denaturation curves, urea also is known to have less effect than GdnHCl on the refractive index and spectral properties of the dyes. Denaturation studies with urea were performed as described for experiments with GdnHCl, ranging urea concentration from 0 M to 8.25 M (Fig. 58, 59, 60, 61 and Table VII).

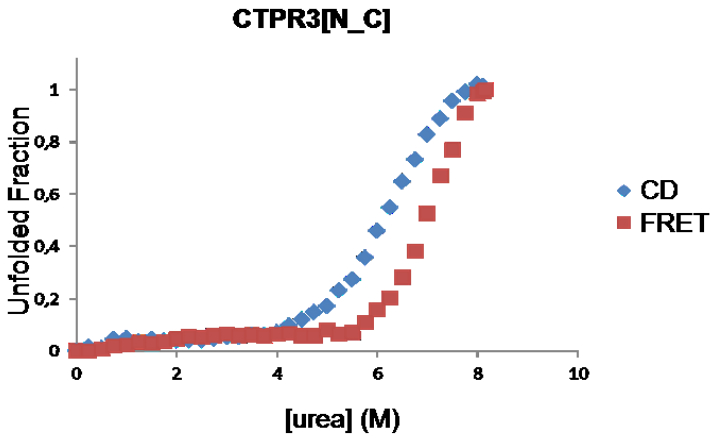

Figure 58 Urea denaturation curves obtained for the doubly-labeled protein CTPR3[N_C] monitoring the unfolding process by CD and FRET.

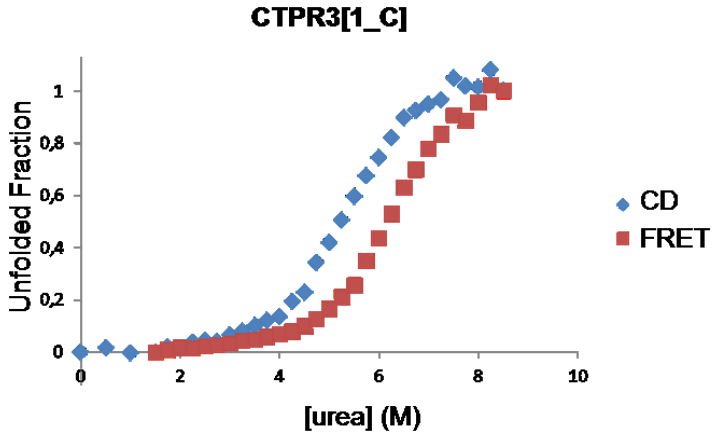

Figure 59 Urea denaturation curves obtained for the doubly-labeled protein CTPR3[1_C] monitoring the unfolding process by CD and FRET.

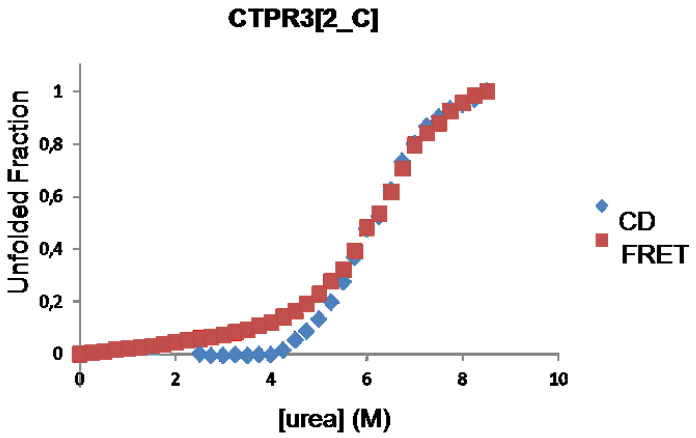

Figure 60 Urea denaturation curves obtained for the doubly-labeled protein CTPR3[2_C] monitoring the unfolding process by CD and FRET.

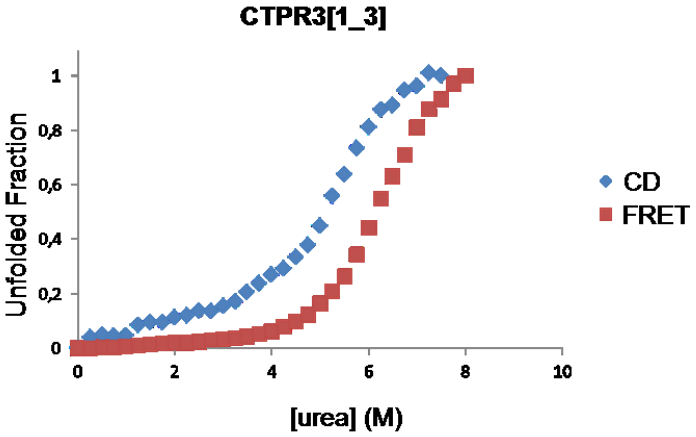

Figure 61 Urea denaturation curves obtained for the doubly-labeled protein CTPR3[1_3] monitoring the unfolding process by CD and FRET.

Doubly-labeled CTPR3 variant	CD transition point	FRET transition point
CTPR3[N_C]	6.25 M	7 M
CTPR3[1_C]	5.25 M	6 M
CTPR3[2_C]	6.5 M	6.25 M
CTPR3[1_3]	5.25 M	6.25 M

Table VII Urea concentration at the CD and FRET transition points for the doubly-labeled CTPR3 proteins.

In CTPR3[1_3], FRET unfolding curve followed CD unfolding curve (transition points were respectively registered at 6.25 M and 5.25 M [urea]) (Fig. 61 and Table VII). This observation is in agreement with the Ising model, by which CTPR proteins are predicted to unfold. In fact, in this construct the labels are placed on internal protein region, which unfolds later. In CTPR3[2_C], FRET and CD unfolding plots are overlapped (transition points respectively at 6.25 M and 6.5 M) (Fig. 60 and Table VII). This observation is in agreement with an increased distance between the fluorophores, respect to the construct CTPR3[1_3], one of which is placed in the C-

terminal region which is a region more prone to unfold. Strange observations come from CTPR3[N_C] and CTPR3[1_C] unfolding plots (Fig 58 and 59). On the basis of what observed for CTPR3[1_3] and CTPR3[2_C], in the constructs CTPR3[N_C] and CTPR3[1_C], in which the distance between the two labels is higher and the labels are positioned in more external regions, the transition monitored by FRET is expected to precede the CD transition. Interestingly, the results show that unfolding transition of CTPR3 variants CTPR3[N_C] and CTPR3[1_C] reported by FRET occurs at higher denaturant concentrations than that measured by CD (Table VII). These data are indicative of a deviation from a simple scenario in which FRET transitions are only function of helical structure loss. Probably, additional phenomena cause a deviation of the unfolding transition plot to higher denaturant concentrations.

CONCLUSIONS

The development of methods for the chemical modification of proteins provides a means by which to characterize and modulate protein structure and function. Different approaches, based on the use of both chemistry and/or molecular biology, have been proposed to prepare proteins specifically modified at multiple sites. However, most of the these technologies are limited by high experimental efforts, low yields, low adaptability to different targets and different kinds of modifications. This PhD thesis reports a useful and easy-to-perform protocol for the synthesis and modification of proteins which overcomes the major limits of the precedent strategies for protein labeling. The semi-synthetic strategy proposed, based on the use of EPL, is generally applicable for the preparation of proteins in which a combination of two different labels is incorporated at any desired position, but could also be adapted to the introduction of more than two probes. The effectiveness of the designed semisynthetic strategy has been assessed through the preparation of four variants of CTPR3, a *consensus* tetratrico peptide repeat protein, labeled with a fluorophores pair and useful for folding studies by fluorescence spectroscopy approaches. The obtained fluorescent variants have been characterized through biophysical studies. The intriguing results of these preliminary structural analysis open the way to further folding studies by single-molecule spectroscopic approaches, such as sm-FRET and FCS, aimed to gain deeper insights into repeat protein folding mechanism and dynamics.

REFERENCES

Alsina J., Yokum S. T., Albericio F., Barany G., Backbone Amide Linker (BAL) Strategy for Nα-9-Fluorenylmethoxycarbonyl (Fmoc) Solid-Phase Synthesis of Unprotected Peptide p-Nitroanilides and Thioesters, 1999, *J. Org. Chem.*, 64, 8761–8769

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, 2005, *J. Biol. Chem.*, 280, 31019-3126

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, 2009, *J. Am. Chem. Soc.*, 131,10800-108001

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

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

Arnold U., Hinderaker M.P., Nilsson B.L., Huck B.R., Gellman S.H., Raines R.T., 2002, Protein prosthesis: a semisynthetic enzyme with a beta-peptide reverse turn, *J. Am. Chem. Soc.*, 124, 8522-8523

Bang D., Pentelute B.L., Kent S.B., Kinetically controlled ligation for the convergent chemical synthesis of proteins, 2006, *Angew. Chem. Int. Ed. Engl.*, 45, 3985-3988

Blanco-Canosa J.B., **Dawson P.E.**, An efficient Fmoc-SPPS approach for the generation of thioester peptide precursors for use in Native Chemical Ligation, 2008, *Ang. Chem. Int. Ed.*, 47, 6851-6855

Botti P., Villain M., Manganiello S., Gaertner H., Native chemical ligation through in situ O to S acyl shift, 2004, *Org. Lett.*, 6, 4861-4864

Bradford M., A rapid and sensitive method for the quantitation of microgram

quantities of protein utilizing the principle of protein dye-binding, 1976, *Anal. Biochem.*, 72,

248-254

Braisted A.C., Judice J.K., Wells J.A., Synthesis of proteins by subtiligase, 1997, *Methods Enzymol.*, 289, 298-313

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

Canne L.E., Ferre- D'Amare A.R., Burley S.K., Kent S.B.H., Total Chemical Synthesis of a Unique Transcription Factor-Related Protein: cMyc-Max, 1995, *J. Am. Chem. Soc.*, 117, 2998–3007

Chang T.K., Jackson D.Y., Burnier J.P., Wells J.A., Subtiligase: a tool for semisynthesis of proteins, *Proc. Natl. Acad. Sci. USA*, 1994, 91, 12544-12548

Chatterjee C., McGinty R.K., Pellois J.P., Muir T.W., Auxiliary-mediated site-specific peptide ubiquitylation, 2007, *Angew. Chem. Int. Ed. Engl.*, 46, 2814-2818

Chen J., Yuan Y., Zhu J., Danishefsky S.J., Native chemical ligation at valine: a contribution to peptide a+nd glycopeptide synthesis, 2008, *Angew. Chem. Int. Ed. Engl.*, 47, 8521-8524

Chen J., Wang P., Zhu J., Wan Q., Danishefsky S.J., A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides, 2010, *Tetrahedron*, 66, 2277-2283

Clerico E.M., Zhuravleva A., Smock R.G., Gierasch L.M., Segmental isotopic labeling of the Hsp70 molecular chaperone DnaK using expressed protein ligation 2010, *Biopolymers*, In printing

Clippingdale C.J., Borrow J.D., Wade J., Peptide thioester preparation by Fmoc solid phase peptide synthesis for use in native chemical ligation, 2000, *Pep. Sci.*, 6, 225-234

Cortajarena A.L., Mochrie S.G., Regan L., Mapping the energy landscape of repeat proteins using NMR-detected hydrogen exchange, 2008, *J. Mol. Biol.*, 379, 617-626

Cortajarena A.L., Lois G., Sherman E., O'Hern C.S., Regan L., Haran G., Non-random-coil behavior as a consequence of extensive PPII structure in the denatured state, 2008, *J. Mol. Biol.*, 382, 203-212

Cotton G.J., Ayers B., Xu R., Muir T.W., Insertion of a Synthetic Peptide into a Recombinant Protein Framework: A Protein Biosensor, 1999, *J. Am. Chem. Soc.*, 121, 1100–1101

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

Crich, D., Banerjee A., Native chemical ligation at phenylalanine, 2007, *J. Am. Chem. Soc.*, 129, 10064-10065

D'Andrea L.D., **Regan L.**, TPR proteins: the versatile helix, 2003, *TRENDS in Biochem. Sci.*, 28, 655-662

Dantas de Araújo A., Palomo J.M., Cramer J., Seitz O., Alexandrov K., Waldmann H., Diels-Alder Ligation of Peptides and Proteins, 2006, *Chem. Eur. J.*, 12, 6095-6109

David R., Richter M.P., Beck-Sickinger A.G., Expressed protein ligation. Method and applications, 2004, *Eur. J. Biochem.*, 271, 663-677

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

Dawson P.E., Muir T.W., Clark-Lewis I., Kent S.B.H., Synthesis of proteins by Native Chemical Ligation, 1994, *Science*, 266, 776-779

deLisle-Milton R., Milton S.C.F., Schnolzer M., Kent S.B.H., 1993, In *Tecniques in Protein Chemistry IV*, ed. RH Angeletti, pp. 257-267, New York: Academic

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, 2000, *Proc. Natl. Acad. Sci. USA*, 97,5179-5184

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

Erlanson D.A., Chytil M., Verdine G.L., The leucine zipper domain controls the orientation of AP-1 in the NFAT.AP-1.DNA complex, 1996, *Chem. Biol.*, 3, 981-991

Evans T.C. Jr., Benner J., Xu M.Q., Semisynthesis of cytotoxic proteins using a modified protein splicing element, 1998, *Prot. Sci.*, 7, 2256-2264

Gaertner H.F., Rose K., Cotton R., Timms D., Camble R., Offord R.E., Construction of protein analogs by site-specific condensation of unprotected fragments, 1992, *Bioconjug. Chem.*, 3, 262-268

Grogan M.J., Pratt M.R., Marcaurelle L.A., Bertozzi C.R., Homogeneous glycopeptides and glycoproteins for biological investigation, 2002, *Annu. Rev. Biochem.*, 71, 593-634

Hackenberger C.P., Friel C.T., Radford S.E., Imperiali B., Semisynthesis of a glycosylated Im7 analogue for protein folding studies, 2005, *J. Am. Chem. Soc.*, 127, 12882-12889

Hackeng T.M., Mounier C.M., Bon C., Dawson P.E., Griffin J.H., Kent S.B., Total chemical synthesis of enzymatically active human type II secretory phospholipase A2, 1997, *Proc. Natl. Acad. Sci. USA*, 94, 7845-7850

Hackeng T.M., Griffin J.H., Dawson P.E., Protein synthesis by Native Chemical Ligation: expanded scope by using straightforward methodology, 1999, *Proc. Natl. Acad. Sci. USA*, 96,10068-10073

Hackeng T.M., Fernández J.A., Dawson P.E., Kent S.B., Griffin J.H., Chemical synthesis and spontaneous folding of a multidomain protein: anticoagulant microprotein S, 2000, *Proc. Natl. Acad. Sci. USA*, 97, 14074-14078

Hagai T., **Levy Y.**, Folding of elongated proteins: conventional or anomalous?, 2008, *J. Am. Chem. Soc.*,130, 14253-14262

Hahn M.E., **Muir T.W.**, Photocontrol of Smad2, a multiphosphorylated cell-signaling protein, through caging of activating phosphoserines, 2004, *Angew. Chem. Int. Ed. Engl.*, 43, 5800-5803

Hahn M.E., **Muir T.W.**, Manipulating proteins with chemistry: a cross section of chemical biology, 2005, *Trends Biochem. Sci.*, 30, 26-34

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

Hangauer M.J., Bertozzi C.R., A FRET-based Fluorogenic Phosphine for Live Cell Imaging with the Staudinger Ligation, 2008, *Angew Chem Int Ed Engl.*, 47, 2394–2397

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, 2001, *Bioorg. Med. Chem. Lett.*, 11, 3091-3094

Hojo H., Onuma Y., Akimoto Y., Nakahara Y., Nakaharay Y., N-Alkyl cysteine-assisted thioesterification of peptides, 2007, *Tetrahedron Lett.*, 48, 25-28

Hondal R.J., Nilsson B.L., Raines R.T., Selenocysteine in native chemical ligation and expressed protein ligation, 2001, *J. Am. Chem. Soc.*, 123, 5140-5141

Ho S.N., Hunt H.D., Horton R.M., Pullen J.K., Pease L.R., Site-directed mutagenesis by overlap extension using the polymerase chain reaction, 1989, *Gene*, 77:51-59.

Huse M., Holford M.N., Kuriyan J., Muir T.W., Semi-synthesis of Hyperphosphorylated Type I TGF-beta Receptor: Addressing the Mechanism of Kinase Activation, 2000, *J. Am. Chem. Soc.*, 122, 8337-8338

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, 2000, *FEBS Lett.*, 468, 155-158

Ingenito R., Bianchi E., Fattori D., Pessi A., Solid Phase Synthesis of Peptide C-Terminal Thioesters by Fmoc/t-Bu Chemistry, 1999, *J. Am. Chem. Soc*., 121, 11369-11374

Javadi Y., **Main E.R.**, Exploring the folding energy landscape of a series of designed consensus tetratricopeptide repeat proteins, 2009, *Proc. Natl. Acad. Sci. USA*, 106, 17383-17388

Johnson E.C.B., Kent S.B.H., Insights into the mechanism and catalysis of the Native Chemical Ligation, 2006, *J. Am. Soc. Chem.*, 128, 6640-6646

Kajander T., Cortajarena A.L., Main E.R., Mochrie S.G., Regan L., A new folding paradigm for repeat proteins, 2005, *J. Am. Chem. Soc.*, 127, 10188-10190

Kajander T., Cortajarena A. L., Regan L., Consensus design as a tool for engineering repeat proteins, 2006, *Methods in Mol. Biol.*, 340, 151-170

Kenner G. W., Towards synthesis of proteins, 1977, *Proc. R. Soc. London*, 353, 441

Kent S.B.H., Total chemical synthesis of proteins, *Chem. Soc. Rev.*, 2009, 38, 338-351

Kent S.B.H., Chemical synthesis of peptides and proteins, *Annu. Rev. Biochem*., 1988, 57, 957-984

Kolb H.C., Finn M.G., Sharpless K.B., Click chemistry: diverse chemical function from a few good reactions, 2001, *Angew. Chem. Int. Ed*., 40, 2004-2021

Kurpiers T., **Mootz H.D.**, Regioselective cysteine bioconjugation by appending a labeled cystein tag to a protein by using protein splicing in trans, 2007, *Angew. Chem. Int. Ed. Engl.*, 46, 5234-5237

Kurpiers T., **Mootz H.D.**, Site-specific chemical modification of proteins with a prelabelled cysteine tag using the artificially split Mxe GyrA intein, 2008, *Chembiochem*, 9, 2317-2325

Lahiri S., Seidel R., Engelhard M., Becker C.F., Photocontrol of STAT6 dimerization and translocation, 2010, *Mol. Biosyst.*, in printing

Laemmli U., Cleavage of structural proteins during the assembly of the head of bacteriophage T4, 1970, *Nature*, 227, 680-685

Li W.Q., Kawakmi T., Aimoto S., Direct preparation of peptide thioesters using an Fmoc solid-phase method, 1998, *Tetrahedron Lett.*, 39, 8669-8672

Lin F.L., HoytH.M., van Halbeek H., Bergman R.G., Bertozzi C.R., Mechanistic Investigation of the Staudinger Ligation, 2005, *J. Am. Chem. Soc.*, 127, 2686–2695

Linn K.M., Derebe M.G., Jiang Y., Valiyaveetil F.I., Semisynthesis of NaK, a Na(+) and K(+) conducting ion channel, 2010, *Biochemistry*, 49, 4450-4456

Lu W.Y., Qasim M.A., Kent S.B.H., Comparative Total Syntheses of Turkey Ovomucoid Third Domain by Both Stepwise Solid Phase Peptide Synthesis and Native Chemical Ligation, 1996, *J. Am. Chem. Soc.*, 118, 8518–8523

Lu W., Gong D., Bar-Sagi D., Cole P.A., Site-specific incorporation of a phosphotyrosine mimetic reveals a role for tyrosine phosphorylation of SHP-2 in cell signaling, 2001, *Mol. Cell.*, 8, 759-769.

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, 2006, *Angew. Chem. Int. Ed. Engl.*, 45, 5218-5221

Macmillan D., Bertozzi C.R., Modular assembly of glycoproteins: towards the synthesis of GlyCAM-1 by using expressed protein ligation, 2004, *Angew. Chem. Int. Ed. Engl.*, 43, 1355-1359

Main E.R.G., Xiong Y., Cocco M.J., D'Andrea L.D., Regan L., Design of stable α-helical arrays from an idealized TPR motif, 2003, *Structure*, 11, 497-508

Main E.R.G., Jackson S.E., Regan L., The folding and design of repeat proteins: reaching a consensus, 2003, *Curr. Opin. Struc. Biol.*, 13, 482-489

Main E.R.G., Lowe A.R., Mochrie S.G.J., Jackson S.E., Regan L., A recurring theme in protein engineering: the design, stability and folding of repeat proteins, 2005, *Curr. Opin. Struc. Biol.*, 15, 464-471

Main E.R.G., Stott K., Jackson S.E., Regan L., Local and long-range stability in tandemly arrayed tetratricopeptide repeats, 2005, *Proc. Natl. Acad. Sci. USA*, 102, 5721-5726

Marinzi C., Offer J., Longhi R., Dawson P.E., An o-nitrobenzyl scaffold for peptide ligation: synthesis and applications, 2004, *Bioorg. Med. Chem.*, 12, 2749-57

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, 1999, *Gene*, 231, 1-13

Metanis N., Keinan E., Dawson P.E., Traceless Ligation of Cysteine Peptides using Selective Deselenization, 2010, *Angew. Chem. Int. Ed. Engl.*, in printing

McGinty R.K., Kim J., Chatterjee C., Roeder R.G., Muir T.W., Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation, 2008, *Nature*, 453, 812-816

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, 2002, *Cell*, 108, 599-614

Mendel D., Cornish V.W., Schultz P.G., Site-directed mutagenesis with an expanded genetic code, 1995, *Annu. Rev. Biophys. Biomol. Struct.*, 24,435-462

Michaelis S., **Blechert S.,** Ring-opening cross-metathesis (ROCM) as a novel tool for the ligation of peptides, 2007, *Chem. Eur. J.*, 13, 2358-2368

Mootz H.D., Split inteins as versatile tools for protein semisynthesis, 2009, *ChemBioChem*, 10, 2579-2589

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

Muir T.W., **Kent S.B.H.**, The chemical synthesis of proteins, 1993, *Curr. Opin. Biotechnol.*, 4, 420-427

Muir T.W., Semisynthesis of proteins by Expressed Protein Ligation, 2003, *Annu. Rev. Biochem.*, 72, 249-289

Muir T.W., Studying protein structure and function using semisynthesis, 2008, *Biopolymers*, 90, 743-750

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 2001, *Cell*, 106, 453-463

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, 2004, *J. Am. Chem. Soc.*, 126, 14004-14012

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

Ottesen J.J., Huse M., Sekedat M.D., Muir T.W., Semisynthesis of phosphovariants of Smad2 reveals a substrate preference of the activated T beta RI kinase, 2004, *Biochemistry*, 43, 5698-5706

Paulick M.G., Wise A.R., Forstner M.B., Groves J.T., Bertozzi C.R., Synthetic analogues of glycosylphosphatidylinositol-anchored proteins and their behavior in supported lipid bilayers, 2007, *J. Am. Chem. Soc.,* 129, 11543-11550

Pentelute B.L., Kent S.B., Selective desulfurization of cysteine in the presence of Cys(Acm) in polypeptides obtained by native chemical ligation, 2007, *Org. Lett.*, 9, 687-690

Perler F.B., InBase, the intein database, 2000, *Nucl. Acids Res.*, 28, 344-345

Perler F.B., Protein splicing mechanisms and applications, 2005, *Life*, 57, 469-476

Peroza E.A., **Freisinger E.**, Tris is a non-innocent buffer during intein-mediated protein cleavage, 2008, *Protein Expr. Purif.*, 57, 217-225

Proft T., Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilization, 2010, *Biotechnol. Lett.*, 32, 1-10

Punna S., Kuzelka J., Wang Q., Finn M.G., Head-to-tail peptide cyclodimerization by copper-catalyzed azide-alkyne cycloaddition, 2005, *Angew Chem Int Ed Engl.*, 44, 2215-2220

Rak A., Pylypenko O., Durek T., Watzke A., Kushnir S., Brunsveld L., Waldmann H., Goody R.S., Alexandrov K., Structure of Rab GDP-dissociation inhibitor in complex with prenylated YPT1 GTPase, 2003, *Science*, 302, 646-650

Ratner V., Kahana E., Eichler M., Haas E., A general strategy for site-specific double labeling of globular proteins for FRET studies, 2002, *Bioconjug. Chem.*, 13, 1163-1170

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, 2004, *Proc. Natl. Acad. Sci. USA*, 101, 6397-6402

Saxon E., **Bertozzi C.R.**, Cell surface engineering by a modified Staudinger reaction, 2000, *Science*, 287, 2007-2010

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

Schnölzer M., **Kent S.B.**, Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease, 1992, *Science*, 256, 221-225

Schwarzer D., **Cole P.A.**, Protein semisynthesis and expressed protein ligation: chasing a protein's tail, 2005, *Curr. Opin. Chem. Biol.*, 9, 561-569

Shafer D.E., Inman J.K., Lees A., Reaction of Tris(2-carboxyethyl)phosphine (TCEP) with maleimide and alpha-haloacyl groups: anomalous elution of TCEP by gel filtration, 2000, *Anal. Biochem.*, 282, 161-164

Shen K., **Cole P.A.**, Conversion of a tyrosine kinase protein substrate to a high affinity ligand by ATP linkage, 2003, *J. Am. Chem. Soc.*, 125, 16172-16173

Shogren-Knaak M.A., Fry C.J., Peterson C.L., A native peptide ligation strategy for deciphering nucleosomal histone modifications, 2003, J. Biol. Chem., 278, 15744-15748

Shogren-Knaak M., Ishii H., Sun J. M., Pazin M. J., Davie J. R., Peterson C. L., Histone H4-K16 acetylation controls chromatin structure and protein interactions, Science 2006, 311, 844–847

Sieber P., Eisler K., Kamber B., Riniker B., Rittel W., Märki F., de Gasparo M., Synthesis and biological activity of two disulphide bond isomers of human insulin: [A7-A11,A6-B7-cystine]- and [A6-A7,A11-B7-cystine]insulin (human), 1978, *Hoppe-Seylers Z. Physiol. Chem.*, 1978, 359, 113-123

Southworth M.W., Adam E., Panne D., Byer R., Kautz R., Perler F.B., Control of protein splicing by intein fragment reassembly, 1998, *EMBO*, 17, 918–926

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* GyrA intein, 1999, *BioTechniques*, 27, 110-120

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

Tam J.P., Yu Q., Methionine ligation strategy in the biomimetic synthesis of parathyroid hormones, 1998, Biopolymers, 46, 319-327

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, 1997, *J. Bacteriol.*, 179, 6378-6382

Tolbert T.J., Wong C.H., New methods for proteomics research: preparation of proteins with N-terminal cysteine for labeling and conjugation, 2002, *Angew. Chem. Int. Edn. Engl.*, 41, 2171-2174

Tofteng A.P., Sørensen K.K., Conde-Frieboes K.W., Hoeg-Jensen T., Jensen K.J., Fmoc solid-phase synthesis of C-terminal peptide thioesters by formation of a

backbone pyroglutamyl imide moiety, 2009, Angew. Chem. Int. Ed. Engl., 48,7411-7414

Tyagarajan K., Pretzer E., Wiktorowicz J.E., Thiol-reactive dyes for fluorescence labeling of proteomic samples, 2003, *Electrophoresis*, 24, 2348-2358

Valiyaveetil F.I., MacKinnon R., Muir T.W., Semi-synthesis and folding of the potassium channel KcsA, 2002, *J. Am. Chem. Soc.*, 124, 9113-9120

Valiyaveetil F.I., Sekedat M., Mackinnon R., Muir T.W., Glycine as a D-amino acid surrogate in the K(+)-selectivity filter, 2004, *Proc. Natl. Acad. Sci. USA*, 101, 17045-17049

Villain M., Gaertner H., Botti P., Native Chemical Ligation with aspartic and glutamic as C-terminal residues: scope and limitations, *Eur. J. Org. Chem.*, 3267-3272

Virdee S., Macmillan D., Waksman G., Semisynthetic Src SH2 domains demonstrate altered phosphopeptide specificity induced by incorporation of unnatural lysine derivatives, 2010, *Chem. Biol.*, 17, 274-284

von Eggelkraut-Gottanka R., Klose A., Beck-Sickinger A.G., Beyermann M., Peptide thioester formation using standard Fmoc-chemistry, 2003, Tetrahedron Lett., 44, 3551-3554

Wan Q., Danishefsky S.J., Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides, 2007, *Angew. Chem. Int. Ed. Engl.*, 46, 9248-9252

Wang D., Cole P.A., Protein tyrosine kinase Csk-catalyzed phosphorylation of Src containing unnatural tyrosine analogues, *J. Am. Chem. Soc.*, 2001, 123, 8883-8886

Weeks S.D., Drinker M., Loll P.J., Ligation independent cloning vectors for expression of SUMO fusions, 2007, *Protein Expr. Purif.*, 53, 40-50

Wells J.A., **Estell D.A.**, Subtilisin--an enzyme designed to be engineered, 1998, *Trends Biochem. Sci.*, 291-297

Wildes D., **Wells J.A.**, Sampling the N-terminal proteome of human blood, 2010, *Proc. Natl. Acad. Sci. USA*, **107**, 4561-4566

Woody R.W., Circular dichroism and conformation of unordered polypeptides, 1992, *Adv. Biophys. Chem.*, 2,37-79

Wu J.W., Hu M., Chai J., Seoane J., Huse M., Li C., Rigotti D.J., Kyin S., Muir T.W., Fairman R., Massagué J., Shi Y., Crystal structure of a phosphorylated Smad2. Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-beta signaling, 2001, J. Mol. Cell., 8, 1277–1289

Xie J., **Schultz P.G.**, Adding amino acids to the genetic repertoire, 2005, Curr. Opin. Chem. Biol., 9, 548-554

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

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

Xu R., Ayers B., Cowburn D., Muir T.W., Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies, 1999, *Proc. Natl. Acad. Sci. USA*, 19, 388-393

Yajima H., **Fujii N.**, Totally synthetic crystalline ribonuclease A., 1981, Biopolymers. 1981, 20, 1859-1867

Yamamoto N., Tanabe Y., Okamoto R., Dawson P.E., Kajihara Y., Chemical synthesis of a glycoprotein having an intact human complex-type sialyloligosaccharide under the Boc and Fmoc synthetic strategies, 2008, *J. Am. Chem. Soc.*, 130, 501-510

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

Yoshihara H.A., Mahrus S., Wells J.A., Tags for labeling protein N-termini with subtiligase for proteomics, 2008, *Bioorg. Med. Chem. Lett.*, 18, 6000-6003

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, 2008, *Protein Sci.*, 17, 736-747

Zhao Z., Ma X., Li L., Zhang W., Ping S., Xu M.Q., Lin M., Protein cyclization enhanced thermostability and exopeptidase-resistance of green fluorescent protein, 2010, *J. Microbiol. Biotechnol.*, 20, 460-466

Zimm B. H., **Bragg, J. K. J.**, 1959, *Chem. Phys.,* 31, 526-535

COMMUNICATIONS

L. De Rosa, A. Romanelli, A. L. Cortajarena, G. Haran, L. Regan, and L. D. D'Andrea, **Semi-synthesis and biophysical characterization of doubly-labeled Consensus Tetratricopeptide Repeat Proteins**, Congresso Nazionale 2010 della Divisione di Chimica dei Sistemi Biololgici, San Vito di Cadore 9-11 Settembre 2010.

L. De Rosa, A. Basile, A. Russomanno, R. Di Stasi, D. Diana, R. Fattorusso, C. Pedone, C. Turco and L. D. D'Andrea, **Design and biological characterization of a proangiogenic peptide reproducing the β-hairpin region 87-100 of PlGF**, 12th Naples Workshop on bioactive peptides, Napoli 04-07 Giugno 2010.

D. Diana, R. Di Stasi, L. De Rosa, C. Pedone, L.D. D'Andrea and R. Fattorusso, **NMR structural characterization of a bioactive β-hairpin peptide and of its interaction with VEGFR-1D2**, 12th Naples Workshop on bioactive peptides, Napoli 04-07 Giugno 2010.

L. De Rosa, A. Romanelli, A. L. Cortajarena, L. Regan, G. Haran and L. D. D'Andrea: **Semi-synthesis of doubly-labeled Consensus Tetratricopeptide Repeat proteins for folding studies**, VIII European Symposium of The Protein Society, Zurigo, 14-18 Giugno 2009 and XIII Congresso Nazionale della Società Chimica Italiana, Sorrento 5-10 Luglio 2009.

L. De Rosa, A. Romanelli, L. Regan, A. L. Cortajarena, G. Haran and L. D. D'Andrea: **Semi-synthesis of Consensus Tetratricopeptide Repeat proteins for folding studies**, 53rd SIB and SCI National Meeting, Riccione, 23-26 Settembre 2008.

L. De Rosa, A. Romanelli, L. Regan, A. L. Cortajarena, G. Haran and L. D. D'Andrea: **Semi-synthesis of Consensus Tetratricopeptide Repeat proteins for folding studies**, 11th Naples Workshop on Bioactive Peptides, Napoli, 24-26 Maggio 2008.

PUBBLICATIONS

D'Andrea L.D., Del Gatto A., De Rosa L., Romanelli A., Pedone C., **Peptide targeting angiogenesis related growth factor receptors**, *Curr. Pharm. Des.* 2009;15(21):2414-29

RESEARCH ACTIVITY IN SCIENTIFIC INSTITUTIONS ABROAD

From November 15th to December 15th 2009 the research activity of Dr. Lucia De Rosa has been carried out at the Dept. of Biophysics and Biochemistry of Yale University, New Haven CT (USA), in the research group of Professor Lynne Regan.