DEVELOPMENT OF A DIMALEIMIDE-BASED PROTEIN LABELLING TECHNIQUE FOR FLUOROGENIC, X-RAY CRYSTALLOGRAPHY AND NMR APPLICATIONS

MIROSLAVA STRMISKOVA

Thesis submitted to the
Faculty of Graduate & Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in Chemistry

Department of Chemistry and Biomolecular Sciences
Ottawa-Carleton Chemistry Institute
Faculty of Science
University of Ottawa

Résumé

Le marquage des protéines par fluorescence est une technique puissante pour visualiser des protéines dans des cellules vivantes, ce qui permet ensuite d'élucider leur localisation et transport intracellulaires, et ultimement, leur fonction au sein de la cellule. Nous avons développé une nouvelle technique de marquage de protéines, basée sur un petit peptide hélicoïdal possédant deux résidus cystéines (dC10), attaché à la protéine d'intérêt. Ce peptide de fusion peut réagir de manière très efficace avec une petite molécule fluorogène, composée d'un fluorophore et d'un core dimaléimide (dM10). Ce dernier assure une spécificité élevée de la réaction de marquage avec les cystéines de dC10, ainsi que l'atténuation quasi-complète de la fluorescence du composé fluorogène via un transfert d'électrons photoinduit (PeT), jusqu'à la réaction des deux groupements maléimide et formation de deux liaisons covalentes avec les résidus cystéine du peptide dC10.

Nos tentatives initiales de marquage intracellulaire ont soulevé l'importance de la sélectivité de la réaction de marquage, qui dépend avant tout de la réactivité du peptide de fusion dC10. Pour améliorer cette réactivité, nous avons conçu une série de séquences mutées de dC10 en utilisant une stratégie rationnelle. Nous avons préparé des librairies de mutants, incluant des combinaisons de mutations sur trois positions spécifiques sur le peptide dC10, et nous avons criblé leur réactivité avec un composé fluorogène. Ainsi, nous avons identifié une nouvelle séquence dC10* qui présente une réactivité 10-fois plus élevée, et nous avons pu démontrer son utilité pour un marquage *in cellulo*. Des études mécanistiques, menées par la suite, ont révélé des raisons possibles pour cette augmentation significative de réactivité.

Cette technique de marquage protéique peut également être appliquée dans le domaine de la biologie structurale. Ces applications, qui sont en ce moment en développement dans le groupe Keillor, incluent un marquage spécifique de protéines par des composés chélatant un ion lanthanide pour la spectroscopie RMN, et également un marquage par des composés portant un atome lourd pour la cristallographie aux rayons X des protéines.

Abstract

Fluorescent protein labelling is a powerful tool for the sensitive visualization of proteins in living cells, allowing the elucidation of their localization, trafficking and ultimately their cellular function. We have developed a novel labelling technique based on the genetic fusion of a protein of interest to a small helical peptide sequence containing two Cys residues (dC10). This tag can undergo an efficient reaction with small fluorogenic labelling agents composed of a fluorophore and a dimaleimide core (dM10) that confers high reaction specificity, and quenches the latent fluorescence through photo-induced electron transfer, until both of its maleimide groups have formed robust covalent bonds with the tag Cys thiol groups.

Our initial efforts at intracellular protein labelling demonstrated the importance of the selectivity of the labelling reaction, which is dependent on the reactivity of the dC10 tag. To that end, we re-engineered the dC10 tag through rational protein design. Mutant libraries were prepared through combinatorial mutation at specific positions of the helical tag sequence, and screened for their fluorogenic reactivity. In this way, we identified a novel sequence for a next-generation dC10 tag that confers 10-fold greater selectivity that we then applied to *in cellulo* labelling. Subsequent mechanistic studies revealed the basis for this dramatic increase in reactivity.

Current applications of this powerful labelling technique, including the site-specific chelation of lanthanide ions for NMR spectroscopy and site-specific covalent heavy-atom labelling for X-ray crystallography, will also be discussed.

Acknowledgements

First and foremost, I would like to express my gratitude to my super-supervisor, Professor Jeffrey Keillor, for taking me on board in his lab at Université de Montréal five years ago. He has always been not only an immense source of knowledge and support to the members of his lab but also of great humor, motivation and good mood. He is an outstanding mentor and truly cares about the success of his students and goes out of his way to support them and help them grow.

My sincere thanks go to my co-supervisor, Professor Natalie Goto, who is a real NMR guru, and has been a great support during the transition period between Université de Montréal and University of Ottawa. I thank her for all the time and effort she dedicated to consulting on my NMR project, and for her guidance and advice with respect to writing and editing of the NMR chapter.

Thanks to the members of the Keillor group, present and past, who know how to create a great ambiance in the lab. I appreciate Hugo's guidance at the beginning of my PhD, and his, Olivier's, Christophe's and Amina's – literally – infectious humor that made my adjustment to a new country and a new lab a pleasant and fun experience. A genuine thank-you goes to Kim, Sam, Dan, Abdullah and Kelvin who made the lab (and office!) the best place to work the past few months of my PhD. Kim, thank you for all our francophone dessert breaks and trips to Première Moisson; Sam and Chris, thank you for the (sometimes much too) open and insightful conversations; Sam, I believe that you truly deserve your own TV show.

Great thanks go to our collaborator, Professor Albert Berghuis (McGill University) who accepted me in his lab for my initial work on the crystallization project; and to Jonathan Blanchet and Michelle McEvoy who have been very resourceful and patient teaching me the art of protein crystallization.

I would like to express my appreciation to the department of Chemistry and Biomolecular Sciences, in particular past and present members of the Goto, Chica, Boddy and Ben labs who have always been friendly to me as a person and helpful in any equipment or research

related questions. A special shout-out goes to Allison, Laura, Tabussom, Alex, Jason, Curtis, Jamie, Luis, Mark, Darija, Mariya and Jennie. I am thankful to Phil Pelletier and Andrew Ochalski for their help in the CAREG common lab and training in microscopy. Cheers to Annette and Josée from the departmental office; and to chemistry and biochemistry undergraduate lab coordinators who have been great bosses to us, TAs.

Last but certainly not the least, a very special thanks goes to my dance mates and teachers for non-science related fun and inspiration; to Sarang, who is the best roomie and friend ever; and to Robert, who has been very supportive and caring during the final year of my PhD, and who helped me to maintain balance in life.

“Droit devant soi on ne peut pas aller bien loin.”

— Antoine de Saint-Exupéry, *Le Petit Prince*

“Straight ahead you can't go very far.”

— Antoine de Saint-Exupéry, *The Little Prince*

Table of contents

List of Abbreviations

Symbol	Description
3D	three-dimensional
Å	angstrőm
FT	Fourier transform
AAC	amino-glycoside acetyl-transferase
AcCoA	acetyl coenzyme A
ACP	acyl carrier protein
$AmSO_4$	ammonium sulfate
ArgN	*E. coli* arginine repressor
Arr	aminoglycoside response regulator of *Pseudomonas aeruginosa*
BirA	*E. coli* biotin ligase
BMRB	Biological Magnetic Resonance Data Bank
BODIPY	boron-dipyrromethene
bp	base pair
BRS	biotin recognition sequence
BSA	bovine serum albumin
CD	circular dichroism
CHES	2-(Cyclohexylamino)ethanesulfonic acid

CMV	cytomegalovirus
CoASH	coenzyme A (free thiol form)
D_2O	deuterium oxide
dCx	dicysteine helix with dimensions of x angstrőms
DIPEA	diisopropylethylamine
dMF	dimaleimide fluorogen
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E	elution fraction
E. coli	*Escherichia coli*
E. faecium	*Enterococcus faecium*
EDA	ethylene diamine
eDHFR	*E.coli* dihydrofolate reductase
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EGFR, ErbB1	epidermal growth factor receptor
Em.	emission wavelength
ER	endoplasmic reticulum
ESI	electrospray ionization
Ex.	excitation wavelength

Fab	Fragment antigen-binding protein
FITC	fluorescein isothiocyanate
FlARe	Fluorogenic Addition Reaction
FlAsH	Fluorescein Arsenical Hairpin binder
FP	fluorescent protein
FRET	Főrster resonance energy transfer
FT	flow-through
FXa	Factor Xa
GB1	B1 domain of Streptococcal protein G
GFP	green fluorescent protein
GSH	glutathione
H2B	histone 2B
hAGT	human O^6-alkylguanine alkyltransferase
HALO tag	haloalkane dehalogenase
HCl	hydrochloric acid
HEK	human embryonal kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	hexahistidine tag
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
HSQC	heteronuclear single quantum coherence

I	induced sample
IDT	Integrated DNA Technologies
IPTG	isopropyl β-D-1-thiogalactopyranoside
JCSG	Joint Center for Strucural Genomics
K_2HPO_4	potassium phosphate dibasic
kDa	kilo Dalton
KH_2PO_4	potassium phosphate monobasic
LAP	lipoic acid ligase acceptor peptide
LB	Luria-Bertrani medium
LC-MS	liquid chromatography-coupled mass spectrometry
LiCl	lithium chloride
LnC01, 2	lanthanide chelating probes 1 and 2
LRMS	low resolution mass spectrometry
LUMO	lowest unoccupied molecular orbital
MAD	multiple-wavelength anomalous diffraction
MALDI	Matrix Assisted Laser Desorption Ionization
MBP	maltose-binding protein
MCS	multi-cloning site
MEM	Minimum Essential Media
MES	2-(*N*-morpholino) ethanesulfonic acid
$MgSO_4$	magnesium sulfate

MIR	multiple isomorphous replacement
mM	millimol per liter
MPA	2-mercaptopropionic acid
MWCO	molecular weight cut-off
NaCl	sodium chloride
NEB	New England Biolabs
NI	non-induced sample
NiNTA	nickel-nitrilotriacetic acid
nm	nanometer
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
nPAGE	native polyacrylamide gel electrophoresis
OD	optical density
P	pellets (insoluble fraction)
p75ICD	intracellular domain of the neurotrophin receptor p75
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PCS	pseudo-contact shift
PDB	Protein Data Bank
PDCA	pyridine - 2,6 - dicarboxylic acid (dipicolinic acid)
PEG	polyethylene glycol

PEI	polyethylene imine
PeT	photoinduced electron transfer
PMSF	phenylmethylsulfonyl fluoride
POI	protein of interest
PpiB	peptidyl-prolyl *cis-trans* isomerase B
PPTase	4'-phosphopantetheinyl transferase
PRE	paramagnetic relaxation enhancement
RCSB	Research Collaboratory for Structural Bioinformatics
RDC	residual dipolar coupling
ReAsH	Red-emitting Arsenical Hairpin binder
RFP	red fluorescent protein
RNA	ribonucleic acid
RNase A, H	ribonuclease A or H
RT	room temperature
S	supernatant (soluble fraction)
S. cerevisiae	*Saccharomyces cerevisiae*
SAD	single-wavelength anomalous diffraction
SDS-PAGE	sodium dodecyl sulfate – polyacrylamide gel electrophoresis
SIR	single isomorphous replacement
SIRAS	single isomorphous replacement with anomalous scattering
SlyD	sensitive to lysis D peptidyl-prolyl *cis-trans* isomerase

TBS	Tris buffer saline
TCEP	tris(2-carboxyethyl)phosphine
TOCSY	total correlation spectroscopy
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer ribonucleic acid
TROSY	transverse relaxation-optimized spectroscopy
Ubi	ubiquitin
UV	ultra-violet
W	unbound proteins (wash fraction)
wt	wild-type

List of Figures

List of Tables

List of Equations

List of Schema

Chapter 1

INTRODUCTION TO PROTEIN LABELLING

1.1. Importance of protein labelling

The human genome was completely sequenced and genes were entirely mapped as of 2012 (*www.genome.gov*). However, given the variability of gene products, such as RNA and proteins, genome sequencing is not the ultimate piece of information necessary to understand the role of each gene, which remains an immense challenge. Different protein labelling techniques are required to identify and assess the role of proteins in a cell and understand their relevance, mostly following protein biosynthesis and localization, interaction with other proteins, and subsequent degradation. Fluorescent labelling of proteins has a certain number of advantages, relying on its sensitivity, limit of detection and ease of detection.

1.1.1. Fluorescence

In general, certain molecules are capable of emitting light by relaxation of an electron from an excited state to the ground state. This phenomenon is called luminescence and is formally divided into two categories - fluorescence and phosphorescence - depending on the nature of the excited state of the molecule. In fluorescence, the electron in the singlet excited state orbital is paired with the second electron in the ground state orbital. In consequence, the excited electron can return to the ground state more rapidly, by emission of a photon. On the other hand, in phosphorescence, the triplet excited state orbital electron has the same spin as the ground state electron, and cannot undergo fast direct relaxation, resulting in much slower emission rates.

More precisely, fluorescence occurs in three stages, as illustrated in **Figure 1.1.** (*left*). First, an electron is promoted from the ground state S_0 to the singlet excited state S_1' after absorption of energy $h\nu_1$. Then, the energy of this excited state S_1' is rapidly partially dissipated, leading to the excited state S_1 of lower energy. Finally, from this state, the electron is relaxed back to the ground state S_0 by emitting a photon of energy $h\nu_2$. Due to the loss of energy between states S_1' and S_1, the emitted energy is smaller than the absorbed energy, hence the frequency ν_1 is always higher than the frequency ν_2, corresponding to a longer wavelength λ_2 of the emitted photon, as illustrated in **Figure 1.1.** (*right*) for fluorescein.

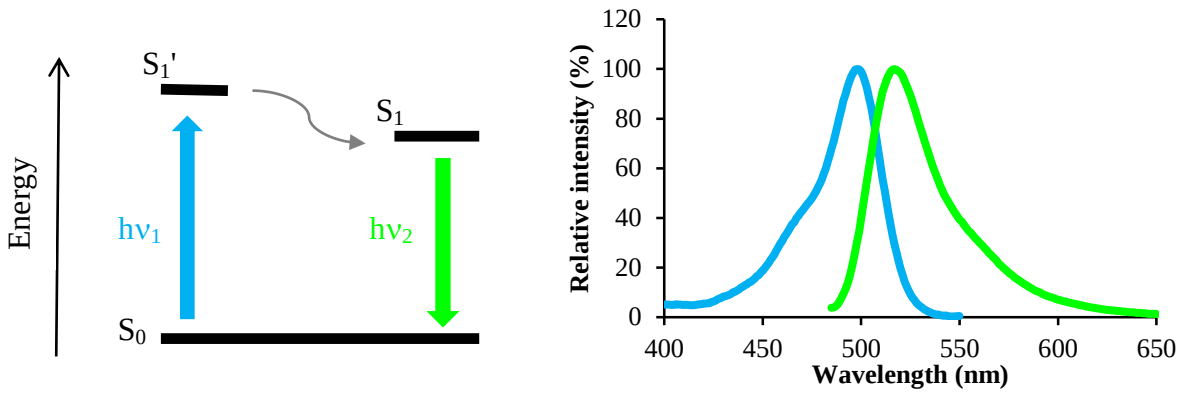


Figure 1.1. Jablonski diagram illustrating the processes involved in fluorescence (*left*) and excitation (*blue*) and emission (*green*) spectrum of fluorescein (*right*) (*www. thermofisher.com*)**.**

1.1.2. Rules and advantages of fluorescence

It is not easy to predict if a chemical compound is fluorescent; however, in general, fluorophores have several common features, including polyaromatic rings, heterocycles [1, 2, 3] and both, electron-donating and electron-withdrawing groups that allow a push-pull structure to be fluorescent [4]. Using fluorescence as a detection method allows achieving a high limit of sensitivity, commonly reaching analyte concentrations of one part per billion, and in ideal cases attaining the concentration limit of one part per trillion. In general, fluorescence methods can perform 1000 to 500 000 times better in terms of detection limit compared to absorption spectrophotometers.

Moreover, many common materials can absorb light, which complicates the measurements using spectrophotometric techniques; however, in the case of fluorescence, autofluorescence of used material or other components of the system is much less frequent. This allows analyte detection with a much higher selectivity than in the case of spectrophotometric measurement.

1.2. Protein labelling techniques

Protein fluorescent labelling techniques rely on the ease of detection that can be performed by a bare eye or a sophisticated fluorescence microscope, depending on the level of resolution that one seeks. The emission characteristics of available fluorophores also offer a large variety of complementary colours that expand the versatility of fluorescent labelling. Lastly, the techniques that are currently used and in development and that will be summarized in this section include use of intrinsically fluorescent proteins or small fluorescent or fluorogenic molecules.

1.2.1. Green fluorescent protein

One of the most widespread methods for the fluorescent labelling of proteins is the genetic fusion of an intrinsically fluorescent protein (FP, [5]) to a protein of interest (POI), where tracking the intrinsic fluorescence of the FP allows to identify the localization of the POI . This was made possible through the isolation and cloning of the native FP from *Aequorea victoria* by Chalfie *et al.* [6]. Since then, remarkable work has been done on studying [7, 8, 9] and improving the stability, brightness and folding of *A. victoria* green fluorescent protein (GFP, **Figure 1.2.**, [8, 9]). A high number of GFP mutants, emitting a wide range of colours [10, 11, 12] have also been developed, allowing the simultaneous use of multiple fluorescent proteins to track several proteins in a cell at the same time, yielding a broad scale of colourful tools for cell biologists.

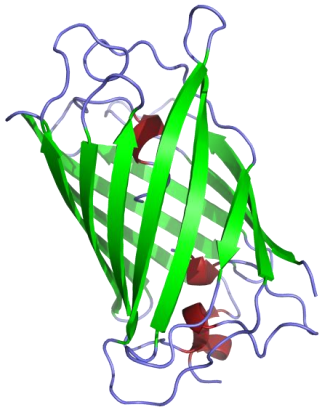


Figure 1.2. Crystal structure of GFP (PDB 1EMA, [5]).

Despite these breakthroughs, this technique still presents several limitations related to the intrinsic properties of the FP: it is a 27-kDa protein that tends to form multimers and aggregate [10, 13, 14], it folds rather slowly [7] and is thought to possibly perturb intracellular trafficking [15]. Although several GFP mutants have been discovered that are strictly monomeric [11, 8], they can still present size-related issues. It has been shown for a number of cases that attaching a FP to a protein of interest lead to its mislocalization [15, 16]. Lastly, the intrinsic function of the fusion protein may perturb the function of the protein of interest, resulting in erroneous localization of the complex.

1.2.2. Small molecules as fluorescent labels

Over the last 15 years, efforts have been made (detailed in review [17]) to overcome some of the limitations of GFP through a combination of genetic modification of the target protein and the use of small molecules. For example, specific amino acids may be introduced into the POI by mutagenesis, or a peptide or protein tag can be fused to its *C*- or *N*-terminus, thereby allowing it to react specifically in a covalent or non-covalent manner with a small molecule.

Several recent reviews summarize and compare the most recently developed small molecule-based labelling methods [17, 18, 19, 20]. Ideally, newly emerging chemical-biology based labelling approaches make use of very small fusion tags that do not perturb the POI, and minimal, non-toxic small molecules that react quickly and specifically to label the fusion tag.

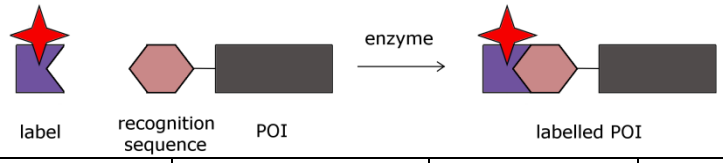
Briefly, the small molecule-based labelling methods can be divided into two groups with respect to the type of interaction between the target protein or tag and the small molecule, depending on whether the tag is recognized and modified by an enzyme, or whether the recognition and reaction is non-enzymatic. Among the widely applied enzyme-mediated methods, transglutaminase [21, 22], sortase [23, 24], phosphopantetheinyl transferase (PPTase) [25, 26, 27], lipoc acid ligase LplA [28] and biotin ligase BirA [29, 30] have been used to modify, respectively, a Q-tag, LPXTG-tag, acyl-carier protein (ACP)-tag, LAP-tag or BRS-tag fused to a POI (**Table 1.1. section 1**). Alternatively, enzymes such as hAGT [31, 32], HALO-tag [33, 34], CLIP-tag [35] and SNAP-tag [31, 36] may be fused to protein of interest and subsequently labelled with a fluorescent irreversible inhibitor designed to

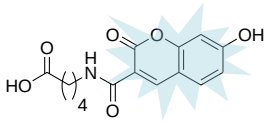
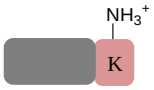
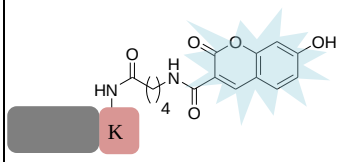
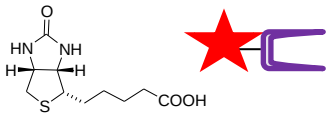
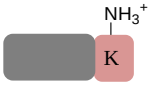
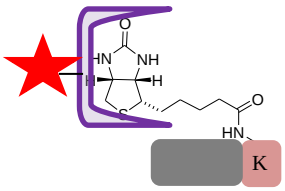
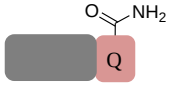
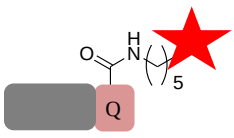
show specificity for the fused enzyme tag (**Table 1.1. section 2**). Non-enzymatic labelling methods generally rely on the introduction of an unnatural amino acid [37] or the special reactivity of genetically encoded sequences of, for example, histidine or cysteine residues (**Table 1.1. sections 3 and 4**), with a fluorescent or fluorogenic probe, or alternatively, on protein splicing (**Table 1.1. section 5**).

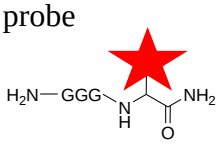
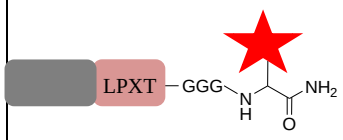
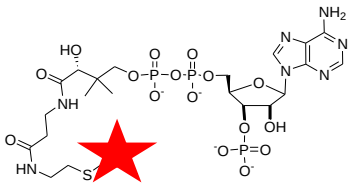
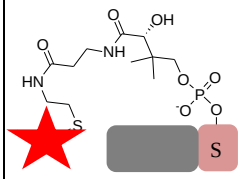
The ultimate goal of these techniques is to present a tool that is at least complementary to, or even better than, the use of FP fusion in terms of specificity, toxicity, cellular perturbation and rapidity.

Table 1.1. Alternative methods for fluorescent (fluorogenic) labelling of proteins.

1. <u>Enzyme-mediated labelling</u>

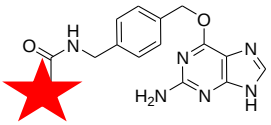
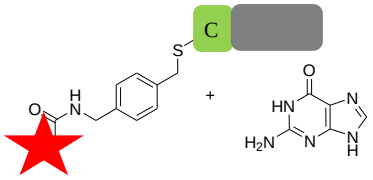
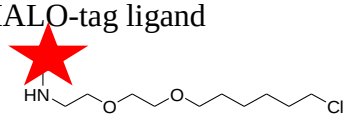
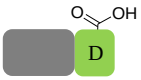
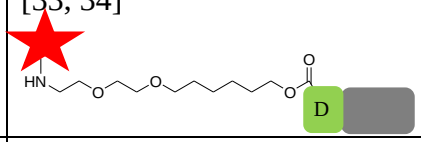
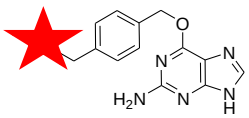
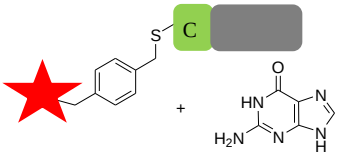
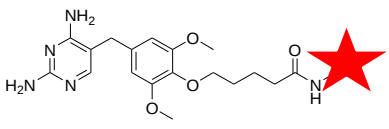
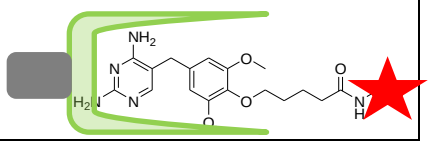


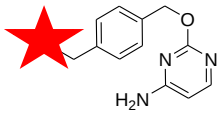
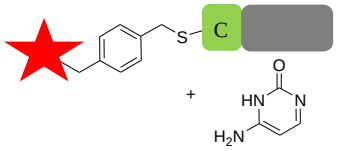
Labelling molecule	***Enzyme***	***Recognition sequence***	***Reference, products***
Coumarin derivative 	LplA W37V	LAP-tag 	[28, 38] 
Biotin and streptavidin attached to a fluorophore 	BirA biotin ligase	BAP-tag 	[29, 30, 39] 
Cadaverine functionalized probe 	Transglutaminase	Q-tag 	[21, 22, 40] 

Oligo-Glycine derivative probe 	Sortase	LPXTG-tag 	[23, 24, 41] 
Coenzyme A probe 	PPTase	ACP-tag (acyl carrier protein) 	[25, 26, 27] 

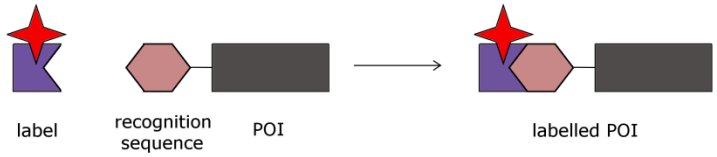
2. Enzyme-substrate analogue recognition-based labelling

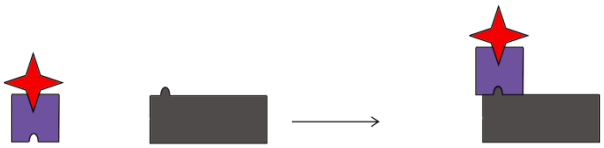
substrate derived label | POI | labelled POI

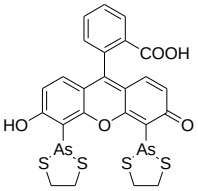
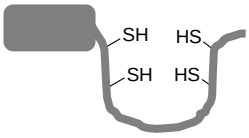
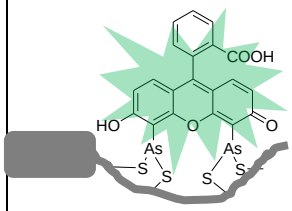
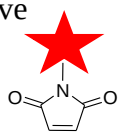
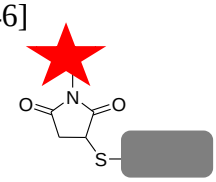
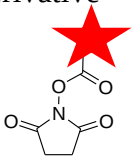
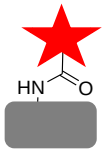
Labelling molecule	*Enzyme*	*Recognition sequence*	*Products, reference*
O6'-Alkylguanine derivative 	hAGT	hAGT-tag 	[31, 32] 
HALO-tag ligand 		HALO-tag 	[33, 34] 
Alkylguanine derivative 		SNAP-tag 	[31, 36, 35] 
Fluorescent MTX derivative 	eDHFR	eDHFR tag 	[42] 

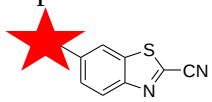
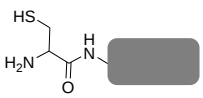
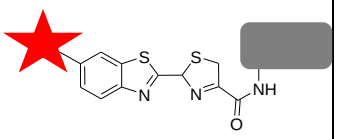
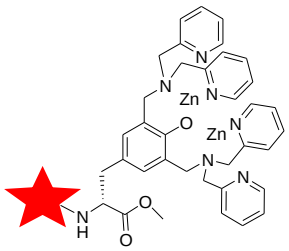
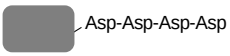
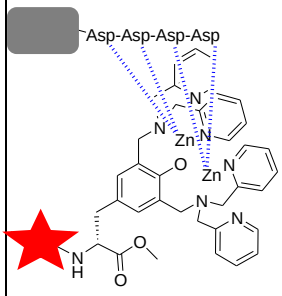
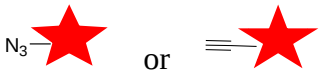
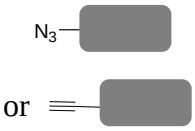
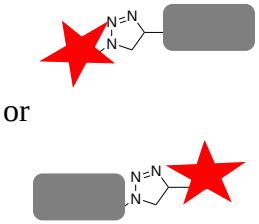
Alkylcytosine derivative 		CLIP-tag 	cytosine, [35] 

3. Small molecule chemical recognition labelling

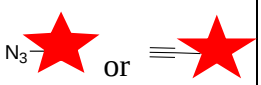
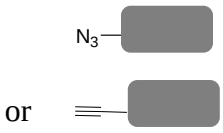
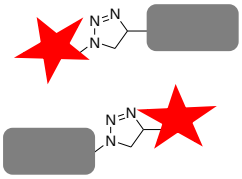
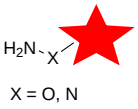
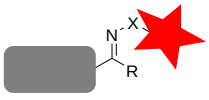


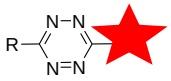
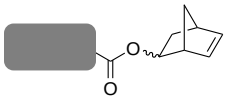
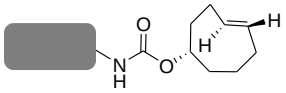
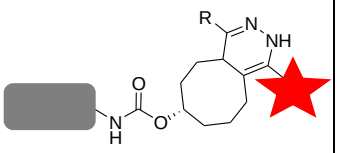
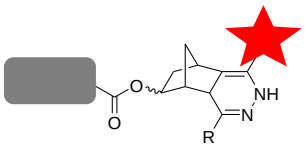


Labelling molecule	***Target***	***Remarks***	***Reference, product***
Metallic arsenic attached to a Fluorescein or Rhodamine fluorophore (FlAsH, ReAsH) 	Tetracysteine tag (CCPGCC) 	Weak bond, compound toxicity	[43, 44, 45] 
Fluorescent Maleimide derivative 	Intrinsic cysteine 	Not specific enough	[46] 
Fluorescent succinimidyl ester derivative 	*N*-terminal amine 	Not specific enough	[47] 

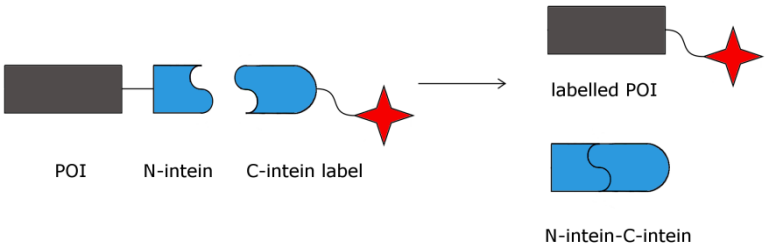
2-cyanobenzothiazole (CBT) attached to a fluorophore 	*N*-terminal cysteine 		[46, 48, 49] 
Ni^{2+} (nitriloacetic acid), Zn^{2+} or Ln fluorophore complex 	Hexahistidine, oligo-aspartate and lanthanide-binding tag 	Elevated cytotoxicity	[50] 
Alkyne (azide) functionalized fluorophore 	Azide (alkyne) 	Cell surface	[51, 52, 53] 

4. <u>Unnatural amino acid mediated labelling</u>

Label	***Target amino acid***	***Remarks***	***Reference, product***
Alkyne/azide derivative 	Azide/alkyne functionalized aminoacids 	Needs a specific tRNA synthase, bioorthogonal	[37] 
Hydrazide carrying fluorophore 	Ketone 	Needs a specific tRNA synthase , cell surface, slow	[54] 

Tetrazine derivative 	Norbornene or *trans*-cyclooctene functionalized amino acids  or 	Needs a specific tRNA synthase, bioorthogonal	[37, 55]  or 

5. <u>Protein splicing</u>



Label	***Recognition sequence***	***Reference***
C-intein attached to a fluorophore	*N*-intein on *C*-terminus of POI	[56, 57]

1.3. FlARe labelling technique

1.3.1. Fluorescent and fluorogenic labelling

The above detailed non-enzymatic, small molecule protein labelling techniques are mostly fluorescent, as opposed to fluorogenic (fluorescence generating), i.e., the small molecule used for labelling is intrinsically fluorescent before reacting and presents a high background. This undesired background signal is usually dealt with by using post-labelling washing steps to remove the excess of unbound fluorescent label, which certainly helps to remove some background fluorescence, but in the case of the Tsien [58, 59] tetracysteine labelling method with arsenic derivative (**Table 1.1. section 3**), for example, it can also dissociate the weak metal-sulfur bond of an already labelled protein and decrease the desired fluorescence of the labelled complex.

Nowadays, many research groups are interested in developing fluorogenic labelling methods to overcome the challenge of a non-specific background signal of the highly fluorescent label, and a number of suitable methods emerged from this effort [60]. In general, the fluorogen activation can be achieved through a) a FRET (Főrster resonance energy transfer) quenching mechanism (**Figure 1.3.,** *(A,B)*) [61, 62], where the unreacted labelling agent contains a fluorophore attached to a quencher through a sequence recognized specifically by the POI or a fusion protein; b) by direct interaction of the fluorophore with its target using the environment sensitivity of the fluorophore (**Figure 1.3.** *(C)*) [63, 64, 65]; or c) by change in spectral properties induced by the covalent modification, as shown in **Figure 1.3.,** *(A) bottom* [66, 67, 68]. Examples of the molecules exhibiting the mentioned phenomena a) and b) are shown in **Figure 1.3.** (B,C), respectively, and FlaRe labelling described in this thesis is an example of phenomenon c).

The Keillor group has developed a fluorogenic protein labelling method that was designed to meet all above mentioned criteria of a good protein labelling method, and also to address the difficulties that other methods encounter. Furthermore, we sought to design a one-step method that would not necessitate introduction of other unnatural functional groups in proteins, and make the design more complex. The resulting Fluorogenic Addition Reaction

(FlARe) that uses a quenched small molecule and a small peptide tag will be described in detail in the following section.

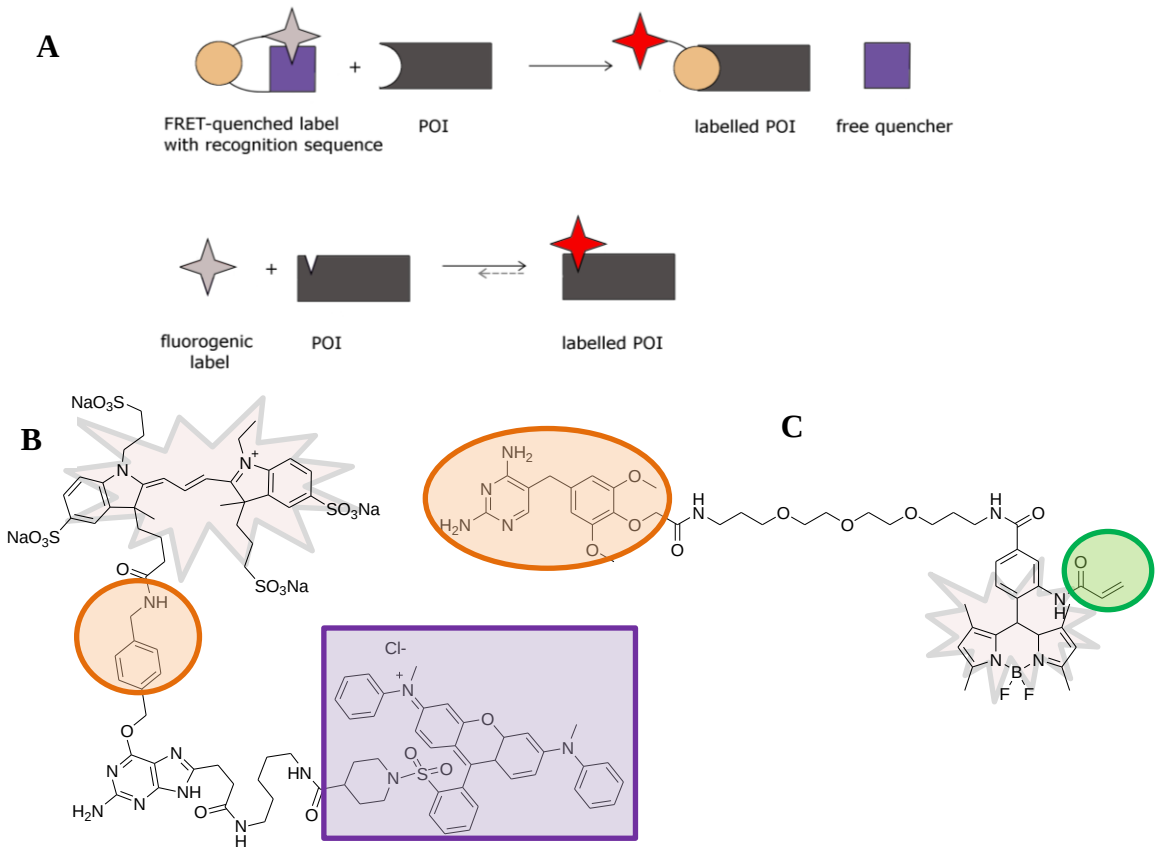


Figure 1.3. Fluorogenic labelling approaches. *(A) Top*: A FRET quenched molecule can form a covalent adduct on the POI through its recognition sequence (*orange sphere*), releasing the quencher (*purple square*). *(A) Bottom*: Direct binding of a fluorogenic label to a POI or a tag can change its spectral properties and activate it. The fluorophore or fluorogen is represented as a *grey* (quenched) or a *red* (unquenched) *star*. *(B):* Structure of CBG-549-QSY7 as an example of FRET-quenched molecule **[62]**. *(C):* Structure of TMP-AcBODIPY as an example of fluorophore quenched by its environment **[67]**. In *green* is shown acrylamide group that allows the BODIPY moiety to bind eDHFR more strongly, causing a fluorescence "turn-on".

1.3.2. General design of FlARe

Maleimides are known to react highly selectively with thiols in near-neutral conditions, and used especially for labelling peptide thiols [69], and they are also known to be able to quench fluorescence in their conjugated form [70]. At the same time, cysteines are relatively

under-represented amino acid residues in naturally occurring proteins, and are often present only in active sites or as tertiary structure features. Based on these three facts, we have developed a new non-enzymatic small-molecule labelling method, based on the spontaneous, uncatalyzed Fluorogenic Addition Reaction between a fluorogenic labelling agent and a peptide fusion tag (**Figure 1.4.**). More specifically, the two cysteines in the short peptide tag sequence react specifically with the two maleimide groups of a fluorogenic molecule comprising a dimaleimide moiety and a fluorophore ([71, 72], see **Figure 1.5.**). This small molecule is not fluorescent due to the spatial proximity of the maleimide groups to the fluorophore, but it becomes fluorescent after the reaction of both its maleimide groups with thiols (**Figure 1.4.**), forming two succinimides. Our peptide tag was likewise designed to contain two cysteine residues separated by two turns (~ 10.8 Å) of a short helical peptide tag sequence; such that their side chain thiol groups are separated by roughly the same distance as that between the maleimide groups of our fluorogenic molecule (~ 7 - 9 Å), as shown in **Figure 1.5.** The helical secondary structure of dC10 tag offers several advantages over a random coil conformation, for example, it maintains the reactive cysteines apart to prevent formation of a disulfide bond, and it exposes the cysteine side chains to the solvent milieu. While the distance of 10.8 Å between two cysteine residues has been determined using the geometry of an α-helix, the possible distances between electrophilic carbons of the dimaleimide were measured on a structure of minimized energy using HyperChem, and molecular modelling. It is important to note that the distance-based design is limited by the orbitals involved in the reaction, where the nucleophilic attack of the thiolate is perpendicular to the C=C bond.

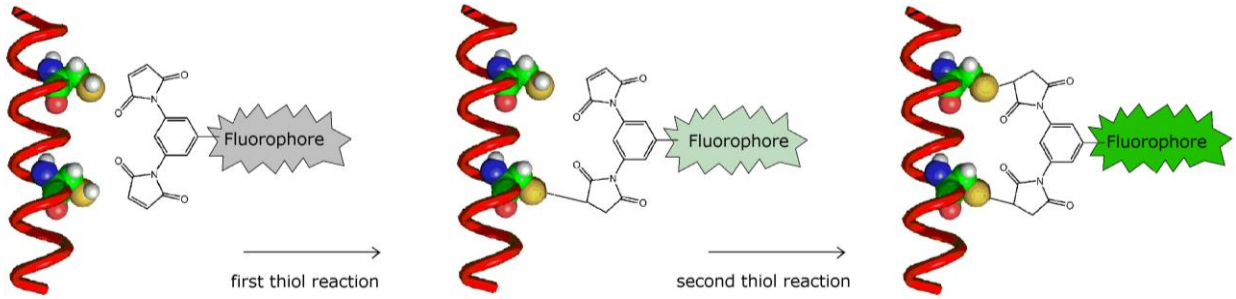


Figure 1.4. Fluorogenic addition reaction scheme. Addition of a first thiol from dicysteine peptide produces a slightly fluorescent complex (*middle*); fluorescence is completely restored only after a second thiol is added on dimaleimide fluorogen molecule (*right*).

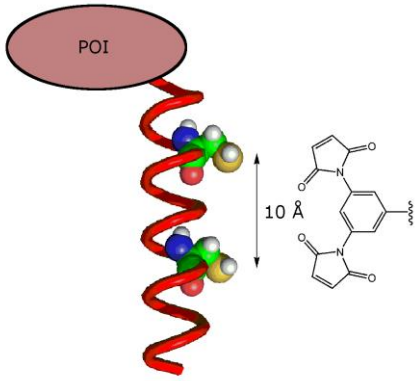


Figure 1.5. Design of FlARe labelling components. Di-cysteine tag (tag in *red*, cysteines in *yellow*) linked to a protein of interest (POI) reacts with a dimaleimide moiety of a fluorogen (*right*) with adequate distance of 10 Å between the two cysteines and between the two maleimide groups.

1.3.3. Design and development of fluorogenic molecules for FlARe

The efficiency of the FlARe method is closely related to the capacity of dimaleimide moieties to quench the fluorescence of the pendant fluorophore (**Figure 1.6.**). In an unreacted fluorogen, the LUMO of maleimide groups is positioned between the ground and the excited state of the attached fluorophore. When an electron of the fluorophore is promoted from the ground state to the excited state, the relaxation occurs through the intermediately positioned LUMO of maleimide, and is non radiative. Such a quenching mechanism is referred to as photoinduced electron transfer (PeT). After a maleimide group has reacted with a thiol, the LUMO of the resulting succinimide is promoted to an energy level that is higher than that of the excited state of the attached fluorophore, allowing a radiative relaxation of the excited state electron directly to the ground state (**Figure 1.6.**).

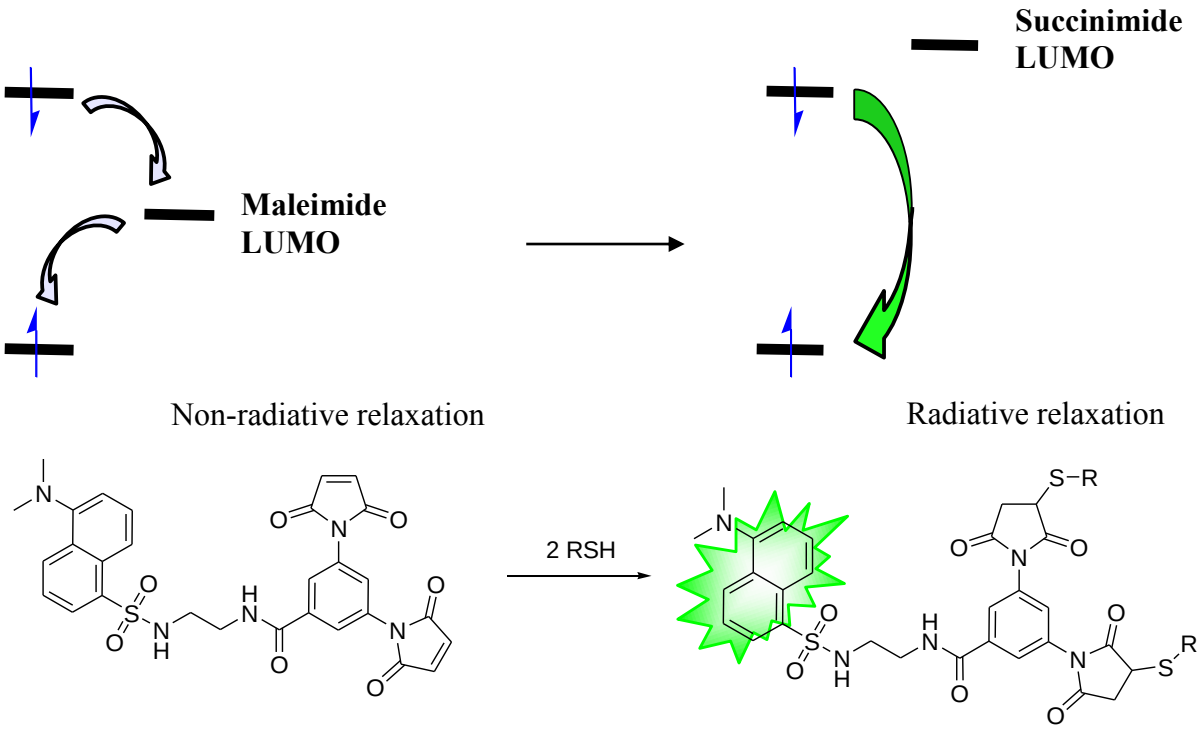
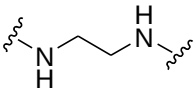
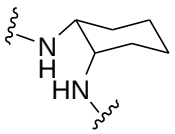
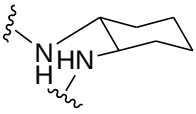


Figure 1.6. Photoinduced electron transfer (PeT) quench mechanism in dimaleimide fluorogen markers. *Left*: LUMO of maleimide is situated between the ground and excited states of the attached fluorophore. *Right*: after reaction with thiols, the LUMO of the resulting succinimide is high in energy and allows a direct radiative relaxation of the excited state electron.

With this quench mechanism and using a dimaleimide fluorogen with a di-cysteine peptide, the design results in a unreacted compound that has a negligible fluorescence, after reaction of one thiol with one maleimide group a very weak fluorescence may be observed, and it is only after both maleimide groups have reacted with two cysteine thiols that the maximum of fluorescence is restored (**Figure 1.4.**). As a demonstration of the quenching efficiency of dimaleimide fluorogens, the quantum yields were previously determined for each isolated and purified intermediate of the thiol addition reaction, namely, for the unreacted compound, the compound that underwent addition of one thiol, and the compound reacted with two molecules of thiol. On the example of dM10-EDA-dansyl **1** (**Table 1.3.**), it was shown that the quantum yield increased for both the unreacted compound and the mono-thiol adduct (quantum yield 0.04 for both), to 0.22 for its di-thiol adduct [72], showing the importance of the reaction of both maleimide groups for recovery of fluorescence.

In previous work, Karine Caron and Virginie Lachapelle have studied the detailed design of dimaleimide fluorogens [71, 72, 73, 74] and found that the efficiency of quenching (and therefore the magnitude of the fluorescence enhancement upon labelling) is dependent on the through-space distance between the maleimide and fluorophore (**Table 1.2.**). Upon reaction of fluorogens shown in **Table 1.2.** with an excess of 2-mercaptopropionic acid, the best quenching efficiency of 5.90 and 4.76 was observed for fluorogens where the dimaleimide moiety was attached to the coumarin fluorophore through 1,2- diaminocyclohexane, positioning the quencher and the fluorophore the closest together amongst all examined spacers [74]. Further variables impacting quenching efficiency, and by consequence fluorescence enhancement, include the conformational rigidity of the system and the relative energies of the excited state orbital (that depends exclusively on the fluorophore properties) and the unoccupied orbital on the maleimide group [72].

Fluorescence enhancement	**Through-space distance (Å)**	**Spacer X**
3.17	15.6 ± 0.9	
4.76	13.4 ± 1.4	
5.90	14.1 ± 2.4	
1.84	16.8 ± 1.5	

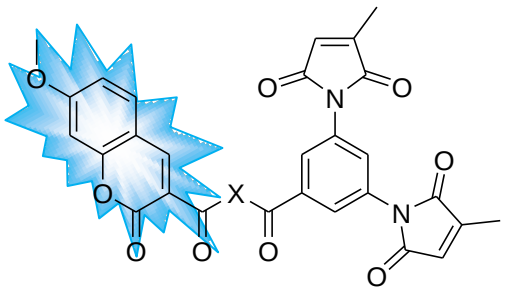
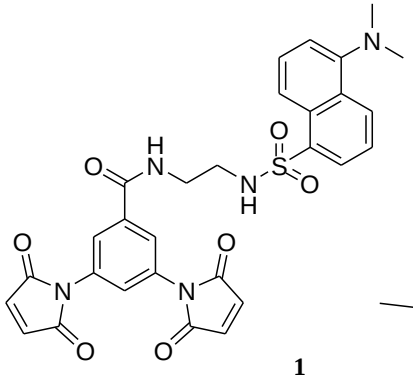


Table 1.2. Spacer-dependent fluorescence enhancement of dM10-coumarin (*right*) upon reaction with excess of 2-mercaptopropionic acid (MPA). (adapted from [74]). Fluorescence was determined as a ratio of fluorescence intensity (peak height), measured at 405 nm before and after reaction, upon excitation at 347 nm.

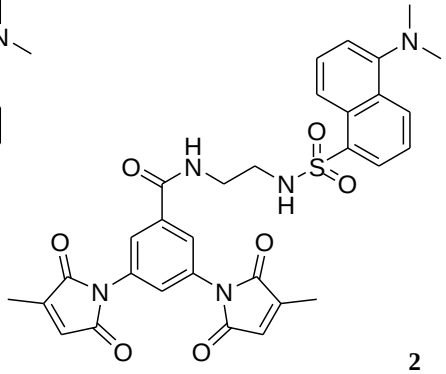
In the Keillor group, we have also investigated the influence of substituents on the maleimide double bond on the reaction rate and selectivity. It was previously established that methyl substituents significantly slow down the reaction of a fluorogen with our di-cysteine peptide tag [73] and, more importantly, this substitution drastically decreases the rate of reaction with glutathione (GSH) that is present in millimolar concentrations inside a cell (**Table 1.3.**). Very recently, further investigative work has been done (Élise DeFrancesco, Kelvin Tsao and [75]) on the maleimide double bond substituent effect on the reaction rate, where alkyl and alkoxy [75] substituents were attached on the maleimide double bond in order to determine which substituent provides the best selectivity for our di-cysteine tag over glutathione or water (unpublished data). So far, the methoxy- substituent seems to be the best candidate for intracellular labelling [75], where selectivity of reaction with the target peptide has to out-compete a high concentration of glutathione. However, more investigation is in progress to determine whether a better substituent can be used amongst those mentioned above.

Table 1.3. Kinetic rate constant attenuation observed for MBP-dC10 and GSH using a methyl-substituted dimaleimide *vs.* an unsubstituted dimaleimide (adapted from [73]). Values were obtained by dividing the rate constant of an unsubstituted dM10-EDA-dansyl **1** (*below left*) by dimethyl dM10-EDA-dansyl **2** (*below middle*), with MBP-dC10 and GSH (*below right*).

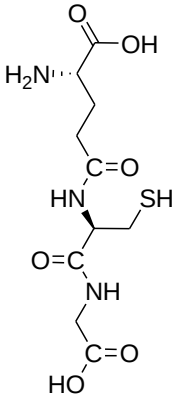
Thiol	**Rate constant attenuation between fluorogens 1 and 2** [73]
MBP-dC10	59
GSH	762



1



2

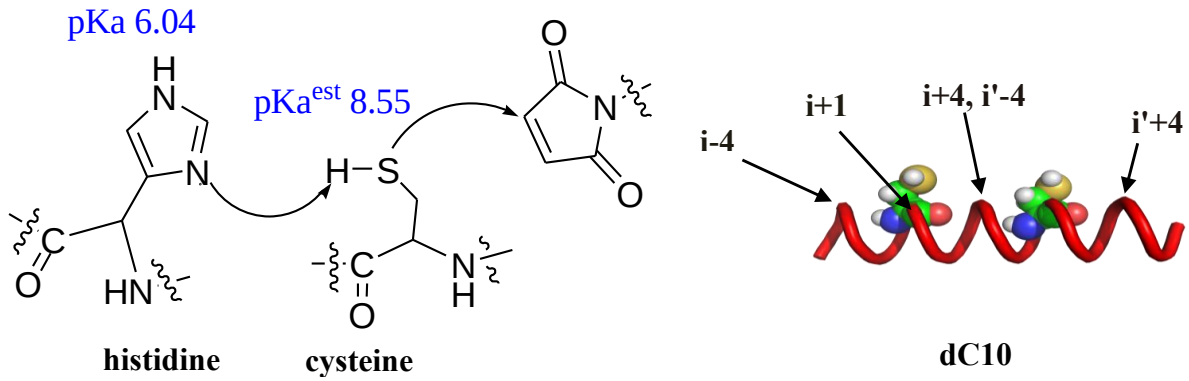


1.3.4. Design and first development of di-cysteine target peptide for FlARe

In the past, a small helix from the protein Fos was used for the first proof of principle of the FlARe labelling reaction [71], but later a *de novo* design of a di-cysteine peptide tag [73] that would allow a complete bioorthogonality with any cellular protein behaviour was proposed. Inspired by a synthetic peptide adopting a stable monomeric helix Ac-EAAAREAAAREAAARQ-NH_2 [76, 77], a *de novo* di-cysteine tag was designed with a certain number of features, for example, the presence of salt bridges between the carboxylate and guanidinium groups of glutamate and arginine, respectively, that lock the peptide in a helical conformation and ensure hydrosolubility, to which were added known *C*-cap and *N*-cap that bring more stability to the small helix by preserving its dipole [78]. This sequence was slightly modified to contain an integer number of turns for better helix stabilization [78]. Two cysteine residues were introduced into the peptide sequence at a relative distance of 10.8 angstrőms at different (*i, i+7*) positions of the sequence, to perfectly complement the design of dimaleimide fluorogen molecules whose reactive maleimide moieties are 10 angstrőms apart, as shown in **Figure 1.5.** The resulting sequences were then tested for predicted helicity by AGADIR algorithm [79, 80, 81] and the best sequence was chosen from this sampling. This di-cysteine tag LSAAE**C**AAREAA**C**REAAARAGGK, referred to as dC10 for "di-Cysteine with a distance of 10 Å", was used as a reference point for the characterisation of most of the fluorogen labelling molecules synthesized in the Keillor lab. Importantly, this dC10 peptide was shown to be helical by CD before and after reaction with fluorogen **dM10-FITC** (**Table 1.4.** *right*, [73]). In addition, other groups designed peptide sequences containing two cysteines and showed them to adopt helical conformation [82].

Additionally, our group tried to enhance the reactivity of dC10 cysteine residues by rational design where histidine residues were introduced in proximity of the reactive thiolates, i.e., one helix turn apart (*i* + 4, *i* - 4) and in neighbouring positions (*i* + 1), as shown in **Scheme 1.1.** Similarly to the mechanism of cysteine proteases [83, 84] it was thought that at the physiological pH of 7.5, nearby histidine side chains could be partially neutral (as opposed to its prevalent protonated state at lower pH due to the pKa of their side chain of 6.04 [85]) and able to deprotonate the nearby cysteine sulfhydryl side chain (estimated pKa of 8.55 in an alanine pentapeptide [86]), which could shift the cysteine thiol/thiolate equilibrium and

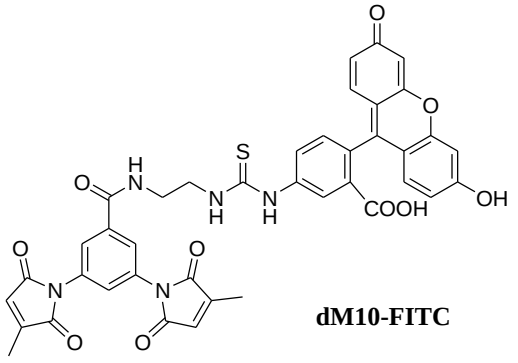
slightly favour more the thiolate form in comparison to the parent dC10 sequence. However, this attempt did not bring any success; for the new dC10 sequences were mostly far less reactive in a FlARe labelling than the parent dC10 sequence (**Table 1.4.** and [73]). In particular, mutants with histidine residues at position 9 were completely unreactive, which was thought to be a consequence of a complete helix destabilization by disruption of the salt bridge between the former arginine 9 and glutamate 5 in dC10 sequence. In general, histidine residues present a low helical propensity [87] and it was thought that this was the reason why all histidine dC10 mutants prepared were less reactive than the parent dC10 sequence.



Scheme 1.1. Evolution of dC10 sequence by rational design. Histidine residues were introduced in nearby positions from reactive cysteines (*right*) in order to enhance the proportion of reactive thiolates that attack the maleimide double bond. Indicated pKa values of histidine and cysteine side chains were estimated from [85, 86].

Table 1.4. Rate constants for reaction of dM10-FITC fluorogen with series of dC10. Adapted from [73].

Helix name	Position of histidine residues with respect to cysteines 6 and 13	k_2 ($M^{-1}s^{-1}$)
dC10		256
dC10-H2	C6, *i* - 4	43
dC10-H7	C6, *i* + 1	33
dC10-H9	C6, *i* + 4 or C13, *i′* - 4	N.R.[a]
dC10-H2H17	C6, *i* - 4 and C13, *i′* + 4	92



All kinetic studies (*left*) were performed at 20°C with 100 µM peptide in 50 mM HEPES (pH 7.5), 10% (*v/v*) DMSO and 100 µM of dM10-FITC fluorogen (*right*). The sample was excited at 495 nm and the fluorescence emission was followed at 525 nm as a function of time. [a] No reaction detected.

Even though no reaction enhancement was achieved through this rationally designed site-specific mutagenesis on dC10 peptide, this first attempt kept our interest and opened the door for more exploration and seeking of a more reactive and stable dC10 sequence.

1.4. Objectives of this thesis

Using small molecules to label proteins in cells is a large field that represents dozens of variations for particular applications, as described in Chapter 1. Despite the variability and richness of the small molecule labelling toolkit, several features still remain a challenge in this field, such as simultaneous labelling of multiple proteins in living cells, use of toxic catalysts for labelling and subsequent washing steps to remove excess of labelling molecule, or slow labelling over which it is difficult to maintain healthy cells. Acknowledging the above mentioned concerns, we will attempt to bring new options in the context of FlARe labelling in meeting the challenges, primarily in the following areas: 1) designing a set of peptide tag / fluorogen pairs that would allow an **orthogonal labelling**; 2) improving the design of the existing peptide tag sequence for a **faster** and **more selective labelling**; 3) applying of our technique for an ***in cellulo* labelling**; 4) **minimizing the peptide** tag to a set of mutations intrinsic to a protein; 5) extending the existing method for protein labelling with heavy metals for **structural biology**. These goals will be detailed and obtained results presented and discussed in the following chapters:

2. Towards an orthogonal FlARe labelling

In this chapter, a new design of di-cysteine peptide tags will be presented based on the helical properties of the first tag dC10, and their complementary dimaleimide fluorogens. The new tags will be characterized with the latest dimaleimide fluorogens dM10 (synthesized at that point in the Keillor group) and their kinetic profile and their potential as orthogonal tags will be determined. At the same time, newly designed and synthesized dimaleimide fluorogens will be used that, according to their design, could be complementary to di-cysteine peptides (other than dC10) and could potentially be used for orthogonal labelling.

3. Road to labelling in complex milieu

We acknowledge that *in vitro* and *in cellulo* (or *in vivo*) labelling represent different challenges related to the component stability (for the former) and to the system complexity (for the latter). In this chapter the labelling reaction, which had been well described *in vitro*, will be tested in more complex environment such as a bacterial lysate, a mammalian cell lysate and ultimately, in live cells. During this work we will speak to the questions of fluorogen selectivity for target peptide over other components of the milieu, such as glutathione or adventitious thiols.

4. Optimization of di-cysteine 10 peptide sequence

Despite the unsuccessful attempt that has been previously done on dC10 sequence optimization in order to obtain a more reactive target peptide, the rational design that was used with newly synthesized fluorogens in the Keillor group will be re-examined and a new plan of attack to obtain a better dC10 peptide sequence will be determined. First, the conformation of dC10 peptide using NMR will be investigated, after which the existing histidine dC10 mutants will be re-characterized with more recent fluorogens, and a small library of point mutants will be created. From there, the mutation positions will be expanded from one to three and the best mutant will be chosen amongst the investigated group. Along the way, the physico-chemical effects that lead to a better reactivity of a newly discovered dC10 peptide sequence will be determined. Ultimately, this new dC10 sequence will be employed for an *in cellulo* protein labelling.

5. Alternative use of dimaleimide-functionalized molecules for protein structure studies – protein NMR

In this chapter we will first introduce the field of protein NMR with emphasis on NMR of large proteins, and then use a newly designed dimaleimide molecules that bear a lanthanide chelating moiety and label proteins that we will study by NMR, in order to see if our dimaleimide probes can be applied in structural biology. At the same occasion, we will investigate a new minimalistic design of a di-cysteine peptide helix that is located on a solvent exposed helix intrinsic to the protein. Finally, we will discuss all results obtained and lessons learnt from this work.

6. Alternative use of dimaleimide-functionalized molecules for protein structure studies – protein X-ray crystallography

Similarly to Chapter 5, here we will briefly introduce the field of protein X-ray crystallography with emphasis on the phase problem and its solving, and propose to use dimaleimide molecules to that end. We will use a new dimaleimide probe that is attached to a palladium and we will investigate if this probe can be useful for simplifying the protein crystallization process and for solving the phase problem in solving a protein crystal structure. Ultimately, we will comment on our observations during this work and on lessons learnt.

N.B: Chapters 5 and 6 refer to subjects substantially different from Chapters 2 - 4 and they will be introduced separately.

Chapter 2

TOWARDS AN ORTHOGONAL FLARE LABELLING

2.1. Introduction

In this chapter, the term "orthogonality" will be used to describe two (or more) labelling events that are mutually exclusive. "Bio-orthogonality" refers to a labelling using a dye that reacts specifically with its target, and not with any other cellular component.

Although it is very easy to label several different proteins at the same time using the GFP-fusion strategy, thanks to a broad range of FP variants of different colours [8, 11, 88], orthogonal labelling is very difficult to achieve using standard small molecule approaches [35]. Enzyme-mediated labelling usually suffers from a poor recognition of orthogonal small molecule and target peptide pairs, but has only very recently been shown to work with protein prenyltransferases [89], and chemical labelling methods have not been optimized yet in order to achieve desired orthogonality. Hence, we propose a new design of the FlARe labelling technique that would allow to achieve an orthogonal and simultaneous protein labelling using dimaleimide fluorogen molecules and corresponding di-cysteine peptide tags.

2.1.1. FlARe orthogonal labelling design using di-cysteine helical peptides

The FlARe labelling technique described in Chapter 1 uses a di-cysteine helical peptide and a dimaleimide coupled fluorophore where the distance between the two cysteine residues on the peptide tag (10.8 Å) and two maleimide moieties of the fluorophore (7 - 9 Å, nominal 10 Å) are very similar. The mutual distance of two cysteine residues was determined using the intrinsic properties of an α-helix, where one helix turn positions two residues 5.4 Å apart [90]. In case of the dimaleimide fluorogen, distances between the two electrophilic carbons were measured on a structure of minimized energy using HyperChem (www.hyper.com). In a FlARe reaction, the first thiol addition on a maleimide is the rate-limiting step and the second thiol addition is an intramolecular reaction that is quasi-instantaneous, owing to the effective concentration for intramolecular nucleophilic attack being ~10^8 M [91], when the geometry is adequate, i.e., not constrained. Therefore, this addition reaction is kinetically favoured only when the dimaleimide fluorogen has an adequate geometry corresponding to the di-cysteine peptide (**Figure 2.1.**). As a result, we can design different pairs of di-cysteine peptide tags (referred to as dCx) and dimaleimide fluorogens (referred to as dMy) where dCx

would preferably react with dMy if x and y represent the same (or similar enough) distance (**Figure 2.1.**).

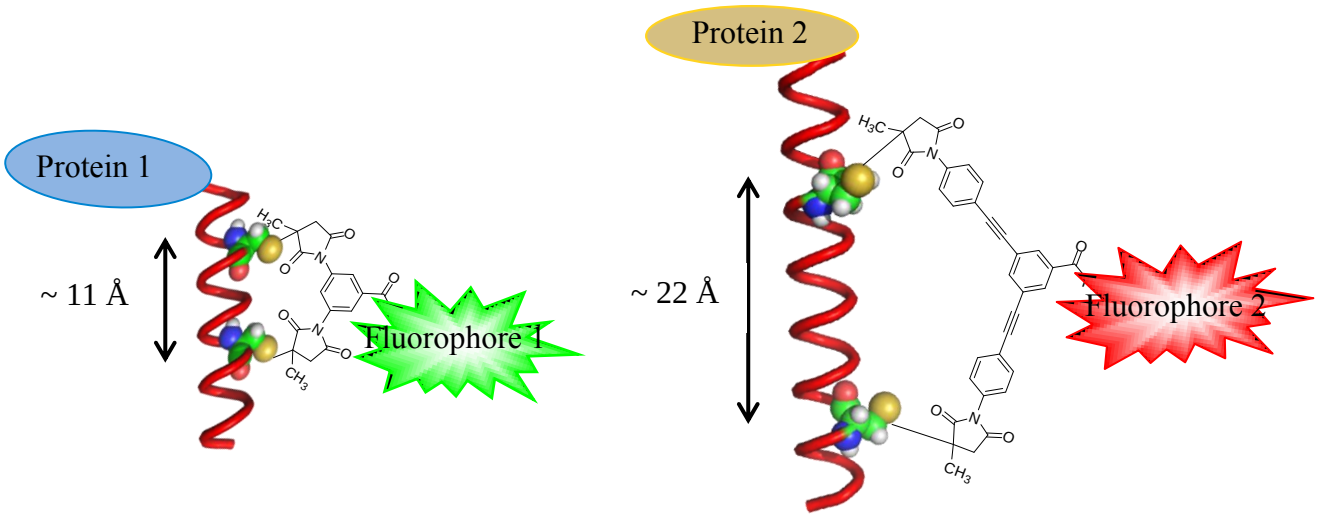


Figure 2.1. Design of FlARe orthogonal labelling. dCx/dCy pairs represented here have ~11 Å (*left*) and ~ 22 Å (*right*).

An advantage that is inherent to the FlARe design of orthogonal labelling is that there is visible and detectable fluorescence only after both maleimides have reacted with thiols. That means, that if there is a cross-reactivity between a non-corresponding pair of dCx and dMy (x and y being a substantially different distance), the second thiol addition reaction will not occur (or will be very slow) and the dMy coupled fluorophore will remain mostly quenched due to the presence of one unreacted maleimide [72]. In general, the reactivity of dCx and dMy pairs is not expected to be perfectly orthogonal, but only kinetically favoured, but the labelling and the resulting fluorescence are expected to be orthogonal, as two thiol addition steps are necessary to restore the latent fluorescence of a dMy fluorogen.

2.1.2. Design of di-cysteine peptide tags for orthogonal labelling

The FlARe labelling technique relies on a helical conformation of di-cysteine peptide tag where the distance between cysteines is determined by inherent properties of an α-helix, where two aligned residues on neighbouring helix turns have a mutual distance of 5.4 Å (see Chapter 1 and [90]). The α-helical secondary structure therefore dictates the distance pattern

between cysteines to be multiples of 5.4 Å (see **Table 2.1.** and **Figure 2.1.**). In accordance with the nomenclature of dC10 (see Chapter 1), we named these peptides dC5, dC15, dC20 and dC25. We also sought to predict their helicity that is the key feature for this distance-induced peptide tag reactivity design (**Table 2.1.**).

Table 2.1. Di-cysteine peptide tag dCx library for FlARe orthogonal labelling. Helical content was predicted using AGADIR [79, 80, 81] at 20°C and ionic strength of 0.1 M.

Name	Primary sequence	Distance between cysteines (Å)	Helical content (%)
dC5	LSAAEA**C**ARE**C**AAREAAARAGGK	5.4	35.1
dC10	LSAAE**C**AAREAA**C**REAAARAGGK	10.8	43.2
dC15	LSAAEA**C**AREAAAREAA**C**RAGGK	16.2	47.8
dC20	LSAAE**C**AAREAAAREAAAR**C**AAAREAAGGK	21.6	69.0
dC25	LSAAE**C**AAREAAAREAAAREAAA**C**RAAGGK	27.0	65.8

Cysteine residues are underlined and in bold

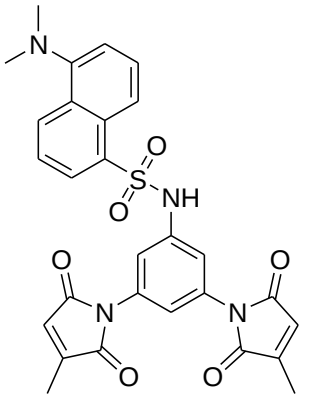
2.1.3. Design of dimaleimide fluorogens for orthogonal labelling

While the geometry of dCx peptides is dictated by their α-helical conformation, the design of dMy fluorogens is not strictly bound to one precise scaffold. dM10 fluorogens have been the most extensively explored molecules in our group ([71, 72, 73, 74], unpublished results Dr. Hugo Lachance) and their design relies on a benzene ring bearing two maleimide groups on positions 3 and 5 (**Table 1.3.** and **Scheme 2.1.**). This simple scaffold that gives the dimaleimide the distance of 10 Å allowed the convergent synthesis of many fluorogens of

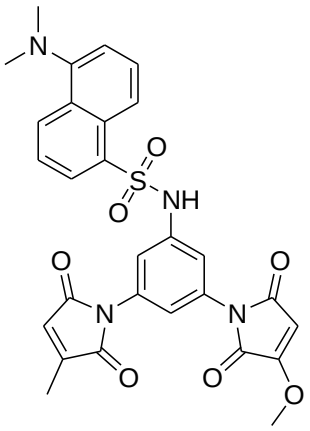
different colours that were described and successfully used for FlARe labelling [71, 72, 73, 74].

The design of other dMy molecules, complementing the corresponding dC5-25 peptides, is somewhat more complex, as the structure of dMy moiety has to meet requirements of rigidity, adequate positioning of the fluorophore and the maleimide groups to allow an efficient quench in the unreacted molecule [72], and the desired distance between maleimides for an optimal reaction with its existing dCx partner. Several designs of different dMy fluorogens have been explored to meet these requirements and corresponding molecules were synthesized by Dr. Christophe Pardin, including **dM17-quinoxaline**, **dM20-dansyl** and **dM28-BODIPY** fluorogens (**Scheme 2.2.**). Previously, Karine Caron prepared a mono-maleimide model compound for dM20-dansyl to evaluate the quenching efficiency of a maleimide group with this ethynylbenzene scaffold. In the mono-maleimide compound, the dansyl group was in a *meta* position with respect to the ethynylbenzene attachment of the maleimide. The obtained weak quenching efficiency of 1.23 of this model mono-maleimide led to conclusion that the dansyl group is too far from the maleimide to produce an efficient quench. Hence, the proposed scaffold of new **dM20-dansyl** has the dansyl fluorophore in *ortho* position with respect to both ethynylbenzene maleimide groups to obtain a stronger quenching (Karine Caron's MSc. Thesis, Université de Montréal, 2009).

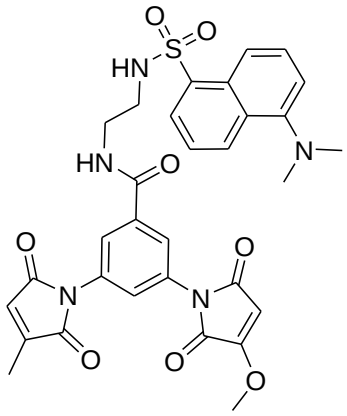
Scheme 2.1. dM10-dansyl fluorogens used in characterization of dCx library (synthesized by Dr. Hugo Lachance and Karine Caron)



dM10-dansyl 1



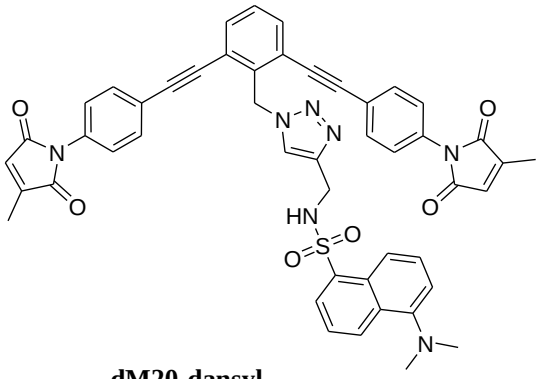
dM10-dansyl 2



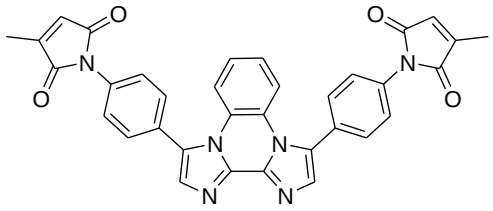
dM10-dansyl 3

In the following sections we will first assess the reactivity and selectivity profile of several dM10-dansyl fluorogens for a potential orthogonal labelling. These fluorogens differ by the substituents on the dimaleimide double bond that yield either a symmetrical dimaleimide with two methyl- substituents, or a non-symmetrical dimaleimide with one methoxy- and one methyl- substituent. As mentioned in Chapter 1, electron donating substituents on the dimaleimide bond are used in order to attenuate the undesired reaction of our fluorogens with glutathione, and it is of a certain interest to investigate non-symmetrical methyl/methoxy- fluorogens. These would on one hand allow the first thiol addition reaction on the methyl-substituted maleimide to happen for our peptide tag and also glutathione, but the second reaction that is truly fluorogenic would be allowed only for a target peptide tag that benefits of a 10^8 M intramolecular effective concentration, and disallowed for glutathione because of its general poor reactivity with a methoxy-substituted maleimide. It is noteworthy that we are not necessarily expecting to achieve an orthogonality between dM/C10 and dM/C15 pairs, but rather between dM/C10 and dM/C20/25, due to the dynamic character of both, helical peptide tags and dimaleimide conformation.

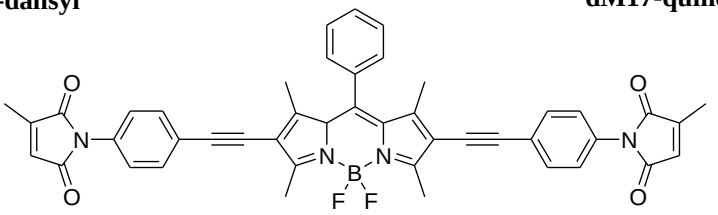
Scheme 2.2. Design of dMy coupled fluorogens (Dr. Christophe Pardin)



dM20-dansyl



dM17-quinoxaline



dM28-BODIPY

2.2. MBP-dC5 to MBP-dC25 peptide mini-library

In this section, we present the results of the orthogonal labelling design described herein. In general, small peptides are not well expressed in *E. coli* [92], hence, a Maltose-Binding Protein (MBP) is used as a test protein and a purification fusion protein for isolation of dCx peptides. MBP is not only easy to express and purify in very high yields, it is also known for its great solubility [93] that is often used as a *N*-terminal tag to enhance the solubility of proteins that suffer from low solubility [94]. Furthermore, it does not contain any cysteine; it is therefore a protein of first choice for first *in vitro* testing.

2.2.1. MBP-dCx library cloning

A small five member library of MBP-dCx proteins was cloned using a commercial vector pMAL-c5x from *New England Biolabs* that contains a multi-cloning site downstream of *malE* gene coding for MBP. A mega-primer, containing coding sequence for the whole dCx peptide was used for amplification of a portion of *malE* gene that was subsequently inserted in pMAL-c5x to yield pMAL-MBP-dC10 plasmid, as detailed in the Experimental section (2.7.2.).

2.2.2. MBP-dCx protein expression

MBP-dCx protein expression was carried out in *E.coli* BL21-Gold(DE3) at 37°C using the standard published protocol (*www.neb.com*) and as detailed in the Experimental section (2.7.3). MBP-dC10 was purified by affinity chromatography on an amylose resin and SDS-PAGE analysis during protein expression confirmed the satisfactory purity of the final protein fraction. According to the Bradford assay, standard yields varied between 40-60 mg of pure protein per litre of expression media. An example of a typical purification profile is shown in **Figure 2.2.** for MBP-dC5 and MBP-dC10. MBP-dCx proteins were stored at 4°C in 50 mM HEPES pH 7.4 buffer containing 1 mM TCEP to reduce dimerization through disulfide bonds.

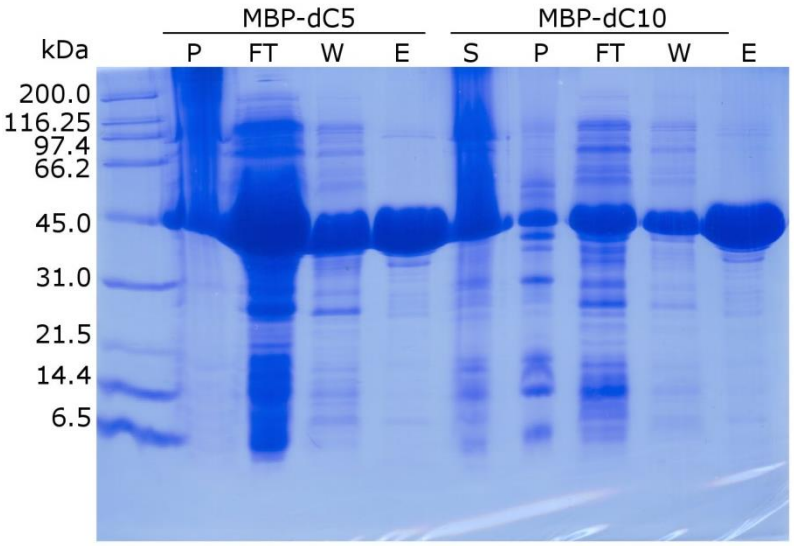


Figure 2.2. SDS-PAGE analysis of MBP-dC5 (45 kDa) and MBP-dC10 (45 kDa) expression. S-soluble protein fraction, P-pellet, FT-flow through from amylose column, W-column wash, E-eluted protein.

2.2.3. Library characterization

The mass of all MBP-dCx proteins was determined and compared the obtained values to the theoretical values predicted by ProtParam (*www.expasy.org*) using the primary protein sequence of each MBP-dCx. Both predicted and experimental values are recorded in **Table 2.2.** In all cases the obtained masses of MBP-dCx proteins correspond well to their respective predicted value.

All MBP-dCx proteins were then subjected to a free thiol quantification test that is based on the reaction of a free thiol with Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid)) that yields a bright yellow color [95]. The Ellman test is also often used to quantify phosphines, thus, prior to subjecting MBP-dCx to the Ellman test, it was necessary to change the sample buffer since the storage buffer contained 1 mM TCEP. TCEP was removed from MBP-dCx solution by a buffer exchange in an Amicon filter and the thiol quantification test was performed immediately thereafter. The resulting TCEP-free samples were subjected to

Bradford protein quantification assay. The results are summarized in **Table 2.2.** and show that there are two free thiols in each molecule of MBP-dCx protein. It is known from the primary sequence that MBP does not contain any cysteines (*www.neb.com*), therefore, the free thiols are the cysteine side chains in their reduced form on dCx peptide tags.

Protein construct	LC-MS (Da)	Predicted mass (Da)	Free thiols / protein
MBP-dC5	44725.8	44725.5	2.3 ± 0.2
MBP-dC10	44725.3	44725.5	2.1 ± 0.2
MBP-dC15	44726.5	44725.5	1.5 ± 0.1
MBP-dC20	45367.7	45366.2	2.1 ± 0.2
MBP-dC25	45382.5	45382.2	1.8 ± 0.2

Table 2.2. Experimental mass and amount of accessible thiols in MBP-dCx constructs. Theoretical mass was determined by prediction using ProtPram (*www.expasy.org*) and a primary sequence of each protein, experimental mass was determined by LC-MS and reactive thiols were quantified by Ellman test [95].

2.3. Labelling of MBP-dCx with dansyl-dM10 fluorogens

Labelling of the MBP-dCx library *in vitro* and kinetic characterization was done first with a series of the three most recently developed and synthesized dM10-dansyl fluorogens (**Scheme 2.1.**) to determine if there is a preference of reactivity of dM10 fluorogens with dC10 peptides, as expected. Labelling reactions were carried out in equimolar concentrations of 50 μM of MBP-dC10 and dM10-dansyl, at 20°C in a Cary-Eclipse Fluorescence Spectrophotometer (Varian). Excitation and emission wavelengths were set at 330 nm and 530 nm, respectively, to observe a maximal signal of dansyl fluorophore. As a control of reactivity with free thiols, a resistance test of all three dM10-dansyl compounds was performed with glutathione (GSH) that mimics the cellular milieu for future applications of dM10 fluorogens in cellular labelling. All labelling reactions were monitored long after completion (200 minutes) in order to determine kinetic parameters with more precision. All reactions were carried out at least in duplicate and second order rate constants were

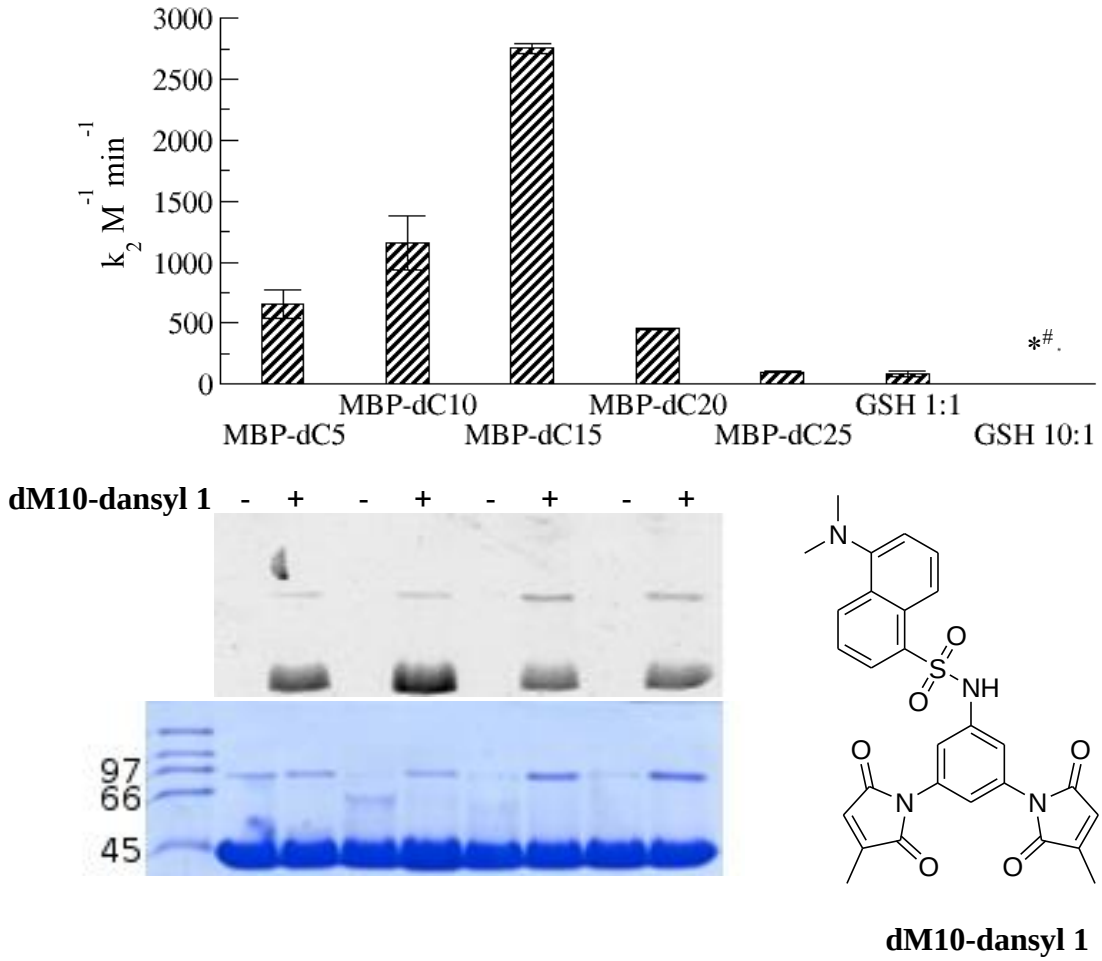
determined individually by fitting using Varian software, from which average values and standard errors were calculated.

2.3.1. Labelling of MBP-dCx library with spacerless dM10-dansyl 1

All second order rate constants of MBP-dCx labelling with **dM10-dansyl 1** are shown in **Figure 2.3**. and suggest a clear preference of reactivity between MBP-dC15 and **dM10-dansyl 1**. MBP-dC10 has half of the maximal detected rate in comparison with MBP-dC15. All other MBP-dCx peptides react significantly slower than MBP-dC10 or MBP-dC15. Even though it would be expected that **dM10-dansyl 1** would react faster with MBP-dC10 than with MBP-dC15 due to a better distance adequacy, there is a clear kinetic profile that favours reaction of **dM10-dansyl 1** with di-cysteine peptides of mutual distance of 10-15 Å.

It is important to note that there was very little reactivity of **dM10-dansyl 1** with glutathione in ratio 1:1 or with 10 equivalents of glutathione.

The brightness of labelled MBP-dCx was also evaluated in a SDS-PAGE gel using a GelDoc imager to give an idea about the integrity of labelled MBP-dCx (**Figure 2.3**.) and brightness of the fluorophore for observation with a naked eye. MBP-dCx were incubated with **dM10-dansyl 1** for two hours at 20°C and the samples were analyzed on a non-reducing SDS-PAGE gel. The preferred reactivity of MBP-dC10 with **dM10-dansyl 1** is striking only by looking at the brightness of the sample, where under UV illumination the band corresponding to labelled MBP-dC10 is the brightest of all fluorescent bands, corresponding to the highest amount of fluorescently labelled MBP-dCx.



dM10-dansyl 1

Figure 2.3. Second order rate constants of MBP-dCx library (*top*) with spacerless dM10-dansyl 1 fluorogen (*bottom right*). * no reaction detected. # first order rate constant. MBP-dC10 (45 kDa) protein samples were migrated on a non-reducing SDS-PAGE gel (*bottom left*) and imaged in a GelDoc (top gel) before Coomassie blue staining (bottom gel). Both labelled (+) and unlabelled (-) protein samples were loaded for comparison. Molecular weight marker was loaded as a reference and band sizes are indicated in kDa (left line).

There was very little dimerization of MBP-dCx before reaction with **dM10-dansyl 1** (present only for MBP-dC5 in a small amount, see Coomassie blue staining in **Figure 2.3**.). A small amount of cross-linked product of molecular weight ~ 90 kDa of two MBP-dCx molecules with the fluorogen was detectable in both Coomassie stain and under UV light, and it was observed mostly for constructs that have longer distance between cysteines in dCx (dC15 and dC20). These tags are far less reactive with the used **dM10-dansyl 1**, as predicted

in the distance-modulated design of dCx, and also provide the necessary access for another molecule of MBP-dCx to react with the second maleimide available, as the ring closing reaction is disfavoured. Hence, the small amount of cross-linking reaction between two MBP-dCx molecules is allowed by the space and time provided by the longer distance between cysteines of dCx. Alternatively, these dimers could possibly be formed directly via one or two disulfide bonds; however, the Coomassie stain of unlabelled samples showed that this dimerization is occurring in a negligible amount in comparison with the dimerization in presence of a fluorogen.

2.3.2. Labelling of MBP-dCx library with spacerless dM10-dansyl 2

Similarly to **dM10-dansyl 3**, **dM10-dansyl 2** fluorogen bears a methoxy- and a methyl-substituent on both maleimide groups. As reported before [74], a methyl- substituent provides more selectivity to the dCx/dMy labelling reaction, compared to a reaction between dMy and a thiol, such as glutathione. Moreover, the methoxy- substituent is a more potent electron donating group that decreases the electrophilicity of the maleimide C=C double bond and therefore make it less reactive for attack of a thiolate, providing more selectivity to dCx di-thiolate attack in comparison with other single thiolates.

Second order rate constants relative to the addition reaction of **dM10-dansyl 2** with MBP-dCx or glutathione are reported in **Figure 2.4.**, from which it is apparent that **dM10-dansyl 2** reacts preferentially with MBP-dC15 and MBP-dC10, and presents a very low potential reactivity with an equimolar concentration or an excess of glutathione. The overall reactivity of this fluorogen is comparable to **dM10-dansyl 1**; however, unlike what would be expected from the methyl/methoxy dimaleimide design (see section 2.1.3), a more stringent profile for this non-symmetrical fluorogen is not observed, in comparison with **dM10-dansyl 1**. A non-symmetrical fluorogen should be kinetically more adequate for a future orthogonal labelling due to the expected lower reactivity of the second thiol addition, but this is not the case for **dM10-dansyl 2** according to the obtained results.

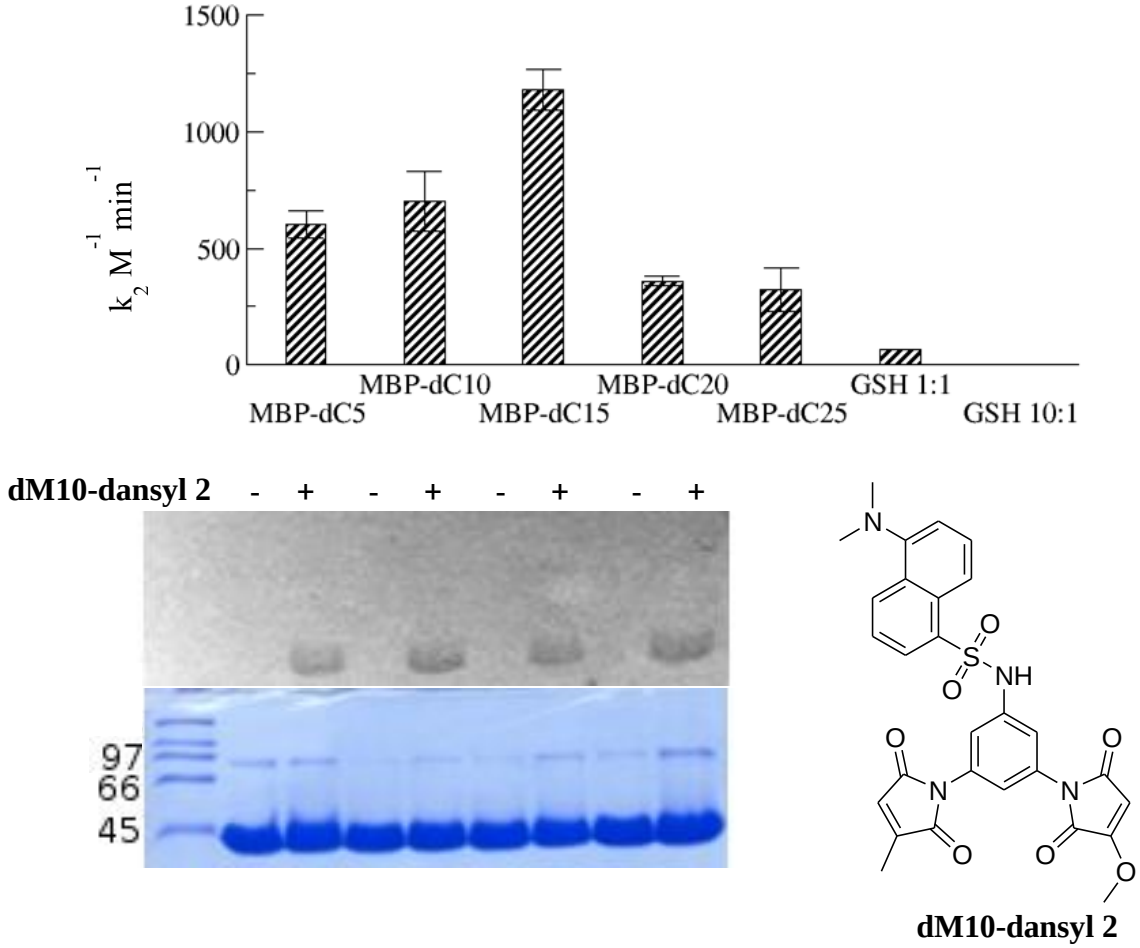


Figure 2.4. **Second order rate constants of MBP-dCx labelling reaction (*top*) with dM10-dansyl 2 (*bottom right*).** * no reaction detected, # first order rate constant. MBP-dC10 (45 kDa) protein samples were migrated on a non-reducing SDS-PAGE gel (*bottom left*) and imaged in a GelDoc (top gel) before Coomassie blue staining (bottom gel). Both labelled (+) and unlabelled (-) protein samples were loaded for comparison. Molecular weight marker was loaded as a reference and band sizes are indicated in kDa (left line).

2.3.3. MBP-dCx library labelling with dM10-dansyl 3

Labelling of MBP-dCx library members with **dM10-dansyl 3** was performed, as previously, in equimolar conditions of 50 µM of protein and fluorogen, in 50 mM HEPES pH 7.4 buffer

supplemented with 1 mM TCEP to keep all thiols reduced. Labelling reaction was followed at 530 nm in a fluorescence spectrophotometer using an excitation wavelength of 330 nm.

Kinetic constants, shown in **Figure 2.5,** were determined using fitting of observed curves corresponding to second order reaction. Between the kinetic constants for MBP-dCx proteins, there is a clear preference of reactivity of **dM10-dansyl 3** for MBP-dC10, giving the reaction with a two-fold higher second order rate constant in comparison with MBP-dC5 or MBP-dC15, and a three-fold higher second order rate constant in comparison with MBP-dC20 or MBP-dC25, resulting in a very satisfactory selectivity of **dM10-dansyl 3** towards MBP-dC10. Furthermore, no reactivity of **dM10-dansyl 3** with a mono-thiol, such as glutathione, even when the thiol was present in a 10-fold excess (**Figure 2.5.**) was detected. It is possible that one maleimide group, possibly bearing a methyl- substituent, underwent a thiolate addition reaction with glutathione, but the second maleimide group, bearing a methoxy- substituent, remained unreactive and allowed the fluorogen molecule to keep its fluorescence quenched, as determined previously **[72]**. These results show that in case of **dM10-dansyl 3** fluorogen, a methoxy- substituent provides more stringency, as expected, to the distance-based reactional design of dCx peptides with **dM10-dansyl 3**. Furthermore, these results are very encouraging for a further development of methoxy-substituted dimaleimide fluorogens that would be very suitable for cellular labelling, where a high level of selectivity of dCx labelling reaction over a simple thiolate attack on dM10-fluorogens is needed in order to achieve a reliable labelling or target proteins, easily distinguishable with a naked eye.

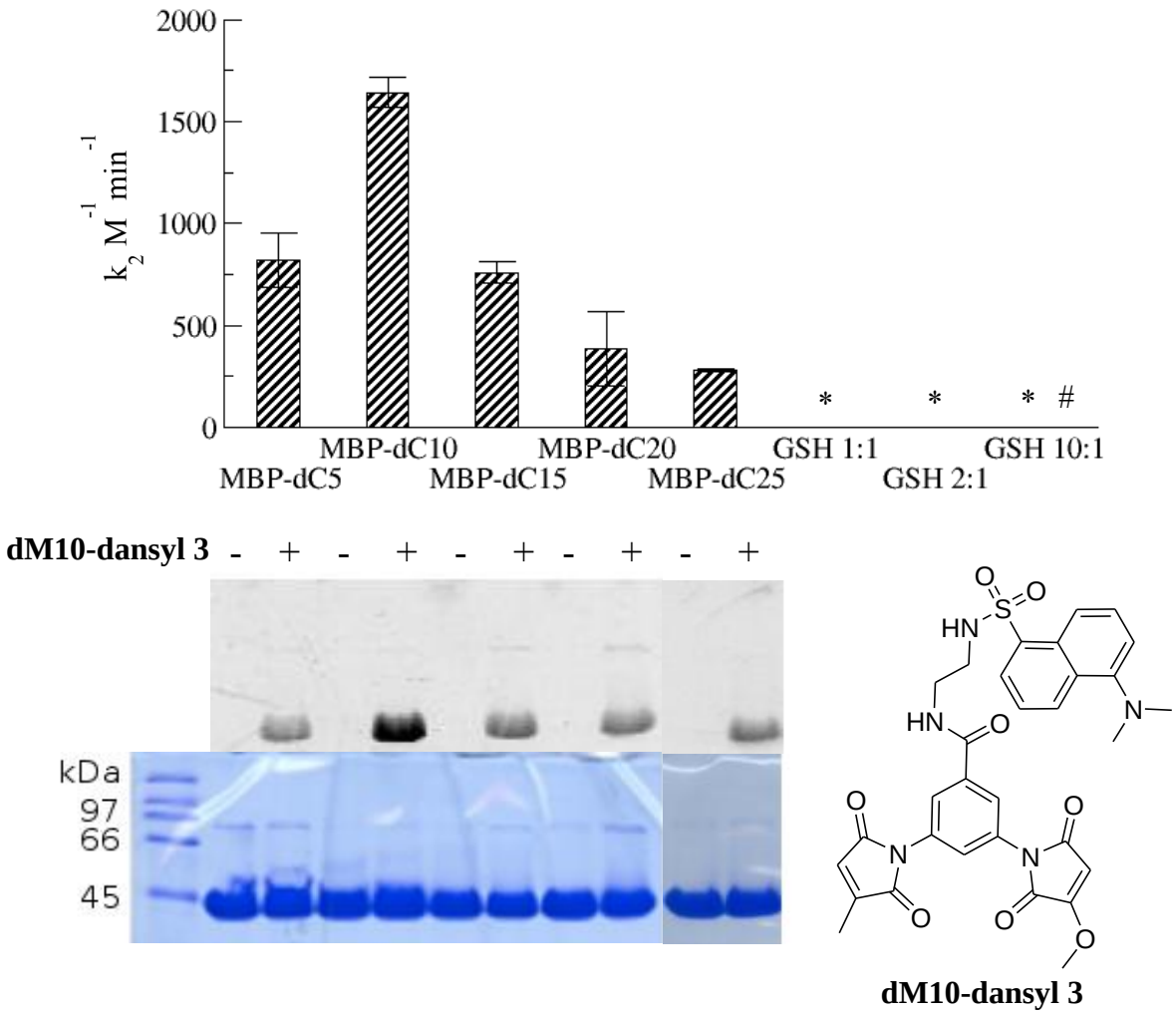


Figure 2.5. Second order rate constants of MBP-dCx labelling reaction (*top*) with dM10-dansyl 3 fluorogen (*bottom right*). * no reaction detected, # first order rate constant. MBP-dC10 (45 kDa) protein samples were migrated on a non-reducing SDS-PAGE gel (*bottom left*) and imaged in a GelDoc (top gel) before Coomassie blue staining (bottom gel). Both labelled (+) and unlabelled (-) protein samples were loaded for comparison. Molecular weight marker was loaded as a reference and band sizes are indicated in kDa (left line).

2.4. Development of dMy fluorogens and labelling of MBP-dCx library

2.4.1. General approach

While the synthetic design of dM10 moiety of fluorogen molecules has been established and achieved on many proof-of principle dM10-fluorogen compounds [71, 72, 73, 74], the design of other dMy moieties appears less obvious, due to the geometry and rigidity requirements of such dMy molecules [72]. Spatial proximity of dimaleimide moieties to the fluorophore is necessary for an optimal quenching efficiency and a relatively high rigidity of a dimaleimide moiety is required to maintain the distance that should confer a higher selectivity to the molecule towards its predesigned dCx partner. However, if this new dMy distance is too large, the maleimides will be too far from the fluorophore to maintain their quenching efficiency. Based on the work by Karine Caron (MSc. Thesis), Dr. Christophe Pardin designed and synthesized three potential candidate fluorogens (**Scheme 2.2**.) that were subjected to the kinetic profiling of MBP-dCx peptide tags.

2.4.2. dM28-BODIPY and dM20-dansyl

Fluorogens **dM20-dansyl** and **dM28-BODIPY** were designed to target specifically dC20 and dC25 peptide tags, respectively. However, peptide helices are dynamic and exist in multistate equilibrium with their unfolded form with higher helix content in the center compared with the helix ends [96, 97, 98]. In that regard, it is not necessarily expected to obtain an exclusive reactivity profile of each studied dMy molecule with one member of the dCx library, especially for greater maleimide distances, but rather a general complementary profile to existing dM10 fluorogen profiles.

After synthesizing **dM28-BODIPY** and **dM20-dansyl**, Dr. Christophe Pardin was able to determine that, despite the very convenient design of the dimaleimide moieties for a reaction with dC20 or dC25, **dM28-BODIPY** exhibits a high background fluorescence and very low fluorescence enhancement (on order of 2 units), probably due to a too great distance between maleimides and the BODIPY core; and **dM20-dansyl**, with a fluorescent enhancement of 4.5, reacts extremely poorly with all MBP-dCx, most likely as a result of a steric hindrance

between the dansyl group and the dCx helix. Therefore, these compounds are not suitable for fluorogenic labelling applications and were not examined further.

2.4.3. dM17-quinoxaline

dM17-quinoxaline has a fluorescence enhancement of 78 and was used for kinetic profiling of MBP-dCx in identical conditions as dM10-dansyl fluorogens, i.e. in 50 mM HEPES pH 7.5 buffer supplemented with 1 mM TCEP, at 20 °C. As displayed in **Figure 2.6,** there is a slight preference of **dM17-quinoxaline** for MBP-dC15, in accordance with the distance-directed reactivity design. Furthermore, **dM17-quinoxaline** reacts five times slower with MBP-dC25 than with its best partner, MBP-dC15. However, the overall reactivity of **dM17-quinoxaline** is far too high to be used in an orthogonal reaction with existing dM10 fluorogens, such as **dM10-dansyl 1-3**, where the best pair reacts still 3-4 times slower than the less favoured pair as per design, **dM17-quinoxaline**/MBP-dC10.

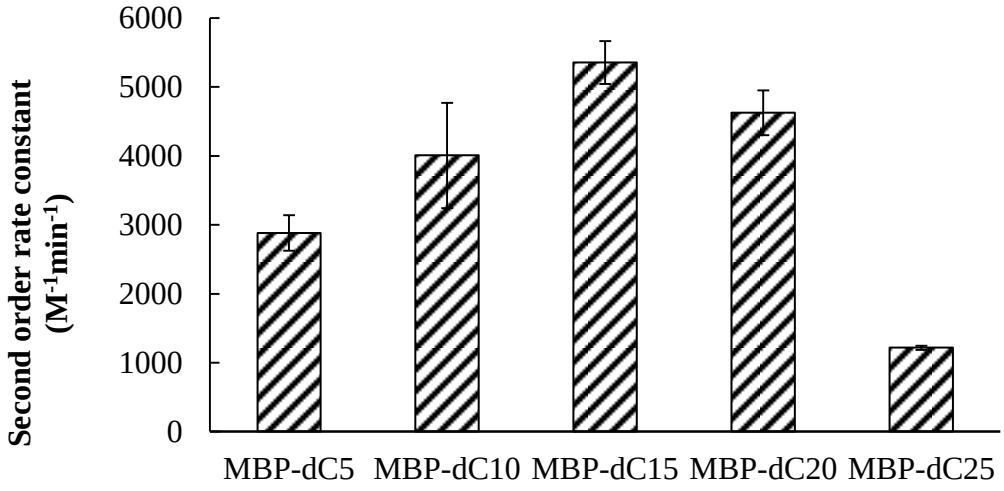


Figure 2.6. Second order rate constants for labelling of MBP-dCx library with dM17-quinoxaline. Rate constants were determined by fitting after completion of reaction of 50 µM MBP-dCx, 50 µM **dM17-quinoxaline** in 50 mM HEPES buffer pH 7.5 and 1 mM TCEP, at 20°C. Fluorescence increase was followed at 425 nm upon excitation at 340 nm.

2.5. Conclusions

Five different di-cysteine peptide tags were designed and cloned, with 5-25 Å distances between cysteine residues, in fusion with a test protein MBP, and were characterized by labelling with three dM10-dansyl fluorogens (**Scheme 2.1.**). A certain degree of selectivity of dM10 fluorogens towards dC10 tags (or dC15 in one case) was observed, which was encouraging for the subsequent design of dMy fluorogens, different from dM10.

Kinetic profiles of these three fluorogens, that differ only by a substituent on the dimaleimide double bond (**dM10-dansyl 1** and **dM10 dansyl 2**) or by a spacer between the fluorophore and the maleimide (**dM10-dansyl 2** and **dM10-dansyl 3**), can be used to make several conclusions in regard of the potential advantage of the mentioned features. It is apparent from the kinetic profiling of MBP-dCx with a dimethyl-dimaleimide (**dM10-dansyl 1**) and methoxy-methyl-dimaleimide (**dM10-dansyl 2**) that the supposed advantage of having a very slowly reactive methoxy-maleimide along with a faster methyl-maleimide in order to get a more selective labelling, is not notable and consistent enough to continue developing this hybrid methoxy-methyl dimaleimide. Moreover, the dimethyl-substituted **dM10-dansyl 1** not only confers experimentally the same level of selectivity as **dM10-dansyl 2**, but also reacts twice as fast with MBP-dC10. **dM10-dansyl 1** is be therefore retained as the best fluorogen for *in vitro* applications.

Similarly, **dM10-dansyl** fluorogen **3** that bears both methoxy and methyl substituents was characterized with the MBP-dCx library. Although it presents the best kinetic profile with MBP-dCx with a distinct selectivity for MBP-dC10, the overall recovered fluorescence after completion of labelling is low in comparison with **dM10-dansyl 1** or **2**, consistent with previous findings in regards of the effect of a spacer between a fluorophore and maleimides in the fluorogen design [74].

With the goal of having at least two pairs of dCx/dMy components for an orthogonal labelling, three different dMy fluorogens were then synthesized by Dr. Christophe Pardin and one was characterized with the small library MBP-dCx. Unfortunately, these new fluorogens do not present all the desired properties of labelling agents. **dM20-dansyl** and **dM28-BODIPY** (**Scheme 2.2.**) present only a low reactivity or very low fluorescence

enhancement, respectively, both features being crucial for FlARe labelling. **dM17-quinoxaline** is more suitable in terms of fluorescence enhancement; however, it has only very little selectivity towards its designed dC15 (or dC20) partner, and mostly reacts with all MBP-dCx tags with a comparable rate that is considerably higher than the rates observed for any dM10-dansyl fluorogens.

2.6. Perspectives

We obtained clear kinetic profiles for a helical tag-based mini library of MBP-dCx for three dM10 fluorogens. Unfortunately, some challenges were experienced, related to inefficient quenching of some dMy fluorogens. It was previously determined that the quenching efficiency of maleimides is dependent on their through-space distance from the fluorophore [74], which inherently limits the design of dMy fluorogens with greater distances between maleimides to complement dC20-dC25 peptide tags. It would be of a certain interest to re-design the **dM17-quinoxaline** fluorogen that had promising quenching capacity, but incompatibly high reactivity. For example, a new fluorogen could be synthesized, bearing two methoxy- substituents on the maleimide moieties, to decrease substantially the reaction rate, which may allow obtaining reaction rates comparable to dM10/dC10 pairs.

In terms of peptide tags, only helical secondary structure motifs were explored, possibly to their full capacity. There is a certain selectivity of dMy fluorogens to their designed dCx partners, but at this stage and with dMy fluorogens currently available, the kinetic profiles are not complementary enough to potentially obtain orthogonality between different dMx/dCx pairs. It would be of a certain interest to explore different secondary structure motifs, such as β-hairpin, that may present increased rigidity and hence would allow a more selective kinetic profile with existing fluorogens. Small β-hairpin peptides have been heavily studied, and their sequences optimized for conformational stability [99, 100]. Thus, provided that introduction of cysteine residues is not detrimental for the β-hairpin integrity; they represent a real potential as peptide tags for FlARe labelling. Despite that, in order to achieve a possible orthogonality of labelling, using α-helical or β-hairpin scaffold, there is

still an acute need of fluorogens that would satisfy the requirement of efficient quenching and of selectivity.

2.7. Experimental section

2.7.1. General Experimental procedures

Expression vector pMAL-c5x and all molecular biology reagents were purchased from *New England Biolabs*; media, buffers and reagents from Bioshop Canada or Sigma Aldrich. DNA primers were purchased from AlphaDNA, Montréal.

2.7.2. Cloning

The dC5, 10, 15, 20 and 25 peptide tag coding sequences were cloned into pMAL vector using the following approach: a fragment of *malE* gene from pMAL vector was amplified by PCR using a forward primer annealing on *BglII* restriction site and a reverse mega-primer containing the reverse coding sequence of desired peptide tag (see **Table 2.3.**). The resulting amplified fragment was inserted into original pMAL-c5x vector using *BglII* and *EcoRI* restriction sites.

Desired clones were selected and confirmed by sequencing at Institut de Recherche en Immunologie et Cancérologie (IRIC), Montréal.

2.7.3. Expression and purification of MBP-dCx proteins

MBP-dCx proteins were over-expressed in *E.coli* strain BL21 (DE3) commonly used in our laboratory or in BL21-Gold(DE3) from Agilent. Transformed bacteria were cultivated in rich expression media (10 g/L biotryptone, 5 g/L yeast extract, 5 g/L sodium chloride and 5% (*w/v*) glucose) supplemented with 100 µg/mL Ampicillin until the optical density at 600 nm reached 0.6, where MBP over-expression was induced by addition of 0.3 mM IPTG. The over-expression was carried out at 37°C for 3 hours. MBP-dCx proteins were purified using a very well established protocol (*www.neb.com*, [94]). Briefly, the cells were harvested

at 3700 g, the bacterial pellet was resuspended in MBP-binding buffer (20 mM Tris-HCl pH 7.5, 200 mM NaCl, 2 mM EDTA) and lysed by sonication (3 x 30 seconds, 50% power), and the lysate was centrifuged at 6400 g for 20 minutes at 4°C to separate soluble and insoluble proteins. The resulting soluble fraction was loaded on an amylose resin pre-equilibrated with MBP-binding buffer, and incubated for at least 2 hours at 4°C, with gentle shaking. The un-bound proteins passed through the column, the resin was washed by 10 mL of MBP-binding buffer and pure MBP-dCx protein fraction was eluted by 10 mL of MBP-binding buffer containing 10 mM of maltose. Protein yield was determined by Bradford assay.

2.7.4. Characterization of MBP-dCx proteins

All MBP-dCx proteins were analyzed by LC-MS at the Regional Mass Spectrometry Centre (Université de Montréal) confirming their expected molecular mass.

2.7.5. Determination of total free cysteines in MBP-dCx

Total reduced thiol amount per protein was determined by Ellman assay [95]. Briefly, all MBP-dCx proteins initially kept in their reduced state (using 1 mM TCEP) were placed in HEPES 50 mM pH 7.5 buffer and their concentration was immediately determined by Bradford assay, as well as the concentration of free thiols using coupled Ellman/cystamine assay [95]. The number of thiols per molecule of protein was calculated by dividing reduced thiol concentration by total protein concentration.

2.7.6. In vitro labelling of dCx containing test proteins by dimaleimide fluorogen molecules

All labelling reactions were carried out at 20°C in 50 mM HEPES pH 7.4, 1 mM TCEP. A non-thiol containing reducing agent was used in order to keep all cysteines of dCx tags reduced and to prevent reactivity of a thiol on dimaleimide fluorogen molecules. Labelling reaction conditions included 50 µM of MBP-dCx, 50 µM of dimaleimide fluorogen dissolved in DMSO. The total DMSO concentration was 2.5% (*v/v*) in all labelling reactions.

Increase of fluorescence resulting from addition reaction of cysteines on maleimide fluorogens was followed by fluorescence spectrometer Cary Eclipse from Varian at 20°C.

For second order rate constant determination the Cary Eclipse software was used, with fitting period 0-200 minutes and reagent concentration 50 μM.

Table 2.3. Oligonucleotides used for cloning of MBP-dCx library

MBP_fw	5′ – CAAAGAT CTGCTGCCG – 3′
MBP-dC5_bw	5′ – GGAATTCCTCATTTGCCACCCGCGCGCGCCGCCGCTTCGCGTGCTGCGCATTCACG TGCACACGCTTCCGCCGCACTCAGCCTTCCCTCGATCCC – 3′
MBP-dC10_bw	5′– GGAATTCCCTACTTTCCTCCAGCTCTAGCTGCAGCTTCTCTGCATGCAGCTTCTCTA GCAGCGCACTCAGCAGCGCTCAGCCTTCCCTCGATCCC – 3′
MBP-dC15_bw	5′ – GGAATTCCTCATTTGCCACCCGCGCGACATGCCGCTTCGCGTGCTGCTGCTTCACG TGCACACGCTTCCGCCGCACTCAGCCTTCCCTCGATCCC – 3′
MBP-dC20_bw	5′ – GGAATTCCTCATTTGCCACCCGCCGCTTCGCGCGCCGCTGCACAGCGCGCCGCCG CTTCGCGTGCTGCTGCTTCACGTGCCGCACATTCCGCCGCACTCAGCCTTCCCTCG ATCCC – 3′
MBP-dC25_bw	5′ – GGAATTCCTCATTTGCCACCCGCCGCGCGGCACGCCGCTGCTTCGCGCGCCGCCG CTTCGCGTGCTGCTGCTTCACGTGCCGCACATTCCGCCGCACTCAGCCTTCCCTCG ATCCC – 3′

Chapter 3

ROAD TO LABELLING IN COMPLEX MILIEU

3.1. Introduction

In vitro labelling presents an intrinsic advantage over *in cellulo* labelling, related to the low complexity of the system. By regulating the chemical composition and physical conditions of the system, one can control and adjust the experimental conditions that would lead to successful labelling. In case of *in cellulo* labelling, however, the conditions are set to accommodate the biochemical processes of the living system, and the chemical composition is given by the living system itself. The challenges that one faces in performing a successful labelling are related to a) the high risk of secondary reactions that can potentially degrade the labelling agent; b) non-specific labelling (bio-orthogonality of successful labelling); c) poorly controllable amount of expressed target protein; d) potential toxicity of the labelling agent and/or the fusion protein or peptide; e) possible limited internalization of the labelling agent inside cells; f) stability of the labelled complex; g) necessity of an extracellular stimulus (photo-activation), etc.

3.1.1. Advances in selective in cellulo protein labelling

Labelling methods based on chemical reactivity that were presented in Chapter 1 bear several intrinsic advantages: they use small compounds that can easily penetrate through the cell membrane, and single amino acids or small peptide tags, and so they are minimalistic in their size. Furthermore, given the fact that these labelling methods are non-enzymatic, interference with the cellular machinery is less probable that in the case of approaches using an altered enzyme as a protein tag. However, other challenges, related to the toxicity of the molecules or additives and the selectivity of the labelling reaction, become very important, as will be presented in the next section.

Several chemical reactions leading to the formation of covalent adducts have been developed [101] since the development of the Huisgen copper-catalyzed [3 + 2] cycloaddition between an azide and a terminal alkyne (**Figure 3.1.**), called the "click reaction" [102, 103, 104]. This reaction was described as being completely bio-orthogonal, i.e. not leading to secondary reactions with any cellular components, as both main reactive components (azide and alkyne groups) are practically absent from living organisms. However, the approach suffered from being dependent on Cu^{I} which made it impractical for live cell labelling due to the cytoxicity

of high concentrations of Cu^{I}. The approach was since developed and improved to be faster (using strained fluorinated alkynes [105], **Figure 3.1.** *bottom*)), and copper-free [52, 53], allowing it to be applied on live cells and whole organisms [106, 107, 108]. A number of other low-toxicity biorthogonal reactions have been developed [109, 110, 111, 112, 113, 114]. Similarly to the click reaction, the inverse-electron-demand Diels-Alder (ieDiels-Alder) reaction, using a tetrazine and a *trans*-cyclooctene, was described [109, 110, 111] as very fast, completely bio-orthogonal and non-toxic, and successfully used for labelling in cell media and cell lysate with yields > 80%. More recently, it was applied for tumour imaging in live mice [115]. Another example of a fast bio-orthogonal reaction is the alkyne-nitrone cycloaddition [114], recently shown to be useful in surface protein labelling [116]. The utility of click chemistry and ieDiels-Alder reactivity has been made possible by recent developments in strategies for metabolically incorporating bioorthogonal functionality into proteins. These include co-translational incorporation of unnatural amino acids (containing small azide or alkyne moiety), pioneered by Schultz *et al.* [117]; or post-translational incorporation of azide-derivatized glycans [105, 51].

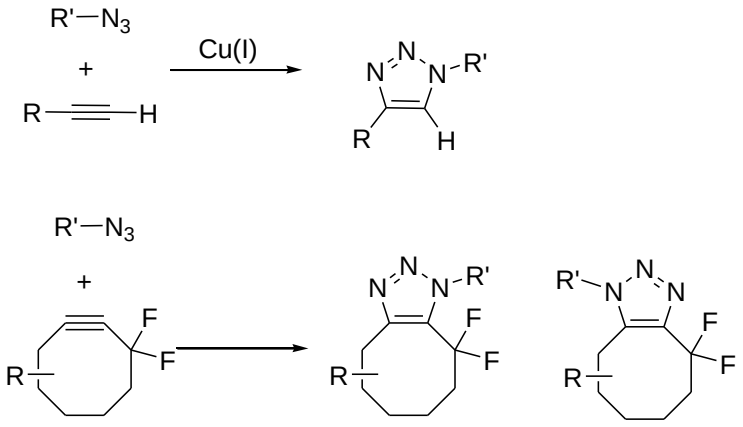


Figure 3.1. Copper (I)-catalyzed (*top*) and difluorinated cyclooctyne strain-promoted click chemistry (*bottom*).

3.1.2. Challenges for FlARe labelling in complex milieu

In the case of *in cellulo* FlARe labelling, the most immediate challenging aspect is related to the selectivity. Dimaleimide fluorogens are able to react with adventitious thiols, such as

glutathione, or with solvent exposed thiols. Previously, our collaborators were able to label a protein expressed on the surface of a cell with **dM10-FITC** fluorogen (**Table 1.4.,** [73]). However, labelling in an intracellular environment can be more challenging due to the low concentration of the protein of interest, or perturbed by many of its components, such as a large excess of glutathione present in mammalian cells (1-10 mM), or any other accessible free thiol that is capable of reacting with a maleimide group. In order to estimate the success of FlARe labelling reaction with currently available fluorogens in such a complex milieu, we first decided to evaluate its success in an *E. coli* lysate and in a HEK293 cell lysate, before proceeding to an intracellular labelling experiment on living cells. Additionally, the work presented in the next section is a good performance test for the best dansyl-derived fluorogens currently available for testing.

3.2. Labelling of MBP-dC10 in *E.coli* lysate

As a first step of the more complex labelling of a well-known and reactive protein that was previously successfully labelled *in vitro* (see Chapter 2), MBP-dC10 was over-expressed and purified according to an established protocol (see page 42) and instead of a dilution of MBP-dC10 stock solution in a buffer, MBP-dC10 was placed in the soluble protein fraction of a lysate from bacteria expressing recombinant guinea pig liver transglutaminase. In a complementary manner, and to subject the labelling reaction to a more "intracellular-like" environment, the same experiment was performed in the presence of 2 mM glutathione. As a proof-of-principle experiment, **dM10-dansyl** fluorogens **1**, **2** and **3** (**Scheme 2.1.**) were used for this application, for the easy visualisation of their dansyl fluorophore with a UV light source upon reaction. Fluorogens were added in an increasing concentration from 0-100 µM along with a constant concentration of 2.5 µM MBP-dC10 and a constant concentration of soluble proteins of 25 µM, in the presence or absence of GSH. It is apparent that only MBP-dC10 is labelled by **dM10-dansyl 3** in these conditions and in the absence of GSH, as demonstrated by a clear fluorescent band at the size of MBP-dC10 (**Figure 3.2.**). However, in the presence of 2 mM GSH, which mimics the intracellular concentration of this reducing agent, the band of labelled MBP-dC10 is significantly lower in intensity, and is clearly seen by a bare eye only when 20 µM of **dM10-dansyl 3** is used. Coomassie staining was

performed to estimate the total protein content and to confirm that all wells contain approximately the same amount of soluble proteins (**Figure 3.2.** *right*).

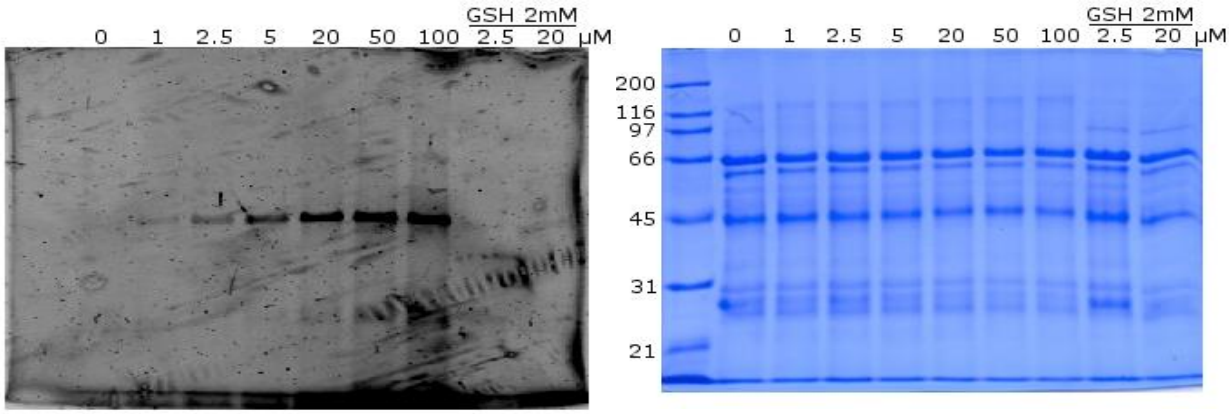


Figure 3.2. Labelling of MBP-dC10 in soluble fraction of *E.coli* cell lysate with dM10-dansyl 3 fluorogen. A ratio of 1/10 of MBP-dC10 (45 kDa) was used with respect to total protein content of *E.coli* soluble proteins, with the addition of the indicated concentration of fluorogen (top of gel, in µM). The labelling reaction was carried out for 2 h at 20°C in the presence or absence of 2 mM GSH, as indicated. The fluorescence of labelled MBP-dC10 was detected by GelDoc using UV excitation for 20 s exposure (*left panel*), after which migrated proteins were stained by Coomassie brilliant blue (*right panel*) to estimate the total protein profile. Protein molecular weight marker was loaded as a reference and the reference band sizes are indicated in kDa (left lane on Coomassie stained gel).

Similar labelling of MBP-dC10 in a pool of soluble *E. coli* proteins was performed with **dM10-dansyl 1** and **dM10-dansyl 2** (**Figure 3.3.**) with increasing concentration of fluorogen from 0-100 µM and 2.5 µM MBP-dC10, in the presence of 1:10 of soluble protein fraction (MBP-dC10 : soluble protein). For both fluorogens, after 2 hours of incubation at 20°C and in absence of GSH, a clear fluorescent band of size of MBP-dC10 appears even for the lowest concentration of fluorogen (2.5 µM). Furthermore, a consistent band increasing in brightness for 25 µM and 100 µM of fluorogen suggests that some amount of fluorogen probably reacted with other components of the lysate, or adventitious thiols. More interestingly, in the presence of 2 mM GSH, a clear fluorescent band of labelled MBP-dC10 for fluorogen concentrations of 25 µM and 100 µM indicates that MBP-dC10 could be labelled by **dM10-dansyl 2** and **1** in the presence of high concentrations of GSH. As

mentioned previously, Coomassie blue stain shows the same protein amount loaded in all wells.

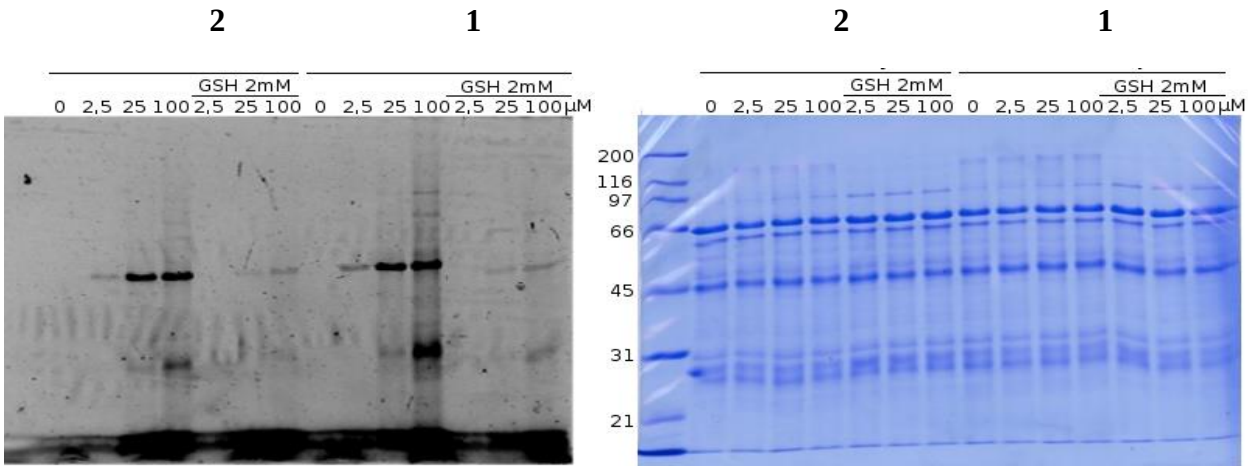


Figure 3.3. Labelling of MBP-dC10 in soluble fraction of *E.coli* cell lysate with dM10-dansyl 2 (*left*) or dM10-dansyl 1 (*right*) fluorogen. A ratio of 1:10 of MBP-dC10 (2.5 µM, 45 kDa) was used with respect to total protein content of *E.coli* soluble proteins, with the addition of the indicated concentration of fluorogen (top of gel, in µM). The labelling reaction was carried out for 2 hours at 20°C in the presence or absence of 2 mM GSH, as indicated. The fluorescence of labelled MBP-dC10 was detected by GelDoc using UV excitation for 10 s exposure (*left panel*), after which migrated proteins were stained by Coomassie brilliant blue (*right panel*) to estimate the total protein profile. Protein molecular weight marker was loaded as a reference and the reference band sizes are indicated in kDa (left line on Coomassie stained gel).

This result is particularly encouraging for the intracellular labelling using fluorogen compounds, as it is showing that it is possible to label specifically one protein in a complex environment, such as a bacterial lysate and detect its fluorescence by a bare eye. However, the fact that an SDS-PAGE gel is used to separate all proteins according to their size means also that the signal from potential parasite reactions (such as with solvent exposed thiols or with GSH) is diluted to the whole length of a migration line. In this case, reaction of a fluorogen with GSH would not be visible as a fluorescent signal in the gel, but only potentially as a decrease of the band that corresponds to MBP-dC10 labelling, as less fluorogen would be available for this desired reaction.

3.3. Evaluation of *in vitro* labelling of solvent exposed thiols

For any kind of labelling distinguishable by a naked eye in a microscope, it is important that the desired reaction is clearly superior to any kind of background reaction. Here we tried to estimate the background reaction of a fluorogen with an untagged protein, such as BSA, in the presence and absence of GSH. Previous kinetic results and labelling in *E. coli* lysate show that, among the dM10-dansyl fluorogens, **dM10-dansyl 1** and **2** are more promising than **dM10-dansyl 3**. The following experiment is therefore limited to **dM10-dansyl 1** and **2**.

A labelling control reaction of 25 µM BSA, pre-treated with 1 mM TCEP, was performed with an increasing concentration of fluorogen from 0 – 250 µM, in the presence and absence of 2 mM GSH. A very low level of background labelling of BSA, and only when 10 equivalents (250 µM) of fluorogen were used, was detected both in the presence and absence of GSH, as witnessed by a weak fluorescent band around 66 kDa (**Figure 3.4.**). As a comparison for the band brightness, similar amount of labelled MBP-dC10 was loaded (see Coomassie stain on *right panel*). The brightness of UV-exposed labelled MBP-dC10 is strikingly superior to the brightness of labelling control with BSA, suggesting that in the case of an intracellular labelling, it should be possible to distinguish clearly between non-target proteins that have an accessible thiol.

By way of comparison, the number of accessible thiols of BSA was determined using the Ellman assay, to ensure that there are some thiols accessible and therefore that BSA is the right protein of choice to perform a control reaction. Ellman assay [95] showed that BSA has in average 0.3 thiols per single protein molecule. Even though this amount seems to be low in comparison to the total number of cysteines of 35 in the primary sequence, it has been previously published that there is one cysteine in a free thiol form per molecule of BSA [118], and experimentally, 0.5 free thiols per molecule of BSA were found [95], which agrees with the number found here.

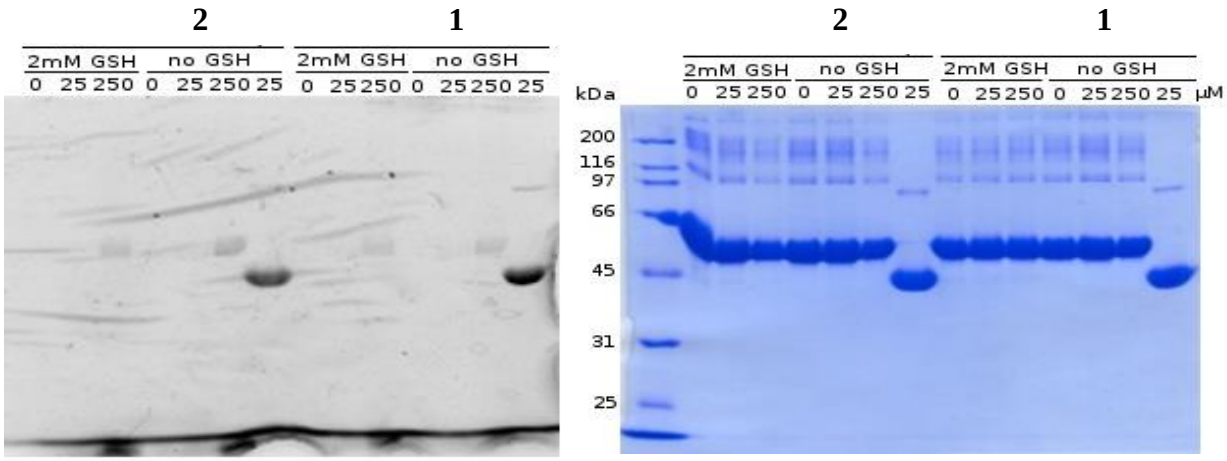


Figure 3.4. Labelling of solvent exposed protein thiols with dansyl-dM10 1 (right of each panel) and dansyl-dM10 2 (left of each panel). 25 μM of BSA (66 kDa) was used along with a 0-250 μM concentration range of each fluorogen, in the presence or absence of 2 mM GSH, as indicated. As an intensity reference, a sample of labelled MBP-dC10 (45 kDa) was included for each fluorogen. The fluorescence of labelled BSA (or MBP-10) was detected by GelDoc using UV excitation over 10 s exposure (*left panel*), after which migrated proteins were stained by Coomassie brilliant blue (*right panel*) to estimate the total protein profile. Protein molecular weight marker was loaded as a reference and corresponding band sizes are indicated in kDa (left lane on Coomassie stained gel).

3.4. Labelling of mNeptune-dC10 in HEK293 cell lysate

The next step towards successful intracellular labelling is testing the reaction in a concentrated mammalian cell lysate. This approach was chosen to increase the amount of dC10-tagged protein expressed in mammalian cells and to facilitate the subsequent detection of fluorescence increase by a plate reader, during the labelling reaction. Furthermore, a test protein that is intrinsically fluorescent was chosen to provide an alternative method of detection, and whose fluorescence is red shifted so as to not interfere with the fluorescence of dansyl or coumarin fluorogens. To this end, a versatile tool for the easy cloning of a dC10-tagged protein in mammalian cells was prepared, namely, the plasmid pJ603-cmyc-MCS-dC10. This plasmid codes for an *N*-terminal cmyc tag that can be used for immunodetection, a *C*-terminal dC10 tag for FlARe labelling and an MCS for the easy

insertion of a gene of interest. For the first mammalian expression of a dC10-tagged protein a new test protein, mNeptune (Ex. 600 nm, Em. 650 nm, [119]), was chosen. Briefly, HEK293 cells were transfected with pJ603-cmyc-mNeptune-dC10 (referred to hereafter as mNep-dC10 for brevity) and after 48 hours of expression, cells were harvested and lysed, divided into several samples, and used for subsequent labelling (for details, see Experimental section on page 59) with the best dansyl compound to date, **dM10-dansyl 1**. In **Figure 3.5.** is shown the fluorescence increase corresponding to the labelling of mNeptune-dC10 with 100 µM and 200 µM **dM10-dansyl 1** (purple and red, *full symbols*). The stable fluorescence signal of mNeptune that over time indicates the excellent stability of the test protein during the time the labelling was performed (*empty symbols*). As a labelling control, cells transfected by an empty vector were used (mock – *blue symbols*). With this simple experiment, a clear mNeptune-dC10 labelling with **dM10-dansyl 1** was achieved where the fluorescence increase is significantly higher (with an initial slope at least three times higher) for cells expressing mNeptune-dC10 (purple and red, *full symbols*) than for mock cells (blue *full symbols*). The unequal signals of mNeptune red fluorescence suggest that the amount of mNeptune in both samples differ, which is surprising due to the fact that both samples originate in one cell culture plate. However, when the red fluorescence signal was used to normalize all data, the same labelling progression curves were obtained for both wells containing mNeptune-dC10, in agreement with the pseudo-first order kinetic curves obtained when the fluorogen is used in high excess in comparison to the mNeptune-dC10 concentration.

Similarly, and as a proof-of-principle of the pJ603-cmyc-MCS-dC10 parent plasmid versatility for labelling applications, a western blot was performed with mNeptune-dC10 HEK293 cell lysate. Clearly, this experiment is not necessary for detection of a fluorescent protein such as mNeptune that can be observed directly using its intrinsic red fluorescence, but it is a necessary proof that the cmyc tag can be used for immunodetection of proteins that are not directly observable.

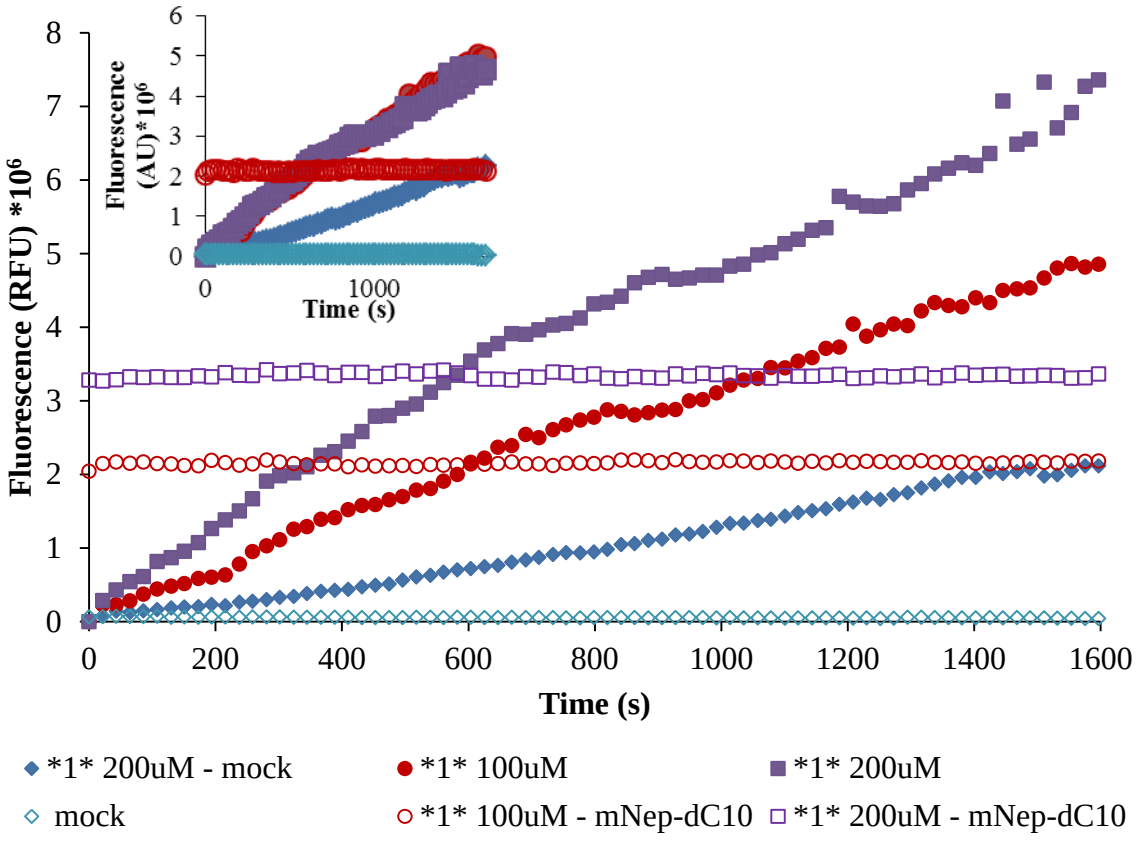
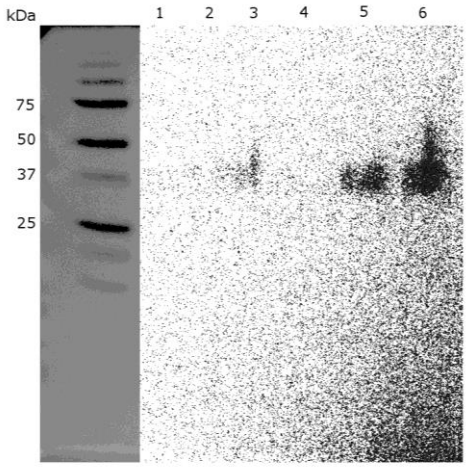
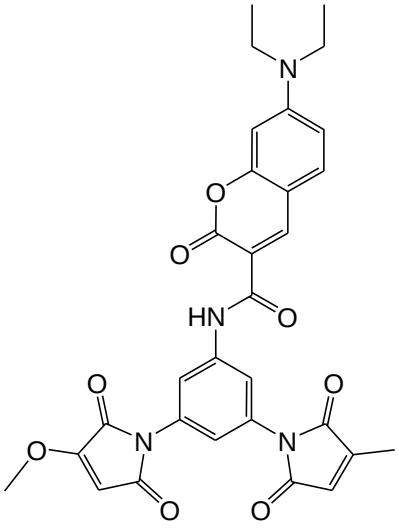


Figure 3.5. Labelling of mNeptune-dC10 in a lysate of HEK293 cells with dM10-dansyl 1 (in legend noted as *1*). Cells were transfected with pJ603-cmyc-mNeptune-dC10 or an empty vector using polyethylene imine (PEI) and protein over-expression was allowed for 48 hours. Cells were detached from plates and lysed and immediately labelled with 100 µM and 200 µM of dM10-dansyl 1. Fluorescence increase was followed at 530 nm upon excitation at 330 nm (*full markers*), and at 650 nm upon excitation at 600 nm (*empty markers*), for FlARe labelling, and mNeptune fluorescence, respectively. In *top left insert*: fluorescence increase corrected by normalization to mNeptune-dC10 red fluorescence.





dM10-coumarin 9

Figure 3.6. Expression levels of cmyc-mNeptune-dC10 (31 kDa) in HEK293 cells after 48 hours detected by Western blot, using an anti-cmyc antibody conjugated to horseradish peroxidase. Amount of DNA used for transfection per one well of a 6-well plate were as follows: 1 – 0.0 µg, 2 – 0.01 µg, 3 – 0.05 µg, 4 – 0.10 µg, 5 – 0.5 µg, 6 – 1.0 g. *Right*: structure of **dM10 coumarin 9** fluorogen.

Using the FlARe labelling design, an intracellular protein can be easily labelled in a mammalian cell lysate. Nonetheless, despite the fact that dansyl-fluorogens are great *in vitro* model compounds, they are not suitable for fluorescence microscopy, as they use wavelengths incompatible with filters of common fluorescent microscopes. At this point, different fluorogens are needed, that are more suitable for widely used fluorescence filters, such as coumarin or BODIPY derivatives, that would also provide an orthogonal signal to the test red fluorescent protein mNeptune. Several coumarin fluorogens were synthesized [75] that take into account the optimized dimaleimide design (with additional methyl- and methoxy- substituents on the maleimide double bond as detailed in Chapter 2) and also the wavelength requirements observation of a maximal fluorescence signal.

Thus, labelling in cell lysate with more cellularly relevant fluorogens, such as **dM10-coumarine 9** (**Figure 3.6.**), was undertaken. Similarly to the previous labelling, mNeptune-dC10 was over-expressed in HEK293 cells and 48 hours after transfection, cells were harvested and lysed, and labelled with **dM10-coumarine 9** (**Figure 3.7.**). The stable red fluorescence of the mNeptune-dC10 transfected cells indicates that the protein is stable and present in the milieu (*empty red symbols*), and that there is no mNeptune-dC10 in mock cells

(*empty blue symbols*). The fluorescence increase at 485 nm over time (*full red symbols*) shows that mNeptune is labelled with **dM10-coumarine 9**, in comparison with mock transfected cells (*full blue symbols*); however, the signal difference is not strikingly high, as it would be desirable for a selective mNeptunde-dC10 labelling. This can be due to a number of reasons: first, a lower concentration of fluorogen (50 µM in comparison with 100 µM of **dM10-dansyl 1**) was used because as seen previously (**Figure 3.5.**), the amount used was sufficient to observe a pseudo-first order kinetics. Secondly, it is possible that the mNeptune-dC10 was expressed at lower levels (as suggested by a lower fluorescence intensity at 650 nm using the same gain), in which case there would be a very little difference between signal from labelled mNeptune-dC10 and background signal from **dM10-coumarine 9**.

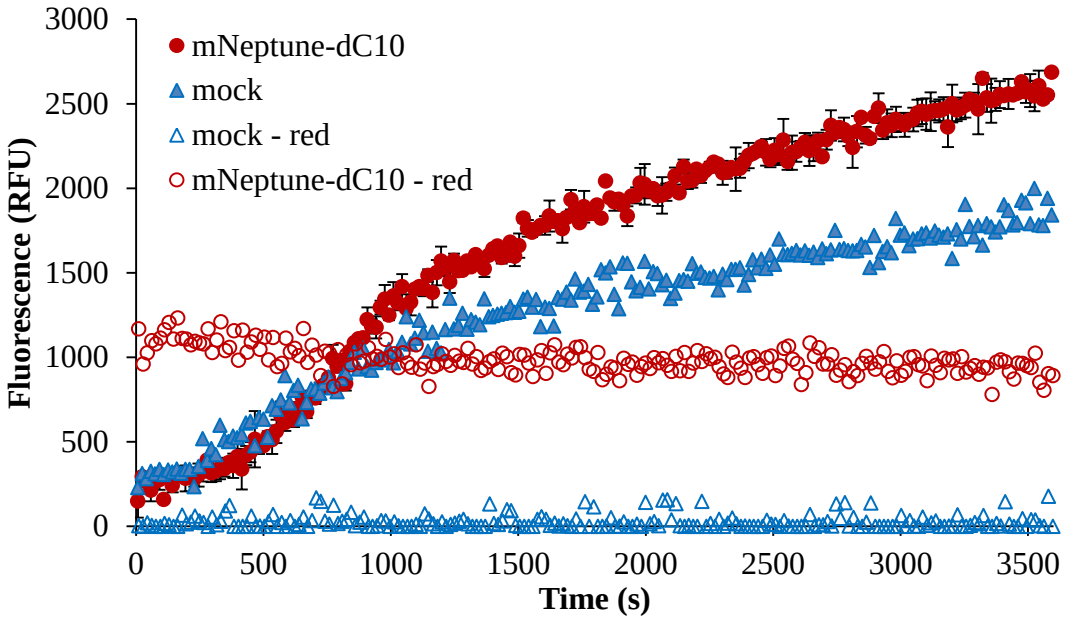


Figure 3.7. Labelling of mNeptune-dC10 in a lysate of HEK293 cells with dM10-coumarine 9. Cells were transfected with pJ603-cmyc-mNeptune-dC10 or an empty vector using PEI and the proteins were expressed for 48 hours. Cells were detached from plates, lysed, and labelled with 50 µM of **dM10-coumarin 9**. Fluorescence increase was followed at 485 nm upon excitation at 420 nm (*full markers*), and at 650 nm upon excitation at 600 nm (*empty markers*), for FlARe labelling, and mNeptune intrinsic fluorescence, respectively. Labelling of mNeptune-dC10 was performed in duplicate and represented error bars correspond to the standard deviation.

Despite the fact that differences noticed between a dC10-tagged protein labelling and background reaction of fluorogen compounds is not consistently high, the FlARe labelling approach is working and could be used for an intracellular labelling. The next section is dedicated to efforts made in obtaining a proof-of-principle intracellular labelling with the FlARe technique.

3.5. Labelling of mNeptune-dC10 in living HEK293 cells

We decided to use a simple approach for labelling an intracellular protein, and a conventional microscope because the presence of an overexpressed protein in the cytoplasm should be rather easy to detect using an epifluorescence microscope, eliminating the need to use a confocal microscope. To that end, HEK293 cells were transfected using Fugene® 6 with pJ603-cmyc-mNeptune-dC10 plasmid and with an empty plasmid as a control, and labelled with **dM10-coumarin 9**. Images were taken at different time points up to 30 minutes after addition of the fluorogen to observe the labelling in real time (**Figure 3.8.**)

In mNeptune-dC10 labelling (*left panel*), increase in green fluorescence intensity corresponds to reaction of **dM10-coumarine 9** for cells that express mNeptune-dC10 (red fluorescence). Similarly, in the same plate, cells that do not express mNeptune-dC10 (and do not have red fluorescence detected in the red filter) have much lower green fluorescence intensity, corresponding to background reaction of **dM10-coumarine 9** with GSH. As a control reaction, mock transfection of HEK293 cells with an empty vector pcDNA3.1(+) and subsequent labelling with **dM10-coumarine 9** (**Figure 3.8.** *right panel*) was performed. These control cells show a green fluorescence corresponding to the background reaction of **dM10-coumarin 9** with GSH; and consistently, its intensity is lower than in the case of mNeptune-dC10 labelling. This correlates very well with the *in vitro* labelling of a cell lysate where **dM10-coumarine 9** reacted with GSH in mock cells and reacted only a little faster with mNeptune-dC10 (**Figure 3.7.**).

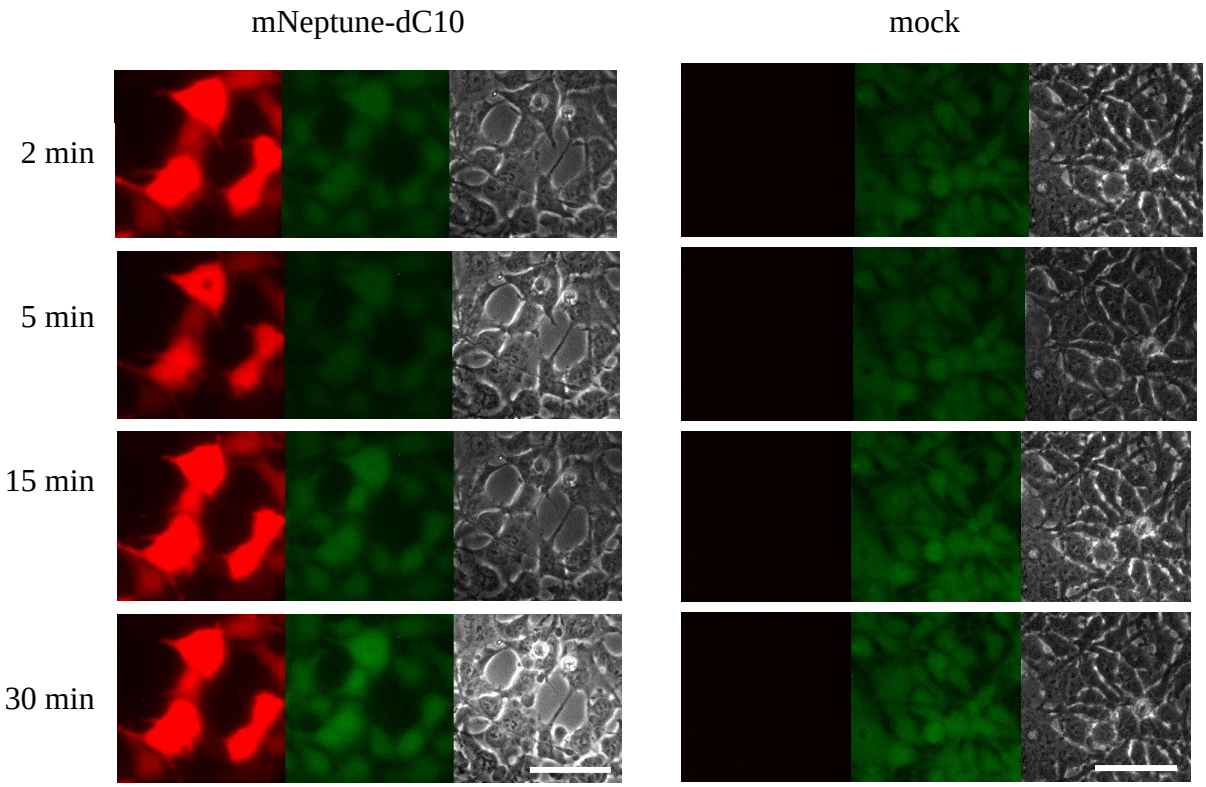


Figure 3.8. *In cellulo* labelling of mNeptune-dC10 with dM10-coumarin 9. Cells were treated with 5 µM of fluorogen and imaged using an inverted epifluorescence microscope in regular intervals until 30 minutes of reaction. *Left*: expression and labelling of mNeptune-dC10, *right*: control labelling with pcDNA3.1(+) vector. Order of filters: red-green-bright field. Images were acquired for 200 ms for red fluorescence (575-640 nm filter), 500 ms for green fluorescence (long-pass 515 nm filter) and 20 ms for bright field. White bar represents 50 µm.

dM10-coumarin 9 is a non-symmetrical fluorogen that is apparently more resistant to GSH thanks to one methoxy- substituent on one of maleimide groups that slows down the reaction with a thiol (see Chapter 2). However, from the results presented above it is apparent that **dM10-coumarin 9** can still be attacked by GSH and cause enough of green fluorescence increase that it can be hard to distinguish a labelled dC10-tagged target protein and background reaction of this fluorogen.

3.6. Conclusions and Perspectives

In this section labelling of MBP-dC10 and mNeptune-dC10 in environments of increasing complexity was carried out to address and evaluate the quality of FlARe labelling inside a mammalian cell. In the absence of GSH, it is possible to label MBP-dC10 in a bacterial cell lysate with **dM10-dansyl 1** and **2**, thanks to their high fluorescence intensity (allowed by a high quenching efficiency of their spacer-less scaffold). Using BSA, we showed that adventitious thiols are not very likely to be labelled by dM10-dansyl fluorogens. In the presence of 2 mM GSH, which mimics intracellular reducing conditions, the labelling of MBP-dC10 was much less efficient, but still prominent enough to be able to distinguish the labelled target protein over the background fluorescence.

In mammalian cell lysate, over-expressed mNeptune-dC10 was labelled and the labelling was clearly predominant over the background reaction of fluorogen **dM10-dansyl 1** with GSH. A much lower selectivity was observed in the case of **dM10-coumarin 9**.

mNeptune-dC10 that is expressed in the cytoplasm of HEK293 cells can be labelled with **dM10-coumarin 9**; however, similarly to results in lysate labelling, this fluorogen can still react with GSH and cause high background fluorescence. A different, less reactive fluorogen is needed for a more successful intracellular labelling where the ratio of fluorescence issued from labelling would be significantly higher than the background reaction of a fluorogen with GSH. A new design for such a fluorogen was done as follows: to decrease the reactivity of the dimaleimide moiety even more, two methoxy- substituted maleimide moieties were used instead of a moderately reactive non-symmetrical methoxy- and methyl- maleimides. This work was done later on in the Keillor group and published with the mNeptune-dC10 test protein prepared and described herein [75].

3.7. Experimental section

3.7.1. Cloning

mNeptune-dC10 mammalian expression plasmid (pJ603-cmyc-mNeptune-dC10) was cloned from a customised vector prepared by DNA2.0 (*www.dna20.com*) containing a cmyc sequence for immunodetection, a multi-cloning site (*AgeI, KpnI, HindIII, BamHI, MfeI, EcoRI, BglII, NotI, XhoI, XbaI*) and a codon optimized dC10 coding sequence, under the control of CMV promoter. mNeptune gene, kindly provided by Professor Robert E. Campbell (University of Alberta), was amplified by PCR using mNep-fw and mNep-bw primers (see **Table 3.1.**) and inserted between c-myc and dC10 coding sequences via *AgeI* and *XhoI* restriction sites. Correct clones for expression of cmyc-mNeptune-dC10 were identified by Sanger sequencing at Génome Québec, Montréal.

Table 3.1. Oligonucleotides

mNeptune_fw	5′- CTAACCGGT ATG GTGAGCAAGGGCGAA - 3′
mNeptune_bw	5′- CCGGCTCGAGCTTATACAGCTCGTCC - 3′

3.7.2. Labelling of MBP-dC10 in E.coli lysate

MBP-dC10 was expressed and purified according to a protocol described in Chapter 2 (page 42) and stored in 50 mM HEPES pH 7.4, 1 mM TCEP. A soluble protein fraction from expression of guinea pig liver transglutaminase [120] was used as a source of *E. coli* proteins that were quantified using Bradford assay. A ratio of 1:10 of MBP-dC10:*E.coli* was used with the mg/mL unit as reference for both samples (0.112 mg/mL MBP-dC10: 1.12 mg/mL *E.coli* soluble proteins corresponding to 2.5 µM:25 µM). For labelling of MBP-dC10 in *E.coli* lysate, all protein components were mixed (and GSH was added for samples where indicated) in 50 mM HEPES pH 7.4, 1 mM TCEP buffer, and the reaction was initiated by addition of 0-100 µM of fluorogen (**dM10-dansyl 1**, **2** or **3**) from a 2 mM stock solution in DMSO. The reaction was carried out at 20°C for 2 hours. All samples were mixed in a 1:1

ratio with Tris-tricine loading buffer without reducing agent (100 mM Tris pH 6.8, 24% (*v/v*) glycerol, 3.5% (*w/v*) SDS, 0.01% (*w/v*) Coomassie brilliant blue) and without boiling before being loaded on a Tris-tricine SDS-PAGE gel. After migration, the in gel fluorescence was evaluated using a GelDoc equipped with a UV lamp. Images were acquired using 10-20 s detection time, as indicated for each fluorogen. After exposure, gels were stained by Coomassie brilliant blue stain.

3.7.3. Control labelling of BSA

Several sources of BSA were tested, among which lyophilized BSA from Sigma-Aldrich showed the most consistent behaviour (expected band size on a SDS-PAGE gel). 30 mg of BSA were dissolved in 50 mM HEPES pH 7.4, 1 mM TCEP and incubated for 30 minutes with a gentle shaking in order to ensure that disulfide bridges would be reduced. After complete dissolution, the protein concentration was determined using Bradford assay and the free thiol concentration using Ellman assay (see page 43), on a TCEP-free sample. The labelling reaction was performed in 50 mM HEPES pH 7.4, 1 mM TCEP at 20°C for 2 hours, where 25 µM of BSA were mixed, in the presence or absence of 2 mM GSH as indicated, with 0-250 µM of **dM10-dansyl 1** or **2** in DMSO. After completion, all samples were mixed with Tris-tricine loading buffer without reducing agent and without boiling, and loaded on a Tris-tricine SDS-PAGE gel. After migration, the in gel fluorescence was evaluated using a GelDoc equipped with a UV lamp. After exposition under UV, gels were stained by Coomassie brilliant blue stain.

3.7.4. HEK293 cell culture and transfection

Human Embryonal Kidney 293 cells (HEK293), kindly provided by Professor Robert N. Ben (University of Ottawa) were grown in MEM minimal media (Life Technologies) supplemented by 10% Fetal Bovine Serum and 1% penicillin and streptomycin, according to the previously published protocol (*www.lifetechnologies.com*, [121]). For cell lysate labelling, cells were transfected using PEI, 16 hours post plating of 6 x 10^5 cells in a 6-well plate. Briefly, PEI was mixed with 100 µL MEM (without serum) and incubated at room temperature for 5 minutes, and a mixture of DNA/MEM (without serum) was added. After incubation for 10 minutes, the DNA/transfection agent complex was added on cells.

For labelling and subsequent fluorescence microscopy detection, cells were transfected using Fugene® 6 (Promega), 16 hours post plating of 6 x 10^5 cells in a 35-mm plate equipped with a microscopy coverslip.

3.7.5. Labelling of mNeptune-dC10 in HEK293 cell lysate

Transfected cells were grown for 24-48 hours to ensure proper protein expression, after which they were washed twice with Ca^{2+}-free and Mg^{2+}-free PBS (HyClone™), harvested using trypsin (HyClone™), and lysed in a 50 mM HEPES pH 7.5, 1 mM TCEP buffer containing 1% (*v/v*) Triton X-100, and with a cycle of freeze-thaw in dry ice, or sonication only. For labelling with **dM10-dansyl 1** one well of a 6-well plate of cells was used for 3 labelling experiments in 150 µL in a black 96-well plate (Greiner), and in case of **dM10-coumarin 9**, one 100-mm plate was used to afford 2 mL of sample of which one labelling experiment in a 96-well plate used 150 µL. The fluorescence increase corresponding to labelling was followed at 530 nm and 485 nm upon excitation at 330 nm and 420 nm, for **dM10-dansyl 1** and **dM10-coumarin 9**, respectively. Fluorescence of mNeptune was followed at 650 nm upon excitation at 600 nm.

3.7.6. Labelling of mNeptune and detection by fluorescence microscopy

For detection of fluorescence by microscopy we used an inverted Zeiss Axio epifluorescence microscope that requires use of cells cultured in a 35 mm plate where the growing surface was replaced by a microscope coverslip. Despite the fact that HEK293 cells are mostly adherent to primary amine-modified surfaces, they also adhere well enough to bare microscope quality glass. 16 hours post-plating, the cells were transfected using Fugene® 6, as detailed above.

The day of labelling, the medium was changed to Opti-MEM and the indicated amount of fluorogen was added, from a 16-mM stock in DMSO. Plates were imaged at post-addition times indicated using specific acquisition times for each filter (200 ms for red fluorescence, 500 ms for green fluorescence and 20 ms for bright field). Detection of fluorescence was carried out with 3-colour inverted Zeiss Axio Observer A1 epifluorescence microscope equipped with the following filters (excitation, emission): Green (450-490 nm, LP 515), Red (534-558 nm, 575-640 nm); and a 120 W Hg fluorescent light source X-cite series 120 for

excitation. Images were acquired with QICAM Fast 1394 CCD monochrome camera and processed with ImageJ.

3.7.7. Western blot

Proteins from the lysate of cells expressing mNeptune-dC10, with an increasing amount of DNA used for transfection, were separated on a Tris-tricine SDS-PAGE gel that was subsequently washed in a transfer buffer for one hour at room temperature. Proteins were transferred to a nitrocellulose membrane in a transfer buffer (25 mM Tris pH 8.5, 0.2 M glycine, 20% (*v/v*) MeOH) for 1 hour at 4°C at 100 V. The membrane was washed in TBS (10 mM Tris pH 7.6, 138 mM NaCl), 5% (*w/v*) milk, 0.1% (*v/v*) Tween-20 overnight prior to incubation with TBS containing 5% (*w/v*) milk, 0.1% (*v/v*) Tween-20 and Anti-cmyc Rabbit antibody (NEB) for 2 hours at 4°C to detect cmyc-mNeptune-dC10, after which the membrane was washed three times with TBS, 5% (*w/v*) milk, 0.1% (*v/v*) Tween-20 solution. Secondary HRP-conjugated Anti-Rabbit antibody was incubated with the membrane for 1 hour at room temperature in TBS, 5% (*w/v*) milk, 0.1% (*v/v*) Tween-20, after which the membrane was washed four times with TBS, 0.1% (*v/v*) Tween-20 alone. A 1:1 mixture of Luminol solution and hydrogen peroxide (both NEB) was used for imaging with a standard CCD camera.

Chapter 4

OPTIMIZATION OF DICYSTEINE 10 PEPTIDE TAG

4.1. Introduction

4.1.1. Structure and reactivity of dC10

The two cysteine residues present in the dC10 peptide are separated by two turns of the alpha-helix, and therefore by ~10 Å (see Chapter 1), to perfectly complement the design of our fluorogen molecules where reactive maleimide moieties are also ~10 Å apart. This di-cysteine tag was used as a reference point for the characterisation of most of the fluorogenic labelling agents synthesized in the Keillor lab, as detailed in [73] and in Chapter 2 of this thesis. As such, the dC10 peptide has proven to be broadly useful. However, we recognise that if the di-cysteine tag were even more reactive, it would offer a kinetic advantage to the labelling reaction, conferring selectivity to the labelling method, relative to the background reaction of dimaleimide fluorogens with adventitious thiols such as glutathione. In this chapter the question of dC10 structure will be first re-addressed, and then its sequence will be evolved to a more reactive di-cysteine tag.

4.1.2. Circular Dichroism

Previously, it was established by circular dichroism (CD) that dC10 peptide adopts overall a largely helical conformation [73] at pH 7.0, before and after labelling with **dM10-FITC** fluorogen (**Table 1.4.**). However, the information provided by CD is an average helical content of the whole peptide, and the extent of helical conformation within the dC10 peptide sequence is not known. The peptide may be sampling helical conformation only locally, such as at its *C*- or *N*- termini or in the central region, all of which would lead to a CD spectrum with a substantial helix signal. Thus, here we will study the conformation of dC10 by solution NMR, which will allow us to obtain information about dC10 secondary structure at both, residue and atomic level, as an important starting point for the evolution of dC10 for higher reactivity.

4.1.3. Choice of system for dC10 NMR studies

dC10 is a 23-amino acid peptide that would be difficult to express alone in bacteria or prepare by synthesis in sufficient concentrations for NMR. To overcome this difficulty, the dC10 tag was fused to a well-characterized protein whose expression in ^{13}C and ^{15}N-labelled

form is easy, inexpensive and quick. In the case of isotope labelled proteins, it is often a challenge to obtain good bacterial growth and final yields even for well-established proteins; therefore, a cell-free expression technique was explored, a complementary strategy to bacterial protein expression that is carried out *in vitro*. This technique uses isotopically labelled amino acids instead of salts and a carbon source; it is therefore more versatile in terms of residues that will be labelled by ^{13}C and ^{15}N than bacterial expression. In this case, the choice of expression system narrowed down our choice of fusion protein with dC10 to the peptidyl-prolyl *cis-trans* isomerase B (PpiB) from *E. coli*, one of the pilot proteins that was successfully expressed in its isotope labelled form, in high yields, using the cell-free expression system [122, 123, 124].

4.2. Structural analysis of dC10 in fusion with Peptidyl-prolyl isomerase B (PpiB)

4.2.1. Cell-free expression versus M9 minimal media expression

As the cell-free expression approach has been previously carried out with success in other laboratories on PpiB [123], we chose to use PpiB as well as our test protein, and prepared an expression plasmid for hexahistidine-tagged H_6-PpiB-dC10 (referred to as PpiB-dC10 in future for the sake of brevity). The detailed cloning approach is described in the Experimental section (see page 104).

As a first step, the S30 ribosome was isolated from *E. coli* BL21 Star (DE3) strain according to the published protocol ([125, 126], see Experimental section on page 103). A total of 96 1-mL aliquots were obtained that were snap-frozen and kept at -80 °C for future use. As an efficiency test, this extract was used to express unlabelled PpiB-dC10 and PpiB-CTG (plasmid donated by Professor Christopher Easton, Australian National University) as a performance reference for PpiB-dC10. The expression of both proteins in parallel was carried out in small reaction volumes of 500 µL, at 37°C and for 24 hours, as protein yields were still increasing after 6 hours of expression. Both proteins were then purified by immobilized metal affinity chromatography. SDS-PAGE gel (**Figure 4.1.** *panel A***)** showed

comparable expression levels for PpiB-CTG and PpiB-dC10 that were confirmed by absorbance at 280 nm. Hence, we can confirm that the extracted S30 is as efficient for PpiB-dC10 expression, as it is for PpiB-CTG. However, the yield of purified protein was variable (**Figure 4.1.** *panel B and C*) and often, the overall amount of expressed protein was too low (see **Figure 4.2.** *panel B*) to be used with the NMR spectrometers currently available to our laboratory.

Seeing this result and bearing in mind that, due to the cost of isotopically labelled components (amino acids, salts and carbon source), an expression of a protein for NMR has to be reliable, a conventional expression of PpiB-dC10 in BL21-Gold(DE3) strain of *E. coli* cells in M9 minimal media was carried out for comparison. We used 250 mL of M9 minimal media supplemented with ^{15}N-labelled ammonium chloride, a conventional approach for a histidine-tagged protein purification, and obtained 8-10 mg of purified protein (**Figure 4.2.** *panel B*) at multiple instances. MALDI analysis also confirmed that the ^{15}N-labelling efficiency of PpiB-dC10 using M9 minimal media is around 92%, which is satisfactory to obtain good NMR spectra.

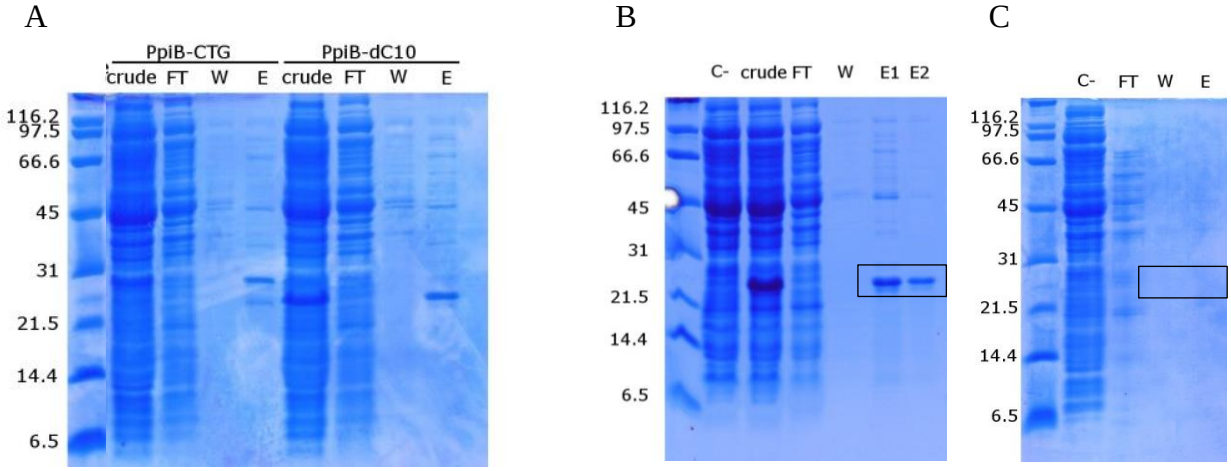
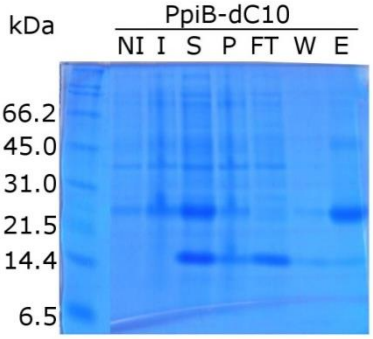


Figure 4.1. SDS-PAGE analysis for PpiB variant cell-free expression. *A*: Expression of PpiB-CTG (24 kDa) and PpiB-dC10 (22 kDa) side by side, *B* and *C*: Expression of PpiB-dC10 on different days; Crude – S30 extract after expression; FT – flow through from NiNTA resin; W – wash from NiNTA resin with sodium phosphate buffer containing 20 mM imidazole; E, E1 or E2 – elution of PpiB variant with sodium phosphate buffer containing 500 mM imidazole. Expected band locations of PpiB-dC10 are indicated by a rectangle. Broad range molecular weight marker was included as a reference and is annotated on the left of each gel (kDa).

A



B

Expression type	Amount of purified PpiB-dC10 (A280)
Cell-free expression / 500 µL	0.1 - 0.2 mg
Bacterial expression in M9 minimal media / 250 mL	8-10 mg

Figure 4.2. SDS-PAGE analysis of *E. coli* expressed PpiB-dC10 and yield comparison. *A*: SDS-PAGE gel of PpiB-dC10 (22 kDa) purification. NI - non induced bacteria, I – induced bacteria, S – supernatant, P – insoluble protein fraction, FT – flow through from NiNTA resin, W – wash with PBS buffer containing 20 mM imidazole, E – eluted protein with PBS containing 250 mM imidazole. *B*: Yields of purified PpiB-dC10 obtained by cell-free and bacterial expression.

In order to obtain reasonable quality NMR spectra of a 23-kDa protein as is required for chemical shift assignment, it is necessary that the sample concentration be approximately 0.5-1 mM in a 300-µL sample volume. Given the yields of the *in vitro* cell free expression system, ~10 mL reaction volume would be required to produce one NMR sample. The technique of cell-free expression usually benefits from small expression volumes that give high protein yields and allow some cost savings due to smaller concentrations of labelled compounds used; however, in this case, the cost savings of a cell-free expressed protein are countered by its lower reliability and therefore, did not present a real advantage. For the rest of work with PpiB-dC10, we opted for a classic bacterial expression in M9 minimal media.

4.2.2. Assignment of PpiB-dC10 backbone

Assignments of amide backbone resonances and α and β carbons of PpiB were previously published by Kariya *et al*. [122] and deposited in the BMRB database (access number 4765). These published assignments were used as a starting point for assignment of PpiB-dC10 backbone resonances, and subsequently, we used a standard suite of triple resonance spectra (namely HNCO, HNCACB, and CBCA(CO)NH spectra) to assign remaining backbone atoms, as well as some dC10 backbone resonances. We were able to assign a few more resonances of dC10 using ^{15}N-TOCSY and ^{15}N-NOESY spectra that allowed the

identification of some of the overlapping alanine signals of the dC10 tag. It is noteworthy that due to the repetitive EAAAR sequence of dC10, the assignment of alanine signals was particularly difficult, as the corresponding peaks often overlapped. Approximately 95% of PpiB-dC10 residues were successfully assigned, as well as PpiB Hα resonances that had not been published previously in BMRB. The final assignment of PpiB-dC10 is listed in **Appendix 1**.

4.2.3. Effect of dC10 on PpiB structure

NMR analysis on PpiB-dC10 and PpiB allows to determine whether if the presence of a dC10 tag on the *C*-terminus of a protein perturbs the protein structure in any way. Two different references were used to address the effect of dC10 on PpiB: First, the obtained PpiB-dC10 backbone assignments were compared to those of PpiB published by Kariya *et al.* [122], and secondly, a sample of PpiB resulting from cleavage of the dC10 tag from PpiB-dC10 using Factor Xa was compared to PpiB-dC10 (for detail, see Experimental section page 106). An overnight reaction gave a quantitative amount of sole PpiB with a minimum of uncleaved PpiB-dC10 remaining, as demonstrated by MALDI spectra (**Figure 4.3.**).

An overlay of PpiB and PpiB-dC10 ^{1}H-^{15}N HSQC spectra shows that most peaks arising from PpiB have the same chemical shift in both forms (**Figure 4.4.**), suggesting that the presence of the dC10 tag on the *C*-terminus of PpiB has little impact on its overall fold and structure.

It is important to note that we have not attempted to acquire an NMR spectrum of a fluorogen-labelled PpiB-dC10 because of poor sample stability, due to protein precipitation at high concentrations after labelling. This spectrum could have been useful for identification of dC10 cysteine signals but our experience with dM10 fluorogens used at that point shows that they are not soluble enough to perform this kind of experiment.

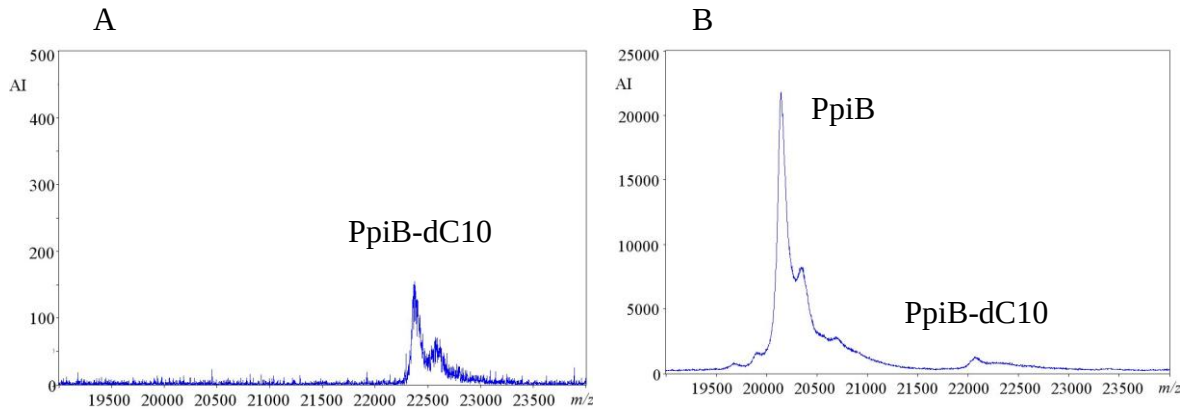


Figure 4.3. MALDI spectra of ^{15}N-labelled PpiB-dC10 (A) and PpiB (B). PpiB was produced by Factor Xa cleavage of dC10 tag from PpiB-dC10 in 20 mM Tris pH 8.0, 100 mM NaCl, 2 mM $CaCl_2$ overnight, at 23°C. Theoretical masses for PpiB-dC10 ^{15}N and PpiB ^{15}N are 22388.7 Da and 20140.3 Da, respectively, according to ProtParam (*www.expasy.org*).

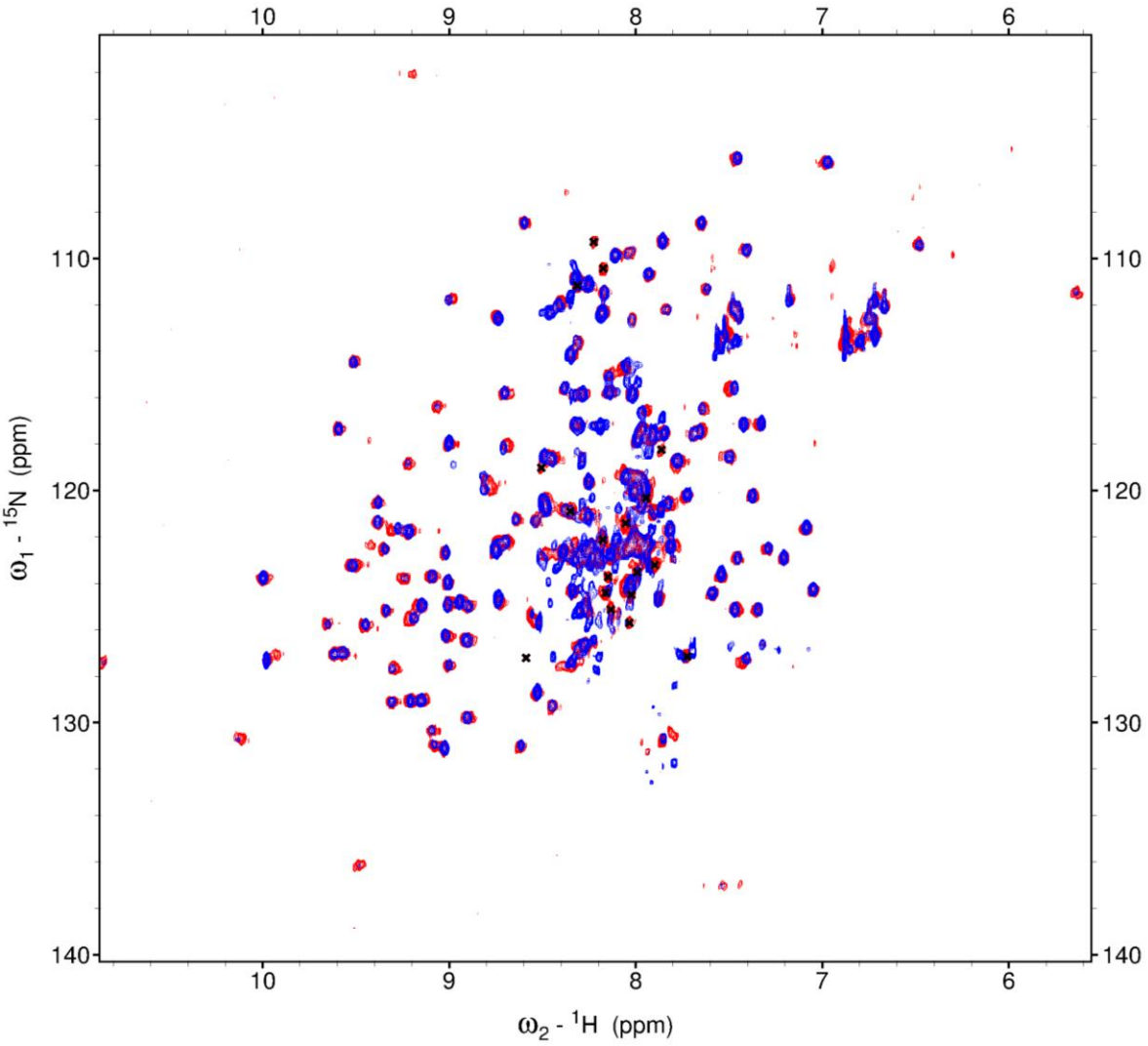


Figure 4.4. Superposition of PpiB (*blue*) and PpiB-dC10 (*red*) ^{1}H-^{15}N HSQC. PpiB sample was produced by cleavage of dC10 tag from PpiB-dC10 by Factor Xa protease overnight at 23°C.

In order to identify any PpiB residues that showed significant chemical shift perturbations from the introduction of the dC10 tag, average amide shift differences were calculated (see equation in **Figure 4.5.,** [127]) using PpiB shifts deposited in BMRB as a reference. *C*-terminal residue E164 of PpiB was not included in this calculation because it is the point of dC10 tag attachment and is expected to be directly affected. As anticipated, most residues showed only small shift differences, with an average shift difference of 0.040 ± 0.045 ppm. Some residues close to the N-terminus (V12-F16) and several residues at the *C*-terminus (E125- V127, T161-E164) of PpiB were significantly affected by the presence of dC10 on the *C*-terminal end of PpiB. As seen in the PpiB structure, all these residues are located at the *N*- and *C*-termini that would be proximal to both the hexahistidine tag on the *N*-terminus and dC10 on the *C*-terminus of PpiB (see **Figure 4.6.**).

$$\Delta\delta_{avg}\,(ppm) = \sqrt{\frac{1}{2}\left(\left(\frac{\Delta\delta^{15}N}{5}\right)^2 + (\Delta\delta^{1}H)^2\right)}$$

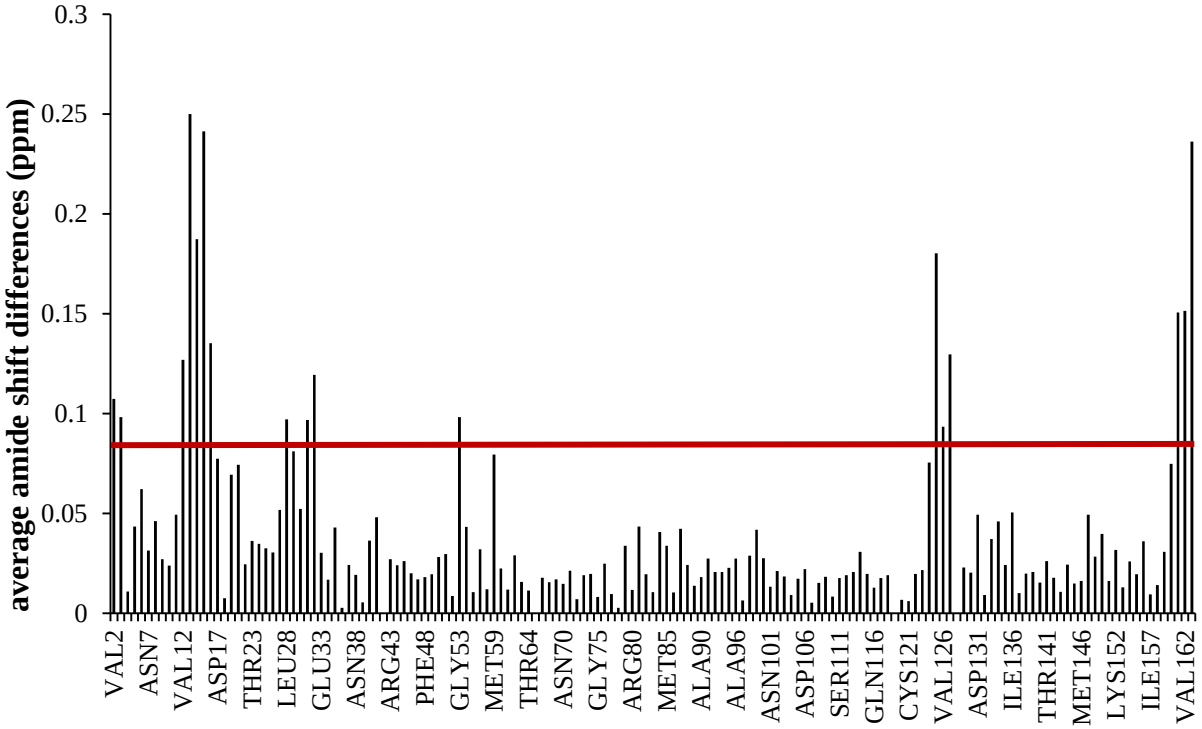


Figure 4.5. Average amide shift differences of PpiB-dC10 and PpiB. Resonance values for PpiB were obtained from BMRB database and those of PpiB-dC10 were determined in this work. The red line indicates an average value of average amide shift differences across the PpiB primary sequence (0.040 ± 0.045 ppm) incremented by the standard deviation, to yield 0.085 ppm. Equation used for average amide shift differences is indicated above.

Thus, the presence of these small changes is not surprising as the chemical environment variation for these residues is most likely due to a change in local environment caused by the presence of the dC10 tag, and not by a change in the global fold of PpiB.

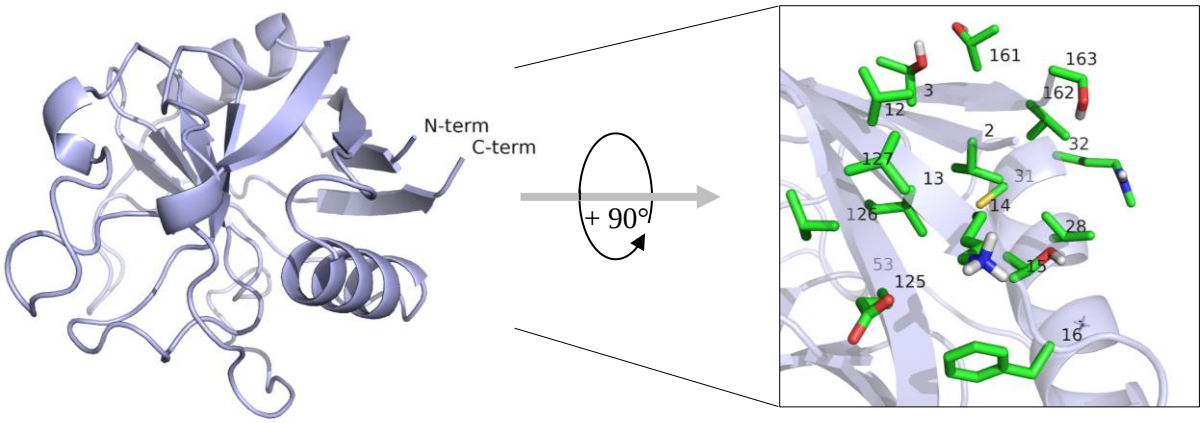


Figure 4.6. Structure of PpiB (*left*) and representation of residues that have a significant average amide shift difference in PpiB-dC10 (*right*). *C*- and *N*-termini (residues 2 and 163) are labelled for clarity and side chains of residues presenting a significant chemical shift change (**Figure 4.5**.) are coloured in *green* (carbon), *blue* (nitrogen), *red* (oxygen), *yellow* (sulphur) and *white* (hydrogen), and identified by their residue number. The inserted figure was obtained by zooming and rotating the structure by + 90°.The dC10 tag is not represented for clarity. PDB : 2NUL

4.2.4. Secondary shift analysis

Protein backbone chemical shifts can be used in a secondary shift calculation to accurately predict secondary structure at a residue level, an analysis that was applied for the dC10 peptide tag. Random coil resonance prediction software CamCoil ([128], *http://www-vendruscolo.ch.cam.ac.uk/camcoil.php*, accessed on March 10th, 2012) was used to generate reference Cα chemical shifts for an unstructured random coil dC10 sequence. These values were subtracted from assigned chemical shifts that were determined above in order to obtain Cα secondary shifts for dC10 (**Figure 4.7.**). Equally, similar Cα secondary shifts were obtained when a different reference values for resonances of random-coil were used (Wishart

et al. [129]). It has been previously determined [130, 131, 129] that positive Cα secondary shifts are typical for α-helical conformation. Here we observe that in dC10, the central region of the tag that encompasses the two Cys residues show Cα secondary shifts that are predominantly positive. This confirms that under the conditions of our experiment, the central region of the dC10 tag adopts an α-helical conformation, as previously suggested by CD experiments (see page 65). It is noteworthy that the absolute values of secondary chemical shifts are smaller than expected for a stable α-helix [132], indicating that segments of dC10 are likely dynamic, but a significant proportion of the population samples an α-helical state. It is notable as well that Cα secondary chemical shifts for *C*-terminal glycines are negative instead of positive, which may reflect a tendency for the terminal GGK sequence to form a *C*-cap, a structural feature that does not have the same characteristics as a regular α-helix. A similar observation has been made by Shen and Bax [133], who determined that the secondary chemical shifts of a *C*-cap can be very close to zero or negative.

Another factor that is likely to influence the degree of secondary structure content of dC10 is the temperature and pH used for NMR spectra acquisition. Conditions of pH 6.2 and 37°C were chosen to accommodate PpiB stability, and to allow comparison of our results to the published assignments that were obtained under these conditions; however, these conditions may decrease the helicity of dC10 attached to PpiB in comparison with conditions used for labelling reactions, namely pH of 7.5 and 20°C. This is substantiated by helix content predictions for dC10 using AGADIR (**Figure 4.8.**) suggesting that while pH changes should not significantly alter helicity, temperature changes can have a dramatic impact. According to this prediction, the helical content of dC10 decreases from 40% helix at 20°C (the temperature of most *in vitro* kinetic and CD experiments that were performed with dC10) to 20% helical content at 37°C (used for NMR spectra acquisition). This may also explain why the values of dC10 Cα secondary shifts were relatively small, and if it had been possible to conduct these experiments at lower temperatures, these values could be expected to increase.

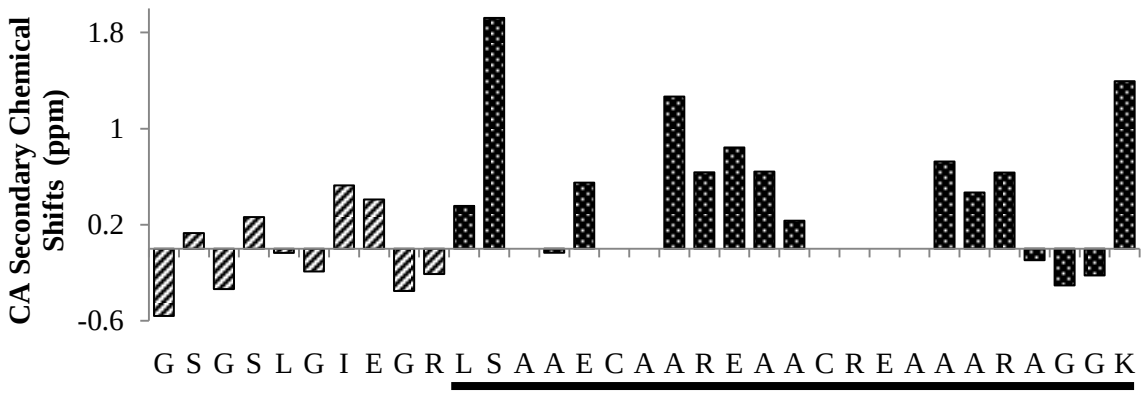


Figure 4.7. Cα Secondary chemical shift analysis on dC10 peptide tag. The sequence represented includes a spacer GSGS and a Factor Xa restriction site before dC10 sequence (*underlined*).

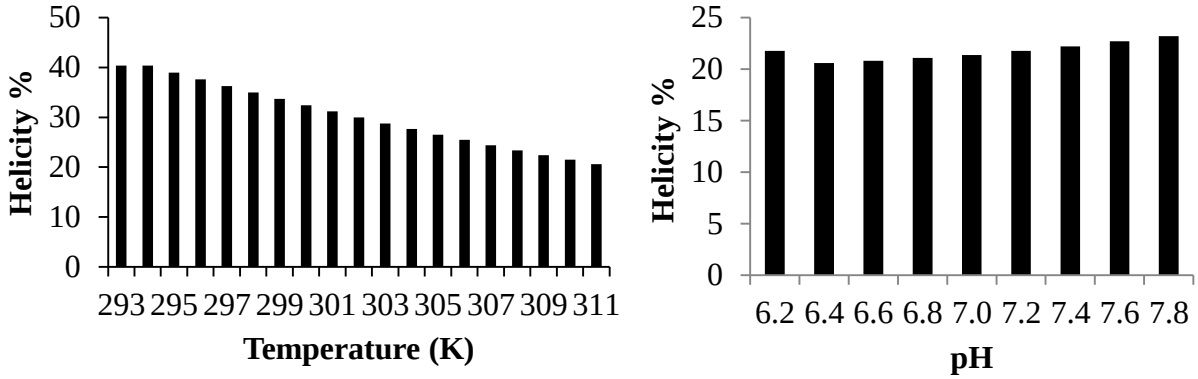


Figure 4.8. Change in helicity with increasing temperature and decreasing pH. *Left*: pH was maintained at 6.2 and helicity was predicted for temperature increasing from 294 K (20°C) to 310 K (37°C). Helicity% for pH 6.2 at 293 K was predicted in a separate calculation; *right*: Temperature was maintained at 310 K and pH was varied from 6.2 to 7.5. All helicity content % were predicted by AGADIR at an ionic strength of 0.1 M.

4.2.5. Conclusion on PpiB-dC10 structure

We studied the structure of dC10 in fusion with PpiB by NMR and we assigned dC10 backbone chemical shifts over most of the length of the tag. The results have shown that a) the attachment of dC10 on the *C*-terminus of PpiB does not affect the structure of the protein it is attached to under the conditions used for NMR, and b) dC10 adopts the largely α-helical conformation predicted by its design, as confirmed by secondary shift analysis on Cα

carbons, even though the higher temperatures used in this study likely had a destabilizing effect on this α-helical structure. The helical structure confirmation of dC10 is an essential starting point for its improvement by mutagenesis in to a more reactive tag that is detailed in the next section.

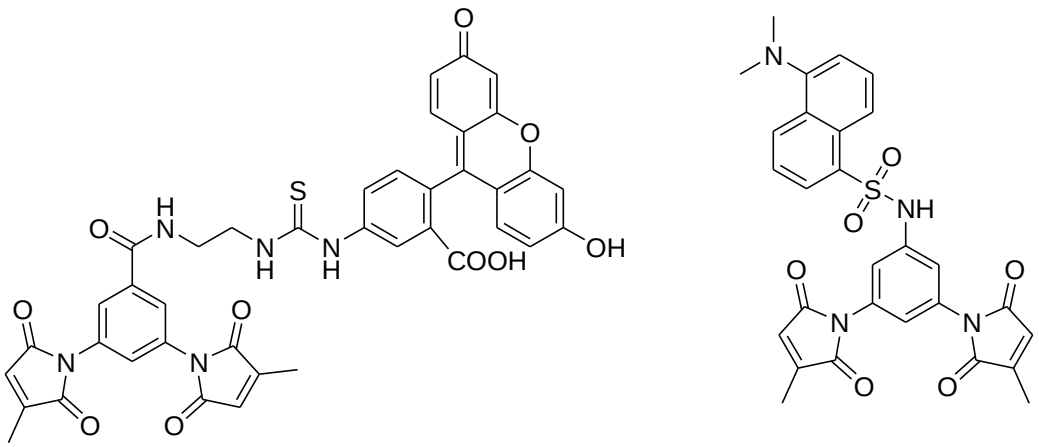
4.3. Preliminary work for dC10 sequence optimization

4.3.1. Introduction

Previous work in the Keillor group attempted to enhance the reactivity of the dC10 tag through rational design, introducing histidine residues in the proximity of the reactive cysteine thiols; however, no improvement in was observed relative to the parent dC10 sequence [73], as detailed in Chapter 1. In this section we describe a more combinatorial approach, wherein libraries presenting a broad variety of residues at specific positions were prepared and screened for the rate of their thiol addition reaction with a fluorogenic molecule.

4.3.2. Preliminary work on dC10 mutants

Previously, there was an attempt to enhance the stability of the reactive thiolates of C13 and C6 [73] in dC10 by introducing a histidine residue in positions adjacent to C6 (i.e. at position A7), or one turn of the helix away from C13 and/or C6 (i.e. at positions R9 or S2). Histidine mutations were chosen since these residues are known to be excellent proton donors and acceptors at physiological pH due to the pK_a of their side chain imidazole group, as demonstrated in their role in protease mechanisms [83, 84]. Some double mutants were also created, where A17 and S2 (one turn away from reactive C6 and C13) were both replaced by histidines. Second order rate constants were then measured for the reaction of these mutant sequences with the **dM10-FITC** fluorogen (**Figure 4.9.**). None of the mutant sequences exhibited better reactivity that the parent dC10, which was attributed to the potential destabilisation of the mutant dC10 helices by the histidine residues (see Chapter 1, page 18 and [87]).



dM10-FITC **dM10-dansyl 1**

L_1SAAE**C**AAREAA**C**REAAARAGGK$_{23}$

Figure 4.9. dM10-FITC and dM10-dansyl 1 (*top*) and dC10 peptide primary sequence (*bottom*).

4.3.3. New characterization of existing dC10 histidine mutants

More recently, we decided to undertake a more combinatorial approach towards dC10 sequence optimization, and as a starting point, we repeated some of the previous experiments using the best fluorogen available at that point, **dM10-dansyl 1** (**Figure 4.9., Figure 4.10** *panel B, white bars*). These original results demonstrated that the double mutant S2H-A17H is more reactive than the parent dC10 (**Figure 4.10.** *panel B*). However, the single mutants S2H, A7H and R9H do not appear to be more reactive than dC10, leading us to infer that only the histidine mutation at position 17 is responsible for the greater reactivity of the double mutant S2H-A17H. To understand the low reactivity of single mutants S2H, A7H and R9H, labelling in presence of secondary structure stabilizing agent was performed and is detailed in the next section.

4.3.4. Labelling of dC10 histidine mutants in presence of secondary structure stabilizing agent

As mentioned above, histidine residues can greatly destabilize the helical conformation of peptides [87]. If this loss of helicity is recovered, the reactivity of dC10 histidine mutants could be restored to the level of dC10 at least, if not benefit further from the ability of the

partially positive charges of histidine side chains at physiological pH to stabilize cysteine thiolates, making these dC10 mutants more reactive. For that purpose, 2,2,2-trifluoroethanol (TFE), a solvent that is known to stabilize secondary structure features in peptides in solution, as studied by Roccatano *et al.* [134], was used. It was previously observed that at low concentrations (below 10% (*v/v*)) TFE interacts with the carbonyl oxygen and with the hydrophobic moieties of proteins [135], whereas at high concentrations, TFE penetrates in the hydrophobic core of proteins and causes its unfolding but it can also locally stabilize secondary structural motifs [135, 136]. The minimal concentrations of TFE that have been reported as having a beneficial effect on the formation of secondary structure motifs (studied by circular dichroism) [137] varies between 30-40% (*v/v*). We performed preliminary labelling tests in 50 mM HEPES, 1 mM TCEP, 30% (*v/v*) TFE with an equimolar concentration of protein and fluorogen of 50 µM, but observed significant precipitation of our protein. The TFE concentration was therefore lowered to 16% (*v/v*) but the same cloudiness in the solution was observed, and it was only using 5% (*v/v*) TFE concentration that no cloudiness, suggesting precipitation, was observed by naked eye, similarly to what was reported by Povey *et al.* [135].

Labelling of all dC10 histidine mutants with **dM10-dansyl 1** was performed in presence of 5% (*v/v*) TFE and corresponding rate constants were determined (**Figure 4.10.** *panel B, black bars*). A 1.5-3 fold increase in reactivity of dC10 histidine mutants was observed across all investigated mutants in the presence of 5% (*v/v*) TFE, which suggests that there may be an α-helical stabilization of dC10 tags. However, this rate enhancement might be also caused by a simple co-solvent effect of TFE. To address this question, a control experiment was performed, where MBP-dC5 reacted with **dM10-dansyl 1** in presence of 7.5% (*v/v*) DMSO (instead of 5% (*v/v*) TFE and 2.5% (*v/v*) DMSO). Indeed, there was a slight increase in reactivity due to presence of 7.5% (*v/v*) DMSO in comparison with usual 2.5% (*v/v*) DMSO (**Figure 4.10.** *panel C*), however this increase was not as prominent as the reactivity increase in presence of TFE in particular, where the second order rate constant increased by a factor of 2.6 in comparison with labelling in 7.5% (*v/v*) DMSO. Hence, a co-solvent effect as an origin of higher reactivity of dC10 mutants in presence of TFE can be ruled out.

A

dC10	L_1SAAE**C**AAREAA**C**REAAARAGGK$_{23}$
dC10-S2H	L_1H_2AAE**C**AAREAA**C**REAAARAGGK$_{23}$
dC10-A7H	L_1SAAE**C**H_7AREAA**C**REAAARAGGK$_{23}$
dC10-R9H	L_1SAAE**C**AAH_9EAA**C**REAAARAGGK$_{23}$
dC10-S2H-A17H	L_1H_2AAE**C**AAREAA**C**REAH_{17}ARAGGK$_{23}$

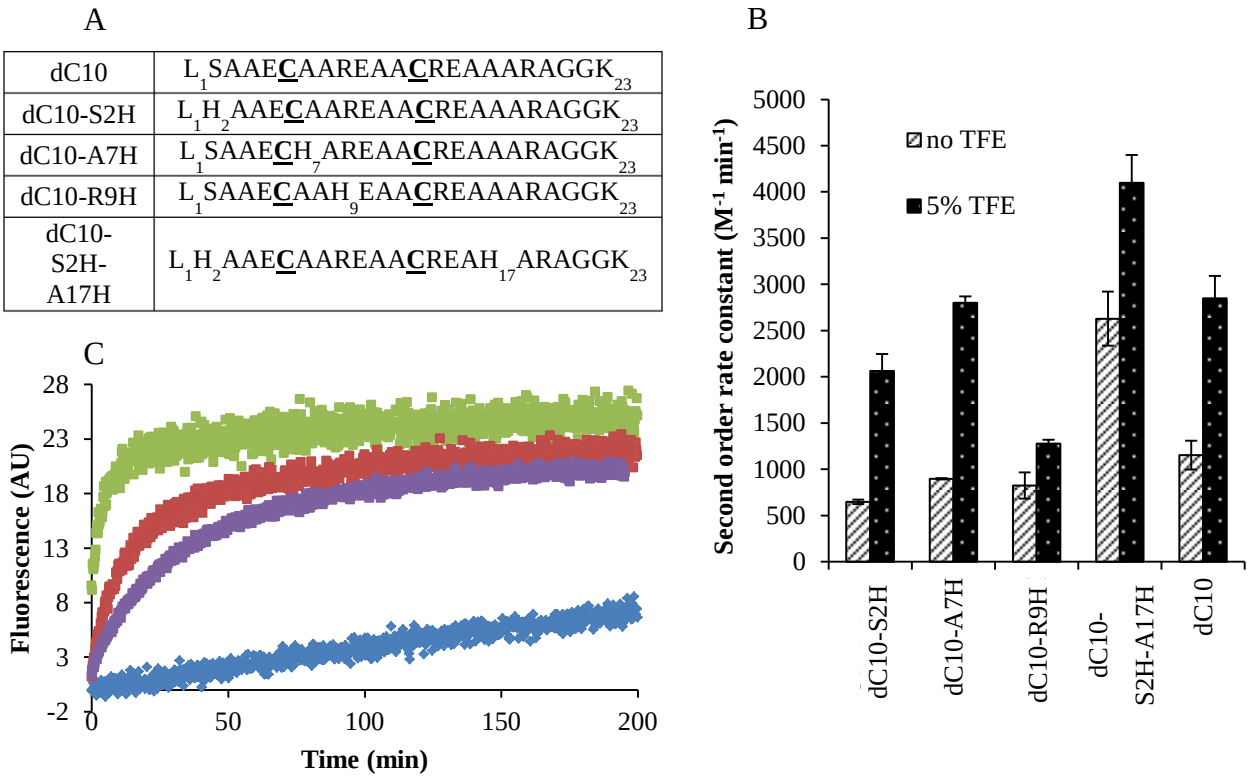


Figure 4.10. Kinetics of labelling reaction on previously investigated histidine dC10 mutants. *Panel A*: Sequences and nomenclature of dC10 histidine mutants; *panel B*: second order rate constants of addition of **dM10-dansyl 1** on MBP-dC10 mutants, n = 3; *panel C*: Fluorescence increase during labelling of MBP-dC5 with **dM10-dansyl 1** was followed in the presence of 2.5% (*v/v*) DMSO (*purple*), 5% (*v/v*) TFE and 2.5% (*v/v*) DMSO (*green*), 7.5% (*v/v*) DMSO (*red*), as well as background reaction of **dM10-dansyl 1** in buffer with 7.5% (*v/v*) DMSO (*blue*), n = 2. Curves were normalized to obtain similar fluorescence intensity at the end of the reaction.

Then again, the presence of TFE may impact the reaction rate by means other than helicity stabilization or co-solvent effect. To address that concern, the reaction between a **dM10-dansyl 1** and a *non-helical* thiol was studied in the presence of TFE. 2-mercaptopropionic acid (MPA) was chosen for this purpose, as it has been previously employed for the characterization of most of fluorogens in the Keillor group. The reaction between **dM10-dansyl 1** and MPA was carried out under pseudo-first order conditions where 1 equivalent of fluorogen was used with 10 equivalents of MPA. As shown in **Figure 4.10.**, the addition of 5% (*v/v*) TFE increases the initial rate of reaction even of the non-peptidic thiol MPA, by a factor of 1.5 (for both kinetics corrected for hydrolysis and non-corrected kinetics).

However, most of the second order rate constants observed for MBP-dC10 histidine mutants were elevated by a factor as high as 2.5 (**Figure 4.10.** *panel B*). This suggests that, along with an effect on fluorescence intensity, TFE truly contributes to mutated dC10 helix stabilization, which increases its reactivity.

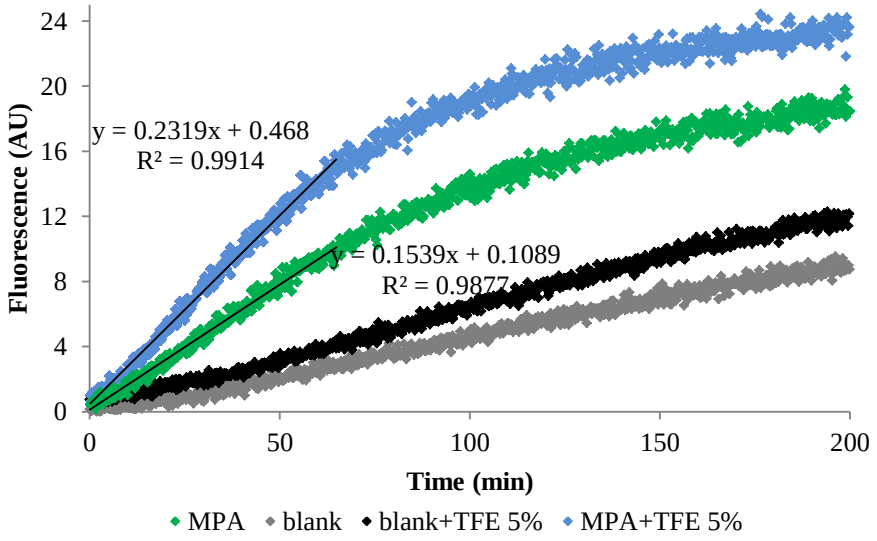


Figure 4.11. Reaction between dansyl-dM10 1 and 2-mercaptopropionic acid. Reaction between **dansyl-dM10 1** and 10 equivalents of 2-mercaptopropionic acid was carried out in 50 mM HEPES pH 7.4, 1 mM TCEP, and with 5% (*v/v*) (*blue line*) or without (*green line*) TFE, as indicated. The background reaction of dansyl-dM10 was carried out in the presence (*black line*) or absence (*grey line*) of 5% (*v/v*) TFE. Fluorescence increase was followed at 530 nm upon excitation at 330 nm and initial slopes of progression curves were determined by linear regression, and are indicated on the figure with their corresponding R^2.

In the ideal case, by bringing all dC10 mutants to the highest point of their helical content using TFE, a clear rate enhancement should be observable, caused by the presence of histidine residues that should stabilize cysteine thiolates (i.e., all rate constants of dC10 histidine mutants in presence of TFE should be higher than the rate constant of parent dC10 in the presence of TFE in **Figure 4.10.** *panel B*). However, the stabilization effect achieved here with TFE is certainly not maximal, due to the precipitation that occurred when concentrations of TFE higher than 5% (*v/v*) were used. In consequence, only a partial increase of the stabilization of dC10 helices was probably obtained. Nevertheless, after eliminating other possibilities for rationalizing the increased reactivity of dC10 mutants in

the presence of TFE, we can draw some conclusions regarding the effect of each mutation on dC10 stabilization and reactivity.

In the case of single mutants S2H and A7H, the second order rate constants increased by a factor of 2.0-2.5 in presence of TFE to reach approximately 2500 $M^{-1}min^{-1}$, a constant similar to parent dC10 in the presence of 5% (*v/v*) TFE. Positions S2 and A7 are one helix turn apart from and adjacent to C6, respectively; they are, theoretically, in close proximity of the reactive thiolate that attacks the maleimide double bond. However, the S2H and A7H mutants react only as well as dC10 in the presence of TFE, and a little less rapidly than dC10 in the absence of TFE, suggesting that mutated histidines at positions 2 and 7 do not represent any marked advantage in the stabilization of the thiolate form of cysteine 6. Furthermore, since S2 is part of the *N*-terminal cap of the dC10 helix that forms a stabilizing hydrogen bond [78], its replacement by another residue may be more destabilizing for the helix than any charge-charge stabilization the new residue may bring, even if the latter is a convenient distance away from cysteine 6 for its stabilization. Similarly, A7 is adjacent to cysteine 6; however, the orientations of the side chains of the two residues are not aligned, as they would be in the case of a residue separated by one turn of the helix, and it is possible that the side chains are pointing in different directions, and not able to interact, which would explain the maximal reactivity of A7H that is comparable to dC10, both in the presence of TFE.

A significant reactivity enhancement for the S2H-A17H double mutant was observed, that is already, in the absence of TFE, twice as reactive as other histidine mutants investigated here, and its rate constant increased by a factor of 1.5 in presence of TFE. We saw previously that the single mutant S2H does not present any advantage in reactivity; therefore it must be the mutation A17H that is crucial for the increased reaction rate of the double mutant S2H-A17H. Position 17 is located one helix turn away from cysteine 13, which may therefore be the cysteine that first attacks the dimaleimide double bond, as it is the only cysteine of dC10 in proximity of A17H. However, this remains a speculation and at this point, there are no data to support this statement.

In the next part of this chapter, the importance of residue 17 is investigated by side-directed mutagenesis where polar and protic residues are introduced in position 17 instead of the

original alanine in dC10 in order to make cysteine 13 thiolate more nucleophilic, and thus more apt to react with a maleimide electrophilic double bond. Increasing this reactivity of dC10 derivatives has the potential to suppress any possible reaction of dimaleimide compounds with glutathione or other thiols present in a cellular environment.

4.4. First library of dC10 A17X single mutants (Library I)

4.4.1. Initial definition of our system

After discovering position 17 as a potential 'hotspot' for reactivity, we decided to expand the variety of residues at that position by creating an eight-member library, substituting A17 with a charged (D, E, H, K, R) or polar (N, Q, S) residue. For that purpose, DNA oligonucleotides were used, containing the degenerate codon VRN that encodes all above mentioned positive or polar residues, plus a glycine residue (see Experimental section on page 107). However, the glycine mutant was of little interest to us, since glycine is known to disrupt helices [87] and it is not able to interact with other residues by charge-charge stabilization. Desired clones were identified by sequencing and corresponding dC10 mutants purified individually.

Plasmids for MBP-dC10 point mutant expression were prepared as described in the Experimental section. MBP-dC10 variants were expressed and purified in a very high yield of 5-15 mg of pure protein from 250-350 mL of expression media.

The reactivity of each single mutant of MBP-dC10 was evaluated through their reaction with **dansyl-dM10 1** ([74] and **Figure 4.9.**). Equimolar concentrations of 50 μM of fluorogen and MBP-dC10 variant were used in pH 7.5 buffer HEPES 50 mM, supplemented with 1 mM of TCEP reducing agent to reduce disulfide bonds that could form during protein purification or storage. Incubation of fluorogen alone in buffer at 28°C shows that reaction between TCEP and dansyl-dM10 can occur (**Figure 4.12.**), but it does so only after a longer period of time than fluorogen addition on MBP-dC10. Hence, for the measurement of rate constants, the rate of background hydrolysis was subtracted, and all reactions were

performed in buffer supplemented with 1 mM TCEP, to ensure that all thiols were reduced and reactive.

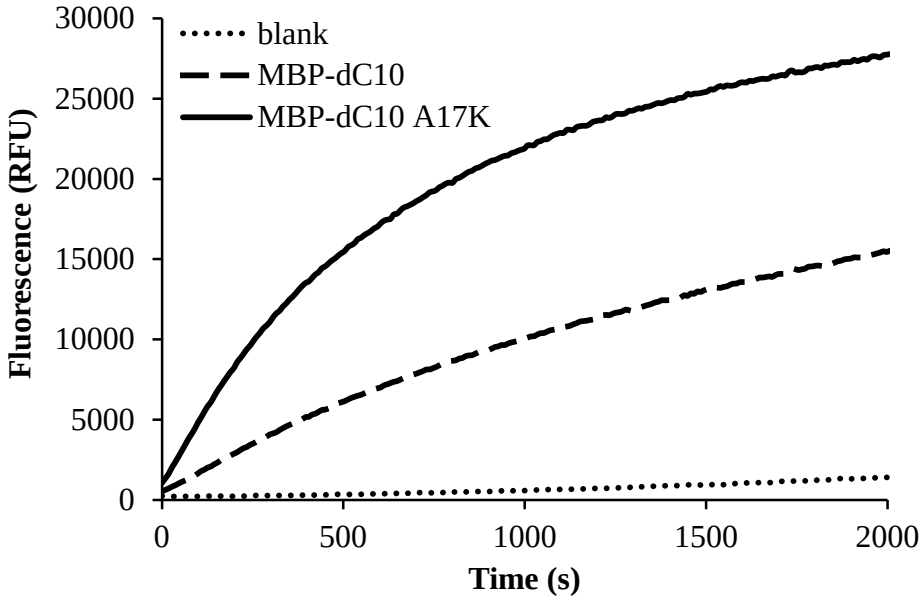


Figure 4.12. Fluorogenic addition reaction kinetics of MBP-dC10 and dM10-dansyl 1. Comparison of background hydrolysis of **dM10-dansyl 1** (*dotted line*) versus reactivity with MBP-dC10 (*dashed line*) and a point mutant MBP-dC10 A17K (*black line*).

Progress curves of the fluorogenic addition reaction were acquired with a plate reader that allowed us to study a large number of samples at the same time; however the reaction temperature needs to be elevated from 20°C used until now to 28°C because the used plate reader does not allow sample cooling. Since the dansyl fluorophore is sensitive to its environment, the reaction of each MBP-dC10 variant led to an end point of slightly different fluorescence intensity. Simple second order kinetic model A+ B $\rightarrow$ C was used to determine the kinetic constant, where reactants A and B were used in equimolar concentrations. The product C is the only fluorescent species in the milieu and its appearance can be directly followed by increase of fluorescence. Use of this model implies that the second thiol addition is extremely fast due to its intramolecular character, that and the second order kinetic constant k_2 depends exclusively on the first thiol addition. This reasonable

approximation allowed us to evaluate the overall dC10 reactivity modulated by the presence of stabilizing residues in the proximity of cysteines C6 and C13.

4.4.2. Effect of different residues on position A17 on dC10 reactivity

The second order rate constants determined for single mutants of MBP-dC10 (**Figure 4.13.**) show that negatively charged (D and E) or polar (N, Q and S) residues fail to significantly improve the reactivity of dC10 thiols. In the case of D and E this is probably due to the low pK_a of their side chain carboxylic acids, resulting in the formation of carboxylates which disfavours the formation of an adjacent thiolate. Conversely, the positively charged residues (H, K and R) show a 1.5- to 2.5-fold enhancement in the second order rate constant. The reactive species in a thiol – maleimide addition reaction is most likely a thiolate [138], normally present in a very small proportion at neutral pH due to higher pK_a of the cysteine side chain (determined to be 8.55 in an alanine pentapeptide [86], or in a range of 7.4 - 9.1 in various peptides [139], or as high as 9.5 for cysteine-like compounds [140, 141]). Therefore, adjacent positively charged residues may stabilize the thiolate form through electrostatic interaction, decreasing the pK_a of the dC10 cysteine side chain and increasing the proportion of reactive thiolate.

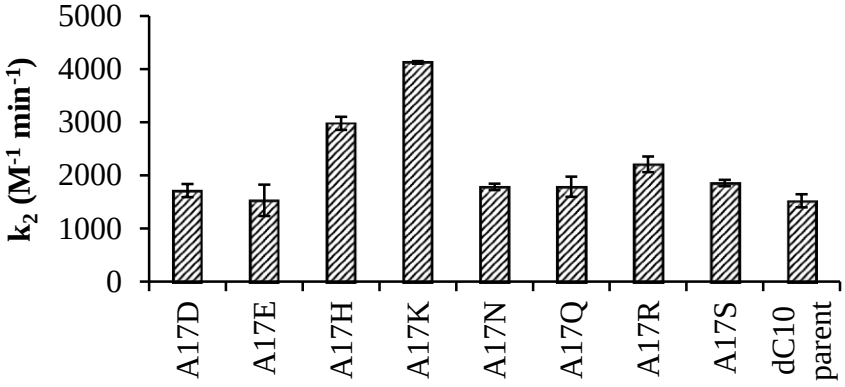


Figure 4.13. Second order rate constants for MBP-dC10 library I. All reactions were performed at equimolar concentration of 50 µM of **dM10-dansyl 1** and MBP-dC10 or variant, in duplicate, on a Synergy H4 plate reader at 28 °C. Fluorescence increase was observed at 515 nm upon excitation at 330 nm.

4.4.3. Effect of A17 point mutations on dC10 helical propensity

A second contribution to be considered in the dC10 – dimaleimide reaction is the helical conformation of the dC10 peptide. It was shown previously [71] that thiols presented in a helical conformation react faster with dimaleimides than simple free thiols. Indeed, the dimensions and geometry of our dimaleimide fluorogens are designed to complement the thiols presented in our di-cysteine helical motif. Thus, we can presume that the propensity of the peptide to adopt a helical conformation is another condition required for the optimal reactivity of both thiols of the dC10 peptide. The ability of each residue to favour a helical conformation varies substantially, and was investigated by Pace *et al.* [87]. While alanine residues are the most prone to form a stable helix, residues such as glycine or proline strongly disfavour a helical conformation. Therefore, replacing an alanine with another residue will inevitably destabilise the dC10 helix to a greater or lesser extent. A scale of propensity for all investigated residues is presented in **Figure 4.14.** and shows that the least reactive mutants are not only unfavourable for thiolate stabilization (D, E, N, Q, S), but also the least prone to stabilise a helix. Furthermore, it would appear that histidine, thought to increase the reactivity of the peptide sequence by favouring formation of thiolate, may also disfavour the reactivity by perturbing the formation of a reactive helix conformation. For example, our results (**Figure 4.13.**) show that while the single histidine mutant A17H reacts faster than the parent dC10 peptide tag, it does not react as fast as the A17K mutant, possibly due to destabilization of the helical conformation required for optimal reactivity with dimaleimide fluorogen.

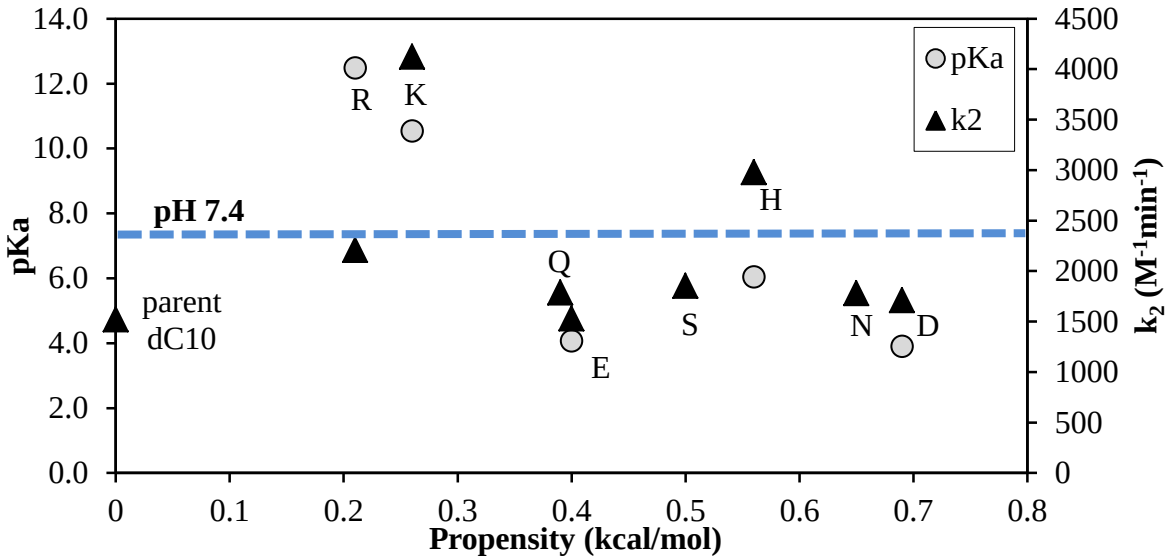


Figure 4.14. Propensity and pKa of individual residues used in MBP-dC10 single mutant library I. Propensity values represent the destabilization energy of a helix of a given residue relative to alanine, as reported in **[87]**. For the single mutant library A17X, the pKa of each residue X is shown (*gray circles*, left vertical axis) along with the measured second order rate constant (*black triangles*, right vertical axis). The dashed line represents physiological pH of 7.4 to emphasize the expected protonation state of each residue.

The results of our initial screen (**Figure 4.12., Figure 4.13.**), show that the best dC10 single mutant is A17K, most likely due to its low helix-destabilizing character and more importantly, to the positive charge on its side chain that may electrostatically decrease the pKa of the adjacent cysteine residue.

4.4.4. Labelling in presence of a secondary structure stabilizing agent

To overcome the unequal helical propensity of some residues, we investigated possible ways how to improve helix content in peptides by using additives. Similarly to the case of single and double histidine mutants (section *4.3.4*,**Figure 4.10.**), we used 2,2,2-trifluoroethanol [142, 143, 137] for this purpose. When used in previously determined concentration of 5% (*v/v*), TFE causes a higher reactivity of all dC10 single mutants from library I by a factor of 1.3-2, but the order of reactivity with respect to the parent dC10 peptide stays roughly the same (**Figure 4.15.**).

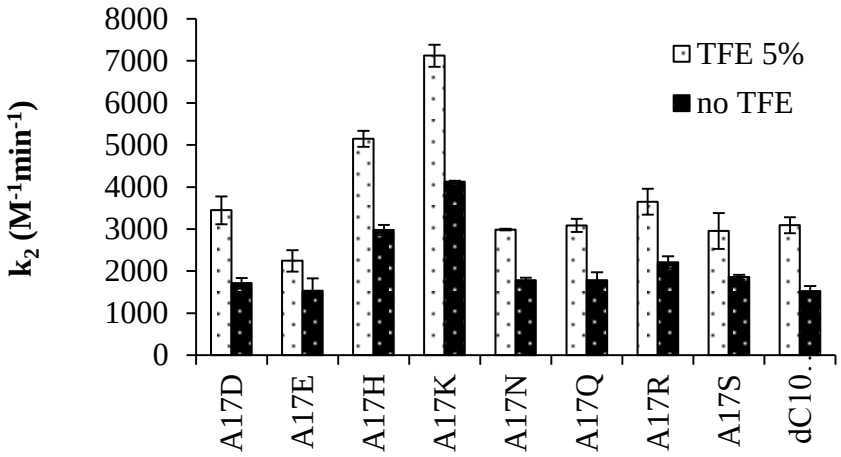


Figure 4.15. Kinetics of addition of MBP-dC10 single mutants on dM10-dansyl 1 in presence (*white bars*) or absence (*black bars*) of TFE (*v/v*) 5%. Reactions were carried out at 28°C in equimolar concentrations of 50 µM of fluorogen and protein, in 50 mM HEPES pH 7.5, 1 mM TCEP. Rate constants were determined as detailed in section *4.4.1.*

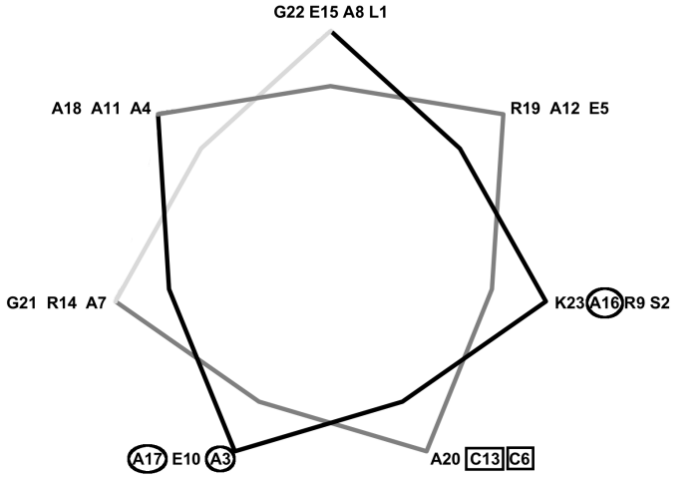


Figure 4.16. Section view of dC10 helix. First loop involves residues L1, S2, A3 and A4 (*black line*), second loop contains residues E5, C6, A7 and A8 (*dark grey line*), etc. Cysteines C6 and C13 are highlighted (*black squares*) as well as mutation sites A3, A16 and A17 (*black circles*).

4.5. Second library of dC10 A16-A17 double mutants (Library II) and third library of dC10 A3-A16-A17 triple mutants (Library III), new dC10* sequence

4.5.1. Double mutant mini-library

Based on the kinetic results obtained for our single mutant library dC10 A17X, mutant residues H, R and K, having side chains likely to be positively charged at neutral pH, were retained for further mutagenesis studies.

Figure 4.16. shows the relative proximity of residues in a helical conformation of dC10, from which two other potential candidates for mutagenesis in close proximity to cysteine 13 are apparent; alanine 16 and arginine 9 are both one turn away from cysteine 13 and could potentially have an effect on its ionisation state and helix stability. Arginine 9 is presumably already engaged in a salt bridge with glutamate 5 [76, 77], thereby stabilizing the helical conformation of dC10. When this arginine residue was mutated to a lysine in the mini-library R9K-A17X (where X is H, K or R), none of the double mutants dC10 R9K-A17X showed any further improvement in reactivity in comparison with their single A17X mutant parent sequences (**Figure 4.17.**). This result suggests that the mutation of residues engaged in salt bridge interactions in dC10 is unlikely to lead to more active dC10 variants. On the other hand, residue alanine 16 is an excellent candidate for mutagenesis because its only stabilizing contribution is its favourable helix propensity. Therefore, we replaced Ala16 with histidine, lysine or arginine in the mini-library A16X-A17K. Screening of this library (see **Figure 4.18.**) confirmed our previous findings with the single mutant library: a reactivity enhancement of up to eight-fold was observed for double mutants relative to the single mutant parent A17K. Double mutant A16H-A17K was not as reactive as double mutants A16K-A17K and A16R-A17K, which is once again consistent with the low helical propensity of histidine (**Figure 4.14.**).

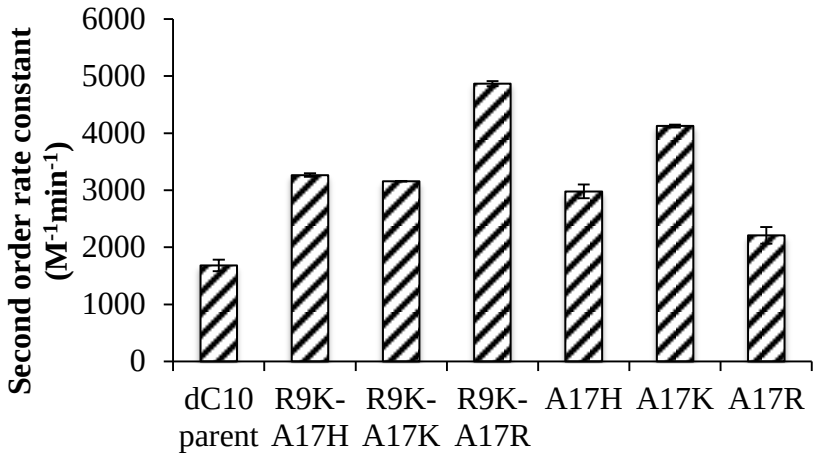


Figure 4.17. Second order rate constants for double mutants R9K-A17X compared to single mutants A17X. All reactions were performed at equimolar concentration of 50 µM of **dM10-dansyl 1** and MBP-dC10 or variant, in duplicate, on a Synergy H4 plate reader at 28 °C. Fluorescence increase was observed at 515 nm upon excitation at 330 nm.

4.5.2. Triple mutant mini-library

At this point in the study, the reactivity of the helical sequence was enhanced by mutations at positions 16 and 17, which are adjacent to cysteine 13. According to **Figure 4.16.**, the only residue in close proximity of cysteine 6 that would be likely to affect reactivity is alanine 3. A library of triple mutants was therefore created by preparing all possible combinations of histidine, lysine and arginine at positions 3 and 16, while retaining lysine at position 17. In total, nine triple mutants were isolated and fully characterised kinetically. As shown in **Figure 4.18.**, four triple mutants (A3K-A16K-A17K, A3K-A16R-A17K, A3R-A16K-A17K and A3R-A16R-A17K) exhibit a nearly 10-fold increase in reactivity towards a dimaleimide fluorogen, with mutant A3K-A16R-A17K reacting 10 times faster than the parent dC10 sequence. This rate enhancement represents a significant breakthrough for our labelling technology where ideally, a highly reactive peptide tag linked to a target protein will greatly favour specific labelling with minimal background reaction. We therefore retained this dC10 triple mutant, named dC10*, as a new pilot sequence for intracellular labelling.

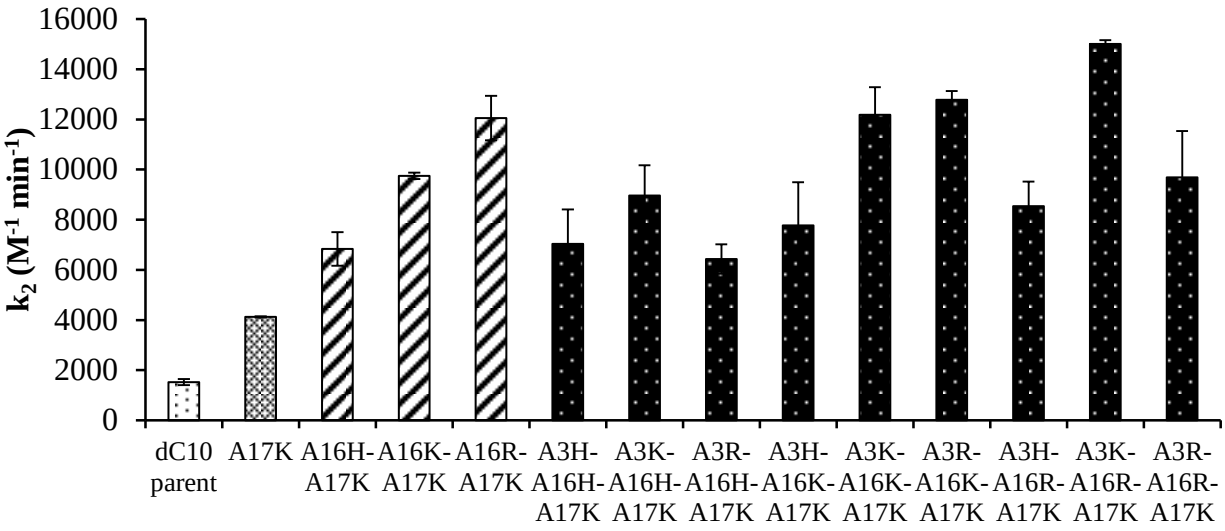


Figure 4.18. Second order kinetic constants determined for MBP-dC10 mutants. Best mutants of Library I A17K (*light gray*), Library II A16-A17 (*gray stripes*) and Library III A3-A16-A17 (*black*) are represented. Parent MBP-dC10 is represented in white. All data were done in duplicate, at least.

4.5.3. Role of residue 3 in dC10

It is generally assumed that when two covalent bonds are formed between two reactive molecules, the intermolecular reaction to form the first covalent bond is slower than the rapid intramolecular reaction to form the second covalent bond [144, 145]. So while it is easy to rationalize the enhanced reactivity of cysteine 13 due to the presence of two thiolate-stabilizing residues (at positions 16 and 17), it is not as obvious how improving the reactivity of cysteine 6 would further enhance the reactivity of the peptide (if the reaction of cysteine 6 were intramolecular). However, this assumes that in all cases cysteine 13 is the first to react with a maleimide moiety of a given dimaleimide fluorogen. Alternatively, enhancing the nucleophilicity of *both* cysteine residues will increase the effective concentration of reactive cysteine residues in the peptide sequence, thereby increasing the overall rate of reaction *regardless* of which cysteine residue reacts first. More specifically, a A3-A16-A17 triple mutant may react faster than its A16-A17 double mutant, because it contains a higher effective concentration of highly reactive cysteines. For example, as shown in **Figure 4.18.**,

the triple mutant A3K-A16R-A17K is slightly more reactive than its parent double mutant A16R-A17K, in support of this hypothesis.

4.6. pH-rate profile and helical propensity profile

Considering the relative reactivity of the members of all three mutant libraries, additional studies were undertaken to attempt to quantify the effect of positively charged residues on the ionisation of the cysteine residue thiol groups. The fluorogenic addition reactions of several selected mutant peptide sequences with a dansyl-dimaleimide fluorogen were studied kinetically over pH 7.50-9.00; higher pH values were practically inaccessible due to protein instability. As shown in **Figure 4.20.**, a linear relationship was observed between the logarithm of the measured second order rate constant and the pH of the milieu, confirming that in all cases, the basic form of the peptide is more reactive. From this plot, one can roughly group the mutant sequences into three groups: the triple mutants that do not contain histidine (**Figure 4.20.** *right*), the triple mutants containing histidine (**Figure 4.20.** *middle*) and the double mutants (**Figure 4.20.** *left*). From this plot it would appear from the horizontal displacement of the lines in the linear region (pH 7.5-8.50 - see below) that the pKa of the fastest mutant (A3K-A16R-A17K) is 0.70 units lower than that of the parent dC10 sequence. However, since we were unable to perform experiments at higher pH, we did not observe any plateau that would have allowed us to determine pKa values for the mutants. Therefore, it is also possible that the increased reactivity of the triple mutant variant is due to an increase in the asymptotic value, without any perturbation of the thiol pKa. For example, the lysine and arginine side chains of the reactive triple mutant may act in concert with the reactive cysteine residue, perhaps to protonate the maleimide group after the nucleophilic attack by the thiolate (**Figure 4.19.** *pathway 2*). However, since we have no additional evidence to support this hypothesis that proposes more complex bifunctional catalysis and the maleimide reprotonation suggested in pathway 2 is likely not a rate-limiting step, we favour the simpler explanation based on increased thiol acidity through local electrostatic effects, and by forming an ammonium thiolate ion pair (**Figure 4.19.** *pathway 1*).

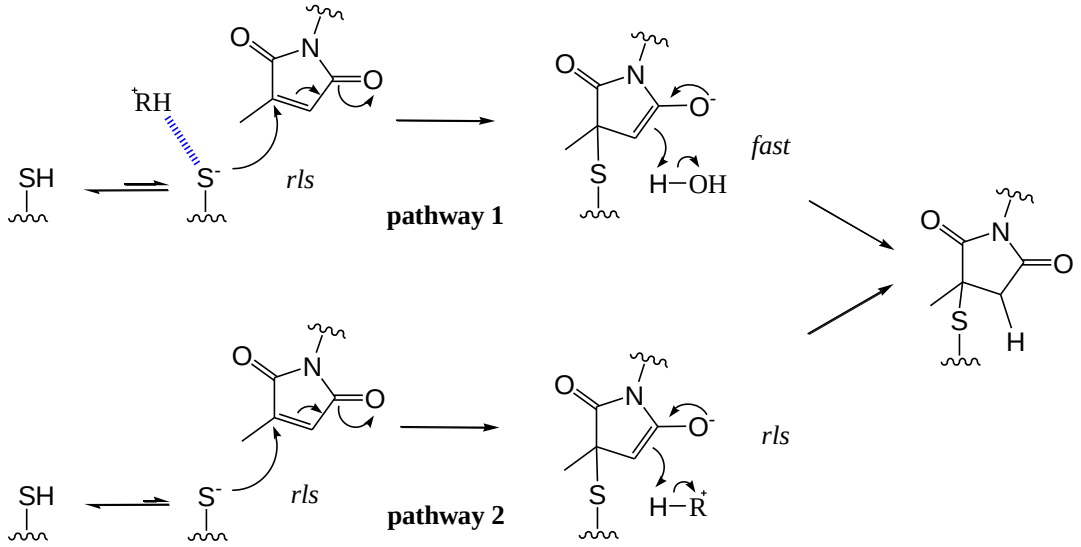


Figure 4.19. Proposed mechanisms for thiolate-maleimide addition reaction. Pathway 1 (*top*) shows the thiolate attack as a rate-limiting step (*rls*), while in pathway 2 (*bottom*) the rate-limiting step is composed of two separate steps: thiolate attack on the maleimide double bond, and a subsequent reprotonation by a positively charged residue RH^+. In pathway 1, the reactive thiolate form S^- is stabilized by forming a salt-bridge (*dashed blue line*) with a positively charged residue ^+RH.

Another pH effect that is important to discuss is the pH dependence of the helical propensity of each residue. We used the online AGADIR algorithm (*http://agadir.crg.es/*) to predict the helical content of the entire dC10 (or variant) sequences at different pH values [80, 81, 79]. As shown in **Figure 4.21.**, the logarithm of the percent global helical content of each dC10 sequence increases with pH, in an almost linear fashion over the pH range 7.50 – 9.00. This predicted increase in helicity may account for a small fraction of the observed increase in reactivity, but the near-unity slopes of the pH-rate profiles shown in **Figure 4.20.** suggest that most of the increased reactivity is due to a single ionisation event.

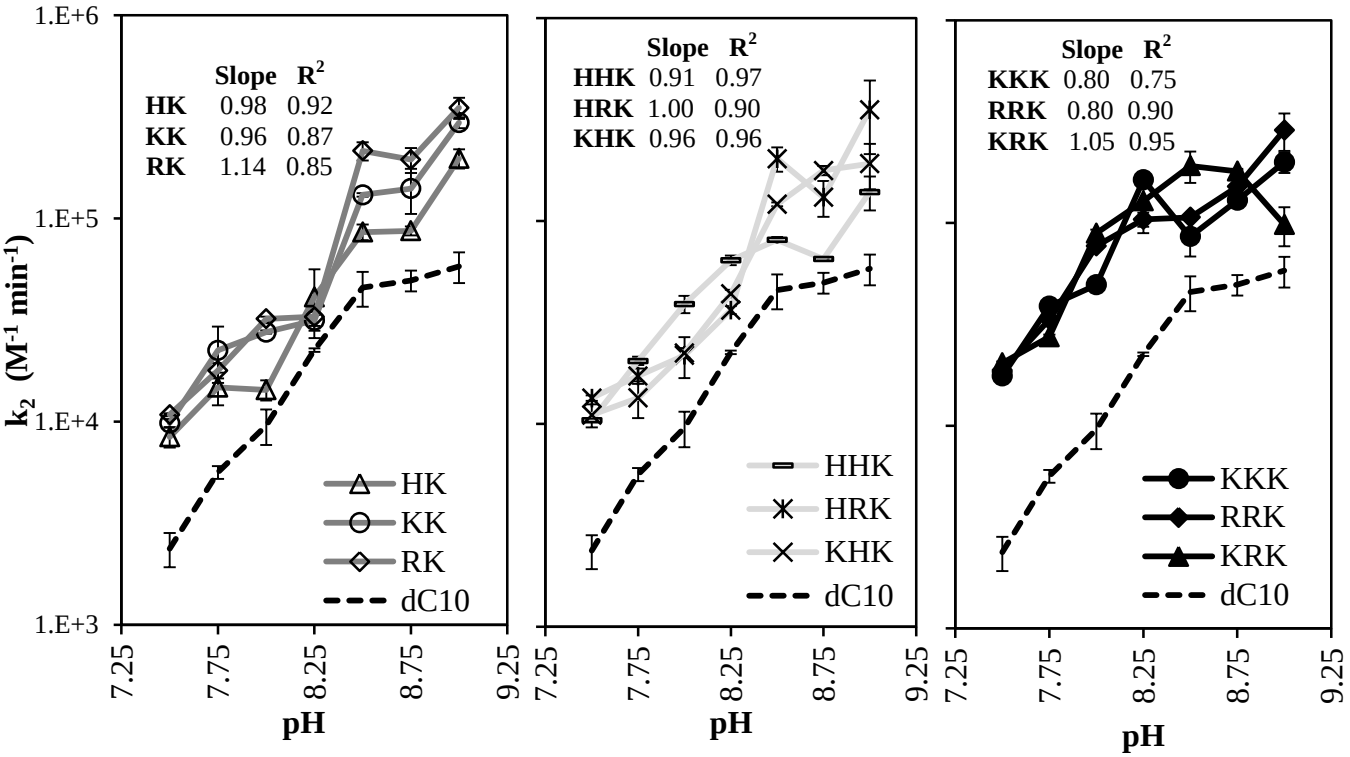


Figure 4.20. Rate-pH diagram for several MBP-dC10 mutants. A pH range between 7.50 and 9.00 was explored to determine the dependence of second order rate constant on pH. Only several mutants from libraries II and III were chosen for pH study: *left*: double mutants (*dark grey line*); *middle*: triple mutants containing histidine and lysine or arginine (*light grey line*); *right*: triple mutants containing lysine and arginine (*black line*), and parent dC10 (*dashed line*). Y-axis is represented in logarithmic scale. Slopes for pH 7.50 – 9.00 are represented in each graph with their corresponding R^2.

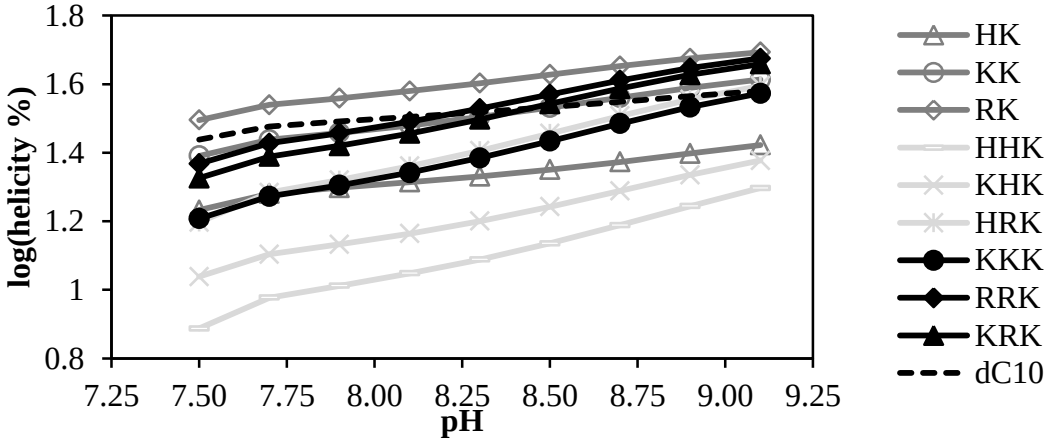


Figure 4.21. Helicity dependence on pH. AGADIR software was used to predict helicity of several MBP-dC10 mutants (identical to **Figure 4.20.**) with increasing pH. The prediction conditions were 301 K and ionic strength 0.1 M.

4.7. Kinetics with *in cellulo* relevant fluorogen dM10-coumarin 9

To demonstrate the importance of the rate enhancement of dC10 for *in cellulo* application, we next evaluated the *in vitro* kinetic behaviour of several dC10 triple mutants with relevant dimaleimide fluorogens. The second order kinetic constants of these reactions are shown in **Figure 4.22.** As a first and best fluorogen for *in cellulo* application available in the Keillor lab at that time we tried **dM10-coumarin 9** (**Figure 4.23.** and [75]). Although **dM10-coumarin 9** is much less reactive than **dM10-dansyl 1** (**Figure 4.22.**), similar relative reactivities were observed *in vitro* among the dC10 triple mutants. Namely, the second order rate constant of the most reactive dC10 mutant A3K-A16R-A17K is 9 times higher than that of the parent dC10 sequence. This increased reactivity is especially important for envisioned intracellular application, in order to compensate for the attenuated reactivity of the fluorogenic labelling agent, which has been shown [73, 75] to increase selectivity.

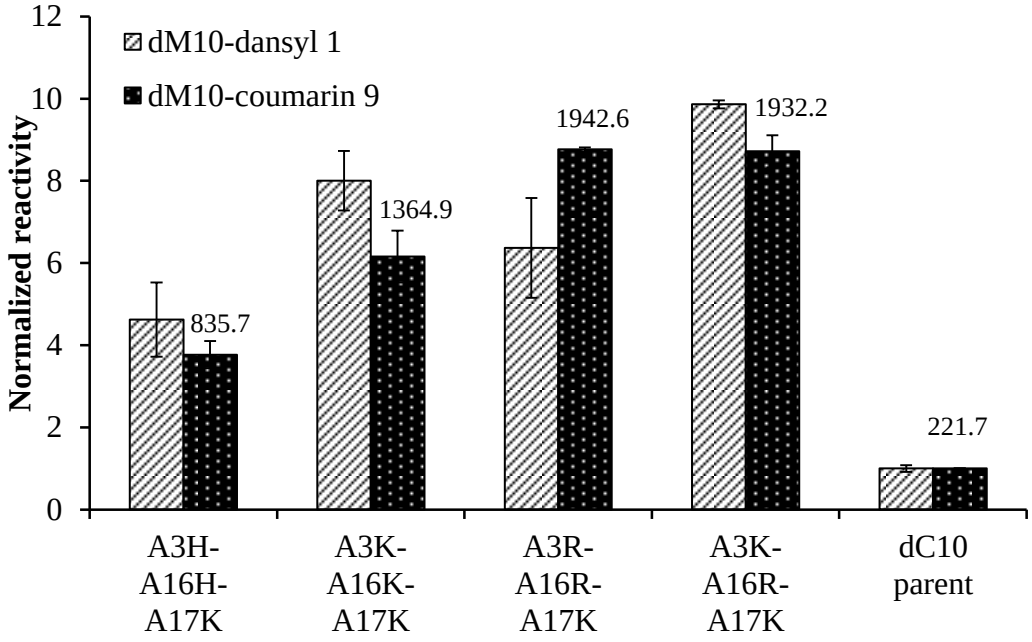
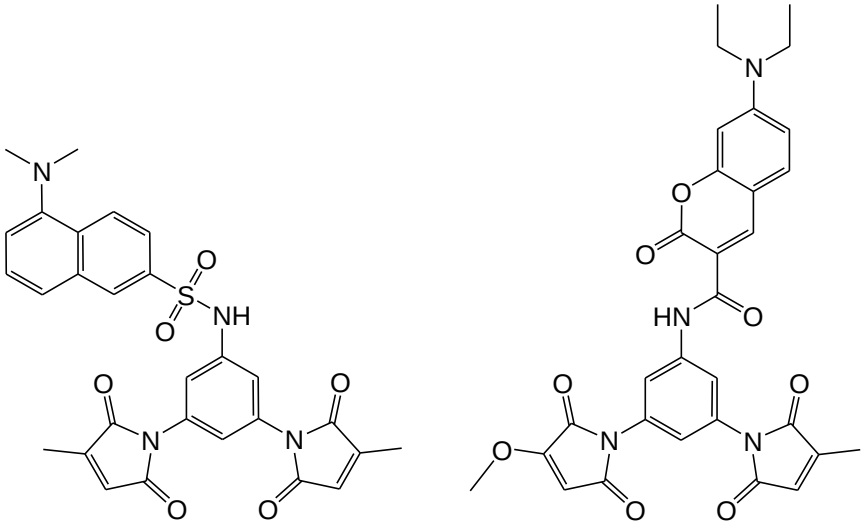


Figure 4.22. Comparision of reactivity of dM10-dansyl 1 and dM10-coumarin 9. Second order rate constants ratios are shown for selected triple mutants of MBP-dC10 and **dM10-dansyl 1** (*grey*), and **dM10-coumarin 9** (*black*). Ratios were obtained by normalizing all rate constants to rate constant of parent MBP-dC10. Absolute values of second order rate constants in $M^{-1}min^{-1}$ are shown for **dM10-coumarin 9**. Results were obtained at least in duplicate. For values of rate constants obtained with **dM10-dansyl 1**, see **Figure 4.18**.



dM10-dansyl 1 **dM10-coumarin 9**

Figure 4.23. Structures of fluorogens used for characterization of dC10 mutants. On the *left* is represented **dM10-dansyl 1** used for *in vitro* kinetics and on the *right* **dM10-coumarin 9** used for *in vitro* kinetics and *in cellulo* labelling.

4.8. Mammalian protein labelling with dC10* in HEK293 cells

4.8.1. Labelling of a cell-surface expressed protein with dC10 - EGFR*

As a first and most immediate application of the newly discovered dC10* sequence, we chose to label a protein expressed on a cell surface. Previously, the Epidermal Growth Factor Receptor (EGFR) bearing a dC10 tag on its *N*-terminus was successfully used by our collaborators for FlARe labelling [73] using **dM10-FITC** fluorogen. Here, **dM10-coumarin 9** was used as it was the best cyan fluorogen available at the time for *in cellulo* applications. We decided to repeat this labelling with **dM10-coumarin 9** and more importantly, to assess how well the new dC10* tag, an order of magnitude more reactive *in vitro*, performs in an *in cellulo* experiment in comparison with dC10.

A previously prepared expression plasmid pcDNA-dC10-EGFR [73] was used, that expresses EGFR with a cmyc-tag, a signal peptide that is responsible for the surface localization of EGFR, and dC10 on its *N*-terminus. This plasmid was altered to express dC10*-EGFR, as detailed in the Experimental section (page 108). **dM10-coumarin 9** fluorogen was used for this live-cell labelling in different concentrations; however, despite observing a time-dependent increasing green fluorescence signal, we were unable to observe a distinct labelling of either dC10-EGFR or dC10*-EGFR (see **Figure 4.24.** for the example of dC10*-EGFR).

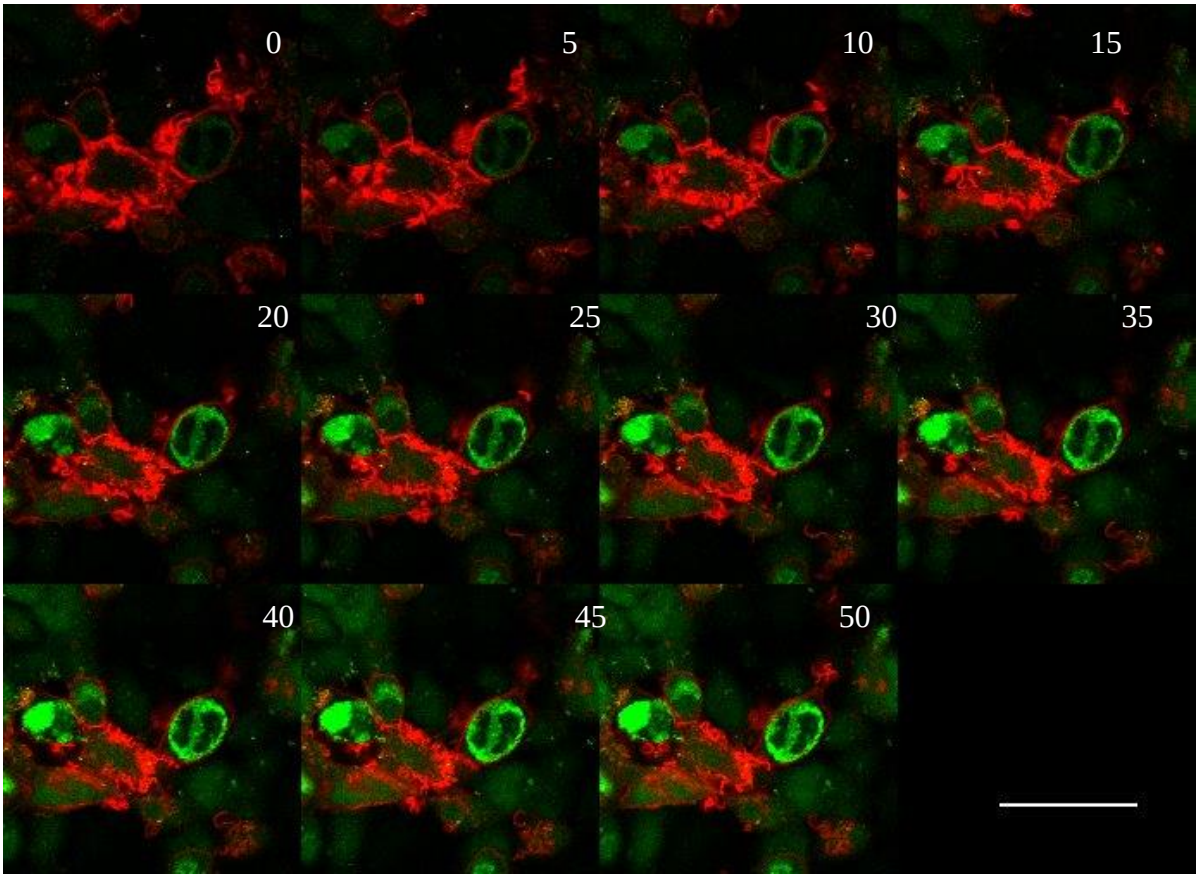


Figure 4.24. Labelling of dC10*-EGFR by dM10-coumarin 9 in living HEK293 cells. 25 µM concentration of fluorogen was used to label dC10*-EGFR (green), labelled beforehand by 2 µg/mL EGF-rhodamine (red). Time-course images were taken by Nikon confocal microscope every 5 minutes from t = 0 min to 50 min, as indicated in top left corner of each image. White scale bar represents 50 µm. Objective used was Fluor 40x. Green fluorescence of **dM10-coumarin 9** was detected with the following parameters: Ex. 488 nm, Em. 525 nm; red fluorescence of EGF-rhodamine was detected with Ex. 561 nm, Em. 595 nm.

To confirm if EGFR variants are expressed, we performed a control experiment where we used a rhodamine-conjugated Epidermal Growth Factor (rhodamine-EGF) that binds specifically to its receptor and can be detected via the red fluorescence of rhodamine. Both dC10-EGFR and dC10*-EGFR were expressed and localized on the cell surface, as confirmed by observing intense red fluorescence (**Figure 4.24**., **Figure 4.25.**) and almost no red fluorescence for mock cells transfected with an empty pcDNA3.1(+) vector. Unfortunately, we were not able to see a colocalization of this red fluorescence and the green fluorescence of **dM10-coumarin 9** that would have labelled dC10-EGFR or dC10*-EGFR (**Figure 4.24.**). More than that, in some cases, the fluorescence signal from dC10*-EGFR labelling by a fluorogen was exclusive from the red fluorescence from direct EGF-rhodamine labelling. Upon binding of its ligand, EGFR undergoes a conformational change and is eventually internalized [146]. It is noteworthy that general increase in green fluorescence, similar to the one shown in **Figure 4.24.**, was obtained when EGFR-dC10-expressing cells were labelled with **dM10-coumarin 9** only, without addition of EGF-rhodamine ligand that would induce a conformational change of EGFR, and potentially hinder the labelling by making the *N*-terminal dC10 tag inaccessible.

We can only speculate that the lack of labelling may have been because of a lower availability of this fluorogen that is more prone to enter the cell because of its more hydrophobic character, as suggested by a slightly increasing green fluorescence signal from inside of the cells. We presume that a fluorogen that bears carboxylic acid groups (such as **dM10-FITC** that was used previously in [73]) that would make the molecule adopt a negative charge, repulsed by the phosphate groups on a cell surface, would be more likely to stay outside of a cell and label dC10-EGFR on a cell surface. However, such a fluorogen is not currently available in the Keillor group.

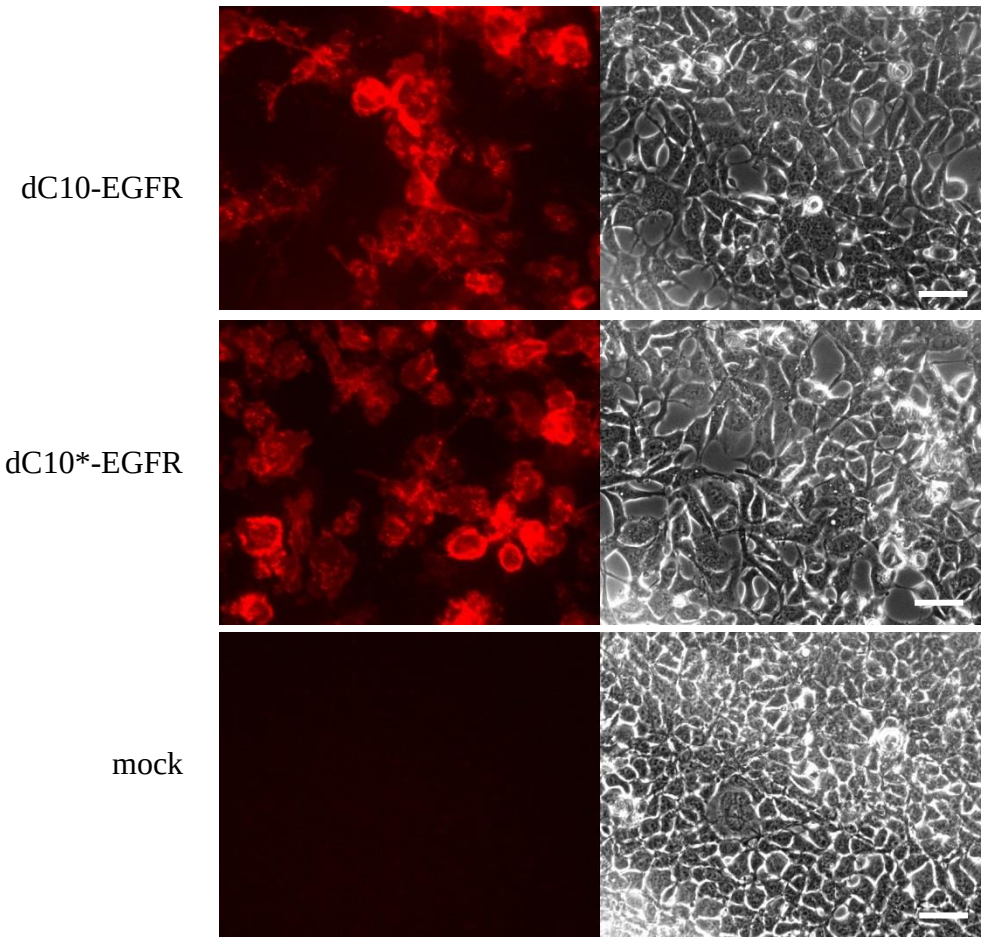


Figure 4.25. Expression control of dC10-EGFR and dC10*-EGFR using EGF-rhodamine. Cells were incubated with 2 µg/mL of EGF-rhodamine for 10 minutes at 37°C and excess of ligand was washed by replacing the media, after which cells were imaged using a Zeiss inverted epifluorescence microscope. Acquisition time: red 500 ms, bright field 50 ms. White bar represents 50 µm.

*4.8.2. Labelling of a protein localized in cell nuclei, using dC10 and dC10**

An important advancement in the fluorogen toolkit was achieved when a dimethoxy-substituted fluorogen **dM10-coumarin 20** (**Figure 4.27.**) was synthesized, as *in vitro* it had no reactivity with GSH over extended periods of time [75]. This new scaffold also caused it

to be the least reactive fluorogen with dC10, where the *in vitro* second order rate constant with MBP-dC10 was determined to be 60 M^{-1} min^{-1} (**Figure 4.27.**), under standard reaction conditions (see Experimental section, page 111). As a comparison to appreciate the slow rate of **dM10-coumarin 20**, the *in vitro* second order rate constants of methyl/methoxy-substituted **dM10-coumarin 9** (**Figure 4.22.**), dimethyl-substituted dM10-EDA-dansyl and unsubstituted dM10-EDA-dansyl were 221 M^{-1} min^{-1}, 190 M^{-1} s^{-1} and 11130 M^{-1} s^{-1}, in their respective experimental conditions [73]. We chose as well a new test protein, histone-2B (H2B), that is localized in the nucleus of a cell [147] and as a result, its labelling and the potential resistance of the fluorogen to glutathione in cytoplasm can be evaluated by simple localisation of observed **dM10-coumarin 20** fluorescence. H2B-dC10 and H2B-dC10* expression plasmid preparation was done by Dr. Christopher M. Clouthier and is detailed in [75].

HEK293 cells were transfected using Lipofectamine with plasmids coding for H2B-dC10 and H2B-dC10* expression and with the empty plasmid pcDNA3.1(+) as a negative control. The proteins were allowed to express for 24-48 hours and were labelled for 10 minutes with 10 µM of **dM10-coumarin 20**. **Figure 4.26.** shows a bright fluorescence of cell nuclei in cells expressing both H2B-dC10 and H2B-dC10*, resulting from **dM10-coumarin 20** labelling, suggesting that only H2B expressing cells were efficiently labelled. Some cells present only a background fluorescence in both nucleus and cytoplasm which would suggest that they were not efficiently transfected, as expected. Control cells that were transfected with an empty vector pcDNA3.1(+) present a negligibly weak fluorescence that is uniform in all cells.

Unfortunately, we do not notice a striking difference between labelling of H2B-dC10 and H2B-dC10*, as we would expect from our *in vitro* tests on **dM10-dansyl 1** and **dM10-coumarin 9**. It may be hard to see a difference in kinetics of dC10 and dC10* using only one time point as we are doing here but our *in vitro* tests suggest as well that the final fluorescence observed after addition of dC10* is significantly superior to the fluorescence intensity of labelled dC10, probably due to the sensitivity of the fluorophore to the change in its environment (data not shown for dC10* but for dC10-A17K in **Figure 4.12.**).

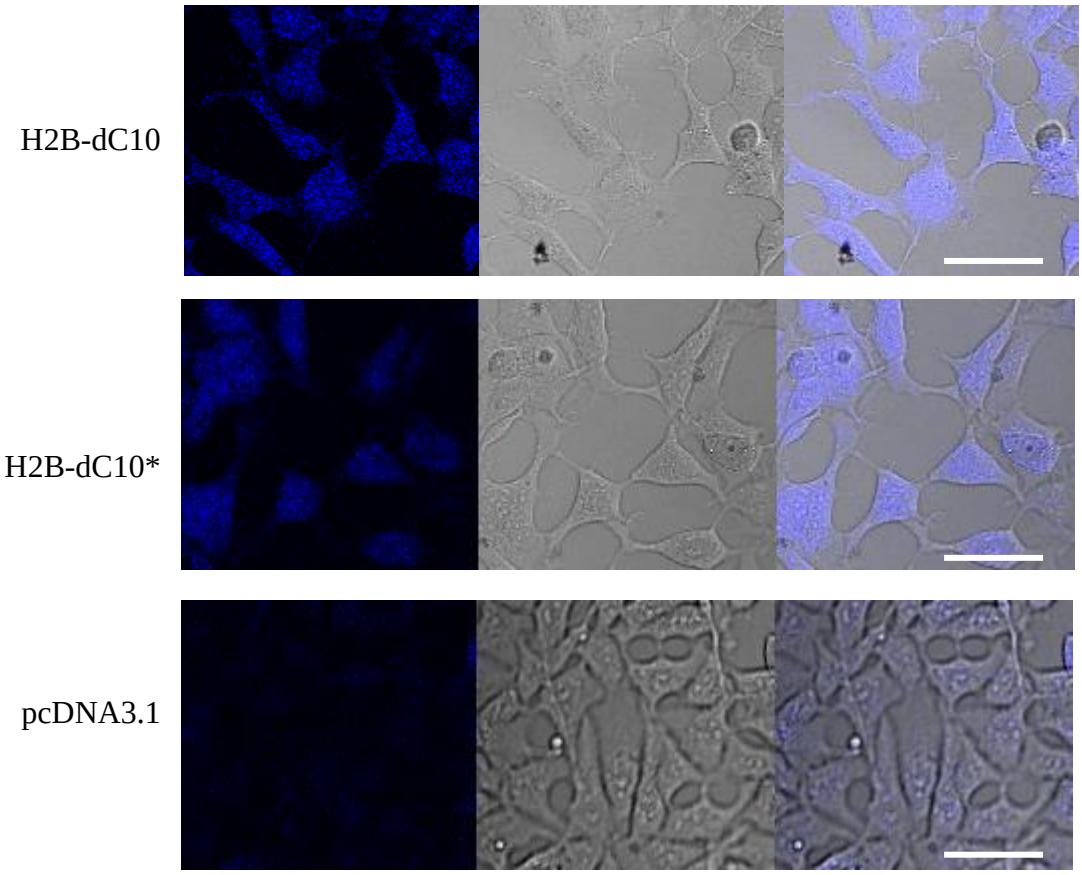
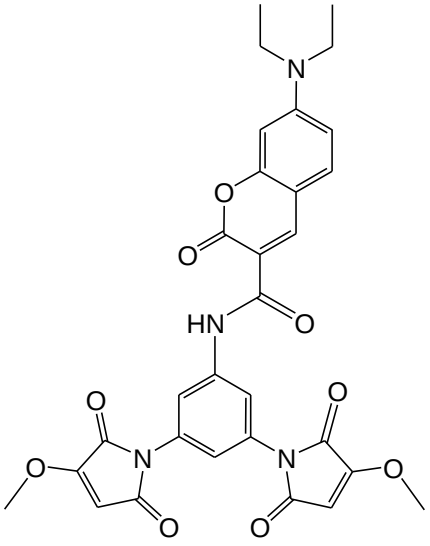


Figure 4.26. H2B-dC10 and H2B-dC10* expressed in HEK293 cells and labelled with 10 µM of dM10-coumarin 20. Cells transfected with plasmids coding for H2B-dC10, H2B-dC10* and with an empty vector pcDNA3.1(+) were labelled with **dM10-coumarin 20** for 10 minutes and imaged using Nikon confocal microscope. Scale bar represents 50 µm.



dM10-coumarin 20

Figure 4.27. dM10-coumarin 20 used for *in cellulo* labelling.

From our *in vitro* results with **dM10-dansyl 1** and **dM10-coumarin 9** we have expected that the labelling of dC10 and dC10* with other coumarin fluorogens would show the same kinetic advantage of dC10* compared to dC10. However, when **dM10-coumarin 20** was tested *in vitro*, surprisingly, it showed only a 2-fold rate enhancement with dC10* (**Figure 4.28.**). The reason for this much more modest rate enhancement remains unexplained and rather problematic for any kind of *in cellulo* demonstration of the kinetic advantage of our new dC10* peptide, considering also the fact that the labelling reaction rate inside a cell is limited by the rate of internalization of fluorogen molecules through the membrane. However, if the internalization of the fluorogen is truly the limiting step, the previously demonstrated advantage of selectivity of **dM10-coumarin 20**, combined with the rate enhancement of dC10*, can still provide a supplemental degree of *selectivity* once the fluorogen penetrates in the cell.

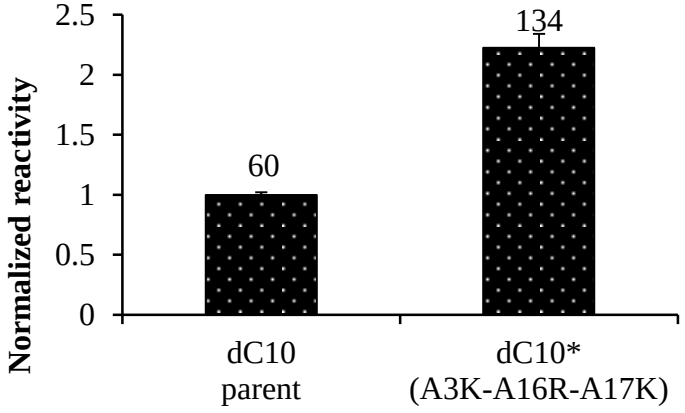


Figure 4.28. *In vitro* reactivity of dM10-coumarin 20 with dC10 and dC10*. Second order rate constants ratios are shown for **dM10-coumarin 20**. Ratios were obtained by normalizing MBP-dC10* rate constants to rate constant of parent MBP-dC10. Absolute values of second order rate constants in $M^{-1}min^{-1}$ are shown for dC10 and dC10*. Results were obtained at least in duplicate.

4.9. Conclusion

We have studied secondary structure of the dC10 tag in solution by NMR and were able to confirm that it adopts an α-helical secondary structure. This fact was essential for us to be able to design a rational approach for dC10 sequence evolution in order to obtain a more reactive peptide for protein labelling.

After evolving dC10, we obtained a substantial improvement of a fluorogenic dimaleimide-based labelling technique. According to our *in vitro* kinetic characterization, this new tag reacts an order of magnitude faster than its parent dC10 tag and promises a much faster and, more importantly, more selective labelling of a POI inside a cell where a fluorogenic molecule is exposed to a large number of potentially reactive thiols. This allows us to explore and develop more diverse, but less reactive, fluorogens that would otherwise exhibit prohibitively low reactivity, unless used with new generation dC10. We believe also that presence of a higher number of charged residues on new dC10 may help the overall solubility of the peptide tag.

For some of the most reactive variant sequences prepared over the course of this work, pH-rate studies were also performed, in an attempt to understand the mechanism of rate enhancement.

Finally, we demonstrated the utility of our best tag sequence in the labelling of a protein expressed in the nucleus, and we attempted the labelling of a new dC10* tag on the surface of live cells.

Lastly, the method used here could be generalized in protein or peptide engineering: we started by using a combinatorial approach to explore a larger ensemble of mutants to draw first conclusions that served as a basis for a more rational and focused second and third step of new dC10 design. We believe that despite being a based on rather simple reasoning on amino acid properties, our rational design is a valid example of focused and efficient peptide engineering.

4.10. Perspectives and other work

It would have been beneficial to obtain a quantification of *in cellulo* labelling of a protein attached to dC10 and dC10*, using, for example, flow cytometry. For instance, we could use an intrinsically fluorescent protein, such as mNeptune (see page 57), attached to a dC10* tag, and label this test protein intracellularly with **dM10-coumarin 20**. Transfected cells should show red fluorescence of the over-expressed FP, allowing to identify the presence of mNeptune-dC10*, and as well cyan fluorescence due to the labelling of dC10* by the fluorogen. We could then determine precisely the ratio of cells expressing mNeptune-dC10* that have been labelled with the fluorogen and compare this number to the ratio obtained with standard mNeptune-dC10. Even with a fluorogen that is only twice as reactive with dC10* tag than with dC10, an advantage in the effectiveness of intracellular labelling using dC10* tag should be seen with this quantitative method.

Lastly, obtaining a more reactive dC10 peptide sequence calls for its direct application in the field of fluorescent protein labelling on other systems, and for its protection as an intellectual property. Our group has carried out additional unpublished work to that end, including:

- Application for a patent to protect the best dC10 mutants, and several other dCx sequences that potentially react as well or better than parent dC10
- FlARe labelling in other cell lines, such as HeLa
- FlARe labelling of proteins with discrete intracellular compartmental localisation, such as ER, cytoskeleton, nucleus or cytoplasm
- Direct comparison of FlARe labelling technique with FlAsH labelling

It is certain that dC10* sequence is the future of the labelling project, and will be introduced to most upcoming applications of FlARe labelling, owing to its high reactivity, promising more selectivity, and potentially providing more solubility to the peptide tag.

4.11. Experimental section

4.11.1. Preparation of S30 extract from E.coli

A single colony of *E. coli* BL21 Star (DE3) strain (courtesy of Professor Christopher Easton, Australian National University) was used to inoculate 5 mL of LB media (without antibiotics). After overnight incubation at 37°C, whole volume of 5 mL was used to inoculate 500 mL of LB media that was grown overnight at 37°C. The next day, the whole volume of 500 mL was used to inoculate 9.5 L of sterilized and pre-warmed Z-media (43 mM KH_2PO_4, 173 mM K_2HPO_4, 1% (*w/v*) yeast extract) in a fermenter, to which was added 100 mL of 1 mg/mL thiamine, 112 mL of sterile 2 M glucose and 2.5 mL Antifoam. The fermenter was set to 395 rpm shaking speed at 37°C and 5 L/min (maximal) airflow, and bacterial growth was followed by OD at 600 nm. After 7 hours of growth and reaching an OD of 3.16, the culture was immediately cooled down on ice and the cells were pelleted by batches of 350 mL at 4300 g for 12 min, at 4°C. The pellets were collected and resuspended on ice and washed in 200 mL of S30α buffer (0.5 mM PMSF, 1 mM DTT, 7.2 mM β-mercaptoethanol in S30 buffer – see below). The cells were centrifuged at 10 000 g for 10 min at 4°C, the supernatant was discarded and the dry cells were centrifuged again at 10 000 g for 10 min. The pellet was snap-frozen on dry ice and kept overnight at -80 °C.

The next day, the cells were thawed on ice and resuspended in 200 mL of S30α buffer, then centrifuged for 12 min at 4300 g and 4°C. The supernatant was discarded and the cells were resuspended a second time in 200 mL of S30α buffer and centrifuged a second time for 12 min at 4300 g. The supernatant was discarded and dry cells were centrifuged a third time for 12 minutes at 4300 g. Finally, the cells were resuspended in 96 mL of S30α buffer (ratio of 1.3 mL of buffer / 1 g pellet) and divided into 15-20 mL fractions that were sonicated for 3 minutes with 1-s pulse. The lysed cells were centrifuged at 30 000 g for 1 h at 4 °C and the supernatant was transferred into two dialysis tubes (Spectrapor n°4 12-14 kDa MWCO) and dialysed against 1.5 L of S30β buffer (1 mM PMSF and 1 mM DTT in S30 buffer – see below) for one hour, at 4°C, three times. The dialysis bags were then transferred into 1 L of a 50% (*w/v*) solution of PEG8000 in S30 buffer (10 mM Tris acetate, 16 mM potassium acetate, 14 mM magnesium acetate, pH 8.25) at 4°C, under very gentle

shaking, and incubated until the volume was decreased by 50% (after 1.5 hours in this case). Then, the supernatant was dialysed against 1.5 L of S30δ solution (1 mM DTT in S30 buffer) for 15 min at 4°C, and finally, against 1 L of S30γ solution (1 mM DTT and 400 mM sodium chloride in S30 buffer) at 4°C overnight.

1 L of S30γ buffer was pre-heated in a water bath at 42 °C and the dialysis tubes containing S30 extract were placed in the pre-heated buffer for 45 min with gentle shaking, which caused a slight white precipitate to form. The extract was dialysed against 1 L of S30 buffer at 4 °C overnight.

The dialysed extract was centrifuged for 45 min at 30 000 g and 4 °C and the resulting supernatant was divided into 1 mL fractions and snap-frozen in liquid nitrogen bath, and kept at -80 °C.

4.11.2. Preparation of PpiB-dC10 expression plasmid

The expression plasmid for PpiB-dC10 was derived from pETMCSI-PpiB-CTG, kindly donated by Professor Christopher Easton (Australian National University). Primers His_6-PpiB_fw and PpiB-GS2-FXa_bw (**Table 4.1.**) were used to amplify the *ppib* gene: we introduced a hexahistidine tag and *NdeI* restriction site on the 5′ end, and a GSGS spacer coding sequence and a Factor Xa coding site on the 3′ end. The resulting amplicon was used as a template for a second amplification and insertion of dC10 coding sequence using primers H6-PpiB_fw and dC10_bw. The final PCR product contains His_6-PpiB-GS2-FXa-dC10 and was inserted in pETMCSI-PpiB-CTG between *NdeI* and *EcoRI* restriction sites.

4.11.3. Cell-free expression of unlabelled PpiB-dC10

Cell-free expression was usually carried out at 37°C for 6-24 hours using the Torizawa *et al.* protocol [126] with a commercial T7 RNA polymerase (Takara). At the end of the expression period, the solution was centrifuged at 30 000 g for 1 hour at 6 °C. The hexahistidine-tagged protein was purified by nickel-affinity column chromatography.

Table 4.1. Oligonucleotides used for PpiB-dC10 cloning

His_6-PpiB_fw	5′-GGAAATCCATATGCATCACCATCACCATCACGTTACTTTCCACACCAATC – 3′
PpiB-GS2-FXa_bw	5′-CCTTCCCTCGATCCCGAGGCTGCCGCTGCCCTCGCTAACGGTCACGCTT – 3′
dC10_bw	5′-GGAATTCCCTACTTTCCTCCAGCTCTAGCTGCAGCTTCTCTGCATGCAGCTTCTCTAGCAGCGCACTCAGCAGCGCTCAGCCTTCCCTCGATCCC - 3′

4.11.4. Expression and purification of PpiB-dC10 in M9 minimal media

BL21-Gold(DE3) *E. coli* cells were transformed with the pET-MCSI-H6-PpiB-dC10 plasmid and a single colony was used for inoculation of 5 mL of LB media supplemented with Ampicillin (100 µM) that was cultured overnight at 37 °C. The pre-culture was briefly centrifuged and the pellet was used to inoculate 250 – 400 mL of M9 minimal media (containing 0.12% (*w/v*) ^{15}N-labelled ammonium chloride as a sole source of nitrogen for ^{15}N-labelled PpiB-dC10, and 0.4% (*w/v*) $^{13}C_6$ uniformly labelled *D*-glucose for doubly labelled PpiB-dC10). When the optical density reached 0.6, the protein expression was induced by addition IPTG to a final concentration of 1 mM and carried out overnight at 28 °C. The cells were harvested by centrifugation at 3700 g for 15 min and resuspended in 15 mL of PBS pH 8.0 (100 mM sodium phosphate, 100 mM NaCl) and frozen at -20 °C. The cells were thawed, 10 mg of lysozyme were added and the lysate was incubated at room temperature for 1 hour, with shaking. Cells were lysed by (2 x 1 min) sonication, after which the nucleic acids were lysed by 30 min incubation at room temperature with 3 µL of DNase and 3 µL of RNase A. The insoluble proteins were separated by centrifugation at 6400g for 20 min and soluble fraction was used for incubation with NiNTA resin for 2 hours at

4 °C with gentle shaking. After incubation, unattached proteins were separated, the resin was washed by an equivalent volume of 50 mM sodium phosphate buffer, pH 6.2 with 50 mM imidazole and PpiB-dC10 was eluted by the same phosphate buffer containing 500 mM of imidazole.

4.11.5. Preparation of PpiB-dC10 sample for NMR

The PpiB-dC10 elution buffer was exchanged by centrifugation for 50 mM sodium phosphate pH 6.2, 1 mM DTT (PpiB NMR buffer) and the protein was concentrated to a final concentration of 0.4 mM, to which deuterium oxide was added to a final amount of 10% (*v/v*). We proceeded as soon as possible to spectra acquisition as the sample seemed to be unstable and slightly precipitate.

Cleavage of dC10 from PpiB was carried out overnight at 23 °C in 20 mM Tris pH 8.0, 100 mM NaCl, 2 mM $CaCl_2$, with protein concentration around 50 µM and in presence of 1 µg of Factor Xa per 500 µL of reaction volume. The completion of the reaction was verified by MALDI. PpiB was thoroughly washed in a 10 kDa MWCO Amicon filter with PpiB NMR buffer to wash off the cleaved dC10 peptide, and concentrated to 0.4 mM, and finally supplemented with 10 % (*v/v*) deuterium oxide.

4.11.6. PpiB NMR spectra acquisition

Simple ^{1}H-^{15}N HSQC spectra on ^{15}N-labelled protein were acquired at 35°C using a Varian INOVA 500-MHz spectrometer at the University of Ottawa, under the same conditions as published previously [122]. Doubly labelled PpiB-dC10 was prepared for 3D spectra acquisition (HNCACB, CBCACONH, HNCO) used for protein backbone assignment that was further improved using 3D ^{15}N-TOCSY and ^{15}N- NOESY spectra acquired on ^{15}N-labelled protein. 3D spectra were acquired using an INOVA 500-MHz spectrometer equipped with a HCN cold probe at the QANUC NMR Facility at McGill University, Montréal. Mixing time for ^{15}N-edited NOESY acquisition was 100 ms.

NMR data were processed using NMRPipe [148] and visualized in NMRDraw [148] and Sparky [149]. Resonance assignment of His_6-PpiB-dC10 was done using Sparky and previously published assignments of PpiB ([122] and BMRB access code 4765).

4.11.7. Backbone resonances assignment and secondary structure analysis

Published backbone assignments for PpiB [122] was used for a comparison with measured PpiB-dC10 and from there determined the combined chemical shift change of PpiB-dC10 versus PpiB, in order to determine the effect of dC10 tag attached on the *C*-terminus of PpiB on its overall structure. For dC10 secondary structure analysis, CamCoil software (*http://www-vendruscolo.ch.cam.ac.uk/camcoil.php*, March 10th 2012) was used for a prediction of Cα chemical shifts of dC10 in a random coil conformation and these were subtracted from assigned Cα chemical shifts from PpiB-dC10 to obtain Cα secondary chemical shifts.

4.11.8. Cloning of MBP-dC10 single, double and triple mutant libraries

All cloning was performed using standard PCR amplification by KOD Xtreme Hot Start DNA polymerase if not stated otherwise, DNA oligonucleotides were purchased at IDT Technologies, and all mutants were identified by Sanger sequencing at the Génome Québec sequencing service (gqinnovationcenter.com). The MBP-dC10 expression plasmid was created using the following approach: a 700 bp 3′ fragment of *malE* gene from pMAL-c5X vector (New England Biolabs) was amplified by PCR using pMAL-fw primer (CAAAGATCTGCTGCCG) and a reverse mega-primer containing a sequence annealing in 3′ of *malE* gene, the reverse coding sequence of dC10 peptide tag, a stop codon and *EcoRI* restriction site GAATTCCCTACTTTCCTCCAGCTCTAGCTGCAGCTTCTCTGCATGC AGCTTCTCTAGCAGCGCACTCAGCAGCGCTCAGCCTTCCCTCGATCCC). The resulting amplified fragment was inserted into original pMAL-c5x vector using *BglII* and *EcoRI* restriction sites and correct clones were identified by restriction analysis and confirmed by sequencing.

A single mutant library I of MBP-dC10 on position A17 was created by site-directed mutagenesis using a degenerate codon VRN and its complementary codon NYB (where V = C, G or T; R = C or T; Y = A or G; B = C, G or T; and N = any of A, T, G, C) that allowed replacement of residue A17 by D, E, G, H, K, N, Q, R or S. The mutant MBP-dC10 A17G was identified but was not characterised further since only polar or positively charged residues at position A17 were desired in this study. A rolling circle PCR was performed with

two primers dC10_A17X_fw and dC10_A17X_bw containing the mentioned VRN codon (**Table 4.2.**). Double mutant library II was created using mutant MBP-dC10 A17K as template, and DNA primers dC10-A16HR_A17K_fw, dC10-A16HR_A17K_bw containing a degenerate CRY codon (where R = C or T and Y = A or G) for creation of double mutants MBP-dC10 A16H-A17K, and MBP-dC10 A16R-A17K (**Table 4.2.**). A pair of standard DNA primers MBP-dC10 A16K-A17K_fw, MBP-dC10 A16K-A17K_bw (**Table 4.2.**) was used to introduce the A16K-A17K mutation. As a control, double mutations R9K-A17H, R9K-A17K and R9K-A17R were prepared to investigate the importance of residue R9 whose side chain is thought to form an ionic bridge with nearby glutamate E5. Simple DNA primers dC10-R9K_fw and dC10-R9K_bw (**Table 4.2.**) were used for rolling circle PCR of parent MBP-dC10 A17H, A17K or A17R plasmids. The triple mutant library III on position A3 was created from previously identified double mutants MBP-dC10 A16H-A17K, MBP-dC10 A16K-A17K and MBP-dC10 A16R-A17K. A degenerate codon CRY was used for introduction of residues H or R on position A3 (**Table 4.2.**) and a pair of standard oligonucleotides was used for A3K mutagenesis (**Table 4.2.**).

Due to some technical issues, the triple mutants A3R-A16K-A17K and A3H-A16R-A17K had to be prepared using a different setup: A new pair of oligonucleotides (dC10_A3R_fw, dC10_A3R_bw) was used to introduce the A3R mutation in MBP-dC10 A16K-A17K by rolling-circle PCR; and a site-overlap extension PCR by Vent Polymerase (New England Biolabs) and primers dC10-A3H_fw and dC10-A3H_bw (**Table 4.2.**) were used to amplify mutated portion of A3H-A16R-A17K that was subsequently inserted into parent pMAL-c5x vector between *BglII* and *EcoRI* restriction sites.

4.11.9. Cloning of ErbB1-dC10 mutants for mammalian expression of EGFR

We used a previously prepared plasmid pcDNA3.1(+) dC10-ErbB1 [73] for mammalian expression of the Epidermal Growth Factor Receptor protein (EGFR) in fusion with a dC10 sequence on its *N*-terminus as template for introduction of a triple mutation A3K-A16R-A17K in the dC10 DNA sequence. One round of rolling circle PCR amplification was used for introduction of A16R-A17K mutations with primers ErbB1-dC10-A16R-A17K_fw and ErbB1-dC10-A16R-A17K_bw (**Table 4.2.**), and the correct resulting plasmid identified by sequencing was used as template for standard PCR amplification of a dC10-A3K-A16R-

A17K gene using ErbB1-dC10-A3K_fw and ErbB1-dC10-A3K_bw primers, subsequently inserted between the *HindIII* and *XhoI* restriction sites in parent pcDNA dC10-ErbB1plasmid. Correct gene insertion was confirmed by Sanger sequencing.

Table 4.2. Oligonucleotides

Name	DNA Sequence 5′ – 3′
dC10-A17X_fw	GCAGAGAAGCTVRNGCTAGAGCTGGAGGAAAGTAGGGAATTCC
dC10-A17X_bw	CCTCCAGCTCTAGCNYBAGCTTCTCTGCATGCAGCTTCTCTAGC
dC10-R9K_fw	GCTGCTGAGTGCGCTGCTAAAGAAGCTGCATGCAGAGAAGC
dC10-R9K_bw	CATGCAGCTTCTTTAGCAGCGCACTCAGCAGCGCTCAGCCTTCC
dC10-A16HR-A17K_fw	GCAGAGAACRYAAAGCTAGAGCTGGAGGAAAGTAGGGAATTCC
dC10-A16HR-A17K_bw	CCTCCAGCTCTAGCTTTRYGTTCTCTGCATGCAGCTTCTCTAGC
dC10-A16K-A17K_fw	GCAGAGAAAAGAAAGCTAGAGCTGGAGGAAAGTAGGGAATTCC
dC10-A16K-A17K_bw	CCTCCAGCTCTAGCTTTCTTTTCTCTGCATGCAGCTTCTCTAGC
dC10-A3HR_fw	GAGGGAAGGCTGAGCCRYGCTGAGTGCGCTGCTAGAGAAG
dC10-A3HR_bw	CAGCGCACTCAGCRYGGCTCAGCCTTCCCTCGATCCCGAG
dC10-A3K_fw	GAGGGAAGGCTGAGCAAAGCTGAGTGCGCTGCTAGAGAAG

dC10-A3K_bw	CAGCGCACTCAGCTTTGCTCAGCCTTCCCTCGATCCCGAG
dC10-A3R_fw	GAGGGAAGGCTGAGCAGAGCTGAGTGCGCTGCTAGAGAAG CTG
dC10-A3R_bw	CAGCGCACTCAGCTCTGCTCAGCCTTCCCTCGATCCCGAG
dC10-A3H_fw	GAGGGAAGGCTGAGCCATGCTGAGTGCGCTGCTAGAGAAGC TG
dC10-A3H_bw	CAGCGCACTCAGCATGGCTCAGCCTTCCCTCGATCCCGAG
ErbB1-dC10-A16R-A17K_fw	GCTGCATGCAGAGAAAGAAAGGCTAGAGCTGGAGGAAAGC TCGAG
ErbB1-dC10-A16R-A17K_bw	TCCTCCAGCTCTAGCCTTTCTTTCTCTGCATGCAGCTTCTCTA GC
ErbB1-dC10-A3K_fw	TCTCAGAGGAGGACCTGAGCAAGGCTGAGTGCGCTGCTAGA GAAGC
ErbB1-dC10-A3K_bw	TAGCAGCGCACTCAGCCTTGCTCAGGTCCTCCTCTGAGATCA GC

* mutated codons underlined

4.11.10. Expression and purification of MBP-dC10 variants

All mutants of MBP-dC10 were expressed in BL21-Gold(DE3) strain of *E. coli* cells and purified in high yields. Transformed cells were grown in 250-350 mL of LB media supplemented with 0.2 % (*w/v*) D-glucose and 100 µM of ampicillin. Expression of recombinant MBP-dC10 (or its variants) was induced with 0.3 mM IPTG at OD ~ 0.6, and was carried out at 37 °C for 3-4 hours with vigorous shaking. Cells were harvested by centrifugation at 4000 g for 15 minutes, resuspended in 10 mL of MBP-binding buffer (Tris 20 mM pH 7.4, NaCl 200 mM, EDTA 1 mM) and stored at -20 °C. Thawed cells were lysed by sonication on ice 2 x 1 minutes, and soluble fraction was separated from insoluble proteins by centrifugation at 6000 g for 15 minutes at 4 °C. The supernatant fraction was

loaded on amylose columns pre-equilibrated in MBP-binding buffer, and incubated at 4 °C with gentle stirring for 2 hours. Unbound proteins were eluted in the flow through fraction and the column was subsequently washed with 5 mL of MBP-binding buffer, and pure MBP-dC10 (or variants) was then eluted by 3 mL of MBP-binding buffer containing 10 mM of maltose. An overnight dialysis was performed into HEPES 50 mM pH 7.5, TCEP 1 mM before kinetic characterisation of MBP-dC10 variants. According to a Bradford assay, a high yield of 5-15 mg of each protein after purification was obtained using this protocol.

4.11.11. Kinetic characterisation of MBP-dC10 variants by fluorogenic reaction with dM10 fluorogens

The reactivity of each MBP-dC10 variant was assessed by determining the second order kinetic constant of fluorogenic labelling reaction with a dansyl-dimaleimide fluorogenic molecule, referred to as **dM10-dansyl 1** (**Figure 4.23.**), synthesized in our laboratory. Initial kinetic studies were performed on MBP-dC10 histidine mutants S2H, R9H, A7H and S2H-A17H on the scale of 500 µL in established conditions described below, using a Cary Eclipse Fluorimeter at 20 °C. The reaction temperature was adjusted to 28 °C only for latter experiments where use of a plate reader that does not allow sample cooling. Libraries I, II and III were assayed using a plate reader: briefly, in a reaction scale of 200 µL, 50 µM MBP-dC10 (or variants) and 50 µM **dM10-dansyl 1** were mixed in HEPES 50 mM pH 7.5 buffer in presence of 1 mM TCEP and the reaction was initiated by addition of fluorogen from a 2 mM stock solution in DMSO. Fluorescence increase was followed at 28 °C far after completion of the reaction (2 hours) by the Synergy H4 (BioTek) plate reader at 515 nm, upon excitation at 330 nm of the dansyl fluorophore of **dM10-dansyl 1**. Second order reaction constants were determined using the initial slope and the final plateau of each kinetic curve. All reactions were performed in duplicates, or quadruplicates.

For pH dependent kinetics, MBP-dC10 variants were dialysed into a buffer of desired pH, containing 50 mM of buffering salt (pH 7.50 - 8.00: HEPES; pH 8.25 – 8.75: Tris-HCl; pH 9.00: CHES) and 1 mM TCEP, overnight at 4°C, with gentle stirring. Fluorogenic reactions were performed in a plate reader as described above, in duplicate. In the case of labelling of MBP-dC10 and its variants with **dM10-coumarin 9** (**Figure 4.23.**), a Fluorimeter Cary Eclipse (Varian) was used: In the same reaction conditions, fluorescence increase was

followed at 28 °C at 485 nm upon excitation at 450 nm. Second order reaction constants were determined by fitting, to the second order equation, an option incorporated in the Eclipse Kinetics software. All reactions were performed in duplicate.

4.11.12. Prediction of peptide helicity content

An online version of AGADIR software (*http://agadir.crg.es/*) [79, 80, 81] was used for dC10 peptide helicity prediction in the pH range of 7.50-9.10, at 301 K and ionic strength of 0.1 M.

4.11.13. Mammalian cell culture, expression of dC10-EGFR and H2B-dC10 variants and in cellulo fluorogenic labelling

HEK293 were grown in MEM minimal media (Life Technologies) supplemented by 10% Fetal Bovine Serum and 1% penicillin and streptomycin, according to the protocol suggested by the manufacturer. For labelling and subsequent fluorescence microscopy detection, cells were transfected using Lipofectamine 2000 or Fugene® 6, 16 hours post plating of $4x10^5$ cells in a 60-mm plate or a 35-mm plate for inverse microscopy. Cells were transfected by pcDNA3.1(+)-dC10-EGFR or H2B-dC10 plasmids, and their dC10* variants and the medium was changed after 45-60 minutes. Expression of test proteins was allowed for 24-48 hours after which cells were labelled using indicated dM10-coumarin fluorogens. **dM10-coumarin 9** was simply added to the media after a media change for Opti-MEM and fluorescence increase was followed using the Nikon NF1 confocal microscope and the Fluor 40x objective, or the Zeiss Axio Observer A1 inverted epifluorescence microscope.

For a complementary labelling of dC10-EGFR with EGF-rhodamine conjugate (Molecular Probes), 2 µg/mL EGF-rhodamine was added to the cell Opti-MEM media and incubated for 10 min at 37°C, followed by a media change and subsequent labelling with **dM10-coumarin 9**, or immediate imaging.

For H2B-dC10 (or dC10*) labelling by **dM10-coumarin 20**, the fluorogen was first diluted to 10 µM (from a 10 mM stock solution in DMSO) in a total amount of 2 mL of Opti-MEM media for microscopy (Life Technologies) with addition of 0.1% (*v/v*) F127 surfactant and DMSO (4 µL to yield a final concentration of 0.3% (*v/v*)), and transfected cells were labelled

by a medium change using this fluorogen containing medium. Labelling was allowed to progress in a 37°C incubator with 5% CO_2 for indicated time, after which the labelling medium was changed for Opti-MEM for microscopy. Cells were imaged using the Nikon NF1 confocal microscope equipped with Apo LWD 25x objective, a filter with a 35 nm bandwidth and excitation and emission at 457.0 nm and 482.0 nm, respectively. Images were converted using NIS-Elements Viewer software and processed using ImageJ.

Chapter 5

ALTERNATIVE USE OF DIMALEIMIDE-FUNCTIONALIZED MOLECULES FOR PROTEIN STRUCTURE STUDIES – PROTEIN NMR

5.1. Introduction

The structure and function of a protein are tightly interlinked and the knowledge of both may allow us to determine the role of a protein in a cellular context. The large majority (90%) of currently known macromolecular structures were determined by X-ray crystallography and only a minor part by NMR (*www.rcsb.org* updated as of June 30th 2015). However, NMR allows structural biologists to study macromolecules as dynamic species, while a crystal is only one rigid form on the structural landscape of a macromolecule. Since approximately 1995 [150], the number of structures solved by NMR grew exponentially until 2007, when the number of newly deposited structures per year reached a maximum and since then it has been in a slight decline. To this day, the number of macromolecular structures determined by NMR and deposited in the Protein Data Bank (*www.rcsb.org*) has reached 11 048.

5.1.1. NMR spectroscopy of large proteins

NMR spectroscopy allows biologists to study a particular protein at atomic resolution and protein supramolecular complexes up to 1 MDa in size, for extremely robust systems. However, NMR techniques are limited by protein size because larger molecules undergo slow molecular reorientation that gives rise to rapid loss of the NMR signal due to efficient relaxation. Also, numerous overlapping cross-peak complicate, if not make impossible, the deconvolution of even simple ^{1}H-^{15}N heteronuclear correlation spectra. Advances in NMR instrumentation and the development of specialized pulse sequences have tremendously improved the sensitivity of NMR experiments [151], and combined with use of special isotope labeling techniques designed for large proteins, it is now possible to study fairly robust proteins up to 50 kDa [152, 153, 154, 155, 156]. Strategies for the specific introduction of [^{1}H,^{13}C]-labelled side chain methyl groups in perdeuterated proteins have substantially extended that upper molecular weight limit, and allowed systems of up to 1MDa to be studied [157, 158], a limit that had never been accessible before.

Along with diverse isotope labelling methods [154, 159, 160] of aliphatic regions that help prevent signal loss caused by dipolar interaction between neighbouring protons,

complementary methods for reduction of peak crowding and overlap using lanthanides have been developed [161]; these will be described in the following sections.

5.1.2. Lanthanides as shift-inducing agents

Lanthanides have unique photophysical and electronic properties that make them valuable for studying protein structure, function and dynamics [162]. Lanthanides are known for inducing pseudo-contact shifts (PCS) and residual dipolar couplings (RDC) in an NMR spectrum, due to their paramagnetic (with the exception of lanthanum and lutetium, which are diamagnetic) and anisotropic properties [163]. Pseudo-contact shifts are easy to measure as they correspond simply to the difference in observed chemical shift between the paramagnetic and the diamagnetic sample. The extent of PCS induced by a lanthanide is closely dependent on its inherent properties (**Figure 5.1.**) and on the distance of observed nuclei from the lanthanide (**Equation 5.1.**). Along with PCS, paramagnetic lanthanides significantly change the relaxation properties of nuclei in their proximity by the paramagnetic relaxation enhancement (PRE) effect, which greatly affects the signal intensity. Generally, PRE causes intensity loss and peak broadening, such that in some cases, it is not possible to observe these signals in a spectrum.

Both PRE and PCS effects are distance-dependent; however, they do not affect the observed nuclei properties to the same extent. As shown in **Equation 5.1.**, the PCS is proportional to $1/r^3$, while PRE strength is proportional to $1/r^6$, where r represents the distance of the observed nucleus from the paramagnetic centre. In other words, an increase in the distance from a lanthanide atom causes the PRE effect to decrease to a greater extent relative to the PCS. This means that there is a distance interval where a nuclear spin could show a PCS, but not PRE, with the PCS potentially being seen in spins that are up to 40 Å away from the paramagnetic centre [164, 165]. For example, dysprosium (Dy) exhibits a strong paramagnetic relaxation enhancement for nuclei that are within 14 Å of the metal centre that could broaden the corresponding peaks beyond the limits of detection, but from 14 Å up to 40 Å, peaks would not undergo a significant PRE, allowing PCSs to be readily observed [165]. For the design of lanthanide-based probes for structure determination, terbium, dysprosium, thulium, ytterbium or europium seem to be best for the observation of PCS, as reviewed by Pintacuda *et al.* [166].

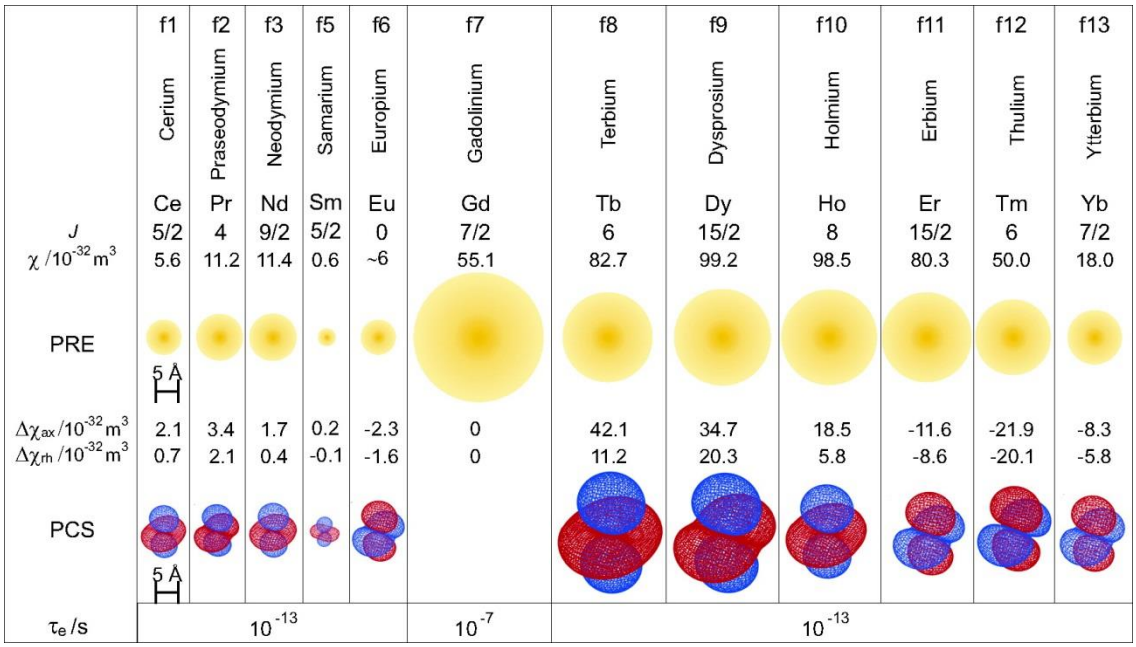


Figure 5.1. Paramagnetic properties of Ln^{3+} ions. Representative isosurfaces are plotted for PCSs by 5 ppm using tensors reported by Bertini *et al.* [167]. Adapted with permission from "NMR Structure Determination of Protein−Ligand Complexes by Lanthanide Labeling", Guido Pintacuda, Michael John, Xun-Cheng Su, and Gottfried Otting, *Acc Chem Res* **2007** 40 (3), 206-212, DOI: 10.1021/ar050087z. Copyright 2007 American Chemical Society.

$$\Delta\delta^{PCS} = \frac{1}{12\pi r^3}\left[\Delta\chi_{ax}(3\cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\varphi\right]$$

Equation 5.1. Equation for PCS in paramagnetic samples. $\Delta\delta^{PCS}$ denotes the difference of chemical shifts between paramagnetic and diamagnetic samples, $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and the rhombic components of the magnetic susceptibility anisotropy ($\Delta\chi$) tensor and r, θ and φ are the polar coordinates of the nucleus with respect of the principal axes of the tensor [168].

5.1.3. Lanthanides in probes for studying protein structure and dynamics

The PCS effect can provide useful information for the study of proteins and/or their interactions. For example, PCS can represent spatial constraints for protein structure determination [169, 170] or for studying protein/protein or protein/ligand complexes [166, 171, 172].

Lanthanide trivalent ions do not have a specific binding motif on a protein surface; therefore, it is necessary to attach a lanthanide-binding moiety to the protein of interest, ideally in a site-specific manner. It was demonstrated in the example of the calcium binding protein calbindin, that various trivalent lanthanides can be bound by its *C*-terminal calcium binding site and induce PCS in the corresponding NMR spectrum [165, 167, 173]. Even though this application worked very well, it is limited to proteins that have a metal binding site that can accommodate a lanthanide ion, and the formed lanthanide-bound state of the protein is only assumed to be isomorphic. Following a similar approach, calmodulin and EF-Vpu binding peptide have been shown to bind Ln^{3+} ions and were successfully used as Tb^{3+} binding tags for the observation of residual dipolar couplings [174, 175].

More recently, small molecules that are able to chelate Ln^{3+} have been designed and used as paramagnetic probes. Several requirements must be considered when a paramagnetic probe is designed, specifically the probe should contain; a) a strong lanthanide chelating moiety; b) site-specific point(s) of attachment on a protein; c) a limited number of flexible bonds; d) no tendency to alter the structure and function in the protein of interest. Usually, cysteine residues are used for the site-specific modification of proteins, due to their infrequent occurrence in protein sequences and relatively reactive sulfhydryl-containing side chain. Generally, Ln^{3+} probes are derived from dipicolinic acid (PDCA – pyridine-2,6-dicarboxylic acid) or from ethylenediaminetetraacetic acid (EDTA [176]) to which a vast number of modifications have been introduced in numerous studies, in order to reinforce the binding affinity for a Ln^{3+} ion and to limit the number of possible stereoisomers [161, 177] formed after ligation.

5.1.4. Example of site-specific protein labelling with a shift-inducting agent

In order to better appreciate the usefulness of lanthanides as shift-inducing agents in protein NMR spectroscopy, we will present in this sub-section a detailed example [170], where such a tool allowed 3D structure determination of ERp29-C protein using PCS. ERp29-C protein was tagged at four different sites using paramagnetic probes described previously [177, 178], chelating Tb^{3+} and Tm^{3+} giving rise to large PCS. Backbone resonances of all four paramagnetically labelled variants of ERp29-C were assigned and PCS were confirmed for over 90% of residues, excluding flexible *N*- and *C*- terminal segments. Then, they used GPS-

Rosetta program, which allows including data from multiple samples using several paramagnetic centres, to calculate the 3D structure of ERp29-C, using PCS data as distance constraints. When this PCS-derived structure was then compared to ERp29-C crystal structure, they both appeared similar (**Figure 5.2.**). However, when compared to an original NOE structure, differences were observed between the PCS-derived structure, crystal structure and the NOE structure, supporting the importance of lanthanide-induced PCS for accurate protein structure determination.

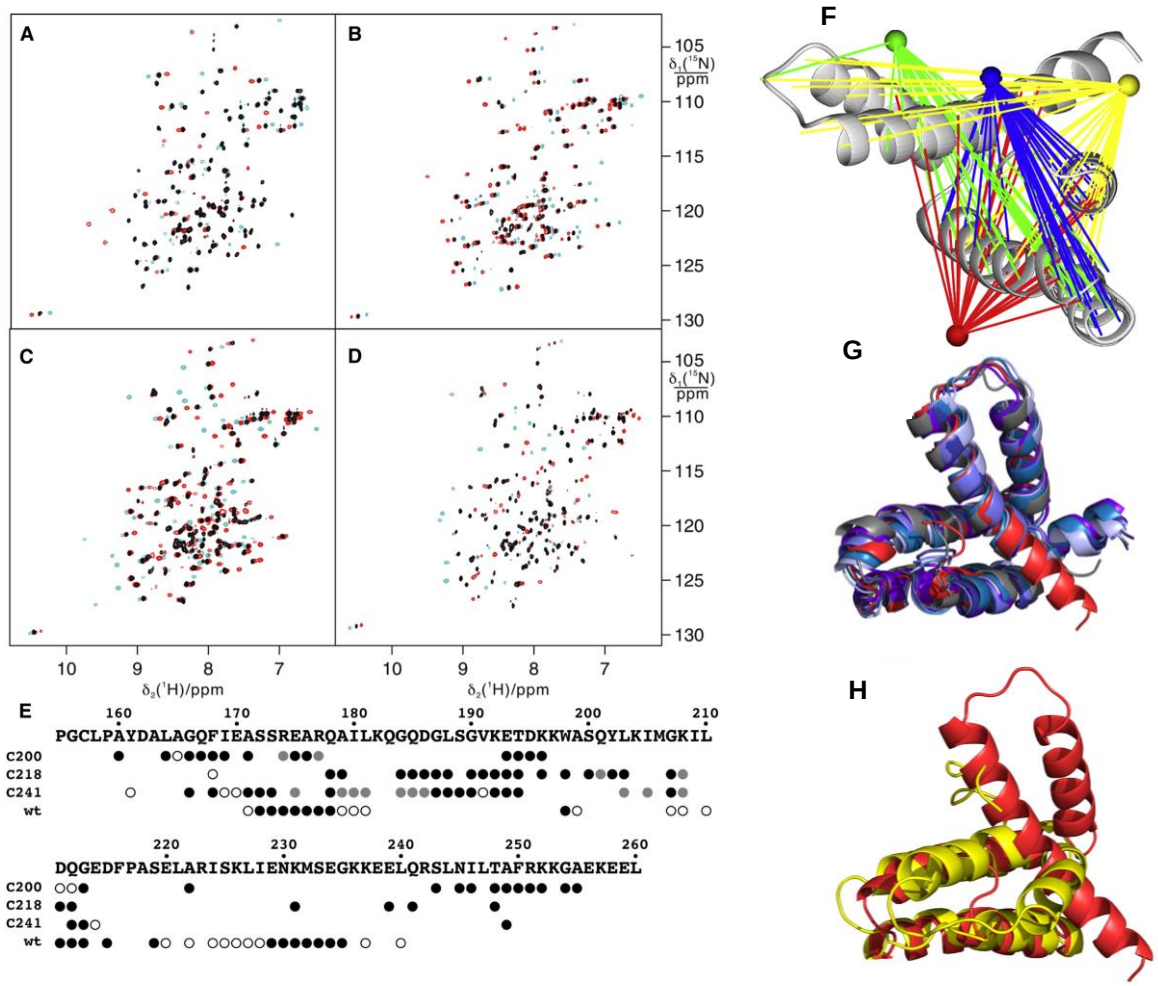
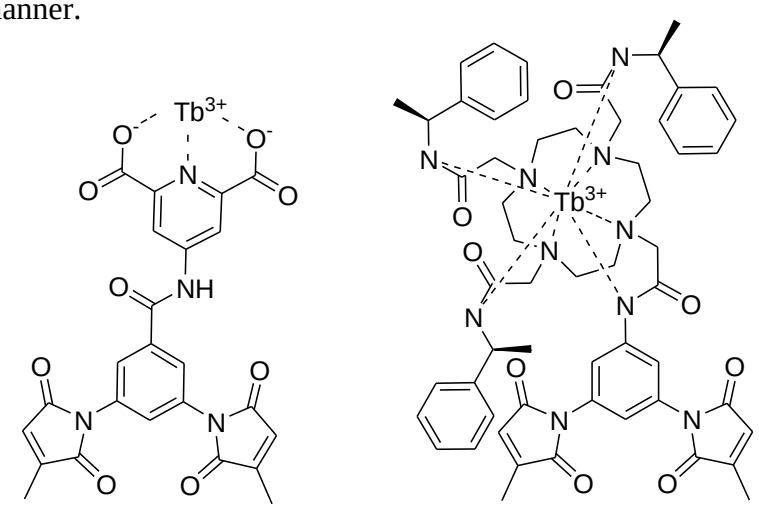


Figure 5.2. Pseudocontact shifts of amide protons in ERp29-C. (*A-D*)*:* Shown are PCS obtained for four variants of ERp29-C with Tb^{3+} (*cyan*), Tm^{3+} (*red*) and Y^{3+} (*black*), and the distribution of observed PCS on the protein sequence (*E*). In (F) are shown schematized effects of lanthanides positioned at four different loci in the protein. (*G-H*) show ERp29-C structure using PCS (G) and the comparison of PCS-derived structure (*red*) with the NOE structure (*yellow*) (*H*). Reprinted from *Structure*, **21**, Yagi, H., Pilla, K.B., Maleckis, A., Graham, B., Huber, T. and Otting, G., "Three-dimensional Protein Fold Determination from Backbone Amide Pseudocontact Shifts Generated by lanthanide Tags at Multiple Sites", 883-890 Copyright 2013, with permission from Elsevier.

5.1.5. Dimaleimide Ln probe design

As detailed in the Introduction, our dimaleimide labelling technology is robust and allows a fast and specific labelling of proteins *in vitro*. Similarly, a number of robust Ln^{3+} chelating tags have been developed and extensively characterized [161, 177, 179, 180], raising the possibility that a dimaleimide-lanthanide probe could be developed for applications in protein NMR. Based on known chelator properties and considers of synthetic feasibility, our group decided to use a dipicolinic acid Ln^{3+} chelator attached to a dM10 moiety through an anilide bond to yield **LnC01** (**Scheme 5.1.**). The synthesis of this first dimaleimide Ln^{3+} probe was performed by Dr. Christophe Pardin (unpublished data) and is not detailed in this thesis.

Later, a different Ln^{3+} chelator was described by Graham *et al.* [177] that has a stronger affinity to an Ln ion and does not form multiple stereoisomers, as it was observed for previously used Ln^{3+} chelators [161]. This Ln^{3+} chelator led us to design a second dM10-lanthanide chelator, referred to as **LnC02** (**Scheme 5.1.**). In the following section both **LnC01** and **LnC02** were tested with a number of test proteins. As the synthesis and availability of paramagnetic probes **LnC01** and **LnC02** was not simultaneous, several experiments were conducted only on **LnC01** and will be presented as such in a chronological manner.



LnC01 Tb **LnC02 Tb**

Scheme 5.1. Lanthanide chelating probes used in this study. Tb^{3+} is shown as an example of a lanthanide chelated by the probes.

5.2. Choice of test protein and tag/tag-free approach is crucial

5.2.1. Test protein choice and design of point of attachment for dM10-lanthanide probes

Similar to dM10-lanthanide probe design requirements, it is important to consider how the probe will be attached to the protein, especially since the dynamic character of peptide tags can have a detrimental effect on the accuracy of distance measurements based on PCSs. Until now, we have used a dC10 tag that was attached to a test protein on its *N*- or *C*-terminus, and the flexibility of this tag had little impact on the way fluorescence was observed. However, for structural applications such as NMR, where the distance of the observed nuclei from the Ln core plays a critical role in the chemical environment, it is important to design a more rigid way of attaching the paramagnetic probe. Hence, a minimalist version of dC10 tag was designed where two amino acids in an intrinsic, solvent-exposed helix of a test protein were mutated to cysteine to yield diCys10. In this chapter, both approaches were used, namely dC10 tag and intrinsic diCys10, and obtained results were compared.

Similarly, PCS can be used as a tool to enhance peak dispersion (and hence reduce overlap) in spectra from large proteins. Therefore, both a "large" protein, MBP, and a "small" protein, ubiquitin, were used as test systems for designed probes.

5.2.2. Flexible dC10 tag

First, a very well characterized and soluble protein, MBP-dC10, which has been extensively studied by solution NMR [181, 182] and used in dimaleimide labelling, was expressed in ^{15}N-labelled form and then derivatized *in vitro* with **LnC01 Eu**. Labelling was performed overnight at room temperature using 4-6 equivalents of probe, and verified by MALDI, where a portion of probe-free MBP-dC10 ^{15}N was detected, and an **LnC01 Eu** – labelled protein was identified via the expected mass shift corresponding to the addition of this paramagnetic probe (**Table 5.1.**).

Table 5.1. Mass analysis of labelling with LnC01 Eu. Shown are expected and experimental masses of unlabelled and **LnC01 Eu**-labelled MBP-dC10 ^{15}N, with the corresponding observed mass difference. The observed and expected mass shifts are indicated in parentheses.

	MBP-dC10 ^{15}N	**MBP-dC10 ^{15}N – LnC01 – Eu^{3+}**
Expected mass (Da)	45254.5	45910.9 (+ 656.4)
Observed masse (Da)	45296.5	45955.8 (+ 659.3)

Along with a Eu^{3+}-chelating probe, Dy^{3+}- and Tb^{3+}-chelating probes were prepared and used in the same manner (referred to as **LnC01 Tb** and **LnC01 Dy**) for labelling of MBP-dC10 and ^{1}H-^{15}N-TROSY spectra were acquired for all variants. As an example, an overlay of a ^{1}H-^{15}N-TROSY spectrum of **LnC01 Tb**- and **LnC01 Eu**-labelled MBP-dC10 is shown on **Figure 5.3.** Unexpectedly, the two spectra are highly similar, with only minor shifts being observed, as indicated on the spectra. **LnC01 Eu** and **LnC01 Tb** should produce PCS in MBP-dC10, oriented in opposite directions from a diamagnetic reference, as suggested by the opposite signs of both the $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ components of their respective magnetic susceptibility anisotropy tensors (**Figure 5.1**.). However, the very small differences observed in **Figure 5.3.** do not allow us to conclude that a LnC01 probe with a paramagnetic lanthanide would cause pseudo-contact shifts in MBP-dC10 NMR spectrum.

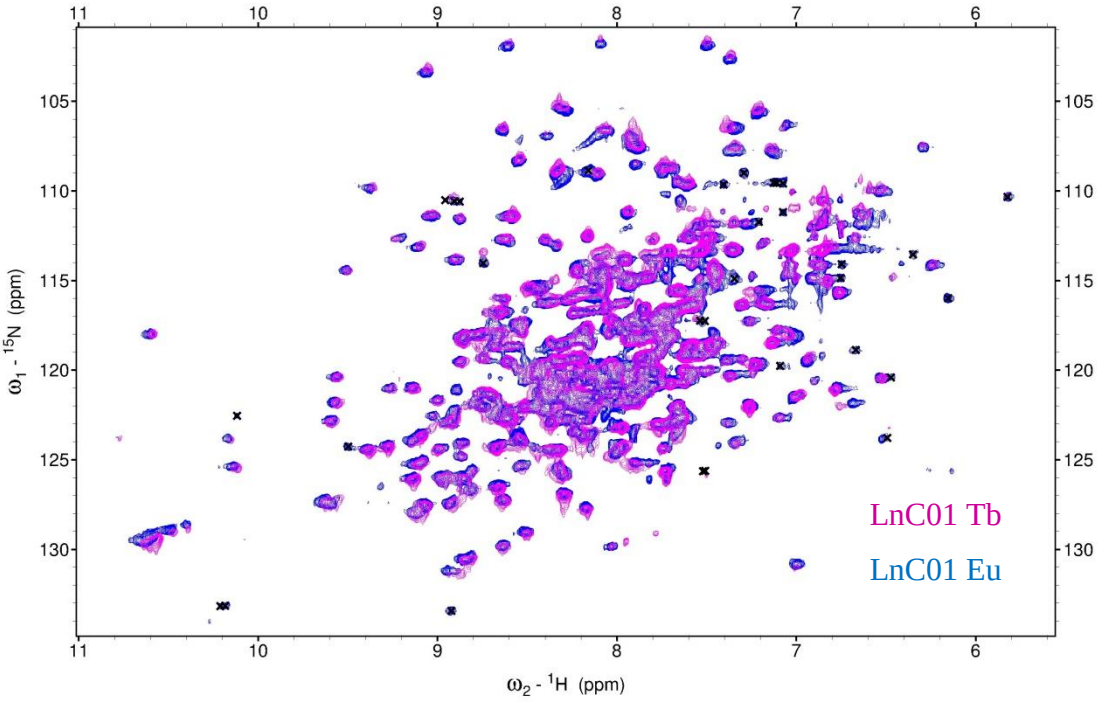


Figure 5.3. Overlay of ^{1}H-^{15}N TROSY spectra of MBP-dC10 ^{15}N labelled with LnC01 Eu (*blue*) and LnC01 Tb (*magenta*). Peaks that may present pseudo-contact shifts are indicated with a black x.

5.2.3. Ubiquitin-dC10 as a "small" test protein

After the first unsuccessful results with MBP-dC10 labelled with a paramagnetic probe, we considered using a smaller protein that would resemble more closely the test proteins commonly used by other groups who work with lanthanide probes, such as *E. coli* arginine suppressor ArgN, ubiquitin, or p75ICD [177, 183]. We turned our efforts to a small and very well studied protein, ubiquitin, and designed a simple hexahistidine-tagged construct, His_6-Ubi-dC10 (referred to as Ubi-dC10 for brevity) especially for this purpose. Ubi-dC10 was expressed in a $^{15}NH_4Cl$ supplemented M9 media and purified on a NiNTA resin. During protein expression, purification and more prominently after elution from NiNTA resin, however, partial cleavage of dC10 was observed on SDS-PAGE gel (**Figure 5.4.**) and by MALDI, despite the use of protease inhibitors. Before attempting an optimization of expression conditions to lower the proportion of cleaved Ubi, it was noticed that cleavage of dC10 was even more prominent after the sample was concentrated, which is undesired and

does not allow us to continue with this construct. We did not explore further if the cleavage was caused by a more exposed spacer sequence, or by a drastic solubility disparity between Ubi and dC10. We can only conclude that this Ubi-dC10 construct was not suitable for any further application using dC10.

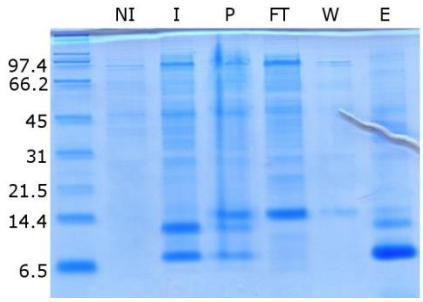


Figure 5.4. SDS-PAGE analysis of His_6-Ubi-dC10 purification process. Ubi (10 kDa) and Ubi-dC10 (13 kDa) bands are present in the induced fraction (I), 14.4-kDa lysozyme used for cell lysis is apparent in the insoluble and unattached fractions (P and FT), and mostly Ubi (10 kDa) with a small amount of Ubi-dC10 (13 kDa) is present in the eluant (E). NI – non induced cells, I – induced cells with 1 mM IPTG, P – insoluble fraction, FT – unattached proteins on NiNTA resin, W – wash, E – eluted proteins. Broad-range protein marker (left line) was used for protein size estimation, indicated in kDa.

5.2.4. Use of an intrinsic helix to label with LnC01 Ln

It is possible that proteins labelled on a flexible dC10 with a paramagnetic probe are not very suitable for NMR studies, likely because of an excessive degree of flexibility of the paramagnetic centre. Acknowledging that fact, our attention was turned to a more minimalist design of protein labelling using an intrinsic di-cysteine helix. A medium sized protein AAC (aminoglycoside acetyl-transferase 6-Ii of *E. faecium* [184]) was chosen, whose dynamics has been extensively studied by NMR, and structure by crystallography in the laboratory of Professor Albert M. Berghuis at McGill University, Montréal [184]. A double mutant R90C-K97C of AAC (referred to as AAC-diCys10 for brevity) was designed where the two newly introduced cysteine residues were 10 Å apart on a solvent exposed helix located outside of the region of high mobility of the protein in the ligand-free state described previously [185].

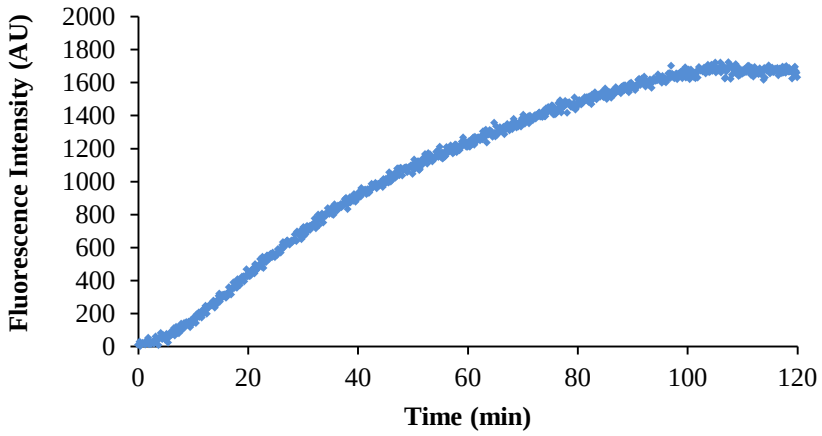


Figure 5.5. Labelling of AAC-diCys10 with dM10-dansyl 1 fluorogen. Equimolar concentrations of 50 µM of protein and fluorogen were used in 50 mM HEPES pH 7.5, 1 mM TCEP buffer, and fluorescence increase was followed at 530 nm upon excitation at 330 nm in a BioTek Synergy H4 plate reader at 28°C.

AAC-diCys10 was expressed and purified by anion-exchange chromatography [186], and characterized by *in vitro* fluorogenic labelling with **dM10-dansyl 1** (**Figure 5.5.**), where the second order rate constant was determined to 378 $M^{-1}min^{-1}$ using the same basic approach as for MBP-dC10 mutant kinetics. This lower rate constant, in comparison with MBP-dC10, may be caused by protein conformational dynamics [185]. The helix containing R90C-K97C was found to be relatively rigid in comparison with the rest of the protein in a wild-type AAC, however, the overall conformational flexibility [185] may have affected the availability of diCys10 for labelling and lowered the rate constant. Furthermore, AAC was shown to adopt a distinct conformation in presence of an excess of its cofactor, acetyl coenzyme A [185]. However, we did not attempt labelling and kinetic characterization of AAC-diCys10 in the presence of an excess of AcCoA because of the significant cost of the AcCoA that would be needed. Isotopically labelled AAC-diCys10 was expressed and purified on an anion exchange column, as published previously [186]. The ^{1}H-^{15}N HSQC spectrum of AAC-diCys10 was recorded without a paramagnetic probe to determine if the two mutations had any effect on the protein structure and fold.

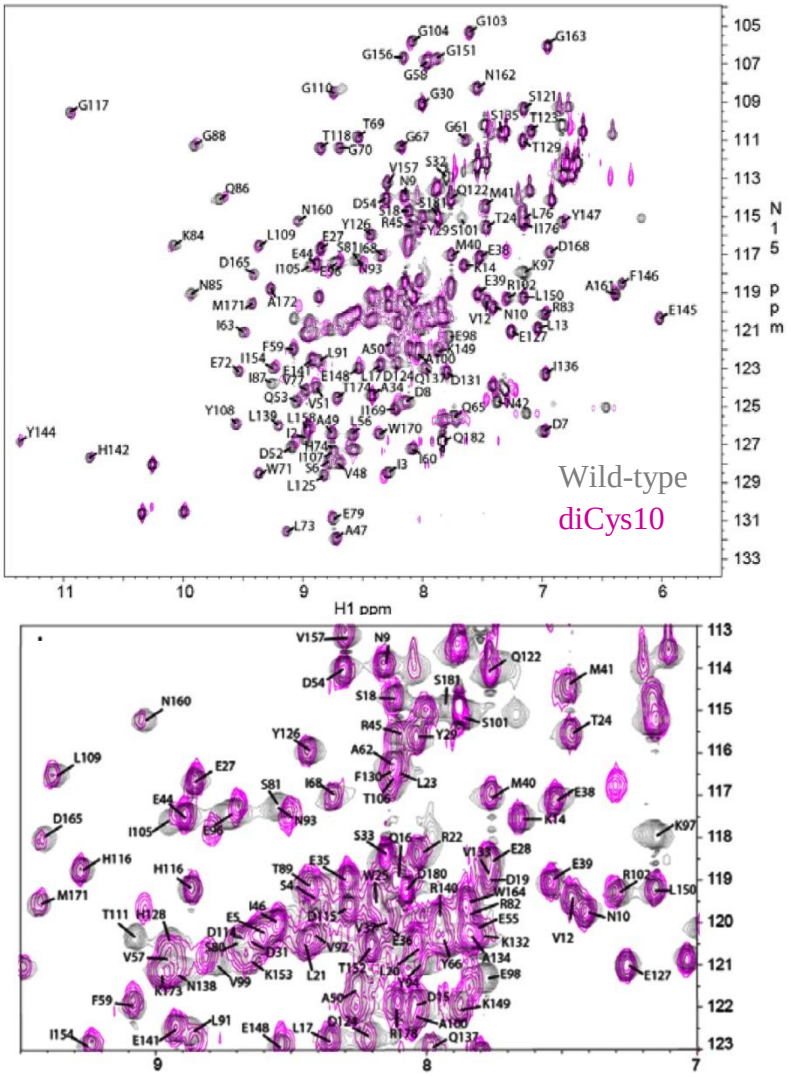


Figure 5.6. Overlay of AAC wild-type (*gray*) and AAC-diCys10 mutant (*purple*) ^{1}H-^{15}N HSQC. Indicated residue assignment corresponds to AAC wild-type. *Top*: entire ^{1}H-^{15}N HSQC spectrum; *bottom*: zoom on central region of spectrum. Adapted by permission from Macmillan Publishers Ltd: "Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme", Freiburger, L.A., Baettig, O.M., Sprules, T., Berghuis, A. M., Auclair, K., Mittermaier, A.K., *Nat Struc Mol Biol*, **18**, 288-294, 2011, doi:10.1038/nsmb.1978, copyright 2011.

An ^{1}H-^{15}N HSQC spectrum of this mutant was very similar to published AAC wild-type spectrum (**Figure 5.6.**) and allowed a tentative assignment of AAC-diCys10 backbone amide resonances. Slight differences were observed for backbone amide resonances of residues that are in proximity of cysteines 90 and 97 (and for cysteine 97 in particular). It is noteworthy that the reliability of this tentative backbone amide assignment is weakest around the sites of mutation.

Average chemical shift differences between wild-type and diCys10 mutant showed that there is little impact of the newly introduced two cysteine residues 90 and 97 on the structure of the AAC (**Figure 5.7.**), and these changes are located mostly in proximity of the two mutated residues.

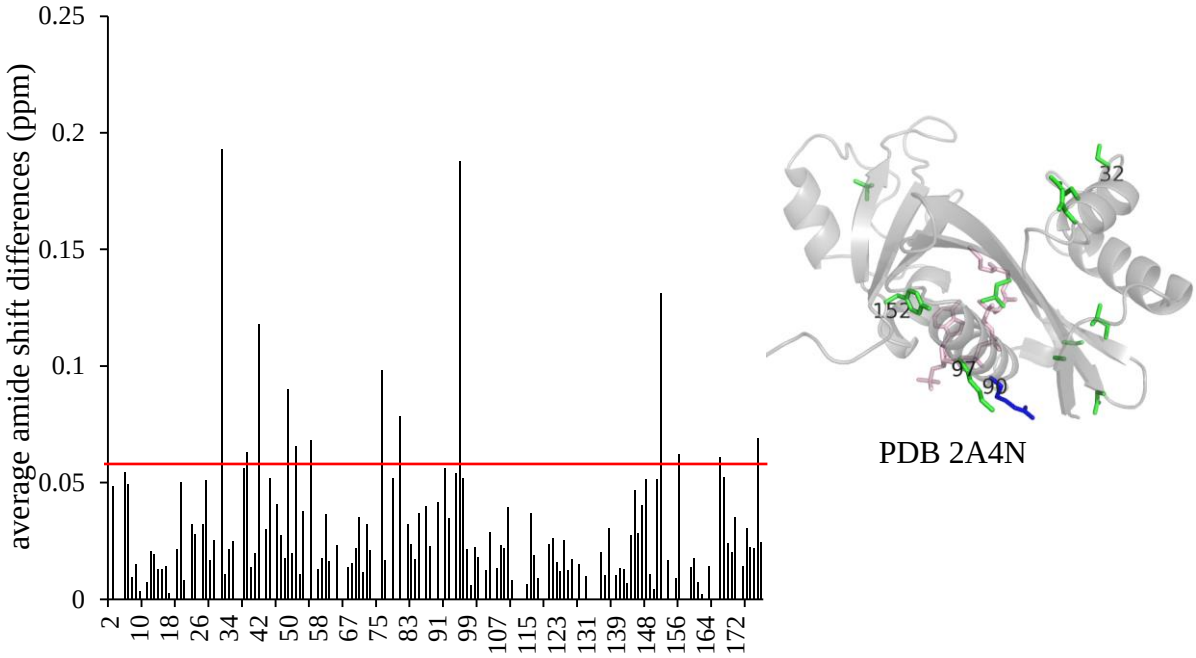


Figure 5.7. Changes in AAC backbone amide chemical shift induced by R90C and K97C mutations. *Left*: Average amide shift differences (see equation in **Figure 4.5.**) are represented for each residue and significantly high differences are delimited by a threshold (*red line*) equal to the average value of all values across the protein sequences plus one standard deviation (0.031 ± 0.030 ppm). *Right:* Wild-type residues that showed significant chemical shift change are shown in a stick representation (*green*), residues with the most prominent change are identified by their sequence number (32, 97, 152), as well as R90C (*blue*) and K97C (*green*) mutations. AcCoA ligand is equally represented (*light pink*).

To confirm that the mutations did not abolish acetyltransferase activity in AAC-diCys10, the activity of this mutant was assessed using a previously published assay [187, 186]. In this experiment, transfer of the acetyl group from AcCoA to kanamycin by AAC gives rise to

CoASH, which reacts with 4,4'-dithiodipyridine to release 4-thiopyridine whose absorbance increase can be followed at 324 nm. When the assay was performed with AAC-diCys10, there was a linear increase in absorbance at 324 nm that was ten times higher than background levels, with an initial slope of 0.00045 AU s^{-1}. Although quantitative measurement of acetyltransferase kinetics for this mutant was beyond the scope of our study, the clear presence of significant acetyltransferase activity in AAC-diCys10 provided good confirmation that the catalytic integrity of the enzyme was largely intact.

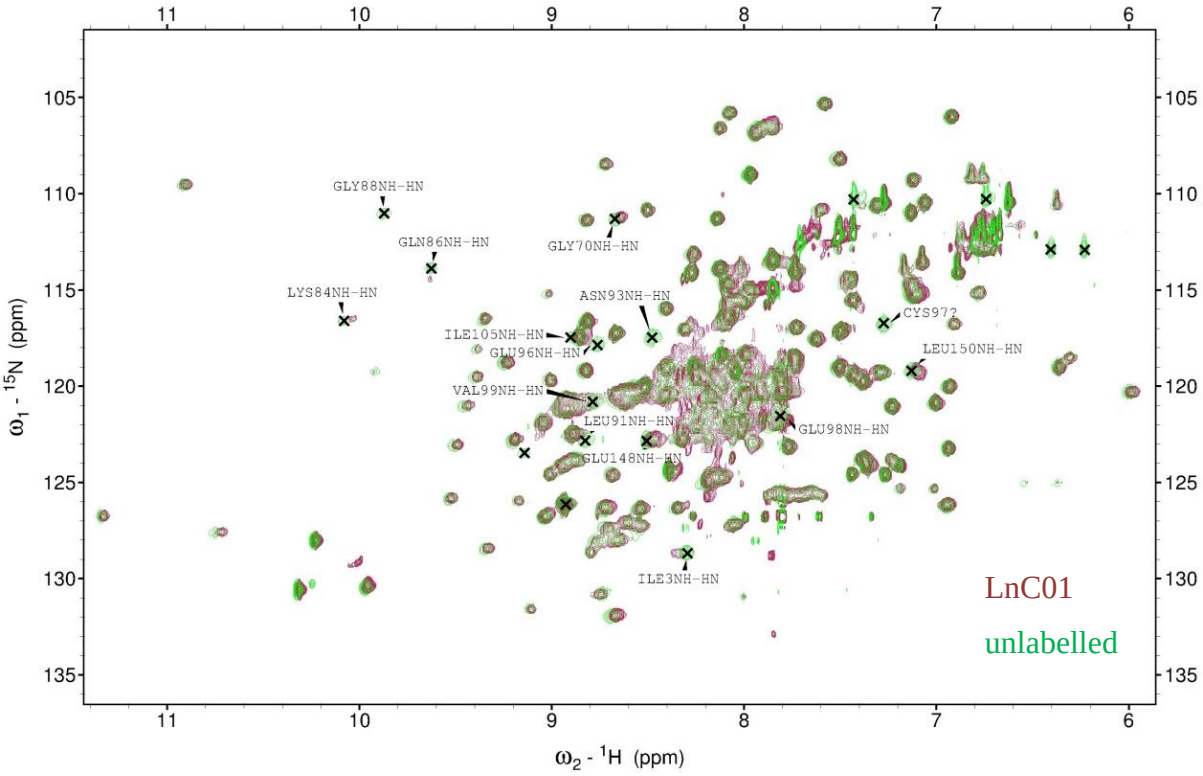


Figure 5.8. Overlay of ^{1}H-^{15}N-HSQC spectra of unlabelled AAC-diCys10 (*green*) and of LnC01-labelled AAC-diCys10 (*burgundy*). Peaks of AAC-diCys10 (tentative assignment based on AAC wild-type spectrum) that have been shifted due to the covalent attachment of **LnC01** are indicated on the spectrum.

Next, AAC-diCys10 was labelled with an excess of **LnC01**, **LnC01 Tb**, **LnC01 Dy** and **LnC01Y**, overnight at 4°C, as detailed in the Experimental section (page 143). An activity assay on **LnC01 Tb** – labelled protein showed an initial slope of 0.00050 AU s^{-1}

(comparable to 0.00045 AU s^{-1} for unlabelled AAC-diCys10). ^{1}H-^{15}N HSQC spectra were recorded at 37°C for each protein with saturating concentrations of AcCoA, to ensure that the protein is in its bound state that produces well resolved NMR spectra [185]. As shown in **Figure 5.8.,** spectra of unlabelled and **LnC01** labelled AAC-diCys10 are highly similar, with only a small number of peaks showing differences, many of which have been tentatively assigned to the helix containing cysteines 90 and 97, in accordance with a different chemical environment caused by a covalent modification. After labelling of AAC-diCys10 with a **LnC01 Ln** probe (Dy or Tb), a slight precipitation of protein sample was observed, along with a decrease of signal in the corresponding ^{1}H-^{15}N HSQC spectra (**Figure 5.9.**). Surprisingly, and similar to MBP-dC10, there are very few, if any, peak shifts, with spectra from Dy or Tb-labeled samples being virtually superimposable with that from the diamagnetic control (Y-labeled). Therefore no PCS appeared to be induced in the paramagnetic Ln-labeled samples.

It was previously determined that dipicolinic acid, despite being a great lanthanide chelator, does not occupy all nine coordination sites that a Ln^{3+} ion has [188]. The six free chelation positions are occupied by exchangeable water molecules, which may cause fluctuations in the lanthanide binding to the PDCA moiety with slight distance variations between the lanthanide and the protein, possibly inducing peak broadening in an NMR spectrum. To ensure that the fluctuations of the lanthanide due to solvation by water is as low as possible, we used two supplemental dipicolinic acid equivalents to modify a **LnC01 Tb** probe and obtained **LnC01 Tb $(PDCA)_2$**, as detailed in the Experimental section (page 145). Unfortunately, **LnC01 Tb $(PDCA)_2$** labelled protein did not show any further improvement in comparison to other previously used lanthanide containing **LnC01** probes (**Figure 5.9.**).

It was shown by other groups that PDCA – derived lanthanide probes often benefit from a neighbouring aspartate or glutamate residue on the surface of a protein, which helps to occupy the remaining Ln^{3+} coordination sites and to reduce the fluctuations of the lanthanide ion [183], allowing more substantial PCS to be observed. Despite using a similar approach with the **LnC01 Tb $(PDCA)_2$** probe, we were not able to achieve further improvement, as it was described by other groups.

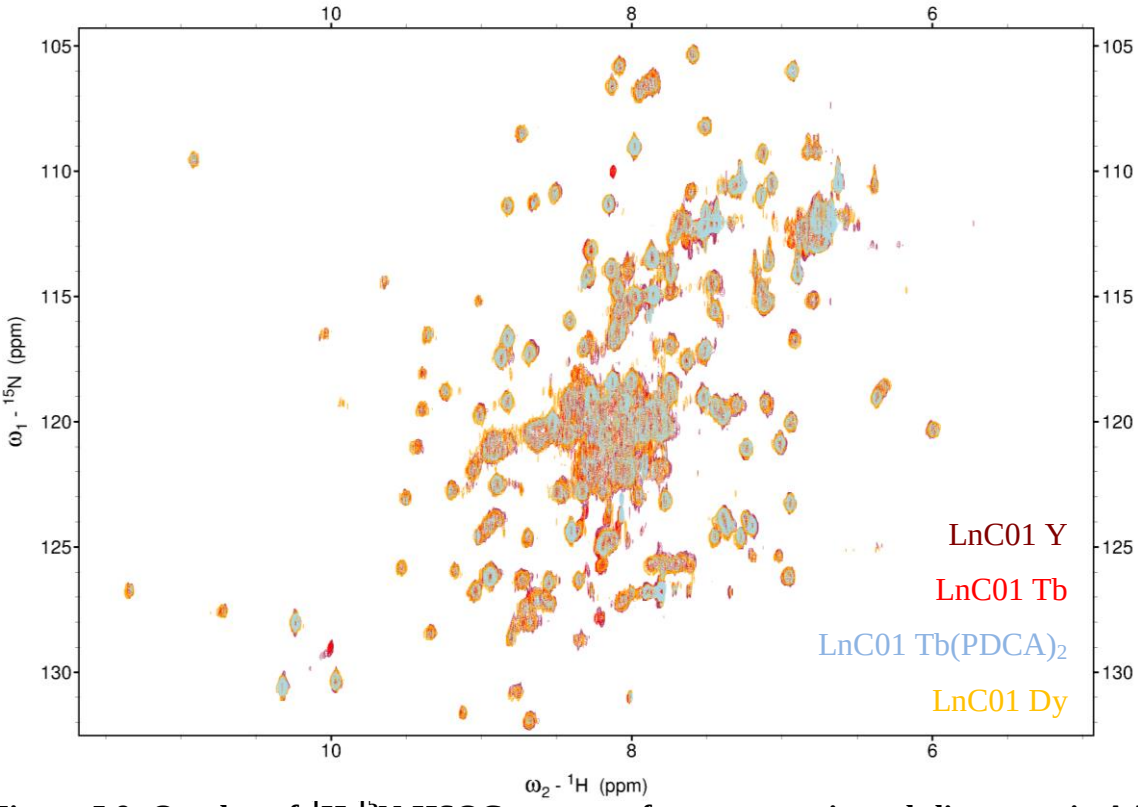


Figure 5.9. Overlay of ^{1}H-^{15}N HSQC spectra of paramagnetic and diamagnetic AAC-diCys10. Protein was labelled with paramagnetic probe **LnC01 Tb** (*red*), **Tb(PDCA)$_2$** (*light blue*), **Dy** (*orange*) and diamagnetic **LnC01 Y** (*maroon*) as a control. Contouring levels were adjusted according to protein concentration, receiver gain and number of transients.

To confirm that the AAC-diCys10 was labelled with LnC01, we turned to the luminescent properties of the PDCA Ln complex, where the PDCA plays the role of an antenna that absorbs light, and transfers it from its $s^3\pi\pi^*$ excited state to the lanthanide's $^5D_4$ excited state, which has a typical emission pattern corresponding to the different electron relaxation events in the *f* orbitals [189]. In particular, Eu and Tb complexes have life-times in the millisecond range, allowing an easy detection [190]. Using this technique, it was determined that the luminescence of **LnC01 Tb** and **LnC01 Tb (PDCA)$_2$** - labelled AAC-diCys10 is similar to the luminescence of free **LnC01 Tb** (**Figure 5.10.** *top*). Presence of robust luminescence signal of **LnC01 Tb** and **LnC01 Tb (PDCA)$_2$** bound to AAC-diCys10 suggests that the lanthanide was still attached to the probe. Nevertheless, the concentration-dependent increase in luminescence of PDCA, was strikingly larger than that for **LnC01** (**Figure 5.10.** *bottom*), suggesting that the photophysical properties of PDCA may have been

changed by the large amide bond-linked dimaleimide substituent that converts it into the **LnC01 Ln** probe. Similar observations were made by Lamture *et al.* [189], where they studied the effect of substituents on the *para* position of the pyridine ring on spectroscopic properties of PDCA Ln; and by Candelon *et al.* [191] who used an azido-substituted PDCA as a luminogenic probe. In the last cited case, the *para*-azido group caused a poor luminescent signal of PDCA probe that was recovered after transformation to the corresponding triazole.

Still, the lack of PCS could be caused by loss of lanthanide ion after labelling. The measured luminescent signal (**Figure 5.10.** top) of the Ln labelled protein could result from soluble aggregates that bind the lanthanide, which would lead to the decrease in NMR signal intensity. Another population of the protein could remain lanthanide-free and soluble, giving rise to the NMR signal. To test that hypothesis, high concentration of **LnC01 Tb** could be added to a control protein sample without dC10 or diCys10, and non-specific PCS or signal broadening of surface residues should be seen. This would validate the utility of **LnC01 Tb** as a PCS agent, and suggest that the sample handling after reaction renders the probe inactive. AAC wild-type protein was used for this purpose and a ^{1}H-^{15}N HSQC spectrum was acquired in the presence of an excess of the paramagnetic probe to see if any non-specific PCS can be triggered by the **LnC01** paramagnetic probe. A ^{15}N-labelled AAC wild-type was prepared in a similar manner as its diCys10 variant, and mixed with an excess of **LnC01 Tb** dissolved in DMSO for ^{1}H-^{15}N HSQC acquisition. Unfortunately, the **LnC01 Tb** probe was not soluble enough in water to be used in a large excess, and AAC-diCys10 precipitates in presence of larger amounts of DMSO. With these limitations, a stable AAC-diCys10 sample could be obtained with a maximum of 3 equivalents of **LnC01 Tb** probe in a solution containing no more than 5% (*v/v*) DMSO. Even in this case, neither non-specific nor specific PCSs were observed (spectrum not shown).

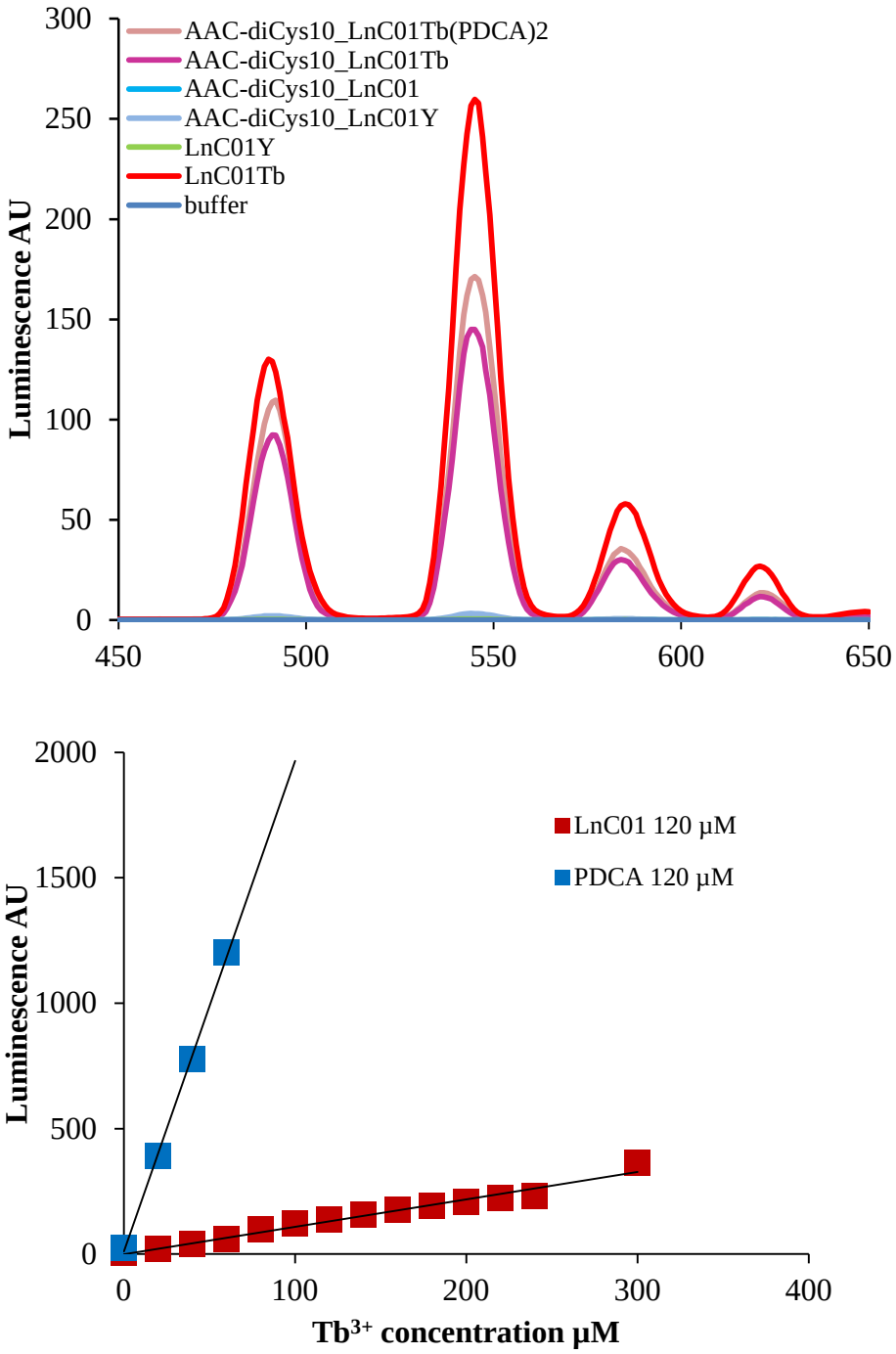
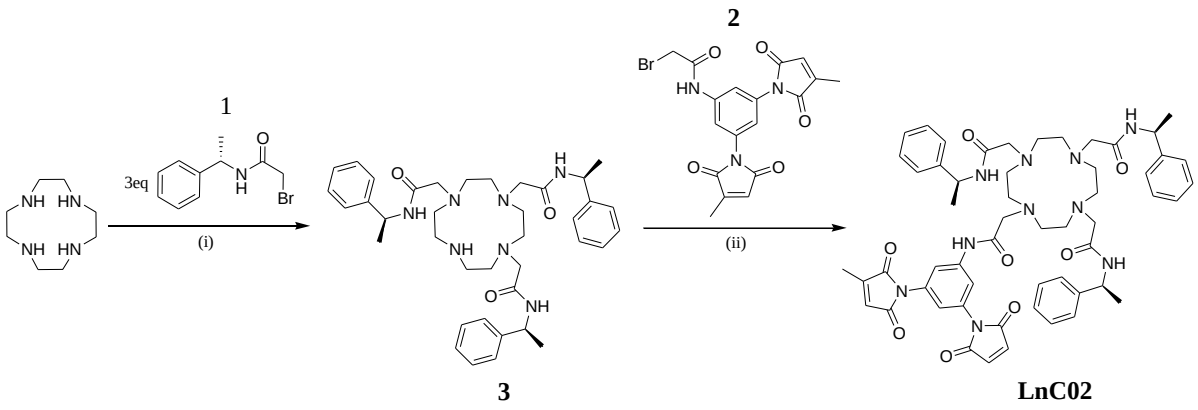


Figure 5.10. Luminescence of LnC01 Tb probe and it derivatives. *Top*: Luminescence properties of free **LnC01 Tb** or attached to AAC-diCys10 protein. Controls of LnC01-labelled protein (*turquoise*), LnC01Y (*green*), LnC01Y-labelled protein (*light blue*) and buffer (*dark blue*) are shown. Emission spectrum was recorded at 450-650 nm upon excitation at 263 nm. *Bottom*: Titration of LnC01 (*red*) and PDCA (*blue*) with Tb^{3+}. Excitation 263 nm, slit 20 nm, Emission 545 nm, slit 10 nm. Lines represent the best fit determined by linear regression.

Finally, it is worth mentioning that even though there are many examples of successful use of PDCA probes in literature, it was noted that in several cases, lanthanide chelating PDCA probes were detrimental for protein solubility and stability [192], similarly to what was observed with **LnC01** probe.

5.2.5. A new paramagnetic probe LnC02

Following the development of LnC01, a diverse selection of lanthanide probes was subsequently described, with a significant number being based on 1,4,7,10-tetraazacyclododecane (cyclen) derivatives containing carboxylic acid groups [177]. Complications arising from the formation of stereoisomers and to the excessive flexibility of the probe were overcome by using bulky chiral substituents [177]. Our group took advantage of this more advanced lanthanide chelator design and proposed a new dimaleimide lanthanide probe, **LnC02** (**Scheme 5.1.**). The synthesis of this new probe followed the previously published synthetic scheme [177]. Briefly, precursor **1** was prepared by the reaction of (*S*)-1-phenylethaneamine with bromoacetyl bromide where the secondary amine selectively displaced the acetyl bromide. A cyclen was then tri-substituted with precursor **1** that ensures proper chelation of a lanthanide, yielding intermediate **3**, after which a bromo-derivative of dimaleimide core **2** was attached to the fourth position on the tri-substituted cyclen, yielding **LnC02** (shown in **Scheme 5.2**.). Unlike Graham *et al*., we did not perform a separation of the tri-substituted cyclen **3** from cyclen substituted with four moieties of **1**, because separation by flash chromatography or by HPLC drastically decreased the isolated yield. Hence, **LnC02** contains a small quantity of cyclen tetra-substituted with **1** that does not contain a maleimide moiety and does not react with proteins containing the two Cys residues required for dimaleimide labeling, and can be therefore removed during the treatment of the labelled protein sample. The lanthanide complexes were prepared as published previously [177], where a mixture of **LnC02** and Ln^{3+} triflate salts (Ln^{3+} being Tb^{3+} or Tm^{3+}, or a non-lanthanide Y^{3+} as a negative control) was heated to reflux overnight in acetonitrile, and the resulting **LnC02 Ln** complex was isolated by precipitation with diethyl ether. All compound analyses agreed acceptably with published findings by Graham *et al.* [177].



Scheme 5.2. Synthesis of LnC02. Conditions: (i) DIPEA, chloroform, RT, o/n, (ii) DIPEA, acetonitrile, RT, 48 h. **1** was synthesized from commercially available material and **2** was synthesized from bromoacetyl bromide and dM10-aniline, prepared in the group by Kelvin Tsao.

Labelling of AAC-diCys10 with excess of **LnC02 Ln** was performed as in the case of **LnC01 Ln**, overnight at 4°C. In the case of **LnC02 Tb**, the sample had a slight pink shade, indicating that Tb was effectively present in solution. Recorded ^{1}H-^{15}N HSQC spectrum suggests that substantial amount of soluble aggregates were formed, leading to a great loss of resolution (**Figure 5.11.**). Alternatively, the observed signal could be attributed to PCS or PRE arising from different conformers, or to strong fluctuations on a microsecond millisecond time scale, which would lead to strong peak broadening. Moreover, a sample labelled with a diamagnetic probe **LnC02 Y** was not very soluble and the recorded spectrum was similar to the one of the paramagnetic sample (not shown). The fact that the paramagnetic spectrum suffered such an extreme loss of quality, including peak broadening and overall signal decrease, suggests that a two-point attachment of a larger hydrophobic molecule such as **LnC02** is detrimental to protein stability, and/or solubility.

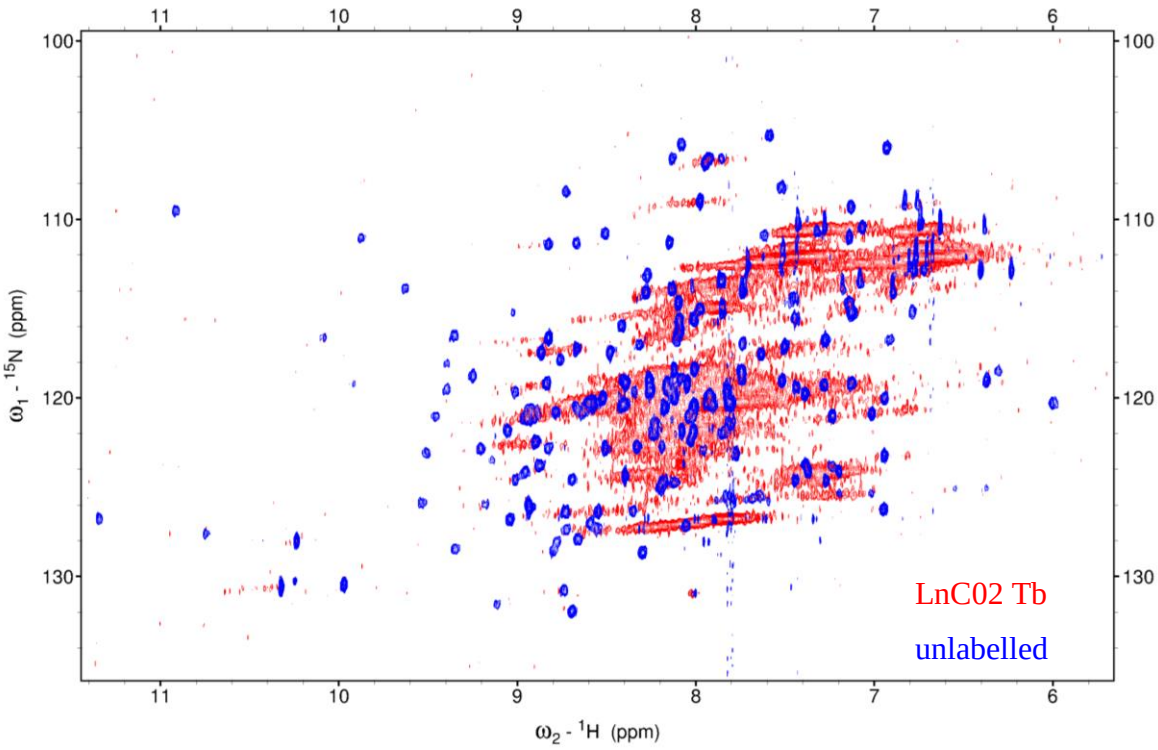


Figure 5.11. Overlay of ^{1}H-^{15}N HSQC spectra of AAC-diCys10 labelled with paramagnetic probe LnC02 Tb (*red*), and unlabelled AAC-diCys10 (*blue*). Contouring level of **LnC02 Tb**-labelled protein was adjusted because of a substantial loss of signal quality.

5.2.6. Testing other di-cysteine proteins

Paramagnetic probes, very similar to **LnC01** and **LnC02**, have been successfully used to induce large and unmistakeable PCSs [177, 179, 180]. However, in our hands, this does not seem to be working for dimaleimide probes that are otherwise identical to probes that have been shown to be effective PCS inducing agents. A common difference between our work and those of previous studies their use of a relatively *small* test protein (most commonly, Ubi, ArgN [177, 193, 166, 165]) in a modest concentration (70-200 µM [177]). Hence, it is possible that each of the protein systems that we have tested was made complicated by precipitation (as observed during the labelling process and sample concentration for NMR), and tag instability. From this time, we chose to re-examine other different diCys10-containing proteins available to us, such as MBP-dC10 and MBP with two cysteine

mutations on an intrinsic helix (S337C-R344C, referred to as MBP-diCys10), with **LnC02**. Similarly to AAC-diCys10, S337 and R344 residues of MBP were chosen as good candidates for mutation for their location on a solvent-exposed helix, allowing direct access for labelling with a dimaleimide probe.

First, we re-examined MBP-dC10 that was labelled with **LnC02 Tb**, **LnC02 Tm** and **LnC02 Y**. Yet again, labelled MBP-dC10 produced very different ^{1}H-^{15}N TROSY spectra, with slight peak broadening, in comparison to unlabelled MBP-dC10; however, there were only minor changes between MBP-dC10 labelled with a diamagnetic **LnC02 Y** and a paramagnetic **LnC02 Tb** probe (**Figure 5.12.**). Certain peaks of **LnC02 Y**-labelled protein spectrum were missing in the paramagnetic sample, indicating that they may have been broadened by PRE of the lanthanide beyond detection. Additionally, small shifts of **LnC02 Tb**-labelled protein were observed that could be attributed to weak PCS (indicated on **Figure 5.12.**).

To help to establish whether the small shifts in **Figure 5.12.** are attributable to PCS, a Tm-chelating LnC02 probe was tested with MBP-dC10, that should show opposite PCS than Tb due to the opposite signs of its magnetic susceptibility anisotropy tensor components, similarly to Eu (**Figure 5.1.**). Sadly, **LnC02 Tm** – labelled protein suffered from severe precipitation that did not allow a spectrum acquisition.

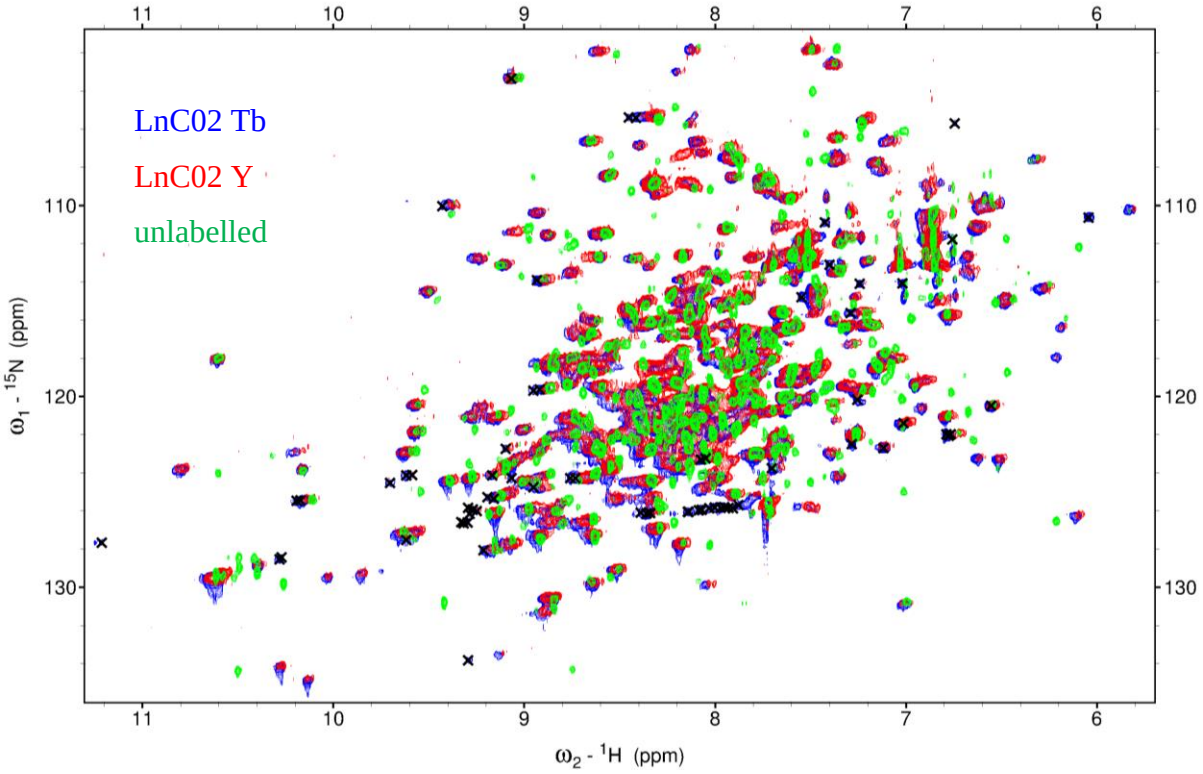


Figure 5.12. Overlay of LnC02 Tb - (*green*), LnC02 Y – labelled (*red*) and unlabelled (*purple*) MBP-dC10 ^{1}H-^{15}N TROSY spectra. Changes between **LnC02 Tb**- and **LnC02 Y**-labelled protein spectra are indicated with black x. Contouring levels were adjusted to reflect differences in protein concentration, receiver gain and number of transients.

MBP containing two Cys residues on a native helix (MBP-diCys10) (**Figure 5.13.** *right*), was first tested *in vitro* with **dM10-dansyl 1** to determine its kinetic characteristics (**Figure 5.13.** *left***)** before acquisition of the respective NMR spectra. MBP-diCys10 mutant was labelled with the fluorogen with a comparable rate to the dC10-tagged MBP, confirming that it is equally a suitable model for dimaleimide-based labelling. In NMR, unlabelled MBP-diCys10 produced a very well resolved spectrum, similarly to unlabelled MBP-dC10; however, after labelling with **LnC02 Tb**, the spectrum suffered from peak broadening, likely caused by PRE (**Figure 5.14.**). Some resonances seem to be affected more than others, suggesting that the broadening is caused by the lanthanide and is distance-dependent, as it is expected for a PRE effect. Alternatively, it could be also due to non-specific interactions of

the protein with the lanthanide or protein conformational exchange induced by the metal. Unfortunately, the peak broadening does not allow an easy assignment of the existing peaks, and even though there are most likely PCSs as well, the overall spectrum stays difficult to interpret and use as a proof-of principle for the method.

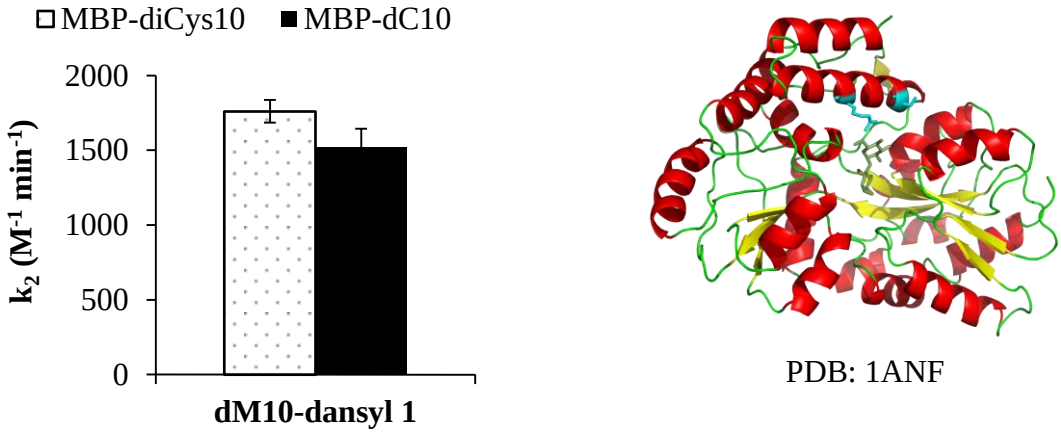


Figure 5.13. Second order rate constants for MBP-diCys10 and MBP-dC10 labelling (*left*), and structure of MBP-diCys10 (*right*). Labelling with **dM10-dansyl 1** was carried out at 20°C in 50 µM equimolar concentrations of protein and, in 50 mM HEPES pH 7.5, 1 mM TCEP. Fluorescence increase was followed at 530 nm upon excitation at 330 nm. *Right*: The S337C-R344C mutations are represented in *cyan* on MBP structure (PDB: 1ANF), along with bound maltose (*olive green*).

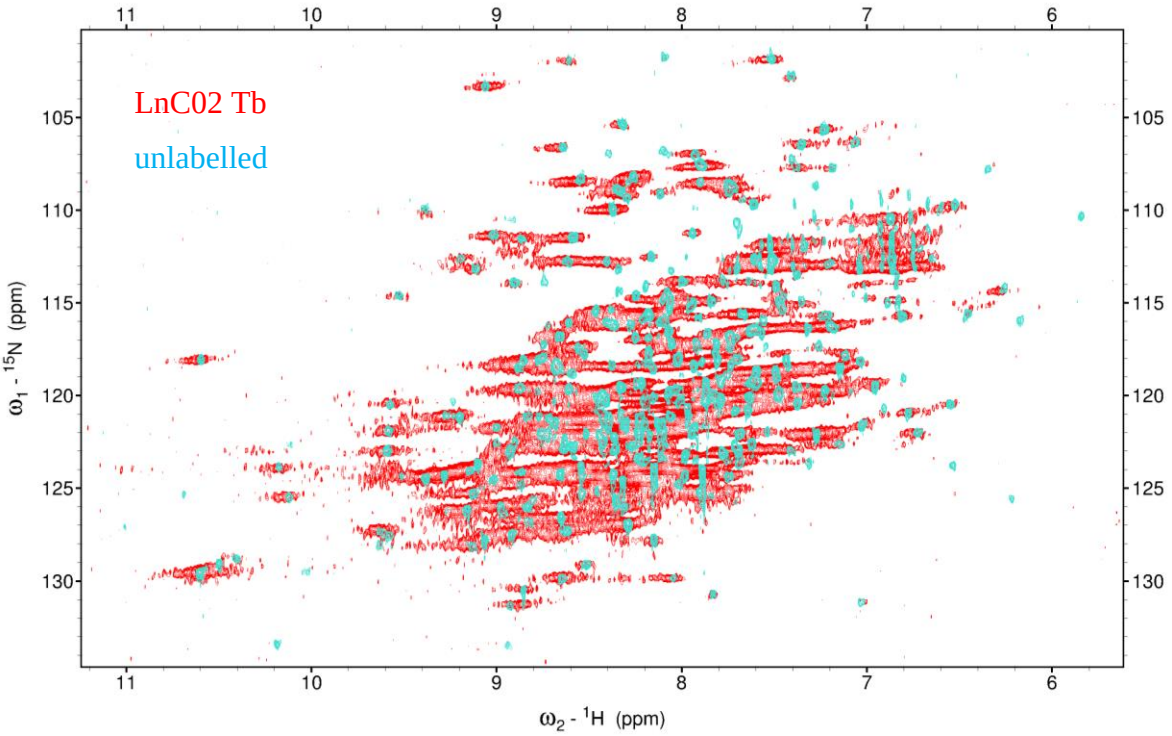


Figure 5.14. Overlay of LnC02 Tb – labelled (*red*) and unlabelled (*cyan*) MBP-diCys10 ^{1}H-^{15}N TROSY spectra. Contouring levels were adjusted to reflect differences in protein concentration, receiver gain and number of transients.

5.3. Conclusion

Two dimaleimide probes containing a lanthanide chelating moiety based on molecules described in the literature [177, 179, 180], **LnC01** and **LnC02**, were synthesized and chelated with different lanthanides (Tb^{3+}, Dy^{3+}, Tm^{3+} and Eu^{3+}) and with diamagnetic Y^{3+}. Several proteins of different sizes, attached to a dC10 peptide, or bearing two cysteine mutations in a solvent exposed helix, were used for dimaleimide-based labelling with paramagnetic probes. It was shown on the example of AAC-diCys10, that catalytic activity of an enzyme modified by two cysteine residues, and labelled with **LnC01 Tb** probe can be retained. It was also shown from chemical shift changes observed in ^{1}H-^{15}N-HSQC spectra that labelling with LnC01 was successful. Unfortunately, after labelling with paramagnetic probes, we were never able to observe consistent and interpretable pseudo-contact shifts that should have been induced by the close proximity of a lanthanide. This may have been a

result of protein stability issues, lower solubility of the probe, and of severe peak broadening as was in case of **LnC02 Tb** – labelled protein.

It is possible that highly concentrated proteins, as used in this project, became unstable after labelling, possibly forming soluble aggregates, and it was observed that even though unlabelled proteins were soluble in NMR-friendly concentrations, some cloudiness and precipitation appeared in labelled samples, and even more so after concentration to ~ 1 mM for NMR spectra acquisition.

It is also possible that the design of dimaleimide probes is not well suited for structural biology, unlike what we would presume for a more 'rigid' design, as the two sites of attachment between the dimaleimide moiety and the protein may impose unfavourable structural restraints that are too rigid to be easily accommodated by a canonical helix. Several successful cases, where a more flexible probe with two sites of attachment to a protein was used for NMR [194, 195], or as a photo-activatable conformational switch [196, 197, 198, 199], show that the general design of this labelling is sustainable; nevertheless, to our knowledge, a dimaleimide probe has never been reported in similar studies.

5.4. Perspectives

To address the hypothesis of destabilization due to attachment of an excessively rigid dimaleimide probe, we could synthesize mono-maleimide analogs of these paramagnetic probes and label corresponding mono-cysteine test proteins. This may address the question of detrimental structure destabilization, potentially caused by dimaleimide probes; however, it would not help the future design of paramagnetic probes, since our technology uses exclusively dimaleimide derivatives. It would rather represent an academic exercise and would help to justify the previous unsuccessful attempts.

In the future, it may be interesting to try different test proteins, for example smaller *highly* soluble proteins with two solvent exposed cysteine mutations, such as GB1 [200] that has been previously well characterized. It would also be beneficial to be able to use lower protein concentration, similar to concentrations used by other groups [177] to ensure stability

of the probe, and avoid potential degradation, protein aggregation or interference. However, in this thesis, a number of test proteins have already been explored with two very different Ln probes that did not prove to be very useful PCS agents for any of the proteins tested. Keeping that in mind, if one specific protein can be found to match one of our dimaleimide probes, it would be a part of a specific, rather than generally applicable, method.

Alternatively, repurposing the existing lanthanides probes for an entirely different application could be potentially useful. For instance, **LnC01** is a luminogenic probe and could be used as such for applications necessitating luminescence signal. In addition, both lanthanide probes could be used chelating gadolinium, for magnetic resonance imaging.

5.5. Experimental section

5.5.1. Cloning

All molecular biology supplies were purchased from New England Biolabs (PCR components, restriction enzymes), Clontech-Takara (nucleic acid purification kits), and all clones were verified by sequencing at Génome Québec at McGill University, Montréal.

AAC-diCys10 (R90C-K97C)

Mutations R90C and K97C were introduced in AAC using a single pair of mega-primers containing both mutations (**Table 5.2.**) in a rolling-circle PCR using the KOD-Xtreme™ Hot Start DNA Polymerase (EMD Millipore). After PCR, the methylated DNA was digested by *DpnI* for 1 hour at 37°C and purified by isopropanol precipitation. 1/100 of the total amount of DNA was used for transformation of DH5α cells and resulting clones were identified by Sanger sequencing.

MBP-diCys10 (S337C-R344C)

Dicysteine mutant of MBP was constructed using the side-overlap PCR approach where a 800 bp portion of *malE* gene of pMAL-c5x was amplified using pMAL_seqF and SCRC_bw primers, and another 300 bp portion was amplified using SCRC_fw and pMAL_seqR

primers (**Table 5.2.**). In both amplicons, the SCRC mutation was introduced with SCRC_fw and SCRC_bw primers. Both amplicons were used to a second round of PCR as mega-primers and templates to complete the synthesis of the complementary strand for the length of 1100 bp that was subsequently inserted into pMAL-c5x via *EcoRI* and *BglII* restriction sites.

Table 5.2. Oligonucleotides used for diCys10 mutants of AAC and MBP. Mutation positions are underlined.

AAC_RCKC_fw	5′ - CCAAATAGGTACTTGCTTAGTCAATTACTTAGAATGTGAAGTAGCTTCC - 3′
AAC_RCKC_bw	5′ - GCTACTTCACATTCTAAGTAATTGACTAAGCAAGTACCTATTTGGTTC - 3′
SCRC_fw	5′ - TCCCGCAGATGTGCGCTTTCTGGTATGCCGTGTGTACTGCGGTGATC - 3′
SCRC_bw	5′ - GATCACCGCAGTACACACGGCATACCAGAAAGCGCACATCTGCGGGA - 3′
pMAL_seqF	5′ - GGACAAGCTGTATCCGTTTAC - 3′
pMAL_seqR	5′ - TGTCCTACTCAGGAGAGCGTTCAC - 3′

5.5.2. Protein expression and purification

^{15}N-labelled MBP-dC10 or MBP-diCys10 was expressed in BL21-Gold(DE3) *E.coli* cells in M9 minimal media supplemented with 0.1% (*w/v*) $^{15}NH_4Cl$ and 100 µM of ampicillin. MBP-dC10 or MBP-diCys10 overexpression was induced by the addition of IPTG to a concentration of 0.3 mM and carried out at 37°C for 4-5 hours, after which cells were harvested by centrifugation at 3700 g for 10 minutes at 4°C. Purification of ^{15}N-labelled protein was carried out using an amylose resin, similarly to the unlabelled protein, as detailed in Chapter 2 (page 42).

^{15}N-labelled AAC or variants were expressed in BL21-Gold(DE3) cells in M9 minimal media supplemented with 0.1% (*w/v*) $^{15}NH_4Cl$ and 100 µM of ampicillin. AAC (or variant) overexpression was induced by addition of IPTG to a concentration of 1 mM and carried out at 28°C overnight, after which the cells were harvested by centrifugation for 20 minutes at 5000 g and 4°C. Cells were washed with 0.85% (*w/v*) NaCl and centrifuged again, and resuspended in buffer A (25 mM HEPES pH 7.4, 2 mM EDTA) supplemented with 200 mM

NaCl. Cells were lysed by sonication on ice and insoluble fraction was separated by centrifugation at 15 000 g for 10 minutes at 4°C. The supernatant fraction was filtered through a 0.45-µm syringe filter and loaded onto a HiPrep Q-Sepharose FF column pre-equilibrated with 25 mM HEPES pH 7.4, 2 mM EDTA, at 4°C. AAC (or variant) was eluted by a gradient of NaCl from 0 – 1 M in buffer A and was liberated from the column at around 60% of elution buffer. AAC containing samples were pooled and the buffer was changed to buffer A by dialysis, or using Amicon filters, at 4°C.

^{15}N-labelled Ubi-dC10 was expressed in BL21-Gold(DE3) cells in M9 minimal media supplemented with 0.1% (*w/v*) $^{15}NH_4Cl$ and 100 µM of ampicillin. Ubi-dC10 overexpression was induced by addition of IPTG to a concentration of 1 mM and carried out at 28°C overnight, after which the cells were harvested by centrifugation for 20 minutes at 5000 g and 4°C. Cells were resuspended in 100 mM sodium phosphate pH 8.0, 300 mM NaCl buffer supplemented with 1 mM PMSF, and lysed by sonication 3 x 1 minute. The insoluble fraction was separated by centrifugation at 15 000 g for 10 minutes at 4°C after which the soluble proteins were incubated for 2 hours at 4°C with NiNTA resin pre-equilibrated with PBS. Unbound proteins were flown-through and resin was washed with PBS containing 20 mM imidazole. Finally, Ubi-dC10 was eluted with PBS containing 250 mM imidazole.

5.5.3. Labelling with LnC01 and LnC02 paramagnetic probes

MBP-dC10 thiols were briefly reduced using 1 mM TCEP and the buffer was subsequently changed to 20 mM Tris-HCl pH 7.6, 1 mM EDTA to avoid interference of TCEP with dimaleimide moieties of the probe over extended periods of time. Immediately after, labelling with paramagnetic **LnC01 Ln** or **LnC02 Ln** probes (or their yttrium variants) was performed using 3-5 equivalents of probe with respect to the protein. Labelling reaction contained 70 µM of freshly reduced MBP-dC10, 3-5 equivalents of probe from a freshly prepared 4 mM stock solution in DMSO, 20 mM Tris-HCl pH 7.6, 1 mM EDTA and was carried out overnight at room temperature for **LnC01 Ln**, or at 4°C for **LnC02 Ln**, with gentle stirring.

Labelling of AAC-diCys10 and MBP-diCys10 was carried out using the same protocol. Protein thiols were briefly reduced using 1 mM TCEP and the buffer was subsequently

changed to 20 mM Tris-HCl pH 7.6, 1 mM EDTA to avoid interference of TCEP with dimaleimide moieties of the probe over extended periods of time. Immediately after, labelling with **LnC01** or **LnC02** (or their paramagnetic or yttrium variants) was performed using 3-8 equivalents of probe with respect to the protein. The labelling reaction contained 70 µM of freshly reduced protein, 3-8 equivalents of probe from a freshly prepared 4-mM stock solution in DMSO, 20 mM Tris-HCl pH 7.6, 1 mM EDTA and was carried out overnight at 4°C with gentle stirring.

5.5.4. AAC-diCys10 activity assay

AAC-diCys10 activity as an acetyl-transferase was verified using the previously published assay [186, 187], where the enzyme catalyzes the transfer of an acetyl group from AcCoA to an aminoglycoside, in this case kanamycin. The liberated CoASH reacts with 4,4'-dithiodipyridine to release 4-thiopyridine whose concentration increase is followed by its absorbance at 324 nm. The assay mixture contained 1.7 mM dithiodipyridine, 125 µM AcCoA, 25 µM kanamycin, 1 mM EDTA in 25 mM MES pH 6.0 and was pre-incubated at 37°C, and the reaction was initiated by addition of 5 µL of enzyme (or buffer in the case of the blank). The mixture was incubated at 37°C and the increase in absorbance of 4-thiopyridine (ε = 19800 $M^{-1}cm^{-1}$) was followed at 324 nm in BioTek Synergy H4 plate reader. Initial slopes were determined by linear regression on the linear portions of the obtained progression curves.

5.5.5. Luminescence measurements

Luminescence of LnC01 Tb probe and AAC-diCys10 labelled with LnC01 Tb was measured in 100 mM sodium phosphate pH 7.5, 2 mM EDTA buffer using Cary-Eclipse fluorimeter with previously determined optimal settings for terbium luminescence detection [201]: total decay 0.2 s, flash 1, delay 0.1 ms, gate time 2.0 ms. Luminescence spectra were acquired at 450 – 650 nm with a 10 nm slit upon excitation at 263 nm with a 20 nm slit. Averaging time was set to 0.1 s, date int. 1 nm with a power of 580 V. All measurements were carried out at 21°C.

Titration of LnC01 and PDCA with terbium was done using the settings detailed above with 120 µM of LnC01 (or PDCA) and by adding $TbCl_3 . 6\ H_2O$ from a 10-mM stock solution in water. For each data point, the solution was allowed 20 minutes for complex formation.

5.5.6. NMR sample preparation and spectra acquisition

All NMR spectra were acquired at 37°C, using the Varian INOVA 500-MHz spectrometer equipped with a HCN triple-resonance RT probe, or the Bruker AVANCE 500-MHz spectrometer equipped with an inverse probe, at University of Ottawa. Comparisons between different spectra were done exclusively between spectra acquired using the same spectrometer. The ^{1}H pulse length was adjusted to obtain a maximal signal and the ^{15}N pulse length was set to 33.0 or 34.0 µs. The number of increments in the indirect dimension was 64 with a spectral width of 2027.4 Hz. All spectra were processed using NMRPipe [148] and analysed with NMRDraw [148] and Sparky [149].

^{15}N-labelled MBP-dC10 or MBP-diCys10 and their paramagnetic variants were concentrated up to 0.7-1.5 mM in 20 mM sodium phosphate pH 7.2 buffer supplemented with 100 µM EDTA, Pefabloc, sodium azide and 9-10% (*v/v*) D_2O *[181]*. ^{1}H-^{15}N-TROSY spectra were acquired at 37°C instead of ^{1}H-^{15}N HSQC to obtain a better peak resolution, as it was previously shown for MBP *[202]*.

^{15}N-labelled AAC-diCys10 and its paramagnetic variants were concentrated up to 0.9-1.3 mM in 100 mM sodium phosphate pH 6.5, 2 mM EDTA, 5 mM AcCoA and 7-10% (*v/v*) D_2O [185].

5.5.7. Synthesis

LnC01 Tb and **LnC01 Eu** probes, and dM10-aniline intermediate were synthesized by Dr. Christophe Pardin and Kelvin Tsao, respectively (unpublished), and their synthesis is not detailed in this thesis.

All chemicals were purchased from Sigma Aldrich, unless specified otherwise, and were used without further purification. Cyclen was purchased from Toronto Research Chemicals, lanthanide and yttrium triflate salts were purchased from Alfa Aesar.

2-bromo-N-(-S)-1-phenylethyl)acetamide (1)

In a 250 mL round bottom flask, 8.14 mL (63 mmol, 2 eq) of (*S*)-1-phenylethanamine was added to 100 mL of anhydrous dichloromethane, under nitrogen, on ice. 2.74 mL (31.5 mmol, 1 eq) of bromoacetyl bromide was added dropwise to the reaction flask with gentle stirring, and maintained at 0°C for 10 minutes. The ice bath was removed and the reaction was allowed to completion for 2 hours at room temperature and under nitrogen. The mixture was washed with 50 mL of 2 N HCl, 50 mL of brine, and dried with $MgSO_4$. After filtration, the solvent was removed by rotatory evaporation to afford a crude white solid in 98% yield. Further purity was achieved by purifying the product by flash chromatography in dichloromethane after which a light fluffy white powder was obtained in 90% yield.

^{1}H NMR (400 MHz, $CDCl_3$): δ 7.35 - 7.24 (m, 5H), 6.80 (s, 1H), 5.10 - 5.03 (apparent p, J = 7.04 Hz, 1H), 3.86 - 3.78 (q, J = 13.56 Hz, 2H), 1.50 – 1.49 (d, J = 6.92 Hz, 3H)

ESI: *m/z* 264, 266 $[M+Na]^+$

2-bromo-N-dM10-acetamide (2)

In a 25-mL round-bottom flask, 199.5 mg (0.64 mmol, 1 eq) of dM10-aniline was dissolved in 5 mL of anhydrous dichloromethane. The mixture was cooled to 0°C in an ice bath and placed under a nitrogen atmosphere. A 10% (*v/v*) solution of bromoacetyl bromide in anhydrous dichloromethane was prepared and 0.55 mL (0.64 mmol, 1 eq) of this solution was added dropwise to the dM10-aniline on ice with stirring. After 10 minutes at 0°C, the ice bath was removed and the reaction was continued at room temperature overnight, under nitrogen. The solution was washed with 7.5 mL of 2 N HCl, 7.5 mL of brine, and dried on $MgSO_4$. After filtration, solvent was removed by rotatory evaporation and under vacuum. The crude product was a light yellow powder obtained in a 98% yield. For best results, the product was purified by flash chromatography in ethyl acetate to afford a light yellow solid in 80% overall yield.

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.39 (s, 1H), 7.60 - 7.60 (d, J = 1.52 Hz, 2H,), 7.27 (s, 1H), 6.47 – 6.46 (d, J = 1.52 Hz, 2H), 3.99 (s, 2H), 2.15 – 2.14 (d, J = 1.56 Hz, 6H)

^{13}C NMR (100 MHz, $CDCl_3$): δ 170.0 (2), 168.9 (2), 163.5, 146.0 (2), 137.9, 132.8 (2), 127.6 (2), 118.7, 115.6 (2), 29.3, 11.2 (2)

ESI: : *m/z* 454, 456 $[M+Na]^+$; 470, 472 $[M+K]^+$

m.p. 174 °C

2,2',2''-(1,4,7,10-Tetraazacyclododecane-1,4,7-triyl)tris-(N-((S)-1-phenylethyl)acetamide) (3)

173 mg (1 mmol, 1 eq) of 1,4,7,10 - tetraazacyclododecane (cyclen) were dissolved in 20 mL of chloroform (pre-dried over molecular sieves) under nitrogen, and 1.374 mL (8.08 mmol, 8.1 eq) of DIPEA was added dropwise. 750 mg (3.1 mmol, 3.1 eq) of **1** was dissolved in 20 mL of dried chloroform and dropwise added in the cyclen mixture, at room temperature and under nitrogen. The reaction was allowed to complete overnight at room temperature. After completion, the solvent was removed by rotatory evaporation to afford an off-white sticky solid in 90% crude yield. Where indicated, further purification of tri-substituted cyclen from tetra-substituted cyclen was achieved by semi-preparatory HPLC, otherwise the crude product was used as is in the next step for the synthesis of LnC02.

^{1}H NMR (400 MHz, $CDCl_3$): δ 8.27 – 8.21 (br s, 1H), 7.54 (br s, 2H), 7.19 – 6.89 (m, 15H), 4.90 – 4.80 (m, 3H), 2.89 – 2.77 (m, 6H), 2.32 – 2.22 (m, 16H), 1.26 – 1.20 (m, 9H)

ESI: *m/z* 656 $[M+H]^+$

LnC02

303 mg (0.46 mmol, 1 eq) of crude **3** was resuspended in 10 mL of anhydrous acetonitrile under nitrogen at room temperature. 800 µL (0.46 mmol, 1 eq) of 10% (*v/v*) solution of DIPEA in acetonitrile was added, after which 200 mg (0.46 mmol, 1 eq) of **2** was added. After 48 hours, the reaction reached completion and the solvent was removed by rotatory evaporation. The crude product was used for complexation with a lanthanide.

For analysis, the crude **LnC02** was dissolved in a minimum of methanol and precipitated by diethyl ether to afford a gray-pink solid in an overall yield of 61%.

^{1}H NMR (400 MHz, MeOD): δ 8.43 (s, 1H), 7.82 – 7.61 (br s, 3H), 7.36 – 7.00 (m, 18H), 6.55 (s, 2H), 4.99 – 4.88 (m, 3H), 3.65 – 3.34 (m, 8H), 2.50 – 1.65 (m, 16 H), 1.64 – 1.00 (m, 15H)

ESI: *m/z* 1007 $[M+H]^+$, 1029 $[M+Na]^+$

m.p. > 200°C

LnC02 Ln and LnC01 Y, Dy complexes

1 eq of crude LnC02 (or LnC01) was dissolved in 15 mL of anhydrous acetonitrile after which 1 equivalent of lanthanide triflate salt (or yttrium salt) was added and the mixture was heated at 80°C overnight. The solvent was removed by rotatory evaporation and the solid residue was resuspended in a minimum of methanol, and precipitated by diethyl ether to afford a light gray solid. A typical yield for LnC02 or LnC01 complexation with a lanthanide or yttrium was 60% and 75%, respectively.

Non-paramagnetic **LnC01 Y** allowed an easy quantification of the percentage of complexation of yttrium by the probe, using the pyridine aromatic NMR signals integration at 8.6597 ppm and 8.3858 ppm, to be 81%.

LnC01 Y: ^{1}H RMN (400 MHz, DMSO-d_6): δ 10.9944 (s, 1H), 8.6597 and 8.3858 (both s, 2H), 8.0140 (s, 2H), 7.6173 (s, 1H), 6.8536 (d, 2H, J = 1.76 Hz), 2.0748 (s, 3H), 2.0708 (s, 3H).

m.p. > 250°C

ESI: **LnC02 Tb**: *m/z* 582 $[M+2H]^{2+}$; **LnC02 Tm** : *m/z* 587$[M+2H]^{2+}$; **LnC02 Y** : *m/z* 547 $[M+2H]^{2+}$

LnC01 Tb (PDCA)$_2$

The **LnC01 Tb (PDCA)$_2$** complex was prepared by refluxing a 1:1 mixture of LnC01 and terbium triflate in anhydrous acetonitrile for 2 hours, after which two equivalents of pyridine-2,6-dicarboxylic acid were added and the mixture was refluxed overnight. The solvent was evaporated and the solid residue was dissolved in a minimum volume of methanol. The formed complex was isolated by precipitation with diethyl ether to afford an off-white solid in 59% yield.

Chapter 6

ALTERNATIVE USE OF DIMALEIMIDE-FUNCTIONALIZED MOLECULES FOR PROTEIN STRUCTURE STUDIES – X-RAY CRYSTALLOGRAPHY

6.1. Introduction

As detailed in Chapter 5 (page 115), 90% of currently known macromolecule structures were determined by X-ray diffraction of crystals, representing nearly 100 000 structures (*www.rcsb.org* consulted on June 30th 2015). Protein structures provide valuable information for a broad variety of research fields that use the structural data for studying protein-protein interactions, enzyme catalysis and allosteric regulation; they guide protein drug and probe design, as well as prediction of protein structure using sequence homology. While most crystal structures provide only a static representation of many possible dynamic conformations that one protein can adopt, there have been some advances in describing the protein conformational landscape using crystallography methods [203, 204, 205, 206]. Recently, this problem has been reviewed by Ronda *et al.* [207].

6.1.1. X-ray crystallography – the work flow

Even though for some systems crystallizing and solving a protein structure has become a routine operation, thanks to the automation of crystallization screens, rapid access to synchrotrons and simplification of structure solving process, it is always useful to understand the general process and problems that can arise, and to have an idea about methods that can help in solving them. Here, instead of detailing the whole process from acquiring a diffraction pattern to obtaining a protein structure (schematized on **Figure 6.1.**), we will focus on one aspect of structure solving that crystallographers inevitably deal with, and the window of opportunity that it represents for the labelling technique developed in the Keillor group.

In the diffraction experiment, an X-ray beam is used to create a diffraction pattern after coming across a crystal of sample of interest. This diffraction pattern is a unique representation of the crystallized molecule at an atomic level. A set of diffraction patterns obtained from differently rotated samples is used to build an electron density map of the macromolecule, as expressed by **Equation 6.1.** In other words, an electron density map ρ (**Equation 6.1.**) is the result of sum of individual contributions to a point *(xyz)* of waves resulting from diffraction from crystallography planes *(hkl).* The amplitude of these waves depends on the number of electrons encountered in the plane and on the phase. Amplitudes

can be measured from the diffraction pattern; however, phases have to be determined independently. This is what represents the phase problem.

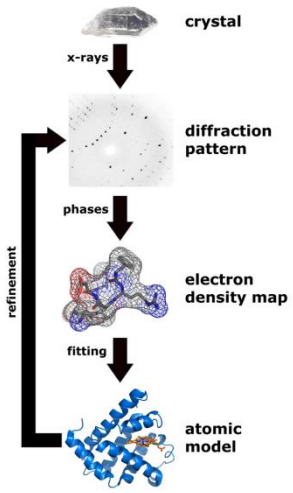


Figure 6.1. General work flow for solving a crystal structure.
Copyright Thomas Splettstoesser (*www.scistyle.com*)

$$\rho(xyz) = 1/V \sum |F_{hkl}| \ \exp(i\alpha_{hkl}) \exp[-2\pi i(hx + ky + lz)]$$

Equation 6.1. Electron density equation. ρ is the electron density, V is the volume of the unit cell, *hkl* are the crystallography planes, α_{hkl} is the phase and $|F_{hkl}|$ is the structure-factor amplitude. The amplitudes can be measured; however the phase has to be determined separately.

Phases have an outstanding importance in the structure determination, as cleverly illustrated in the Book of Fourier Transforms by Kevin Cowtan (**Figure 6.2.**), where amplitudes of a Fourier transform of object A (cat), mixed with phases of transformed object B (duck), resulted, after reverse transformation exclusively in object B (duck).

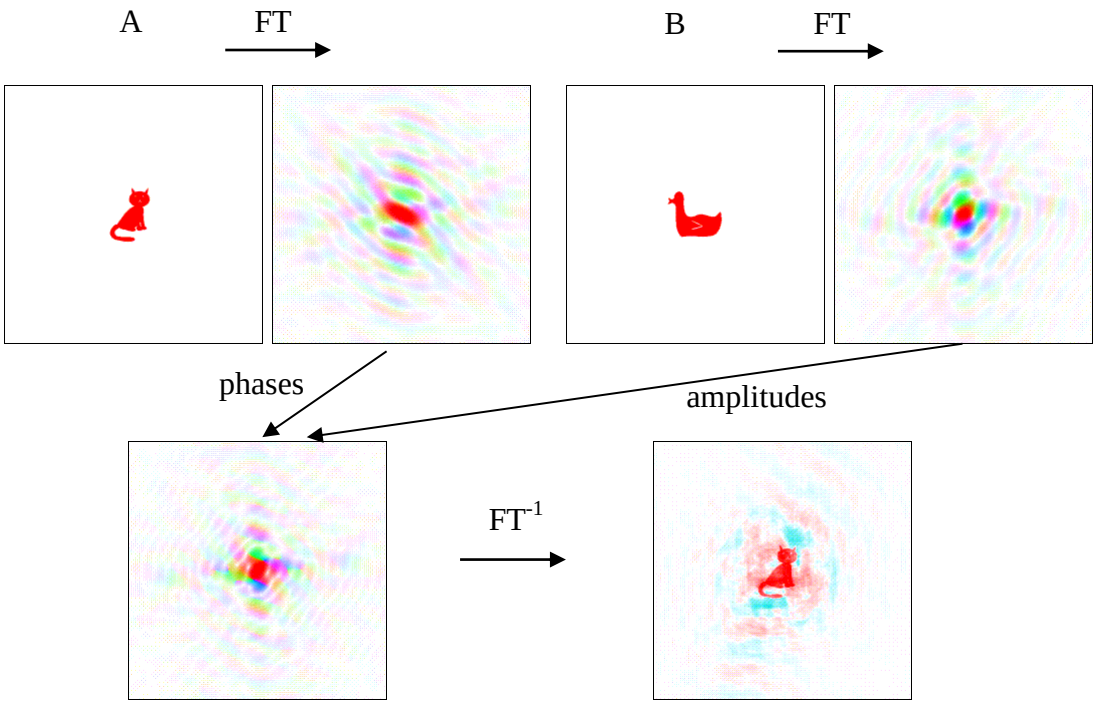


Figure 6.2. Illustration of the importance of phases. *Top*: the diffraction pattern (Fourier transform - FT) of objects A (cat) and B (duck). *Bottom left*: Diffraction pattern derived from the phases of the diffraction pattern of A and amplitudes of the diffraction pattern of B. *Bottom right*: Image that would give rise to this mixed diffraction pattern. Pictures are adapted with permission from the Book of Fourier Transforms by Professor Kevin Cowtan (*http://www.ysbl.york.ac.uk/~cowtan/fourier/fourier.html*, consulted July 30th 2015).

A considerable number of approaches to tackle the phase problem have emerged and they were compiled and reviewed by Taylor [208, 209], and Cowtan [210]. In the next section, we will briefly describe and compare the existing methods and then present a design of the dimaleimide-based labelling as another potential solution to the phase determination.

6.1.2. Solutions to the phase problem

Over the years, many solving approaches arose, ranging from direct methods, to molecular and isomorphous replacement. Direct methods are suitable for small molecules up to 2000 atoms and therefore not very useful for proteins.

Molecular replacement

Molecular replacement is used for structures that have a certain degree of similarity to another structure, for example, for a protein, a sequence homology of > 30% is generally required, as well as structural similarity and resolution [211]. The term "molecular replacement" was first used by M. G. Rossmann in a review [212] that compiled the existing articles to that date where this approach was used [213]. G. Scapin [211] reviewed molecular replacement and its importance in phase problem solving, underlining the fact that molecular replacement is a preferred method of choice for more than 70% of protein crystal structures [211]. Briefly, a search model (structure or fragment previously determined in high resolution) is fitted in the structural elements of the studied molecule, and a transformation to the model is applied in order to obtain a perfect correspondence with the diffraction data. This transformation includes six independent variables corresponding to a translational transformation (three coordinates) and rotational transformation (three more coordinates), and can have different features, such as Patterson [214, 215] or Crowther [216] rotation functions. One of the early applications of molecular replacement and Patterson refinement was the determination of the structure of fragment antigen-binding (Fab) proteins using five different Fab molecules as search models [217]. Since then, more sophisticated phased rotation and translation function were implemented in the cutting-edge molecular replacement programs MOLREP [218, 219] and *Phaser* [220] that are implemented in most recent and used software suites.

Isomorphous replacement

A frequently used method to solve the phase problem is the isomorphous replacement, consisting in introducing a reactive group containing a heavy metal in the protein structure that causes a measurable change in the diffraction pattern without disturbing the protein structure. The two structures, with and without the heavy metal, should be as similar as possible; for that reason this method is referred to as *isomorphous*. Initially, such changes were achieved by soaking the protein crystal in heavy metal salt solution, creating potentially one or multiple metal binding sites on the protein [221, 222].

A single isomorphous replacement (SIR) limits the stoichiometry of heavy metal to protein to 1 : 1. This is usually achieved by a specific or covalent modification of the protein by a heavy metal containing probe. Comparison between two isomorphous diffraction patterns leads to a close estimation of a phase. However, usually, multiple crystals of singly-modified proteins bearing different heavy metals at different positions, giving a "multiple isomorphous replacement" (MIR), are necessary in order to obtain a more precise phase. An alternative to using MIR is to combine a single isomorphous experiment SIR with a subsequent anomalous scattering (SIRAS).

Isomorphous replacement quite often suffers from multiple problems related to the fact that the two systems are not exactly isomorphous in the protein structure, solvation by water and ions, and orientation in unit cell. This can be overcome by using multi-wavelength anomalous scattering (MAD) where the diffraction data are collected at several wavelengths.

6.1.3. Methods for isomorphous incorporation of heavy metals

Non-covalent modification

Over the years, multiple different methods have been employed to incorporate an electron-rich, X-ray scattering atom in a protein crystal, ranging from low specificity binding of small ions to highly specific ligand analogue binding, or to protein covalent modifications. At first, halide anions [223], mono- or divalent cations [224, 225] or noble gases [226] were used for a rapid cryosoaking of protein crystals. Alternatively, quick soaking in solutions of heavy metal salts, such as Au, Pt, or Hg, provided a satisfying degree of derivatization [227, 228]. Additional more strongly binding molecules containing iodine or bromine, such as triiodoisophthalic and tribromoisophthalic acid derivatives, were synthesized and used as a more potent source of signal scattering [229, 230]. Binding of small anions and cations to proteins is problematic to predict, with the exception of metal-binding proteins, hence, an attempt to rationalize the choice of heavy metal compounds for protein crystallization was led by Agniswamy *et al.* [231] (and similar work was done by Joyce *et al*. [232]), where over 40 compounds of interest for crystallography were screened in different pH and buffer conditions with select peptides. Twenty-one compounds were identified as being most promising and consistent and have been compiled in heavy atom reactivity tables.

A more interesting design was proposed by Fütterer *et al*. [233], generalizable for the case of protein-fatty acyl complexes, where they synthesized a non-hydrolysable analogue of myristoyl-CoA presenting an electron rich iodine atom at the end of the fatty acid chain. They used this analogue, along with a selenomethionine substitution, for crystallization and structure solution of the *N*-myristoyl transferase Nmt1p of *S. cerevisiae*. Similarly, iodine-substituted carbohydrate analogs were used for crystallization and solution of the structure of maltodextrin (maltose)-binding protein [234]. More recently, selenium oligonucleotides were shown to be useful for DNA- or RNA-binding protein structure determination, as shown on the proof-of-principle example of RNase H [235].

Covalent modification

Even though the number of cases where weakly bound heavy metals was successfully incorporated in a crystal seems to be high, it is without a doubt tempting to incorporate an electron rich atom in a protein in a covalent and fully controlled fashion. Such an attempt was made by Xie *et al.* [236], using the well-established unnatural amino acid incorporation method to integrate *p*-iodo-*L*-phenylalanine in the bacteriophage T4 lysozyme. With a simpler methodology, the incorporation of selenomethionine had been achieved and used in methionine-rich proteins [237, 238]. Similarly, aza-tryptophan was used for the unprecedented crystallization of bacteriophage λ lysozyme, presenting a potential site for derivatization by a heavy metal atom [239].

A heavy metal probe design was attempted by Purdy *et al*. [192] where a lanthanide chelating probe was incorporated into a protein via a single covalent attachment on a cysteine residue. Despite some minor success, for many of the proteins tested, the lack of solubility and difficulty in crystal formation suggested that the method may not be well suited for high-throughput crystallography.

6.1.4. General approach of our method

The non-generalized methods for incorporation of a heavy atom in a protein crystal represents a window of opportunity for the dimaleimide-based protein labelling technology, as the labelling reaction kinetics and conditions have been broadly studied. Moreover, using a two-point-attachment labelling may bring more rigidity to the system, allowing a more isomorphous positioning of the heavy metal, which is crucial to obtain a high quality diffraction pattern. A dimaleimide-derived palladium containing probe **dM10-Pd** was designed (**Figure 6.3.**) and synthesized for this purpose by Dr. Christophe Pardin, taking inspiration from previous work by McNamara *et al.* [240]. The proposed scaffold was chosen for the ease of synthesis and facile attachment to a dimaleimide moiety, and also because it is symmetrical and may not lead to potential stereoisomers after reacting with a di-cysteine helix. Initially, we proposed to use a previously crystallized and well-studied protein [184] where two cysteine residues will be introduced on a solvent exposed helix. The residues were carefully chosen so that their modification does not interfere with formation of protein oligomers [184], crystallization and with protein dynamics [185]. In a first phase, the protein will be expressed and labelled with **dM10-Pd** and crystallized. After that, in a complementary fashion to *in vitro* labelling and similarly to work published by Sun *et al.* [227, 228], we will also attempt to crystallize the unlabelled protein and then label the crystal with **dM10-Pd**.

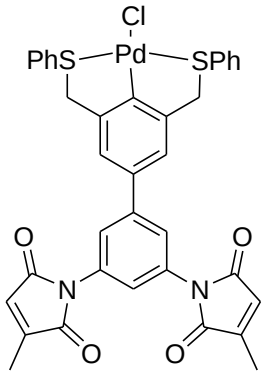


Figure 6.3. dM10-Pd (Dr. Christophe Pardin)

Here we will apply the dimaleimide labelling technique to introduce a palladium atom into a protein in covalent manner, to obtain an easy and user-friendly method for solving the phase problem. Two test proteins are used for this purpose: a protein bearing an intrinsic di-cysteine helix, AAC-diCys10 (see Chapter 5, page 124), and a dC10-tagged maltose-binding protein, MBP-dC10, used for FlARe *in vitro* testing (see Chapter 2, page 24).

Size-exclusion chromatography and crystallization screens were performed in the laboratory of Prof. Albert M. Berghuis (McGill University), in collaboration with Michelle McEvoy, and later with Jonathan Blanchet.

6.2. Crystallization of AAC-diCys10 labelled with dM10-Pd probe

6.2.1. Purification of AAC and variants and labelling with dM10-Pd

AAC-diCys10 was expressed and purified by anion exchange chromatography as detailed in Chapter 5 (page 142), with the only difference being that the expression was carried out in 1 L of LB medium instead of in minimal medium. Typically, a 40 mg yield of pure protein was obtained according to a Bradford assay. AAC-diCys10 was labelled with the **dM10-Pd** probe *in vitro*, before a final purification step by size-exclusion chromatography as described in the Experimental section (page 167). Labelling by **dM10-Pd** was confirmed by MALDI where a mass shift satisfyingly close to the expected value was observed (**Figure 6.4.** *left*). Sadly, the Pd-labelled protein showed a tendency to precipitate, which is a limiting factor for its handling and stability. The elution profile from size-exclusion chromatography (**Figure 6.4.** *right*) showed that AAC-diCys10 and wild-type retention times are similar (indicated by a # in **Figure 6.3.**), indicating that both proteins are in a similar oligomeric state, as it was previously studied in detail for the wild-type protein by Freiburger *et al.* [185]. Most of the Pd-labelled AAC-diCys10, however, is eluted much faster than either wild-type or unlabelled protein, suggesting that a higher oligomeric state, or possibly soluble aggregates were formed after labelling. For different batches of AAC-diCys10 labelled with **dM10-Pd**, the ratio of the first and second peak (peak shoulder) varied slightly. We collected fractions corresponding to the main peak, and separately fractions corresponding to the minor peak

(peak shoulder) of similar retention time as AAC-diCys10 or AAC wild-type, and used both separately for crystallization screens.

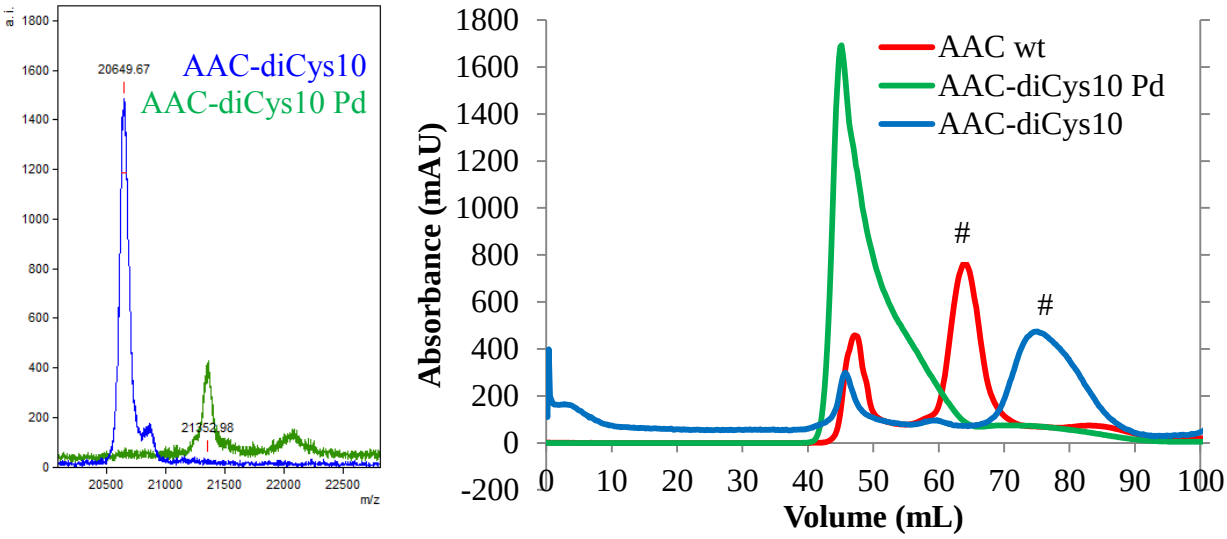


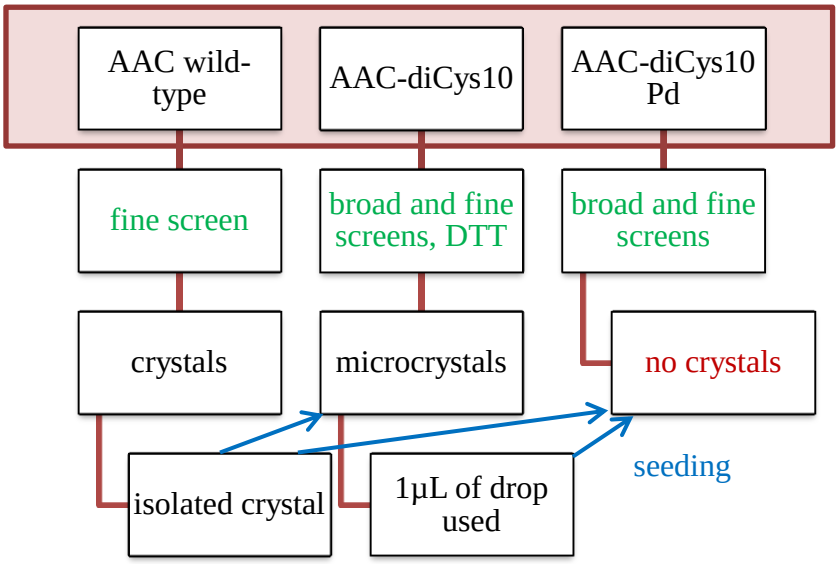
Figure 6.4. Labelling and purification of AAC variants. *Left*: MALDI analysis of unlabelled and labelled AAC-diCys10. *Right*: Purification of AAC wild-type (*red*), unlabelled (*blue*) and labelled (*green*) AAC-diCys10 by size exclusion chromatography. # denotes elution peaks of AAC wild-type and AAC diCys10.

6.2.2. Crystallization screens for AAC and variants

The crystallization of wild-type AAC was successfully performed previously in the laboratory of Professor Albert M. Berghuis at McGill University. Conditions previously established for wild-type AAC [184] were therefore used in the Berghuis laboratory for the preparation, storage and crystallisation of the AAC-diCys10 Pd protein. First, a series of Index, JCSG Core I-IV, $AmSO_4$ and PEG II broad screens were performed at 22°C and 4°C at protein concentrations of 7 mg/mL and 10 mg/mL and in the presence of CoASH as a ligand. Unfortunately, none of the screened conditions gave potentially promising conditions for crystallization, as the protein precipitated in large part. We decided to crystallize wild-type AAC and unlabelled AAC-diCys10 in parallel with the Pd-labelled protein, to

determine whether the instrumentation we used could have an impact on AAC crystallization.

New broad screens JCSG+, Cryo I-II and PEGs I at 22°C were thus prepared where both wild-type and Pd-labelled protein were set up in a sitting drop configuration in the same well in order to use exactly same conditions for both proteins. A fine screen of ammonium sulphate ($AmSO_4$) in a hanging drop configuration was prepared as well where $AmSO_4$ concentration and pH were varied between 1.6 and 2.6 M, and 5.0 - 6.3, respectively (for details, see Experimental section, page 170). The overall approach and work-flow are summarized in **Scheme 6.1**.



Scheme 6.1. Workflow for attempted AAC-diCys10 Pd crystallization. Crystallization conditions (*green*) were screened for three constructs (*in red rectangle*) and crystals or microcrystals of AAC wild-type or double mutant were used for seeding (*blue*) of Pd-labelled protein.

Fine needle-like crystals were obtained for wild-type protein in an $AmSO_4$ fine screen, in the presence of 2.2 and 2.4 M $AmSO_4$, pH 6.0 (**Figure 6.5.** *top*). One crystal of wild-type was isolated and used for seeding of AAC-diCys10 and AAC-diCys10 Pd to promote their crystallization in fine screens of $AmSO_4$ and the broad screens of JCSG Core I-IV, Opti-Salt and Index screen, as detailed in the following text. AAC-diCys10 contains two solvent exposed cysteine residues that may potentially hinder the formation of a correct crystal by

forming disulfide bridges; hence, we examined crystallization conditions in the presence and in absence of a reducing agent DTT in broad screens (JCSG Core I-IV). Microcrystals of unlabelled protein were obtained in conditions 100 mM CHES, 200 mM NaCl, pH 9.5, 1.6 M $AmSO_4$ (JCSG II – A2), both with and without 4 mM DTT (**Figure 6.5.** *bottom*, panels *B* and *C*); however, AAC-diCys10 Pd did not show any promising improvement in forming crystals by seeding in any of fine screen or broad screen conditions tested.

Next, seeding by AAC-diCys10 microcrystals was attempted, where a volume of 1 µL of AAC-diCys10 microcrystals was used for preparation of a seeding solution and used for screening Pd-labelled protein crystallization conditions in a fine screen using conditions close to the ones where AAC-diCys10 microcrystals were obtained, and in broad screens. Unfortunately, no Pd-labelled AAC-diCys10 crystals were obtained by seeding with a wild-type crystal, or AAC-diCys10 microcrystals.

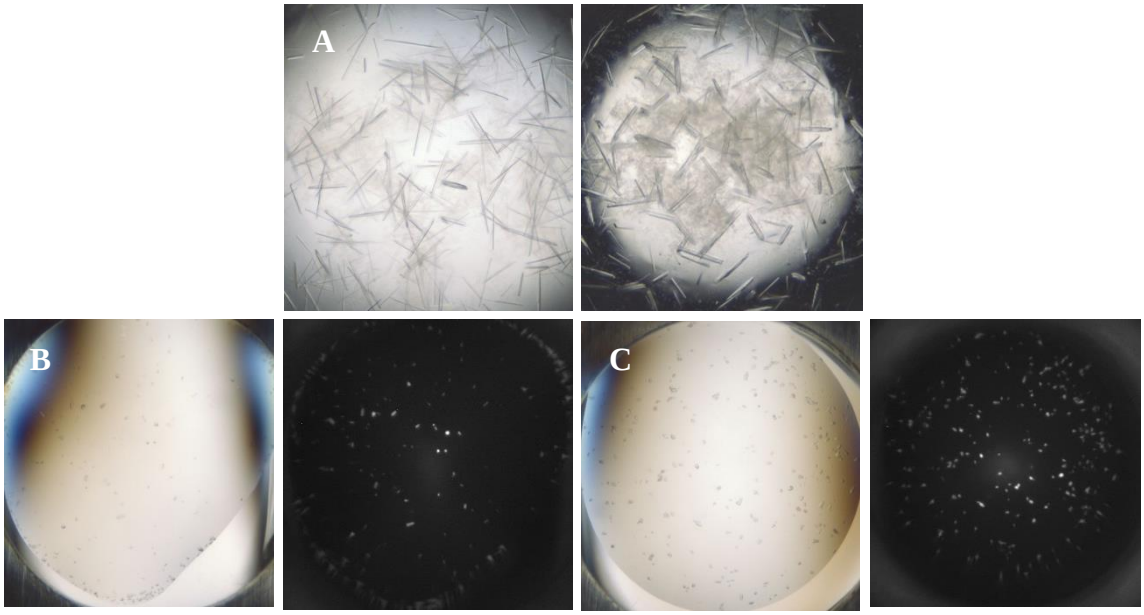


Figure 6.5. Crystals of AAC wild-type (*top*) and microcrystals of AAC-diCys10 (*bottom*). Crystals were obtained in $AmSO_4$ fine screen (*panel A*) and in JCSG Core II broad screen with and without DTT (*panels B, C*, respectively), both at 22°C. Crystallization drops were imaged in bright field (*panel A* and *panels B, C left*) and under UV (*panels B, C right*).

It seems that, contrary to the design and purpose of dM10-Pd probe, instead of promoting protein crystallization and helping to solve a problem inherent to analyzing protein crystals, **dM10-Pd** is not suitable for the crystallization of AAC-diCys10 in any way. One may hypothesize that the **dM10-Pd** probe hinders protein solubility and promotes aggregation, as suggested by the elution profile of the size-exclusion chromatography, and as observed for several other dimaleimide-based probes (see Chapter 5). While it was possible to obtain microcrystals of the AAC-diCys10, both in reducing and non-reducing conditions, we have never been able to achieve a better crystallization of this double mutant either. The solubility problem of a labelled test protein could be circumvented by using a more soluble protein from the start, which will be described in the next section.

6.3.“New” test protein for crystallization - MBP-dC10

Since AAC-diCys10 did not seem to be soluble and stable enough after labelling with **dM10-Pd**, which probably led to the impossibility of obtaining a crystal, we decided to switch to a potentially more soluble test protein. For this application, we turned our attention back to MBP-dC10, which has been well characterized in our laboratory for labelling purposes. Also, its crystal structure has been well described, both in *apo*-form and bound to different carbohydrates [234, 241, 242]. Unlike for protein NMR applications, where the observation of PCS is distance-dependent (page 116) and the point of attachment of the probe on the protein is crucial for a suitable PCS detection, in crystallography, the relative localization of the heavy metal with respect to the protein core is somewhat less critical, as long as it is homogeneously conserved across the whole crystal. Therefore, it is, in theory, possible to opt for a dC10-tagged test protein, such as MBP-dC10, instead of a di-cysteine mutant AAC-diCys10.

6.3.1. Preparation of Palladium-labelled MBP-dC10

MBP-dC10 was expressed and purified as detailed previously, and labelled with **dM10-Pd** as detailed in the Experimental section (page 167). The formation of **dM10-Pd** - labelled protein was verified by MS. MBP-dC10 Pd was subsequently purified on an anion-exchange column as suggested by previous work to enhance the capacity of crystal formation [234],

and on a size-exclusion column to remove any insoluble aggregates (**Figure 6.6.** *top*). Several fractions of size-exclusion column were analyzed on a native gel to test if the protein is folded or if it forms aggregates, where the MBP-dC10 Pd migrated close to correctly folded Arr protein (49 kDa) used as a reference (**Figure 6.6.** *bottom*), suggesting that MBP-dC10 Pd remained soluble and in a monomeric state.

6.3.2. MBP-dC10 Pd crystallization

Broad screen conditions were set for crystallization of 10 mg/mL and 17 mg/mL MBP-dC10 Pd, both in *apo*-form and in presence of 10 mM of its ligand maltose [234]. Screening conditions included Index screen, Classics, JCSG+, all three at 22°C and 4°C, and JCSG Core I-IV, as detailed in the Experimental section (page 170). After several days of incubation most wells where 10 mg/mL protein was used remained clear, indicating that the protein concentration may have not been high enough to cause a precipitation or crystallization. Interestingly, one well showed presence of crystals shortly after the beginning of incubation at 22°C (**Figure 6.7.**); however, these crystals dissolved before they could be characterized or used for seeding.

Unfortunately, at this stage of the project, we ran out of time before we were able to obtain a larger and stable crystal of MBP-dC10 Pd. Though the crystalline form obtained from JCSG IV screen is small and dissolved within a short period of time, it gives an indication that MBP-dC10 Pd can potentially be crystallized and it is a positive evolution for applicability of **dM10-Pd** probe in crystallography.

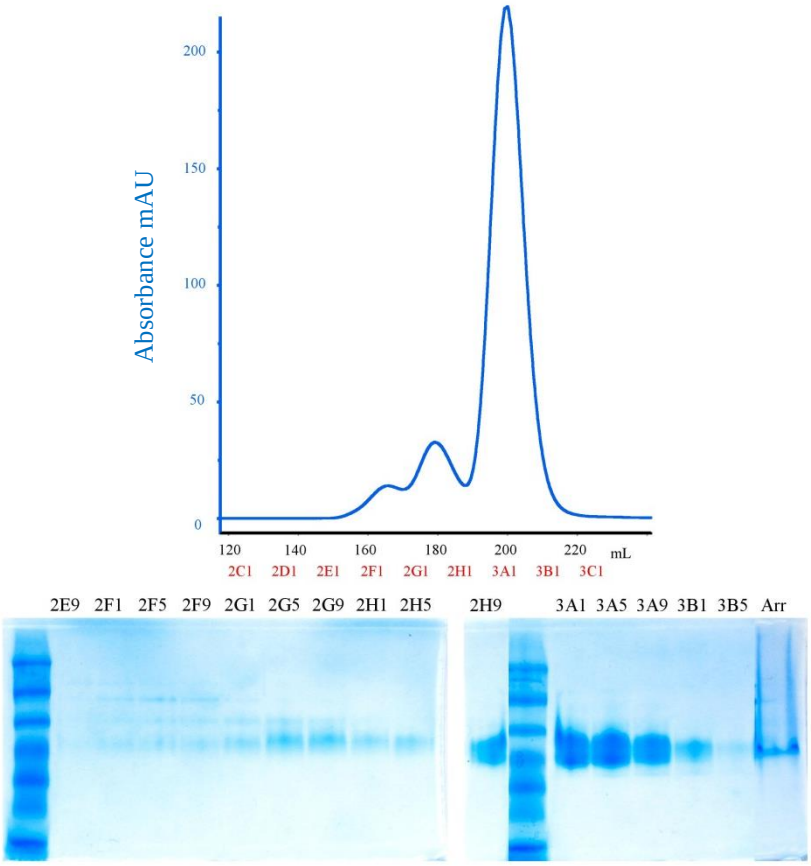


Figure 6.6. Purification of MBP-dC10 Pd (45 kDa) and nPAGE. *Top*: Preparative size exclusion chromatography on Superdex 200 column. *Bottom*: native PAGE of selected fractions. A molecular weight marker was used for comparison between the two gels and a 49 kDa protein Arr was used as an approximate size reference.

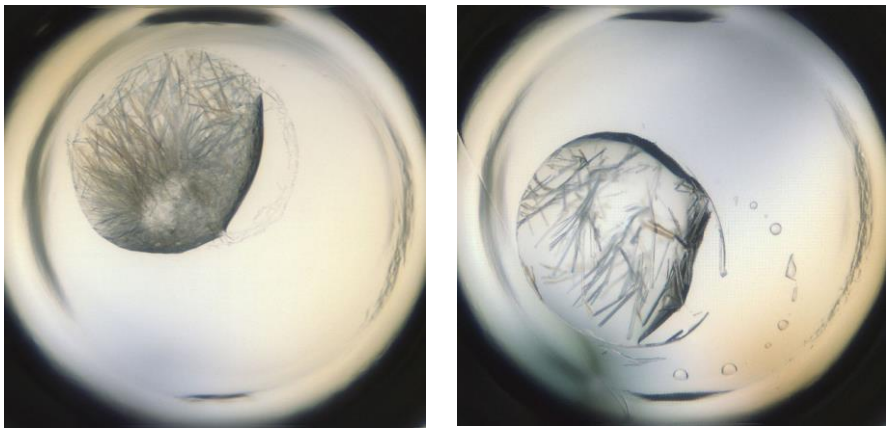


Figure 6.7. Crystals obtained from MBP-dC10 Pd screening. Small needle-shaped crystals were obtained in both *apo*-form (*left*) and in presence of 10 mM maltose (*right*) in 100 mM CAPS pH 10.5, 200 mM lithium sulfate with 2 M ammonium sulfate as precipitant, at 22°C (JCSG Core IV, A1). Protein concentration was 10 mg/mL.

6.4. Conclusions and Perspectives

6.4.1. Conclusion and ongoing work

In this project, we tried to apply a well-established dimaleimide-based protein labelling technique to the field of X-ray crystallography. A palladium containing dimaleimide probe **dM10-Pd** was designed and synthesized (Dr. Christophe Pardin) for that purpose, and used for labelling of two di-cysteine helix containing proteins, AAC-diCys10 and MBP-dC10. It was observed that in the case of AAC-diCys10, where the cysteine residues are localized in an intrinsic helix of the protein, the protein suffered from severe solubility issues, after labelling. This is probably related to the hydrophobicity of **dM10-Pd**, and resulted in the inaptitude of the protein to form crystals, under various screening conditions. The unlabelled AAC-diCys10 produced small crystals; however, these could not be improved by further screening.

The second test protein, MBP-dC10, potentially more soluble and stable than AAC-diCys10, seems to maintain its integrity better, even after labelling with **dM10-Pd**, as demonstrated by its elution profile in a size-exclusion chromatography. After a brief screening of crystallization conditions, small crystals were obtained in both *apo*- and bound form of the protein, suggesting that the covalent modification probably did not perturb the molecular arrangement process in a growing crystal. After this first successful screening, it is important to obtain a crystal of a satisfying size and quality. Sadly, we ran out of time before finalizing the screening of MBP-dC10 Pd crystallization conditions. Nonetheless, this project is still ongoing and could be successfully completed with the work plan proposed below.

Once more, a seeding strategy, similar to the one used for AAC-diCys10, can be used to obtain a MBP-dC10 Pd crystal that would produce an interpretable diffraction pattern. In parallel, a second labelling strategy could be tried, where a crystal of unlabelled MBP-dC10 could be soaked in solution of **dM10-Pd** that would label MBP-dC10 in its crystalline form.

6.4.2. Different approach for a heavy metal probe?

During this whole work we were constantly faced with the problem of the low solubility of the dimaleimide probes, which often caused our test proteins to precipitate. Additionally, **dM10-Pd** is the first probe of its kind that has been tested for protein labelling and crystallization. It would be of a certain interest to review the probe design and propose a probe that would be smaller and/or more water-soluble. It was shown [229, 230, 233, 234] that a slightly higher electron density, provided by probes containing iodine or bromine, is also very beneficial for isomorphous replacement. A smaller, iodine- or bromine-substituted dimaleimide probe could therefore be designed and tested, or alternatively, the **dM10-Pd** water-solubility could be increased by attachment of a PEG [243] or sulfonic acid [244] moiety, as suggested in **Figure 6.8.**

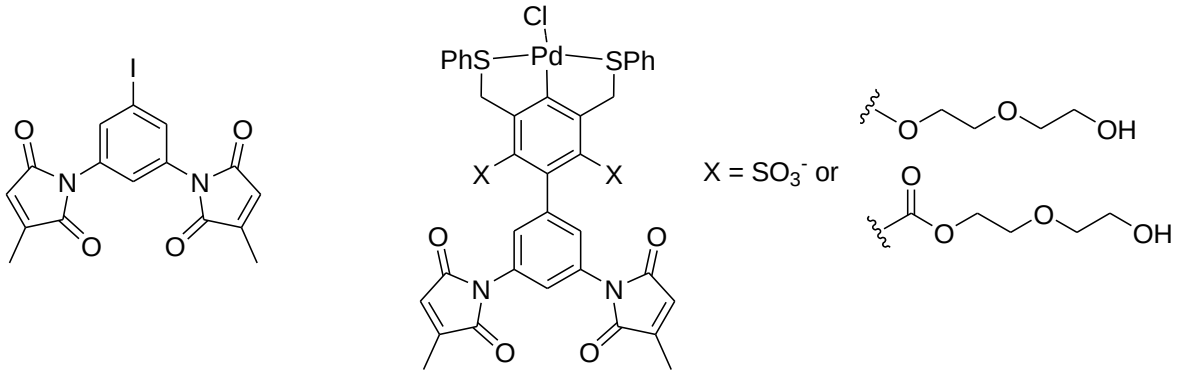


Figure 6.8. Proposed structures for other probes for X-ray crystallography.

On the protein side, both intrinsic di-cysteine helix and dC10-tag strategies have already been tried with the pilot probe **dM10-Pd**, where the dC10-tag approach proved itself as more sustainable for attachment of **dM10-Pd** and for maintaining protein stability and solubility. However, a more soluble and smaller probe could be tested even with a diCys10-modified test protein that has so far proven to be less stable, such as AAC-diCys10. After then, it may be of a certain interest to determine the structure of unlabelled MBP-dC10 using X-ray crystallography, to determine whether the peptide tag is really suitable for crystallisable proteins.

In general, this project has merely been an opening door to new opportunities for applications of dimaleimide-based protein labelling, showing the true limitations of the design for applications necessitating high protein concentration. Above all, the purpose of this presented chapter is to point to new directions where this approach could be valuable, and to assess the qualities and limitations of FlARe labelling.

6.5. Experimental section

6.5.1. Protein expression, purification and labelling with dM10-Pd probe

The **dM10-Pd** probe was designed and synthesized by Dr. Christophe Pardin and the synthesis is not detailed in this thesis.

AAC-diCys10 mutant was cloned, expressed and purified as detailed in the previous chapter (page 141) with the sole difference being that *E. coli* BL21-Gold (DE3) cells were grown in LB medium and not a minimal medium.

MBP-dC10 protein was cloned, expressed and purified as detailed in Chapter 2 (page 42) with the only difference that the protein was expressed in a volume of 1 L of medium in order to get a protein yield adequate for extensive crystallisation screens.

Purified proteins were briefly incubated with 1 mM TCEP to reduce any potential disulphide bonds, and then the buffer was changed to 25 mM HEPES pH 7.4, 2 mM EDTA or 50 mM HEPES pH 7.4, for AAC-diCys10 and MBP-dC10, respectively. Proteins were labelled in 70 µM concentration with 3 – 5 equivalents of **dM10-Pd** probe dissolved in DMSO (from a freshly prepared 4 mM stock) in their respective buffer, overnight at 4°C with gentle stirring. Labelling efficiency was determined by MALDI (AAC-diCys10 Pd) or ESI (MBP-dC10 Pd).

MBP-dC10 Pd in particular was subsequently purified on a HiPrep Q-Sepharose Fast Flow anion exchange column pre-equilibrated in 10 mM Tris-HCl pH 7.2 buffer with 0.02% (*w/v*) sodium azide, at 4°C, and eluted with a gradient of 0.0 - 0.3 M KCl.

Solutions of labelled proteins were concentrated to 2-3 mL using Amicon filters of 10-kDa MWCO and the subsequent purification of these proteins was finalized in the laboratory of Prof. Albert M. Berghuis by size-exclusion chromatography on a HiLoad Superdex 200 26/60 or a Superdex 75 16/60 column in 25 mM HEPES pH 7.4, 2 mM EDTA (AAC-diCys10 Pd) or 20 mM MES pH 6.2 (MBP-dC10 Pd).

6.5.2. Crystallization screens for AAC and variants

Crystallization screens were purchased from Qiagen (JCSG Core I, II, III and IV, JCSG+, PEGs I and II, Opti-Salt, Classics), Hampton Research (Index screen) and Rigaku (Cryo I-II) and each plate included 96 different pH, buffer, additive and precipitant conditions. A seed bead kit was purchased from Hampton Research and used according to the manufacturer's protocol. Twenty-four-well VDXm fine screening plates were prepared manually for a hanging drop configuration, and 96-well MRC broad screening plates were prepared in a sitting drop configuration using a crystallization robot.

Broad screens of pH, buffer, precipitant and additives, combined in JCSG Core I-IV and Index screen at 4°C, and Index screen at 22°C were prepared, where the protein was used from a stock solution of 10 mg/mL and 7 mg/mL of AAC-diCys10 Pd supplemented with 5 mM of CoASH ligand. In a sitting drop setup on 96-well MRC plate, 1 µL of protein was mixed with 1 µL of reservoir solution. Later, broad screens JCSG+, Cryo I-II, PEGs I and a fine screen of $AmSO_4$ at 22°C were prepared, where both AAC wild-type and Pd-labelled AAC-diCys10 were set up in the presence of 5 mM CoASH, from a 7 mg/mL stock solution. An $AmSO_4$ fine screen was set up in a hanging drop configuration and included 24 conditions created by the combination of 1.6 - 2.6 M of $AmSO_4$ precipitant (by intervals of 0.2 M) and 100 mM of citrate buffer at pH 5.6, 6.0, 6.3 (where the first row was a blank with no buffer).

In a parallel manner, unlabelled AAC-diCys10 was subjected to fine screening conditions of $AmSO_4$ at 22°C and 4°C, in the same 24 combinations of pH and precipitant conditions as wild-type and Pd-labelled protein (see above), with one drop for wild-type protein as a positive control.

6.5.3. Crystallization induced by seeding with wild-type AAC crystal

One crystal of AAC wild-type was isolated from well C4 of the $AmSO_4$ fine screen (C4 buffer conditions: 2.2 M $AmSO_4$, 100 mM citrate buffer pH 6.0) and placed in the C4 well buffer that was used for preparation of a seeding solution according to the kit instructions. The final wild-type seed solution was diluted in 1:10 ratio in the C4 well buffer before use. The wild-type seeding solution was used for inducing crystallization of both unlabelled and Pd-labelled AAC-diCys10. First, unlabelled AAC-diCys10 was screened for crystallization in an $AmSO_4$ fine screen with seeding, using the same 24 conditions as detailed above. The following ratios of protein-reservoir-seeds for crystallization drops were used (in µL): 3-1-0.5, 1-3-0.5, 2-2-0.5, both at 22°C and 4°C in a hanging drop configuration. The protein concentration was at 7 mg/mL concentration with 2 mM CoASH.

Later, unlabelled AAC-diCys10 was screened for crystallization with seeding in presence of a 4 mM concentration of reducing agent DTT. Briefly, broad screens JCSG Core I-IV in a sitting drop configuration were set up at 22°C where ratios of protein-reservoir-seeds and presence of DTT was varied (in µL): 1-1-0.5 with no DTT, 1-1-0.5 with 4 mM DTT and 1.5-0.5-0.5 with 4 mM DTT. Protein concentration was set to 5 mg/mL in presence of 2 mM of CoASH.

Next, both unlabelled and Pd-labelled AAC-diCys10 were screened in parallel for crystallization with seeding using a fine screen of $AmSO_4$ in conditions close to JCSG Core II A2 well where AAC-diCys10 microcrystals were obtained. Two 24-well plates, one at 22°C and one at 4°C, were prepared where the concentration of $AmSO_4$ precipitant was varied from 1.00, 1.25, 1.50 to 2.0 M and the pH of 100 mM CHES buffer was varied from 8.3, 9.5 to 10.3. These twelve conditions were in presence of 0.2 M of NaCl, and were replicated in presence of 0.4 M of NaCl to give 24 different conditions per plate in total. Two drops of unlabelled protein and two drops of Pd-labelled AAC-diCys10 were used per well, giving a total of four drops per each well, each protein using both 1/10 and 1/100 dilution of wild-type seeding solution. The proteins were used in 5 mg/mL concentration with 2 mM of CoASH.

Finally, different additives were screened for crystallization of AAC-diCys10 Pd using the 96 Opti-Salt screening conditions at 22°C. Pd-labelled protein was used at 5 mg/mL in the presence of 2 mM CoASH, using a 1/10 volume ratio of wild-type seeding solution. Three drops were set per each well, with different ratios of protein-reservoir-seeds solutions.

6.5.4. Crystallization of AAC-diCys10 Pd induced by seeding with AAC-diCys10 microcrystals

A volume of 1µL of the drop of AAC-diCys10 crystallized in presence of 4 mM DTT (drop A2 from JCSG II plate at 22°C) was used for preparation of a seeding solution, according to the kit standard protocol. A 24-well fine screen of $AmSO_4$ (1.00, 1.25, 1.50 and 2.00 M $AmSO_4$ horizontally, 100 mM CHES pH 8.3, 9.5, 10.3 vertically, all in presence of 0.2 M NaCl, all of these twelve conditions replicated with addition of 5% (*v/v*) glycerol to give 24 different conditions) was used for crystallization of AAC-diCys10 Pd at 22°C, where the following ratios in µL were used for crystallization drops: protein-reservoir-seeds: 2-2-1 and 3-1-1. Four drop were set per each well, where for each ratio of protein-reservoir-seeds the protein was in *apo*-form (with 8.4 mg/mL of protein) and in the presence of 2 mM of CoASH (with 8.0 mg/mL of protein).

Ultimately, broad screening conditions from a 96-well Index screen plate were used for AAC-diCys10 Pd crystallisation induced by seeds of wild-type protein (drop 1) and seeds of unlabelled AAC-diCys10 (drop 2), obtained from well C4 of an $AmSO_4$ fine screen and from well A2 of JCSG Core II broad screen, respectively. AAC-diCys10 Pd was used in 8.0 mg/mL concentration with 2 mM of CoASH.

6.5.5. Crystallization screens for MBP-dC10 Pd

Purified MBP-dC10 Pd from fractions 2H7-3B3 from the size-exclusion chromatography was concentrated to 10 mg/mL and used for 96-well plate screens in a sitting drop setup, where one protein drop was in *apo*-state and one drop contained 10 mM of maltose as ligand. Screening conditions were Classics, Index screen, both at 22°C and 4°C with drops composed of 0.1 µL of protein and 0.1 µL of reservoir solution, and JCSG Core I-IV at 22°C where drops were constituted of 0.2 µL of protein and 0.2 µL of reservoir solution.

Later on, JCSG+ screens at 22°C and at 4°C were performed with a 17 mg/mL of protein concentration (drop 1), 10 mg/mL in *apo*-state (drop 2) and in bound state with 5 mM maltose (drop 3). Similarly, a PEGs I screen was performed using the same drop composition as JCSG+, with the exception of rows E-H, where 10 mg/mL protein was used for drop 1 due to lack of stock at higher concentration (drops 1 and 2 are identical for these rows).

Chapter 7

CONCLUSIONS AND FUTURE DIRECTIONS

In this thesis, further improvements and more diverse applications of a currently developed fluorogenic protein labelling method are presented. Five different objectives were established at the beginning of this thesis, and the success of their completion will be reviewed and discussed in detail in this last chapter. In a complementary manner, future guidelines will be suggested, that may drive this research further to other exciting directions.

7.1. Objective 1: Orthogonal FlARe labelling

7.1.1. Achieved results and conclusions

In accordance with a previously established design of a protein fluorogenic labelling using a small peptide tag, we proposed and cloned test protein MBP with five different di-cysteine peptide tags, with 5-25 Å distances between the side chain thiol groups. We characterized these five constructs by labelling with three of the most recently synthesized dM10-dansyl fluorogens. A certain selectivity of dM10 fluorogens towards dC10 tags was observed, which was encouraging for the subsequent design of dMy fluorogens, different from dM10 and complementary to other dCx peptides from the group of five used sequences.

With the goal of obtaining at least two pairs of dCx/dMy components for an orthogonal labelling, three different dMy fluorogens were then synthesized by Dr. Christophe Pardin. We found that, unfortunately, two of these new fluorogens did not present all the desired properties for us to be able to use them as labelling agents, such as a high fluorescence enhancement, probably due to the spatial orientation of the dimaleimide groups with respect to the fluorophore. The last fluorogen, **dM17-quinoxaline**, is more suitable in terms of fluorescence enhancement; however, it has only very little selectivity towards its designed dC15 (or dC20) partner, and mostly reacts with all MBP-dCx tags with a comparable rate.

7.1.2. Future work for a more successful orthogonal labelling

A satisfying kinetic profile for a MBP-dCx mini-library was obtained for dM10-dansyl fluorogens. Unfortunately, some challenges were encountered with the design of corresponding dMy fluorogens, most of which presented inefficient fluorescence quenching. The design of novel dMy fluorogens is inherently limited by the obligatory spatial orientation of the fluorogen and dimaleimide moieties [74] which complicates further the

future development of this family of compounds. However, new fluorophore scaffolds are still under development (reviewed by Terai *et al.* [245]) and may represent new openings for dimaleimide fluorogen design.

On the other hand, one can easily think of other approaches for designing di-cysteine peptide tags. We have observed that there could be a certain degree of selectivity with dMy fluorogens with *helical* peptide tags; however, other secondary structure motifs have not been explored yet. It would be of a certain interest to explore β-hairpin secondary structure motive, which presents a different rigidity and hence may allow a more selective kinetic profile with existing fluorogens. Small β-hairpin proteins and peptides have been heavily studied [246, 247, 248, 249, 250], and this pioneer work represents a solid base for us to design a series of di-cysteine β-hairpin sequences and seek a selective profile in kinetics of labelling with dimaleimide fluorogens.

Despite this possible development of di-cysteine sequences, there is still an acute need for fluorogens that would satisfy the requirement of efficient quenching and of certain selectivity towards any couple of di-cysteine peptides.

7.2. Objective 2: FlARe labelling in complex milieu

7.2.1. Achieved results in labelling

Through a set of increasingly complex milieu, such as a bacterial lysate, a mammalian cell lysate and ultimately, living mammalian cells, we were able to establish the conditions and limitations of the FlARe labelling technique at each step, and drive the optimization of fluorogen design to limit its reactivity with glutathione. Ultimately, we succeeded in labelling an intracellular protein in living mammalian cells. The development of fluorogen molecules and testing in an intracellular context continued further in the group after the work presented here, and resulted in obtaining the best fluorogen for *in cellulo* labelling [75].

7.2.2. Future development

Most of the currently available dimaleimide fluorogens are intended for intracellular labelling, as it was set as an initial goal. However, it is equally of interest to label a protein exposed on a cell surface, for which fluorogens with a negative charge (such as **dM10-FITC** in [73]) are more suitable because of their cell-impermeability. More importantly, there are fewer methods for cell-surface labelling [73, 251, 252, 253] and hence, this would be an occasion to demonstrate the uniqueness and complementarity of FlARe labelling as an advantageous tool for cell biologists. Therefore, the successful labelling of a cell-surface protein with a new fluorogen of a complementary colour would be beneficial for the overall method development.

7.3. Objective 3: Optimization of dC10 peptide sequence

7.3.1. New optimized dC10 sequence

First, the dC10 secondary structure was studied by solution NMR and we were able to confirm that dC10 adopts an α-helical secondary structure. This was an essential validation leading to the opportunity of rational design for dC10 sequence evolution with the goal of obtaining a more reactive peptide for protein labelling.

After evolving dC10 using a rational approach, we obtained a substantial improvement of a fluorogenic dimaleimide-based labelling technique. According to our *in vitro* kinetic characterization, this new tag reacted an order of magnitude faster than its parent dC10 tag and promised a much faster and, more importantly, more selective labelling of a protein of interest inside a cell where a fluorogenic molecule is exposed to a large number of potentially reactive thiols. This allowed exploring and developing more diverse, but less reactive, fluorogens that would otherwise exhibit too low reactivity, unless used with new generation dC10. We believe also that the presence of a higher number of charged residues in a new dC10 sequence may help the overall solubility of the peptide tag.

Finally, we demonstrated the utility of our best tag sequence dC10* in the labelling of a protein expressed in the nucleus, and we attempted labelling of a new dC10* tag on the

surface of live cells. Overall, obtaining a more reactive dC10* peptide sequence that is universally as good as, and more often better than the original dC10 sequence, suggests that dC10* should naturally replace dC10 for all labelling experiments where a selectivity and side reactivity of fluorogens may represent a challenge.

7.3.2. Future work

New optimized dC10* sequences showed its potential *in vitro* and proved itself to be useful for intracellular labelling. Nonetheless, it would have been beneficial to obtain a quantification of *in cellulo* labelling of a protein attached to dC10 and dC10*, using, for example, flow cytometry, where we would be able to determine an exact ratio of cells containing dC10 or dC10* via an attachment to a red fluorescent protein, and the labelling efficiency of FlARe using a cyan fluorogen. An alternative would be to use a high-throughput screening method, such as an autonomous plate reader coupled to a microscope that would allow showing a true kinetic advantage of dC10* and obtaining a quantitative result with statistical significance. After this demonstration, the optimization of dC10* could be considered as finalized, due to the number of variable residues limited by the helical character of the tag, and more effort could be potentially directed to different peptide sequences.

In accordance with the design of a minimalist fusion peptide, one could suggest a design of a shorter peptide than the current 23 residue long dC10 (or dC10*), maintaining the desired distance between cysteine residues. This shorter peptide length would allow using a truly thorough approach for sequence optimization, allowing variations of every residue close to reactive cysteines and screening large libraries of mutants with currently available high-throughput techniques. This may be beneficial especially for applications that require high solubility and an even more stringently minimalist perturbation of the protein of interest.

7.4. Objective 4: Use of dimaleimide-functionalized molecules for protein structure studies – protein NMR

7.4.1. Achieved results

Two dimaleimide probes containing a lanthanide chelating moiety based on molecules found in literature [177, 179, 180], were synthesized and several proteins of different sizes, attached to a dC10 peptide, or having two cysteine mutations in a solvent exposed helix, were used for dimaleimide-based labelling with these probes. It was demonstrated that the used probes were attached to the protein and also that attachment of such a probe does not hinder the protein enzymatic activity. However, no clear and consistent pseudo-contact shifts could be observed. It was hypothesized that this may be due to protein instability induced by labelling and by the high protein concentration necessary for NMR spectra acquisition, or by the inherent design of a two point attachment of our probes.

7.4.2. Future directions and suggestions

In the future, it would be certainly interesting to try different test proteins, for example smaller *highly* soluble proteins with two solvent exposed cysteine mutations, such as GB1 [200] that has been previously well characterized. It would also be beneficial to be able to use lower protein concentration, similar to values used by other groups [177] to ensure stability of the probe, and avoid potential degradation or interference.

Similarly, to address the hypothesis of destabilization due to attachment of an excessively rigid dimaleimide probe, we could synthesize mono-maleimide analogs of these paramagnetic probes and label corresponding mono-cysteine test proteins. This may address the question of detrimental structure destabilization, potentially caused by dimaleimide probes; however, it would not help the future design of paramagnetic probes, since our technology uses exclusively dimaleimide derivatives. It would rather represent an academic exercise and would help to justify the previous unsuccessful attempts.

7.5. Objective 5: use of dimaleimide-functionalized molecules for protein structure studies – protein X-ray crystallography

7.5.1. Achieved progress

In this project, a palladium containing dimaleimide probe was designed and synthesized (Dr. Christophe Pardin) for X-ray crystallography applications. Two di-cysteine helix containing proteins, AAC-diCys10 and MBP-dC10 were labelled and their crystallization was attempted. In the first case, crystals of Pd-labelled protein could not be obtained in any of broad screens, and in the second case, small crystals were obtained after a first couple of screens. This suggests that MBP-dC10 could be more suitable for labelling dimaleimide heavy metal probe and crystallization.

7.5.2. Ongoing work

A first immediate task is to obtain MBP-dC10 Pd crystals of better size and quality. A fine-screen can be used to map the conditions where small crystals were transiently obtained in order to obtain more stable crystals. A seeding strategy can be used for that purpose in order to obtain a crystal that produces a consistent diffraction pattern. Additionally and in accordance with the original goals, unlabelled MBP-dC10 should be crystallized and labelling with Pd probe of the MBP-dC10 crystal, similar to well-known cryosoaking, should be attempted, as a complementary experiment to the *in vitro* labelling.

7.5.3. Future directions

During this whole work the low solubility of **dM10-Pd** probe and labelled test protein represented a constant challenge. This ongoing issue may be an incentive to review the probe design and attempt to accommodate the requirements for a higher solubility. In agreement with previous work, a smaller, iodine- or bromine-containing dimaleimide probe could be designed and tested, potentially containing a PEG moiety to increase its water-solubility and avoid aggregation previously encountered with the palladium probe.

7.6. Final word

Working on a method that may one day become a standard tool for biologists and biochemists is a challenging but an extremely rewarding task. The progress that has been achieved and presented in this thesis may seem lesser; however, in the general scope of a method development, it can be seen as a stepping stone for the future growth of these methods. This work is still ongoing and many new challenges are encountered every day, especially related to the new potential and exciting applications that have not been explored before. Hence, obstacles and complications are expected, but not accepted as showstoppers.

Appendix 1

Assignment of PpiB-dC10

Group	Atom	Nuclei	Shift
VAL2	CA	13C	60.471
VAL2	CB	13C	35.281
VAL2	CO	13C	174.661
VAL2	HA	1H	4.915
VAL2	HB2	1H	1.696
VAL2	HG1%	1H	0.926
VAL2	HG2%	1H	0.801
VAL2	HN	1H	9.212
VAL2	NH	15N	125.609
THR3	CA	13C	61.157
THR3	CB	13C	70.11
THR3	CO	13C	174.129
THR3	HA	1H	5.107
THR3	HB	1H	3.752
THR3	HN	1H	9.096
THR3	NH	15N	123.758
PHE4	CA	13C	54.625
PHE4	CB	13C	39.335
PHE4	CO	13C	174.221
PHE4	HA	1H	4.961
PHE4	HB2	1H	3.38
PHE4	HB3	1H	2.865
PHE4	HN	1H	9.6
PHE4	NH	15N	127.058
HIS5	CA	13C	54.13
HIS5	CB	13C	26.77
HIS5	CO	13C	174.276
HIS5	HA	1H	5.483
HIS5	HB2	1H	3.444
HIS5	HB3	1H	3.305
HIS5	HN	1H	9.002
HIS5	NH	15N	126.434
THR6	CA	13C	60.241
THR6	CB	13C	72.5
THR6	CO	13C	174.609
THR6	HG2%	1H	1.365
THR6	HN	1H	7.949
THR6	NH	15N	117.336
ASN7	CA	13C	54.635
ASN7	CB	13C	35.834
ASN7	CO	13C	176.278
ASN7	HA	1H	4.655
ASN7	HD21	1H	6.217
ASN7	HN	1H	9.393
ASN7	NH	15N	120.637
HIS8	CA	13C	56.685
HIS8	CB	13C	32.626
HIS8	CO	13C	174.931
HIS8	HA	1H	4.583
HIS8	HB2	1H	3.337
HIS8	HB3	1H	2.396
HIS8	HN	1H	8.798
HIS8	NH	15N	119.727
GLY9	CA	13C	43.801
GLY9	CO	13C	173.371
GLY9	HA2	1H	4.816
GLY9	HA3	1H	3.903
GLY9	HN	1H	7.404
GLY9	NH	15N	109.747
ASP10	CA	13C	54.416
ASP10	CB	13C	41.7
ASP10	CO	13C	175.057
ASP10	HA	1H	5.74
ASP10	HB2	1H	2.649
ASP10	HB3	1H	2.201
ASP10	HN	1H	8.944
ASP10	NH	15N	124.871
ILE11	CA	13C	61.246
ILE11	CB	13C	42.362
ILE11	CO	13C	175.557
ILE11	HB	1H	1.792
ILE11	HD1%	1H	0.591
ILE11	HG2%	1H	0.715
ILE11	HN	1H	8.695
ILE11	NH	15N	122.318
VAL11	HG12	1H	1.048
VAL12	CA	13C	61.664
VAL12	CB	13C	32.946
VAL12	CO	13C	174.07
VAL12	HA	1H	4.438
VAL12	HB	1H	1.905
VAL12	HG1%	1H	0.93
VAL12	HG2%	1H	0.806
VAL12	HN	1H	8.9
VAL12	NH	15N	129.88
ILE13	CA	13C	58.758
ILE13	CB	13C	41.279
ILE13	CO	13C	175.662
ILE13	HA	1H	5.005
ILE13	HD1%	1H	0.504
ILE13	HG12	1H	1.173
ILE13	HG13	1H	1.483
ILE13	HN	1H	9.577
ILE13	NH	15N	127.134
LYS14	CA	13C	53.7
LYS14	CB	13C	36.708
LYS14	CO	13C	173.546
LYS14	HA	1H	5.194
LYS14	HN	1H	8.907
LYS14	NH	15N	126.566
THR15	CA	13C	60.621
THR15	CB	13C	69.528
THR15	CO	13C	175.782
THR15	HA	1H	4.33
THR15	HB	1H	3.69
THR15	HG2%	1H	1.122
THR15	HN	1H	8.704
THR15	NH	15N	115.93
PHE16	CA	13C	54.135
PHE16	CB	13C	37.335
PHE16	CO	13C	175.34
PHE16	HA	1H	5.449
PHE16	HN	1H	8.478

PHE16	NH	15N	122.843
ASP17	CA	13C	58.508
ASP17	CB	13C	39.03
ASP17	CO	13C	176.805
ASP17	HA	1H	5.439
ASP17	HB2	1H	2.556
ASP17	HN	1H	9.308
ASP17	NH	15N	121.809
ASP18	CA	13C	55.425
ASP18	CB	13C	40.803
ASP18	CO	13C	176.618
ASP18	HA	1H	4.552
ASP18	HB2	1H	2.675
ASP18	HN	1H	8.747
ASP18	NH	15N	112.677
LYS19	CA	13C	55.547
LYS19	CB	13C	34.325
LYS19	CO	13C	177.417
LYS19	HA	1H	4.542
LYS19	HN	1H	7.494
LYS19	NH	15N	118.665
ALA20	CA	13C	50.272
ALA20	CB	13C	18.877
ALA20	CO	13C	176.958
ALA20	HA	1H	4.733
ALA20	HB%	1H	1.329
ALA20	HN	1H	7.583
ALA20	NH	15N	124.529
PRO21	CA	13C	66.507
PRO21	CB	13C	31.71
GLU22	CA	13C	59.47
GLU22	CB	13C	28.947
GLU22	CO	13C	181.158
GLU22	HA	1H	4.111
GLU22	HN	1H	9.993
GLU22	NH	15N	123.83
THR23	CA	13C	66.769
THR23	CO	13C	179.969
THR23	HA	1H	4.135
THR23	HN	1H	9.505
THR23	NH	15N	123.31
VAL24	CA	13C	68.296
VAL24	CB	13C	31.414
VAL24	CO	13C	178.426
VAL24	HB	1H	2.192
VAL24	HG1%	1H	1.323
VAL24	HG2%	1H	1.064
VAL24	HN	1H	9.455
VAL24	NH	15N	125.914
LYS25	CA	13C	60.462
LYS25	CB	13C	32.234
LYS25	CO	13C	176.606
LYS25	HA	1H	3.837
LYS25	HB2	1H	1.894
LYS25	HN	1H	7.799
LYS25	NH	15N	122.439
ASN26	CA	13C	56.657
ASN26	CB	13C	41.055
ASN26	CO	13C	177.45
ASN26	HA	1H	4.119
ASN26	HB2	1H	2.848
ASN26	HB3	1H	2.696
ASN26	HN	1H	7.637
ASN26	NH	15N	117.499
PHE27	CA	13C	61.252
PHE27	CB	13C	40.184
PHE27	CO	13C	178.352
PHE27	HA	1H	4.585
PHE27	HB2	1H	2.938
PHE27	HB3	1H	3.212
PHE27	HN	1H	7.953
PHE27	NH	15N	120.767
LEU28	CA	13C	57.992
LEU28	CB	13C	41.377
LEU28	CO	13C	178.084
LEU28	HA	1H	3.726
LEU28	HB2	1H	1.934
LEU28	HD1%	1H	0.146
LEU28	HD2%	1H	0.644
LEU28	HG	1H	1.038
LEU28	HN	1H	9.384
LEU28	NH	15N	121.503
ASP29	CA	13C	57.887
ASP29	CB	13C	39.152
ASP29	CO	13C	179.123
ASP29	HA	1H	4.354
ASP29	HB2	1H	2.238
ASP29	HB3	1H	1.902
ASP29	HN	1H	8.387
ASP29	NH	15N	122.727
TYR30	CA	13C	61.895
TYR30	CB	13C	38.112
TYR30	CO	13C	178.502
TYR30	HA	1H	4.099
TYR30	HB2	1H	2.939
TYR30	HB3	1H	2.572
TYR30	HN	1H	7.361
TYR30	NH	15N	120.295
CYS31	CA	13C	64.476
CYS31	CB	13C	27.561
CYS31	CO	13C	179.265
CYS31	HA	1H	4.105
CYS31	HB2	1H	3.739
CYS31	HB3	1H	2.938
CYS31	HN	1H	8.324
CYS31	NH	15N	117.22
ARG32	CA	13C	59.806
ARG32	CB	13C	30.611
ARG32	CO	13C	179.927
ARG32	HA	1H	3.963
ARG32	HD2	1H	3.213
ARG32	HN	1H	9.029
ARG32	NH	15N	122.778
GLU33	CA	13C	56.606
GLU33	CB	13C	29.8
GLU33	CO	13C	178.057
GLU33	HA	1H	4.274
GLU33	HB2	1H	1.986
GLU33	HB3	1H	1.857
GLU33	HG2	1H	2.411
GLU33	HG3	1H	2.154
GLU33	HN	1H	7.993
GLU33	NH	15N	117.906
GLY34	CA	13C	45.642
GLY34	CO	13C	177.001
GLY34	HA2	1H	4.262
GLY34	HA3	1H	3.82
GLY34	HN	1H	7.641
GLY34	NH	15N	108.569

PHE35	CA	13C	61.762
PHE35	CB	13C	40.814
PHE35	CO	13C	176.105
PHE35	HA	1H	4.104
PHE35	HN	1H	8.193
PHE35	NH	15N	123.081
TYR36	CA	13C	57.831
TYR36	CB	13C	37.948
TYR36	CO	13C	176.142
TYR36	HA	1H	4.408
TYR36	HN	1H	7.457
TYR36	NH	15N	112.429
ASN37	CA	13C	54.283
ASN37	CB	13C	36.278
ASN37	CO	13C	176.171
ASN37	HA	1H	3.938
ASN37	HB2	1H	2.947
ASN37	HB3	1H	2.752
ASN37	HN	1H	7.2
ASN37	NH	15N	122.999
ASN38	CA	13C	54.677
ASN38	CB	13C	38.868
ASN38	CO	13C	175.923
ASN38	HA	1H	4.159
ASN38	HN	1H	9.359
ASN38	NH	15N	122.603
THR39	CA	13C	60.089
THR39	CB	13C	72.891
THR39	CO	13C	174.092
THR39	HA	1H	5.11
THR39	HN	1H	7.447
THR39	NH	15N	105.769
ILE40	CA	13C	58.963
ILE40	CB	13C	40.198
ILE40	CO	13C	174.593
ILE40	HA	1H	5.76
ILE40	HD1%	1H	0.755
ILE40	HN	1H	8.029
ILE40	NH	15N	109.847
PHE41	HA	1H	4.199
PHE41	HN	1H	8.33
PHE41	NH	15N	123.09
HIS42	CA	13C	56.677
HIS42	CB	13C	31.898
HIS42	HN	1H	7.952
ARG43	CA	13C	55.415
ARG43	CB	13C	32.794
ARG43	CO	13C	173.113
ARG43	HA	1H	4.709
ARG43	HD2	1H	2.946
ARG43	HN	1H	7.035
ARG43	NH	15N	124.402
VAL44	CA	13C	62.426
VAL44	CB	13C	35.51
VAL44	CO	13C	173.378
VAL44	HB	1H	1.643
VAL44	HN	1H	9.095
VAL44	NH	15N	130.468
ILE45	CA	13C	60.741
ILE45	CB	13C	40.571
ILE45	CO	13C	173.147
ILE45	HA	1H	4.328
ILE45	HB	1H	1.9
ILE45	HD1%	1H	0.906
ILE45	HD2%	1H	0.703
ILE45	HG2	1H	1.073
ILE45	HG3	1H	1.317
ILE45	HN	1H	8.335
ILE45	NH	15N	127.619
ASN46	CA	13C	54.501
ASN46	CB	13C	37.066
ASN46	CO	13C	174.894
ASN46	HA	1H	4.249
ASN46	HB2	1H	2.961
ASN46	HB3	1H	2.788
ASN46	HN	1H	9.216
ASN46	NH	15N	129.164
GLY47	CA	13C	46.006
GLY47	CO	13C	177.574
GLY47	HA2	1H	4.059
GLY47	HA3	1H	3.746
GLY47	HN	1H	7.623
GLY47	NH	15N	111.458
PHE48	CA	13C	57.821
PHE48	CB	13C	38.812
PHE48	CO	13C	173.098
PHE48	HA	1H	4.273
PHE48	HB2	1H	3.192
PHE48	HB3	1H	2.915
PHE48	HN	1H	7.908
PHE48	NH	15N	117.748
MET49	CA	13C	55.127
MET49	CB	13C	35.654
MET49	CO	13C	173.939
MET49	HA	1H	5.139
MET49	HB2	1H	1.883
MET49	HN	1H	8.446
MET49	NH	15N	118.734
ILE50	CA	13C	59.838
ILE50	CB	13C	40.755
ILE50	CO	13C	172.927
ILE50	HA	1H	4.927
ILE50	HG2	1H	1.498
ILE50	HG3	1H	1.688
ILE50	HN	1H	7.833
ILE50	NH	15N	112.173
GLN51	CA	13C	54.197
GLN51	CB	13C	32.432
GLN51	CO	13C	172.922
GLN51	HA	1H	4.933
GLN51	HN	1H	9.299
GLN51	NH	15N	129.183
GLY52	CA	13C	45.707
GLY52	CO	13C	174.479
GLY52	HA2	1H	4.632
GLY52	HA3	1H	4.379
GLY52	HN	1H	8.173
GLY52	NH	15N	112.561
GLY53	CA	13C	46.138
GLY53	CO	13C	172.531
GLY53	HN	1H	8.353
GLY53	NH	15N	107.195
GLY54	CA	13C	44.855
GLY54	CO	13C	171.639
GLY54	HA2	1H	4.687
GLY54	HN	1H	9.186
GLY54	NH	15N	102.126
PHE55	CA	13C	56.818
PHE55	CB	13C	42.242

PHE55	CO	13C	172.864
PHE55	HN	1H	9.419
PHE55	NH	15N	122.415
GLU56	CA	13C	54.355
GLU56	CB	13C	29.928
GLU56	CO	13C	174.758
GLU56	HN	1H	8.632
GLU56	NH	15N	121.285
PRO57	CA	13C	65.088
PRO57	CB	13C	31.246
GLY58	CA	13C	44.601
GLY58	CO	13C	176.303
GLY58	HA2	1H	3.6
GLY58	HA3	1H	4.178
GLY58	HN	1H	9.524
GLY58	NH	15N	114.573
MET59	CA	13C	54.701
MET59	CB	13C	26.417
MET59	CO	13C	175.179
MET59	HA	1H	4.061
MET59	HN	1H	8.132
MET59	NH	15N	115.109
LYS60	CA	13C	55.185
LYS60	CB	13C	32.859
LYS60	CO	13C	175.132
LYS60	HA	1H	4.555
LYS60	HB2	1H	1.64
LYS60	HN	1H	7.084
LYS60	NH	15N	121.672
GLN61	CA	13C	56.912
GLN61	CB	13C	28.919
GLN61	CO	13C	175.706
GLN61	HA	1H	4.622
GLN61	HN	1H	9.043
GLN61	NH	15N	131.278
LYS62	CA	13C	56.88
LYS62	CB	13C	33.502
LYS62	CO	13C	175.968
LYS62	HA	1H	4.217
LYS62	HN	1H	7.416
LYS62	NH	15N	127.476
ALA63	CA	13C	52.879
ALA63	CB	13C	19
ALA63	CO	13C	176.364
ALA63	HA	1H	4.386
ALA63	HB%	1H	1.437
ALA63	HN	1H	8.54
ALA63	NH	15N	128.84
THR64	CA	13C	60.598
THR64	CB	13C	71.806
THR64	CO	13C	177.669
THR64	HG%	1H	1.184
THR64	HN	1H	8.189
THR64	NH	15N	112.406
LYS65	CA	13C	55.526
LYS65	CB	13C	32.905
LYS65	CO	13C	175.264
LYS65	HN	1H	7.954
LYS65	NH	15N	119.595
GLU66	CA	13C	56.19
GLU66	CB	13C	28.87
GLU66	CO	13C	175.935
GLU66	HN	1H	7.952
GLU66	NH	15N	119.96
PRO67	CA	13C	62.224
PRO67	CB	13C	33.172
ILE68	CA	13C	59.22
ILE68	CB	13C	41.461
ILE68	CO	13C	177.012
ILE68	HA	1H	4.526
ILE68	HN	1H	8.31
ILE68	NH	15N	113.714
LYS69	CA	13C	54.985
LYS69	CB	13C	32.77
LYS69	CO	13C	175.517
LYS69	HN	1H	7.938
LYS69	NH	15N	122.338
ASN70	CA	13C	53.291
ASN70	CB	13C	37.581
ASN70	CO	13C	176.173
ASN70	HA	1H	4.029
ASN70	HN	1H	10.863
ASN70	NH	15N	127.528
GLU71	CA	13C	55.677
GLU71	CB	13C	30.344
GLU71	CO	13C	176.321
GLU71	HA	1H	4.328
GLU71	HN	1H	8.288
GLU71	NH	15N	127.365
ALA72	CA	13C	55.332
ALA72	CB	13C	19.141
ALA72	CO	13C	177.001
ALA72	HA	1H	4.337
ALA72	HB%	1H	1.998
ALA72	HN	1H	9.222
ALA72	NH	15N	121.875
ASN73	CA	13C	52.427
ASN73	CB	13C	34.867
ASN73	CO	13C	177.141
ASN73	HB2	1H	3.329
ASN73	HN	1H	8.007
ASN73	NH	15N	115.893
ASN74	CA	13C	52.952
ASN74	CB	13C	38.81
ASN74	CO	13C	174.863
ASN74	HA	1H	4.492
ASN74	HB2	1H	3.335
ASN74	HN	1H	8.489
ASN74	NH	15N	120.815
GLY75	CA	13C	45.851
GLY75	CO	13C	176.529
GLY75	HA2	1H	3.738
GLY75	HA3	1H	4.122
GLY75	HN	1H	8.099
GLY75	NH	15N	109.961
LEU76	CA	13C	54.779
LEU76	CB	13C	41.54
LEU76	CO	13C	174.58
LEU76	HA	1H	4.422
LEU76	HB2	1H	1.979
LEU76	HD1%	1H	0.467
LEU76	HD2%	1H	1.202
LEU76	HN	1H	7.81
LEU76	NH	15N	121.74
LYS77	CA	13C	55.072
LYS77	CB	13C	35.33
LYS77	CO	13C	177.261
LYS77	HN	1H	8.246
LYS77	NH	15N	123.071

ASN78	CA	13C	54.574
ASN78	CB	13C	37.513
ASN78	CO	13C	176.722
ASN78	HA	1H	4.56
ASN78	HB2	1H	2.635
ASN78	HN	1H	9.659
ASN78	NH	15N	125.839
THR79	CA	13C	60.951
THR79	CB	13C	70.41
THR79	CO	13C	175.769
THR79	HA	1H	4.337
THR79	HB	1H	4.783
THR79	HG2%	1H	1.494
THR79	HN	1H	8.284
THR79	NH	15N	115.967
ARG80	CA	13C	58.477
ARG80	CB	13C	29.709
ARG80	CO	13C	175.752
ARG80	HB2	1H	1.728
ARG80	HD2	1H	3.303
ARG80	HG2	1H	1.136
ARG80	HN	1H	9.34
ARG80	NH	15N	125.241
GLY81	CA	13C	45.345
GLY81	CO	13C	176.69
GLY81	HA2	1H	4.814
GLY81	HA3	1H	3.468
GLY81	HN	1H	9.061
GLY81	NH	15N	116.485
THR82	CA	13C	61.832
THR82	CB	13C	72.377
THR82	CO	13C	173.52
THR82	HA	1H	6.146
THR82	HG%	1H	1.495
THR82	HN	1H	8.168
THR82	NH	15N	111.584
LEU83	CA	13C	54.564
LEU83	CB	13C	44.063
LEU83	CO	13C	173.43
LEU83	HN	1H	8.015
LEU83	NH	15N	122.412
ALA84	CA	13C	49.333
ALA84	CB	13C	24.719

ALA84	CO	13C	176.586
ALA84	HB%	1H	0.694
ALA84	HN	1H	8.13
ALA84	NH	15N	122.937
MET85	CA	13C	52.787
MET85	CB	13C	30.388
MET85	CO	13C	177.528
MET85	HN	1H	7.628
MET85	NH	15N	116.528
ALA86	CA	13C	51.683
ALA86	CB	13C	19.676
ALA86	CO	13C	174.475
ALA86	HB%	1H	1.4
ALA86	HN	1H	8.405
ALA86	NH	15N	127.672
ARG87	CA	13C	55.847
ARG87	CB	13C	31.897
ARG87	CO	13C	177.017
ARG87	HA	1H	4.173
ARG87	HN	1H	8.059
ARG87	NH	15N	114.723
THR88	CA	13C	60.951
THR88	CB	13C	68.504
THR88	CO	13C	176.165
THR88	HA	1H	4.72
THR88	HN	1H	8.01
THR88	NH	15N	112.738
GLN89	CA	13C	58.88
GLN89	CB	13C	28.852
GLN89	CO	13C	173.517
GLN89	HB2	1H	2.072
GLN89	HN	1H	8.289
GLN89	NH	15N	118.594
ALA90	CA	13C	49.742
ALA90	CB	13C	17.807
ALA90	CO	13C	176.791
ALA90	HA	1H	4.14
ALA90	HB%	1H	1.464
ALA90	HN	1H	8.056
ALA90	NH	15N	124.122
PRO91	CA	13C	65.474
PRO91	CB	13C	30.353
PRO91	HA	1H	4.217

HIS92	CA	13C	54.346
HIS92	CB	13C	28.713
HIS92	CO	13C	177.491
HIS92	HA	1H	4.689
HIS92	HN	1H	8.058
HIS92	NH	15N	119.445
SER93	CA	13C	57.752
SER93	CB	13C	66.22
SER93	CO	13C	175.096
SER93	HA	1H	4.474
SER93	HN	1H	6.474
SER93	NH	15N	109.505
ALA94	CA	13C	54.358
ALA94	CB	13C	21.12
ALA94	CO	13C	175.533
ALA94	HA	1H	4.376
ALA94	HB%	1H	1.509
ALA94	HN	1H	8.623
ALA94	NH	15N	131.176
THR95	CA	13C	59.812
THR95	CB	13C	70.157
THR95	CO	13C	176.072
THR95	HA	1H	4.809
THR95	HB	1H	4.374
THR95	HG2%	1H	1.516
THR95	HN	1H	8.993
THR95	NH	15N	111.853
ALA96	CA	13C	51.636
ALA96	CB	13C	24.029
ALA96	CO	13C	172.391
ALA96	HB%	1H	0.652
ALA96	HN	1H	8.038
ALA96	NH	15N	122.655
GLN97	CA	13C	57.603
GLN97	CB	13C	31.173
GLN97	CO	13C	179.008
GLN97	HA	1H	4.721
GLN97	HN	1H	7.95
GLN97	NH	15N	116.71
PHE98	CA	13C	54.965
PHE98	CB	13C	43.609
PHE98	CO	13C	175.274
PHE98	HA	1H	4.705

PHE98	HB2	1H	3.203
PHE98	HN	1H	7.772
PHE98	NH	15N	118.781
PHE99	CA	13C	55.211
PHE99	CB	13C	44.571
PHE99	CO	13C	172.357
PHE99	HB2	1H	3.206
PHE99	HN	1H	9.593
PHE99	NH	15N	117.399
ILE100	CA	13C	59.495
ILE100	CB	13C	40.071
ILE100	CO	13C	172.771
ILE100	HN	1H	9.222
ILE100	NH	15N	118.871
ASN101	CA	13C	55.73
ASN101	CB	13C	39.933
ASN101	CO	13C	177.155
ASN101	HA	1H	4.871
ASN101	HN	1H	9.002
ASN101	NH	15N	127.562
VAL102	CA	13C	62.172
VAL102	CB	13C	30.376
VAL102	CO	13C	173.814
VAL102	HA	1H	4.563
VAL102	HG1%	1H	0.76
VAL102	HN	1H	7.788
VAL102	NH	15N	118.927
VAL103	CA	13C	59.51
VAL103	CB	13C	35.991
VAL103	CO	13C	174.09
VAL103	HA	1H	4.45
VAL103	HB	1H	2.268
VAL103	HG1%	1H	0.899
VAL103	HN	1H	7.849
VAL103	NH	15N	117.647
ASP104	CA	13C	55.154
ASP104	CB	13C	39.6
ASP104	CO	13C	173.131
ASP104	HN	1H	8.26
ASP104	NH	15N	119.709
ASN105	CA	13C	51.707
ASN105	CB	13C	39.498
ASN105	CO	13C	175.027
ASN105	HA	1H	5.157
ASN105	HN	1H	8.446
ASN105	NH	15N	129.359
ASP106	CA	13C	56.913
ASP106	CB	13C	40.343
ASP106	CO	13C	175.087
ASP106	HA	1H	5.16
ASP106	HN	1H	8.283
ASP106	NH	15N	122.692
PHE107	CA	13C	57.373
PHE107	CB	13C	37.277
PHE107	CO	13C	177.971
PHE107	HA	1H	4.592
PHE107	HN	1H	7.491
PHE107	NH	15N	115.708
LEU108	CA	13C	54.565
LEU108	CB	13C	42.601
LEU108	CO	13C	175.914
LEU108	HA	1H	4.564
LEU108	HN	1H	7.277
LEU108	NH	15N	122.603
ASN109	CA	13C	52.912
ASN109	CB	13C	38.614
ASN109	CO	13C	175.634
ASN109	HA	1H	5.238
ASN109	HN	1H	7.417
ASN109	NH	15N	117.184
PHE110	CA	13C	59.04
PHE110	CB	13C	40.338
PHE110	CO	13C	176.646
PHE110	HA	1H	4.352
PHE110	HN	1H	8.902
PHE110	NH	15N	125.146
SER111	CA	13C	57.657
SER111	CB	13C	64.794
SER111	CO	13C	176.083
SER111	HA	1H	4.348
SER111	HN	1H	9.541
SER111	NH	15N	123.296
GLY112	CA	13C	45.654
GLY112	CO	13C	173.433
GLY112	HA2	1H	3.644
GLY112	HA3	1H	3.137
GLY112	HN	1H	5.634
GLY112	NH	15N	111.615
GLU113	CA	13C	55.055
GLU113	CB	13C	27.143
GLU113	CO	13C	170.865
GLU113	HN	1H	7.95
GLU113	NH	15N	120.002
SER114	CA	13C	56.364
SER114	CB	13C	65.692
SER114	CO	13C	175.042
SER114	HA	1H	4.705
SER114	HN	1H	7.725
SER114	NH	15N	120.328
LEU115	CA	13C	59.012
LEU115	CB	13C	41.614
LEU115	CO	13C	174.147
LEU115	HN	1H	8.836
LEU115	NH	15N	123.421
GLN116	CA	13C	57.286
GLN116	CB	13C	28.598
GLN116	CO	13C	179.418
GLN116	HA	1H	4.348
GLN116	HN	1H	8.388
GLN116	NH	15N	115.724
GLY117	CA	13C	45.808
GLY117	CO	13C	177.385
GLY117	HA2	1H	4.14
GLY117	HA3	1H	3.23
GLY117	HN	1H	7.926
GLY117	NH	15N	110.813
TRP118	CA	13C	62.282
TRP118	CB	13C	30.638
TRP118	CO	13C	176.066
TRP118	HA	1H	4.395
TRP118	HA2	1H	3.141
TRP118	HA3	1H	2.988
TRP118	HN	1H	7.878
TRP118	NH	15N	124.799
GLY119	CA	13C	45.665
GLY119	CO	13C	176.778
GLY119	HN	1H	7.531
GLY119	NH	15N	137.063
TYR120	CA	13C	57.627

TYR120	CB	13C	40.373
TYR120	CO	13C	174.411
TYR120	HA	1H	5.271
TYR120	HB2	1H	3.768
TYR120	HB3	1H	3.611
TYR120	HN	1H	8.373
TYR120	NH	15N	120.94
CYS121	CA	13C	59.679
CYS121	CB	13C	28.387
CYS121	CO	13C	175.526
CYS121	HA	1H	5.301
CYS121	HB2	1H	3.152
CYS121	HB3	1H	2.9
CYS121	HN	1H	9.915
CYS121	NH	15N	127.164
VAL122	CA	13C	62.752
VAL122	CB	13C	31.832
VAL122	CO	13C	173.7
VAL122	HA	1H	4.337
VAL122	HB	1H	1.975
VAL122	HG1%	1H	0.844
VAL122	HN	1H	9.484
VAL122	NH	15N	136.204
PHE123	CA	13C	54.734
PHE123	CB	13C	42.026
PHE123	CO	13C	170.913
PHE123	HN	1H	7.988
PHE123	NH	15N	119.469
ALA124	CA	13C	50.877
ALA124	CB	13C	23.718
ALA124	CO	13C	172.566
ALA124	HA	1H	3.946
ALA124	HB%	1H	0.545
ALA124	HN	1H	7.454
ALA124	NH	15N	125.192
GLU125	CA	13C	54.511
GLU125	CB	13C	33.232
GLU125	CO	13C	174.426
GLU125	HA	1H	4.952
GLU125	HN	1H	8.133
GLU125	NH	15N	115.82
VAL126	CA	13C	63.615
VAL126	CB	13C	32.794
VAL126	CO	13C	174.083
VAL126	HA	1H	4.011
VAL126	HN	1H	8.998
VAL126	NH	15N	124.911
VAL127	CA	13C	61.41
VAL127	CB	13C	32.75
VAL127	CO	13C	176.626
VAL127	HA	1H	4.506
VAL127	HN	1H	9.248
VAL127	NH	15N	123.851
ASP128	CA	13C	54.721
ASP128	CB	13C	43.54
ASP128	HA	1H	4.037
GLY129	CA	13C	45.868
GLY129	CO	13C	175.139
GLY129	HA2	1H	5.005
GLY129	HA3	1H	4.89
GLY129	HN	1H	8.395
GLY129	NH	15N	112.014
MET130	CA	13C	56.26
MET130	CB	13C	29.499
MET130	CO	13C	177.213
MET130	HN	1H	9.014
MET130	NH	15N	124.034
ASP131	CA	13C	56.762
ASP131	CB	13C	39.016
ASP131	CO	13C	177.939
ASP131	HA	1H	4.329
ASP131	HN	1H	9.004
ASP131	NH	15N	118.085
VAL132	CA	13C	66.127
VAL132	CB	13C	31.123
VAL132	CO	13C	178.743
VAL132	HA	1H	3.534
VAL132	HB	1H	2.406
VAL132	HN	1H	7.347
VAL132	NH	15N	125.267
VAL133	CA	13C	67.114
VAL133	CB	13C	31.078
VAL133	CO	13C	178.37
VAL133	HA	1H	4.512
VAL133	HN	1H	7.442
VAL133	NH	15N	122.99
ASP134	CA	13C	56.404
ASP134	CB	13C	40.115
ASP134	CO	13C	177.115
ASP134	HA	1H	4.335
ASP134	HN	1H	8.488
ASP134	NH	15N	118.569
LYS135	CA	13C	59.085
LYS135	CB	13C	32.608
LYS135	CO	13C	177.833
LYS135	HA	1H	4.107
LYS135	HB2	1H	2.008
LYS135	HN	1H	7.531
LYS135	NH	15N	123.799
ILE136	CA	13C	65.363
ILE136	CB	13C	37.766
ILE136	CO	13C	175.288
ILE136	HN	1H	8.228
ILE136	NH	15N	122.647
LYS137	CA	13C	58.07
LYS137	CB	13C	31.633
LYS137	CO	13C	175.986
LYS137	HA	1H	4.79
LYS137	HN	1H	8.087
LYS137	NH	15N	114.799
GLY138	CA	13C	44.648
GLY138	CO	13C	174.833
GLY138	HA2	1H	4.308
GLY138	HA3	1H	3.656
GLY138	HN	1H	6.965
GLY138	NH	15N	105.946
VAL139	CA	13C	61.89
VAL139	CB	13C	33.287
VAL139	CO	13C	175.531
VAL139	HA	1H	4.176
VAL139	HB	1H	2.446
VAL139	HG1%	1H	0.96
VAL139	HG2%	1H	1.058
VAL139	HN	1H	7.328
VAL139	NH	15N	117.202
ALA140	CA	13C	52.996
ALA140	CB	13C	19.124
ALA140	CO	13C	175.51
ALA140	HA	1H	4.381

ALA140	HB%	1H	1.435
ALA140	HN	1H	8.346
ALA140	NH	15N	124.331
THR141	CA	13C	59.365
THR141	CB	13C	73.385
THR141	CO	13C	179.525
THR141	HA	1H	5.017
THR141	HG%	1H	1.093
THR141	HN	1H	8.352
THR141	NH	15N	114.259
GLY142	CA	13C	45.575
GLY142	CO	13C	174.229
GLY142	HA2	1H	4.309
GLY142	HA3	1H	3.844
GLY142	HN	1H	8.592
GLY142	NH	15N	108.536
ARG143	CA	13C	55.473
ARG143	CB	13C	31.94
ARG143	CO	13C	171.447
ARG143	HA	1H	4.899
ARG143	HN	1H	8.271
ARG143	NH	15N	121.135
SER144	CA	13C	57.022
SER144	CB	13C	63.607
SER144	CO	13C	177.391
SER144	HA	1H	4.541
SER144	HN	1H	8.487
SER144	NH	15N	120.465
GLY145	CA	13C	47.035
GLY145	CO	13C	174.659
GLY145	HA2	1H	4.042
GLY145	HA3	1H	3.607
GLY145	HN	1H	9
GLY145	NH	15N	118.943
MET146	CA	13C	55.404
MET146	CB	13C	31.963
MET146	CO	13C	175.386
MET146	HN	1H	8.897
MET146	NH	15N	126.652
HIS147	CA	13C	56.245
HIS147	CB	13C	32.122
HIS147	CO	13C	175.648
HIS147	HA	1H	4.527
HIS147	HD2	1H	7.159
HIS147	HN	1H	8.119
HIS147	NH	15N	122.731
GLN148	CA	13C	54.635
GLN148	CB	13C	31.801
GLN148	CO	13C	174.643
GLN148	HA	1H	4.526
GLN148	HN	1H	8.758
GLN148	NH	15N	122.351
ASP149	CA	13C	55.615
ASP149	CB	13C	39.812
ASP149	CO	13C	174.963
ASP149	HA	1H	4.572
ASP149	HN	1H	8.704
ASP149	NH	15N	118.133
VAL150	CA	13C	59.408
VAL150	CB	13C	33.662
VAL150	CO	13C	173.377
VAL150	HA	1H	4.484
VAL150	HB	1H	1.866
VAL150	HG1%	1H	0.913
VAL150	HN	1H	8.539
VAL150	NH	15N	121.404
PRO151	CA	13C	64.04
PRO151	CB	13C	32.759
LYS152	CA	13C	58.827
LYS152	CB	13C	32.415
LYS152	CO	13C	176.713
LYS152	HA	1H	5.016
LYS152	HN	1H	8.261
LYS152	NH	15N	126.683
GLU153	CA	13C	54.403
GLU153	CB	13C	31.109
GLU153	CO	13C	176.935
GLU153	HA	1H	4.497
GLU153	HN	1H	8.019
GLU153	NH	15N	120.157
ASP154	CA	13C	56.289
ASP154	CB	13C	41.1
ASP154	HN	1H	8.242
ASP154	NH	15N	122.973
VAL155	CA	13C	62.177
VAL155	CB	13C	31.913
VAL155	CO	13C	176.724
VAL155	HA	1H	4.319
VAL155	HB	1H	2.338
VAL155	HG1%	1H	1.016
VAL155	HN	1H	10.121
VAL155	NH	15N	130.813
ILE156	CA	13C	60.99
ILE156	CB	13C	43.147
ILE156	CO	13C	174.87
ILE156	HA	1H	4.523
ILE156	HB	1H	1.477
ILE156	HN	1H	8.324
ILE156	NH	15N	126.856
ILE157	CA	13C	62.722
ILE157	CB	13C	36.827
ILE157	CO	13C	176.525
ILE157	HA	1H	4.472
ILE157	HN	1H	9.293
ILE157	NH	15N	127.719
GLU158	CA	13C	58.973
GLU158	CB	13C	30.474
GLU158	CO	13C	174.418
GLU158	HA	1H	4.187
GLU158	HN	1H	9.071
GLU158	NH	15N	131.123
SER159	CA	13C	57.619
SER159	CB	13C	64.739
SER159	CO	13C	177.346
SER159	HA	1H	4.536
SER159	HN	1H	7.857
SER159	NH	15N	109.353
VAL160	CA	13C	60.539
VAL160	CB	13C	35.636
VAL160	CO	13C	172.137
VAL160	HA	1H	5.348
VAL160	HB	1H	1.95
VAL160	HG1%	1H	0.797
VAL160	HN	1H	7.821
VAL160	NH	15N	120.604
THR161	CA	13C	61.148
THR161	CB	13C	71.279
THR161	CO	13C	175.837
THR161	HA	1H	4.772

THR161	HB	1H	4.036
THR161	HG2%	1H	1.276
THR161	HN	1H	8.758
THR161	NH	15N	122.859
VAL162	CA	13C	61.907
VAL162	CB	13C	33.576
VAL162	CO	13C	173.344
VAL162	HA	1H	4.577
VAL162	HN	1H	9.154
VAL162	NH	15N	129.112
SER163	CA	13C	58.074
SER163	CB	13C	64.795
SER163	CO	13C	175.218
SER163	HA	1H	4.784
SER163	HB2	1H	3.939
SER163	HN	1H	9.164
SER163	NH	15N	125.101
GLU164	CA	13C	56.586
GLU164	CB	13C	30.52
GLU164	CO	13C	174
GLU164	HA	1H	4.396
GLU164	HN	1H	8.741
GLU164	NH	15N	124.89
GLY165	CA	13C	45.232
GLY165	CO	13C	176.823
GLY165	HA2	1H	3.815
GLY165	HA3	1H	3.71
GLY165	HN	1H	8.264
GLY165	NH	15N	111.415
SER166	CA	13C	58.413
SER166	CB	13C	63.972
GLY167	CA	13C	45.354
GLY167	CO	13C	175.084
GLY167	HA2	1H	3.932
GLY167	HN	1H	8.466
GLY167	NH	15N	112.432
SER168	CA	13C	58.396
SER168	CB	13C	63.877
SER168	HA	1H	4.424
SER168	HB2	1H	3.82
SER168	HN	1H	8.196
SER168	NH	15N	117.325
LEU169	CA	13C	55.317
LEU169	CB	13C	42.444
LEU169	CO	13C	174.478
LEU169	HA	1H	4.328
LEU169	HB	1H	1.593
LEU169	HD1%	1H	0.815
LEU169	HN	1H	8.271
LEU169	NH	15N	125.35
GLY170	CA	13C	45.71
GLY170	CO	13C	176.716
GLY170	HA2	1H	3.932
GLY170	HN	1H	8.331
GLY170	NH	15N	110.959
ILE171	CA	13C	61.387
ILE171	CB	13C	38.819
ILE171	CO	13C	174.276
ILE171	HA	1H	4.127
ILE171	HB	1H	1.851
ILE171	HD1%	1H	0.869
ILE171	HN	1H	7.914
ILE171	NH	15N	121.106
GLU172	CA	13C	56.888
GLU172	CB	13C	30.109
GLU172	CO	13C	176.422
GLU172	HA	1H	4.216
GLU172	HG2	1H	2.268
GLU172	HN	1H	8.565
GLU172	NH	15N	125.569
GLY173	CA	13C	45.453
GLY173	CO	13C	176.422
GLY173	HN	1H	8.51
GLY173	NH	15N	112.361
ARG174	CA	13C	55.808
ARG174	CB	13C	30.997
ARG174	CO	13C	173.809
ARG174	HA	1H	4.323
ARG174	HD2	1H	3.17
ARG174	HN	1H	8.004
ARG174	NH	15N	122.107
LEU175	CA	13C	55.485
LEU175	CB	13C	42.514
LEU175	CO	13C	176.333
LEU175	HA	1H	4.402
LEU175	HB2	1H	1.648
LEU175	HB3	1H	1.565
LEU175	HD1%	1H	0.849
LEU175	HN	1H	8.177
LEU175	NH	15N	124.496
SER176	HA	1H	4.419
SER176	HB2	1H	4.127
SER176	HB3	1H	3.967
SER176	HN	1H	8.524
SER176	NH	15N	119.212
ALA177	HA	1H	4.16
ALA177	HB%	1H	1.453
ALA177	HN	1H	8.606
ALA177	NH	15N	127.248
ALA178	CA	13C	53.101
ALA178	CB	13C	18.654
ALA178	HA	1H	4.325
ALA178	HB%	1H	1.425
ALA178	HN	1H	8.033
ALA178	NH	15N	125.771
GLU179	CA	13C	56.734
GLU179	CB	13C	30.807
GLU179	HA	1H	4.263
GLU179	HN	1H	7.933
GLU179	NH	15N	120.509
ALA182	CA	13C	53.186
ALA182	CB	13C	19.001
ALA182	HA	1H	4.223
ALA182	HB%	1H	1.457
ALA182	HN	1H	8.132
ALA182	NH	15N	125.173
ARG183	CA	13C	57.112
ARG183	CB	13C	29.887
ARG183	HA	1H	4.186
ARG183	HD2	1H	2.975
ARG183	HN	1H	8.344
ARG183	NH	15N	121.009
GLU184	CA	13C	57.159
GLU184	CB	13C	29.943
GLU184	CO	13C	176.977
GLU184	HA	1H	3.914
GLU184	HN	1H	8.316
GLU184	NH	15N	111.17
ALA185	CA	13C	53.366

ALA185	CB	13C	19.028
ALA185	CO	13C	174.498
ALA185	HN	1H	8.196
ALA185	NH	15N	125.529
ALA186	CA	13C	52.948
ALA186	CB	13C	18.879
ALA186	CO	13C	177.545
ALA186	HN	1H	8.054
ALA186	NH	15N	123.365
CYS187	HN	1H	8.046
CYS187	NH	15N	118.602
ARG188	HA	1H	4.199
ARG188	HD2	1H	3.209
ARG188	HN	1H	8.146
ARG188	NH	15N	123.806
GLU189	HA	1H	4.141
GLU189	HG2	1H	2.048
GLU189	HN	1H	8.195
GLU189	NH	15N	122.241
ALA190	HA	1H	4.162
ALA190	HB%	1H	1.441
ALA190	HN	1H	8.023
ALA190	NH	15N	124.588
ALA191	CA	13C	53.467
ALA191	CB	13C	19.104
ALA192	HA	1H	4.235
ALA192	HB%	1H	1.431
ALA192	HN	1H	7.894
ALA192	NH	15N	123.305
ARG193	CA	13C	56.635
ARG193	CB	13C	30.246
ARG193	HA	1H	4.172
ARG193	HD2	1H	3.179
ARG193	HN	1H	8.053
ARG193	NH	15N	121.554
ALA194	HA	1H	4.229
ALA194	HB%	1H	1.469
ALA194	HN	1H	7.994
ALA194	NH	15N	123.624
GLY195	CA	13C	45.454
GLY195	CO	13C	178.227
GLY195	HA2	1H	3.978
GLY195	HN	1H	8.233
GLY195	NH	15N	109.417
GLY196	CA	13C	45.456
GLY196	CO	13C	174.708
GLY196	HA2	1H	3.958
GLY196	HN	1H	8.176
GLY196	NH	15N	110.518
LYS197	CA	13C	57.374
LYS197	CB	13C	33.885
LYS197	CO	13C	173.202
LYS197	HA	1H	4.182
LYS197	HB2	1H	1.823
LYS197	HD2	1H	1.69
LYS197	HG2	1H	1.377
LYS197	HN	1H	7.741
LYS197	NE1	1H	2.984
LYS197	NH	15N	127.241

References

[1] I. Johnson et M. T. Z. Spence, Molecular Probes Handbook, I. Johnson et M. T. Z. Spence, Éds., *Life Technologies Corporation*, **2010**.

[2] J. R. Lakowicz, Principles of Fluorescence, *Springer Science & Business Media*, **2007**.

[3] G. G. Guilbault, Practical Fluorescence, G. G. Guilbault, Éd., *Marcel Dekker, Inc*, **1990**.

[4] S. Pang, D. Jang, W. S. Lee, H.-M. Kang, S.-J. Hong, S. K. Hwang et K.-H. Ahn, «The effect of a "push-pull" structure on the turn-on fluorescence of photochromic thio-ketone type diarylethenes.,» *Photochem Photobiol Sci*, 14, 14, 765-774, **2015**.

[5] M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien et S. J. Remington, «Crystal structure of the Aequorea victoria green fluorescent protein.,» *Science*, 273, 1392-1395, **1996**.

[6] M. Chalfie, Y. Tu, G. Euskirchen, W. W. Ward et D. C. Prasher, «Green fluorescent protein as a marker for gene expression.,» *Science*, 263, 802-805, **1994**.

[7] J. Krasowska, M. Olasek, A. Bzowska, P. Clark et B. Zielgus-Kutrowska, «The comparison of aggregation and folding of enhanced green fluorescent protein (EGFP) by spectroscopic studies,» *J Spectroscopy*, 24, 343-348, **2010**.

[8] N. C. Shaner, G. H. Patterson et M. W. Davidson, «Advances in fluorescent protein technology.,» *J Cell Sci*, 120, 4247-4260, **2007**.

[9] X. Shu, N. C. Shaner, C. A. Yarbrough, R. Y. Tsien et S. J. Remington, «Novel chromophores and buried charges control color in mFruits.,» *Biochemistry*, 45, 132, 9639-9647, Aug **2006**.

[10] G. S. Baird, D. A. Zacharias et R. Y. Tsien, «Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral.,» *Proc Natl Acad Sci*

U S A, 97, 122, 11984-11989, **2000**.

[11] R. E. Campbell, O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird, D. A. Zacharias et R. Y. Tsien, «A monomeric red fluorescent protein.,» *Proc Natl Acad Sci U S A*, 99, 112, 7877-7882, **2002**.

[12] R. Heim et R. Y. Tsien, «Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer.,» *Curr Biol*, 6, 12, 178-182, **1996**.

[13] H. Mizuno, A. Sawano, P. Eli, H. Hama et A. Miyawaki, «Red fluorescent protein from Discosoma as a fusion tag and a partner for fluorescence resonance energy transfer.,» *Biochemistry*, 40, 18, 2502-2510, **2001**.

[14] U. Lauf, P. Lopez et M. M. Falk, «Expression of fluorescently tagged connexins: a novel approach to rescue function of oligomeric DsRed-tagged proteins.,» *FEBS Lett*, 498, 11, 11-15, **2001**.

[15] C. Stadler, E. Rexhepaj, V. R. Singan, R. F. Murphy, R. Pepperkok, M. Uhlén, J. C. Simpson et E. Lundberg, «Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells.,» *Nat Methods*, 10, 14, 315-323, **2013**.

[16] L. M. Costantini, M. Fossati, M. Francolini et E. L. Snapp, «Assessing the tendency of fluorescent proteins to oligomerize under physiologic conditions.,» *Traffic*, 13, 15, 643-649, **2012**.

[17] D. Jung, K. Min, J. Jung, W. Jang et Y. Kwon, «Chemical biology-based approaches on fluorescent labeling of proteins in live cells.,» *Mol Biosyst*, 9, 15, 862-872, **2013**.

[18] C. P. Toseland, «Fluorescent labeling and modification of proteins.,» *J Chem Biol*, 6, 13, 85-95, **2013**.

[19] M. Rashidian, J. K. Dozier et M. D. Distefano, «Enzymatic Labeling of Proteins:

Techniques and Approaches.,» *Bioconjug Chem,* **2013**.

[20] M. Z. Lin et L. Wang, «Selective labeling of proteins with chemical probes in living cells.,» *Physiology (Bethesda),* 23, 131-141, **2008**.

[21] S. K. Oteng-Pabi, C. Pardin, M. Stoica et J. W. Keillor, «Site-specific protein labelling and immobilization mediated by microbial transglutaminase.,» *Chem Commun (Camb),* **2014**.

[22] C. Gnaccarini, W. Ben-Tahar, A. Mulani, I. Roy, W. D. Lubell, J. N. Pelletier et J. W. Keillor, «Site-specific protein propargylation using tissue transglutaminase.,» *Org Biomol Chem,* 10, 127, 5258-5265, **2012**.

[23] R. Parthasarathy, S. Subramanian et E. T. Boder, «Sortase A as a novel molecular "stapler" for sequence-specific protein conjugation.,» *Bioconjug Chem,* 18, 12, 469-476, **2007**.

[24] M. W. Popp, J. M. Antos, G. M. Grotenbreg, E. Spooner et H. L. Ploegh, «Sortagging: a versatile method for protein labeling.,» *Nat Chem Biol,* 3, 111, 707-708, **2007**.

[25] Z. Zhou, P. Cironi, A. J. Lin, Y. Xu, S. Hrvatin, D. E. Golan, P. A. Silver, C. T. Walsh et J. Yin, «Genetically encoded short peptide tags for orthogonal protein labeling by Sfp and AcpS phosphopantetheinyl transferases.,» *ACS Chem Biol,* 2, 15, 337-346, **2007**.

[26] N. Johnsson, N. George et K. Johnsson, «Protein chemistry on the surface of living cells.,» *Chembiochem,* 6, 11, 47-52, **2005**.

[27] L. Vivero-Pol, N. George, H. Krumm, K. Johnsson et N. Johnsson, «Multicolor imaging of cell surface proteins.,» *J Am Chem Soc,* 127, 137, 12770-12771, **2005**.

[28] J. D. Cohen, P. Zou et A. Y. Ting, «Site-specific protein modification using lipoic acid ligase and bis-aryl hydrazone formation.,» *Chembiochem,* 13, 16, 888-894, **2012**.

[29] M. Howarth, K. Takao, Y. Hayashi et A. Y. Ting, «Targeting quantum dots to surface proteins in living cells with biotin ligase.,» *Proc Natl Acad Sci U S A*, 102, 121, 7583-7588, **2005**.

[30] M. Howarth, W. Liu, S. Puthenveetil, Y. Zheng, L. F. Marshall, M. M. Schmidt, K. D. Wittrup, M. G. Bawendi et A. Y. Ting, «Monovalent, reduced-size quantum dots for imaging receptors on living cells.,» *Nat Methods*, 5, 15, 397-399, **2008**.

[31] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel et K. Johnsson, «A general method for the covalent labeling of fusion proteins with small molecules in vivo.,» *Nat Biotechnol*, 21, 11, 86-89, **2003**.

[32] A. Keppler, M. Kindermann, S. Gendreizig, H. Pick, H. Vogel et K. Johnsson, «Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro.,» *Methods*, 32, 14, 437-444, **2004**.

[33] G. V. Los et K. Wood, «The HaloTag: a novel technology for cell imaging and protein analysis.,» *Methods Mol Biol*, 356, 195-208, **2007**.

[34] Promega, «HaloTag Interchangeable Labeling Technology,» **2006**.

[35] A. Gautier, A. Juillerat, C. Heinis, J. I. R. Corrêa, M. Kindermann, F. Beaufils et K. Johnsson, «An engineered protein tag for multiprotein labeling in living cells.,» *Chem Biol*, 15, 12, 128-136, **2008**.

[36] N. B. Cole, «Site-specific protein labeling with SNAP-tags.,» *Curr Protoc Protein Sci*, 73, p. Unit 30.1, **2013**.

[37] K. Lang et J. W. Chin, «Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins.,» *Chem Rev*, 114, 19, 4764-4806, **2014**.

[38] C. Uttamapinant, K. A. White, H. Baruah, S. Thompson, M. Fernández-Suárez, S. Puthenveetil et A. Y. Ting, «A fluorophore ligase for site-specific protein labeling inside living cells.,» *Proc Natl Acad Sci U S A*, 107, 124, 10914-10919, **2010**.

[39] E. de Boer, P. Rodriguez, E. Bonte, J. Krijgsveld, E. Katsantoni, A. Heck, F. Grosveld et J. Strouboulis, «Efficient biotinylation and single-step purification of tagged transcription factors in mammalian cells and transgenic mice.,» *Proc Natl Acad Sci U S A*, 100, 113, 7480-7485, **2003**.

[40] C.-W. Lin et A. Y. Ting, «Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells.,» *J Am Chem Soc*, 128, 114, 4542-4543, **2006**.

[41] S. Tsukiji et T. Nagamune, «Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering.,» *Chembiochem*, 10, 15, 787-798, **2009**.

[42] L. W. Miller, J. Sable, P. Goelet, M. P. Sheetz et V. W. Cornish, «Methotrexate conjugates: a molecular in vivo protein tag.,» *Angew Chem Int Ed Engl*, 43, 113, 1672-1675, **2004**.

[43] B. A. Griffin, S. R. Adams et R. Y. Tsien, «Specific covalent labeling of recombinant protein molecules inside live cells.,» *Science*, 281, 15374, 269-272, **1998**.

[44] B. A. Griffin, S. R. Adams, J. Jones et R. Y. Tsien, «Fluorescent labeling of recombinant proteins in living cells with FlAsH.,» *Methods Enzymol*, 327, 565-578, **2000**.

[45] S. R. Adams, R. E. Campbell, L. A. Gross, B. R. Martin, G. K. Walkup, Y. Yao, J. Llopis et R. Y. Tsien, «New biarsenical ligands and tetracysteine motifs for protein labeling in vitro and in vivo: synthesis and biological applications.,» *J Am Chem Soc*, 124, 121, 6063-6076, **2002**.

[46] H. Ren, F. Xiao, K. Zhan, Y.-P. Kim, H. Xie, Z. Xia et J. Rao, «A biocompatible condensation reaction for the labeling of terminal cysteine residues on proteins.,» *Angew Chem Int Ed Engl*, 48, 151, 9658-9662, **2009**.

[47] I. Amitani, B. Liu, C. C. Dombrowski, R. J. Baskin et S. C. Kowalczykowski, «Watching individual proteins acting on single molecules of DNA.,» *Methods*

Enzymol, 472, 261-291, **2010**.

[48] G. Liang, H. Ren et J. Rao, «A biocompatible condensation reaction for controlled assembly of nanostructures in living cells.,» *Nat Chem,* 2, 11, 54-60, **2010**.

[49] D. P. Nguyen, T. Elliott, M. Holt, T. W. Muir et J. W. Chin, «Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation.,» *J Am Chem Soc,* 133, 130, 11418-11421, **2011**.

[50] N. Soh, «Selective Chemical Labeling of Proteins with Small Fluorescent Molecules Based on Metal-Chelation Methodology,» *Sensors,* 8, 1004-1024, **2008**.

[51] S. J. Luchansky, S. Goon et C. R. Bertozzi, «Expanding the diversity of unnatural cell-surface sialic acids.,» *Chembiochem,* 5, 13, 371-374, **2004**.

[52] J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli et C. R. Bertozzi, «Copper-free click chemistry for dynamic in vivo imaging.,» *Proc Natl Acad Sci U S A,* 104, 143, 16793-16797, **2007**.

[53] B. R. Varga, M. Kállay, K. Hegyi, S. Béni et P. Kele, «A non-fluorinated monobenzocyclooctyne for rapid copper-free click reactions.,» *Chemistry,* 18, 13, 822-828, **2012**.

[54] H. Zeng, J. Xie et P. G. Schultz, «Genetic introduction of a diketone-containing amino acid into proteins.,» *Bioorg Med Chem Lett,* 16, 120, 5356-5359, **2006**.

[55] K. Lang, L. Davis, J. Torres-Kolbus, C. Chou, A. Deiters et J. W. Chin, «Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction.,» *Nat Chem,* 4, 14, 298-304, **2012**.

[56] R. Borra, D. Dong, A. Y. Elnagar, G. A. Woldemariam et J. A. Camarero, «In-cell fluorescence activation and labeling of proteins mediated by FRET-quenched split inteins.,» *J Am Chem Soc,* 134, 114, 6344-6353, **2012**.

[57] H. D. Mootz, «Split inteins as versatile tools for protein semisynthesis.,» *Chembiochem,* 10, 116, 2579-2589, **2009**.

[58] B. R. Martin, B. N. G. Giepmans, S. R. Adams et R. Y. Tsien, «Mammalian cell-based optimization of the biarsenical-binding tetracysteine motif for improved fluorescence and affinity.,» *Nat Biotechnol,* 23, 110, 1308-1314, **2005**.

[59] K. Stroffekova, C. Proenza et K. G. Beam, «The protein-labeling reagent FLASH-EDT2 binds not only to CCXXCC motifs but also non-specifically to endogenous cysteine-rich proteins.,» *Pflugers Arch,* 442, 16, 859-866, **2001**.

[60] M. P. Bruchez, «Dark dyes-bright complexes: fluorogenic protein labeling.,» *Curr Opin Chem Biol,* 27, 18-23, **2015**.

[61] C. Jing et V. W. Cornish, «A fluorogenic TMP-tag for high signal-to-background intracellular live cell imaging.,» *ACS Chem Biol,* 8, 18, 1704-1712, **2013**.

[62] X. Sun, A. Zhang, B. Baker, L. Sun, A. Howard, J. Buswell, D. Maurel, A. Masharina, K. Johnsson, C. J. Noren, M.-Q. Xu et J. I. R. Corrêa, «Development of SNAP-tag fluorogenic probes for wash-free fluorescence imaging.,» *Chembiochem,* 12, 114, 2217-2226, **2011**.

[63] G. Lukinavičius, L. Reymond, E. D'Este, A. Masharina, F. Göttfert, H. Ta, A. Güther, M. Fournier, S. Rizzo, H. Waldmann, C. Blaukopf, C. Sommer, D. W. Gerlich, H.-D. Arndt, S. W. Hell et K. Johnsson, «Fluorogenic probes for live-cell imaging of the cytoskeleton.,» *Nat Methods,* 11, 17, 731-733, **2014**.

[64] A. E. Ondrus, H.-l. D. Lee, S. Iwanaga, W. H. Parsons, B. M. Andresen, W. E. Moerner et J. Du Bois, «Fluorescent saxitoxins for live cell imaging of single voltage-gated sodium ion channels beyond the optical diffraction limit.,» *Chem Biol,* 19, 17, 902-912, **2012**.

[65] C. A. Telmer, R. Verma, H. Teng, S. Andreko, L. Law et M. P. Bruchez, «Rapid, specific, no-wash, far-red fluorogen activation in subcellular compartments by targeted

fluorogen activating proteins.,» *ACS Chem Biol*, 10, 15, 1239-1246, **2015**.

[66] E. Prifti, L. Reymond, M. Umebayashi, R. Hovius, H. Riezman et K. Johnsson, «A fluorogenic probe for SNAP-tagged plasma membrane proteins based on the solvatochromic molecule Nile Red.,» *ACS Chem Biol*, 9, 13, 606-612, **2014**.

[67] W. Liu, F. Li, X. Chen, J. Hou, L. Yi et Y.-W. Wu, «A rapid and fluorogenic TMP-AcBOPDIPY probe for covalent labeling of proteins in live cells.,» *J Am Chem Soc*, 136, 112, 4468-4471, **2014**.

[68] Y. Hori, T. Norinobu, M. Sato, K. Arita, M. Shirakawa et K. Kikuchi, «Development of fluorogenic probes for quick no-wash live-cell imaging of intracellular proteins.,» *J Am Chem Soc*, 135, 133, 12360-12365, **2013**.

[69] D. G. Smyth, O. O. Blumenfeld et W. Konigsberg, «Reactions of N-ethylmaleimide with peptides and amino acids.,» *Biochem J*, 91, 13, 589-595, **1964**.

[70] G. Kokotos et C. Tzougraki, «Synthesis and study of substituted coumarins. A facile preparation of D,L-o-tyrosine,» *J Heterocycl Chem*, 23, 87-92, **1986**.

[71] S. Girouard, M.-H. Houle, A. Grandbois, J. W. Keillor et S. W. Michnick, «Synthesis and characterization of dimaleimide fluorogens designed for specific labeling of proteins.,» *J Am Chem Soc*, 127, 12, 559-566, **2005**.

[72] J. Guy, K. Caron, S. Dufresne, S. W. Michnick, W. G. Skene et J. W. Keillor, «Convergent preparation and photophysical characterization of dimaleimide dansyl fluorogens: elucidation of the maleimide fluorescence quenching mechanism.,» *J Am Chem Soc*, 129, 139, 11969-11977, **2007**.

[73] J. Guy, R. Castonguay, N. B. Campos-Reales Pineda, V. Jacquier, K. Caron, S. W. Michnick et J. W. Keillor, «De novo helical peptides as target sequences for a specific, fluorogenic protein labelling strategy.,» *Mol Biosyst*, 6, 16, 976-987, **2010**.

[74] K. Caron, V. Lachapelle et J. W. Keillor, «Dramatic increase of quench efficiency in

"spacerless" dimaleimide fluorogens.,» *Org Biomol Chem,* 9, 11, 185-197, **2011**.

[75] Y. Chen, C. M. Clouthier, K. Tsao, M. Strmiskova, H. Lachance et J. W. Keillor, «Coumarin-based fluorogenic probes for no-wash protein labeling.,» *Angew Chem Int Ed Engl,* 53, 150, 13785-13788, **2014**.

[76] S. Marqusee, V. H. Robbins et R. L. Baldwin, «Unusually stable helix formation in short alanine-based peptides.,» *Proc Natl Acad Sci U S A,* 86, 114, 5286-5290, **1989**.

[77] G. Merutka et E. Stellwagen, «Positional independence and additivity of amino acid replacements on helix stability in monomeric peptides.,» *Biochemistry,* 29, 14, 894-898, **1990**.

[78] S. Penel, R. G. Morrison, R. J. Mortishire-Smith et A. J. Doig, «Periodicity in alpha-helix lengths and C-capping preferences.,» *J Mol Biol,* 293, 15, 1211-1219, **1999**.

[79] V. Muñoz et L. Serrano, «Elucidating the folding problem of helical peptides using empirical parameters.,» *Nat Struct Biol,* 1, 16, 399-409, **1994**.

[80] V. Muñoz et L. Serrano, «Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides.,» *J Mol Biol,* 245, 13, 275-296, **1995**.

[81] V. Muñoz et L. Serrano, «Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence.,» *J Mol Biol,* 245, 13, 297-308, **1995**.

[82] D. G. Flint, J. R. Kumita, O. S. Smart et G. A. Woolley, «Using an azobenzene cross-linker to either increase or decrease peptide helix content upon trans-to-cis photoisomerization.,» *Chem Biol,* 9, 13, 391-397, **2002**.

[83] L. Hedstrom, «Serine protease mechanism and specificity.,» *Chem Rev,* 102, 112, 4501-4524, **2002**.

[84] A. J. Beveridge, «A theoretical study of the active sites of papain and S195C rat trypsin: implications for the low reactivity of mutant serine proteinases.,» *Protein Sci,* 5, 17, 1355-1365, **1996**.

[85] D. R. Lide, Handbook of Chemistry and Physics, D. R. Lide, Éd., CRC Press, **2003-2004**.

[86] R. L. Thurlkill, G. R. Grimsley, J. M. Scholtz et C. N. Pace, «pK values of the ionizable groups of proteins.,» *Protein Sci,* 15, 15, 1214-1218, **2006**.

[87] C. N. Pace et J. M. Scholtz, «A helix propensity scale based on experimental studies of peptides and proteins.,» *Biophys J,* 75, 11, 422-427, **1998**.

[88] N. C. Shaner, P. A. Steinbach et R. Y. Tsien, «A guide to choosing fluorescent proteins.,» *Nat Methods,* 2, 112, 905-909, **2005**.

[89] Y. Zhang, M. J. Blanden, C. Sudheer, S. A. Gangopadhyay, M. Rashidian, J. L. Hougland et M. D. Distefano, «Simultaneous Site-Specific Dual Protein Labeling Using Protein Prenyltransferases.,» *Bioconjug Chem,* 26, 112, 2542-2553, **2015**.

[90] L. Pauling, R. B. Corey et H. R. Branson, «The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain.,» *Proc Natl Acad Sci U S A,* 37, 14, 205-211, **1951**.

[91] A. J. Kirby et P. W. Lancaster, «Structure and Efficiency in Intramolecular and Enzymic Catalysis. Catalysis of Amide Hydrolysis by the Carboxy-group of Substituted Maleamic Acids,» *J Chem Soc, Perkin Trans 2,* 1206, **1972**.

[92] Y. Cheng et D. J. Patel, «An efficient system for small protein expression and refolding.,» *Biochem Biophys Res Commun,* 317, 12, 401-405, **2004**.

[93] P. Riggs, «Expression and purification of recombinant proteins by fusion to maltose-binding protein.,» *Mol Biotechnol,* 15, 11, 51-63, **2000**.

[94] P. Riggs, «Expression and purification of maltose-binding protein fusions.,» *Curr Protoc Mol Biol*, Chapter 16, **2001**.

[95] C. K. Riener, G. Kada et H. J. Gruber, «Quick measurement of protein sulfhydryls with Ellman's reagent and with 4,4'-dithiodipyridine.,» *Anal Bioanal Chem*, 373, 14-5, 266-276, **2002**.

[96] B. Fierz, A. Reiner et T. Kiefhaber, «Local conformational dynamics in alpha-helices measured by fast triplet transfer.,» *Proc Natl Acad Sci U S A*, 106, 14, 1057-1062, **2009**.

[97] B. H. Zimm et J. K. Bragg, «Theory of the Phase Transition between Helix and Random Coil in Polypeptide Chains,» *J Chem Phys*, 31, 526-535, **1959**.

[98] H. Qian et J. A. Schellman, «Helix-Coil Theories: A Comparative Study for Tinite Length Polypeptides,» *J Phys Chem*, 96, 3987-3994, **1992**.

[99] Z. Cheng, M. Miskolzie et R. E. Campbell, «In vivo screening identifies a highly folded beta-hairpin peptide with a structured extension.,» *Chembiochem*, 8, 18, 880-883, **2007**.

[100] M. T. Pastor, M. López de la Paz, E. Lacroix, L. Serrano et E. Pérez-Payá, «Combinatorial approaches: a new tool to search for highly structured beta-hairpin peptides.,» *Proc Natl Acad Sci U S A*, 99, 12, 614-619, **2002**.

[101] M. F. Debets, J. C. M. van Hest et F. P. J. T. Rutjes, «Bioorthogonal labelling of biomolecules: new functional handles and ligation methods.,» *Org Biomol Chem*, 11, 138, 6439-6455, **2013**.

[102] R. Huisgen, G. Szeimies et L. Mobius, «Chemische Berichte-Recueil,» *Chem. Ber.*, 100, 2494, **1967**.

[103] H. C. Kolb, M. G. Finn et K. B. Sharpless, «Click Chemistry: Diverse Chemical Function from a Few Good Reactions.,» *Angew Chem Int Ed Engl*, 40, 111, 2004-

2021, **2001**.

[104] V. V. Rostovtsev, L. G. Green, V. V. Fokin et K. B. Sharpless, «A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes.,» *Angew Chem Int Ed Engl,* 41, 114, 2596-2599, **2002**.

[105] N. J. Agard, J. A. Prescher et C. R. Bertozzi, «A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems.,» *J Am Chem Soc,* 126, 146, 15046-15047, **2004**.

[106] S. T. Laughlin, J. M. Baskin, S. L. Amacher et C. R. Bertozzi, «In vivo imaging of membrane-associated glycans in developing zebrafish.,» *Science,* 320, 15876, 664-667, **2008**.

[107] H. Koo, S. Lee, J. H. Na, S. H. Kim, S. K. Hahn, K. Choi, I. C. Kwon, S. Y. Jeong et K. Kim, «Bioorthogonal copper-free click chemistry in vivo for tumor-targeted delivery of nanoparticles.,» *Angew Chem Int Ed Engl,* 51, 147, 11836-11840, **2012**.

[108] Y. Brudno, R. M. Desai, B. J. Kwee, N. S. Joshi, M. Aizenberg et D. J. Mooney, «In vivo targeting through click chemistry.,» *ChemMedChem,* 10, 14, 617-620, **2015**.

[109] M. L. Blackman, M. Royzen et J. M. Fox, «Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity.,» *J Am Chem Soc,* 130, 141, 13518-13519, **2008**.

[110] J. Schoch, M. Wiessler et A. Jäschke, «Post-synthetic modification of DNA by inverse-electron-demand Diels-Alder reaction.,» *J Am Chem Soc,* 132, 126, 8846-8847, **2010**.

[111] J. Yang, Y. Liang, J. Šečkutė, K. N. Houk et N. K. Devaraj, «Synthesis and reactivity comparisons of 1-methyl-3-substituted cyclopropene mini-tags for tetrazine bioorthogonal reactions.,» *Chemistry,* 20, 112, 3365-3375, **2014**.

[112] Z. Yu et Q. Lin, «Design of spiro[2.3]hex-1-ene, a genetically encodable double-

strained alkene for superfast photoclick chemistry.,» *J Am Chem Soc,* 136, 111, 4153-4156, **2014**.

[113] J. Moran, M. C.S. et J. Pezacki, «Strain-promoted 1,3-dipolar ccycloaddition of diazo compounds with cyclooctynes,» *Can J Chem,* 89, 148-151, **2011**.

[114] D. A. MacKenzie, A. R. Sherratt, M. Chigrinova, L. L. W. Cheung et J. P. Pezacki, «Strain-promoted cycloadditions involving nitrones and alkynes--rapid tunable reactions for bioorthogonal labeling.,» *Curr Opin Chem Biol,* 21, 81-88, **2014**.

[115] R. Rossin, P. R. Verkerk, S. M. van den Bosch, R. C. M. Vulders, I. Verel, J. Lub et M. S. Robillard, «In vivo chemistry for pretargeted tumor imaging in live mice.,» *Angew Chem Int Ed Engl,* 49, 119, 3375-3378, **2010**.

[116] C. S. McKay, J. A. Blake, J. Cheng, D. C. Danielson et J. P. Pezacki, «Strain-promoted cycloadditions of cyclic nitrones with cyclooctynes for labeling human cancer cells.,» *Chem Commun (Camb),* 47, 136, 10040-10042, **2011**.

[117] L. Wang et P. G. Schultz, «Expanding the genetic code.,» *Angew Chem Int Ed Engl,* 44, 11, 34-66, **2004**.

[118] V. Schnaible, S. Wefing, A. Resemann, D. Suckau, A. Bücker, S. Wolf-Kümmeth et D. Hoffmann, «Screening for disulfide bonds in proteins by MALDI in-source decay and LIFT-TOF/TOF-MS.,» *Anal Chem,* 74, 119, 4980-4988, **2002**.

[119] M. Z. Lin, M. R. McKeown, H.-L. Ng, T. A. Aguilera, N. C. Shaner, R. E. Campbell, S. R. Adams, L. A. Gross, W. Ma, T. Alber et R. Y. Tsien, «Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals.,» *Chem Biol,* 16, 111, 1169-1179, **2009**.

[120] S. M. F. G. Gillet, R. A. Chica, J. W. Keillor et J. N. Pelletier, «Expression and rapid purification of highly active hexahistidine-tagged guinea pig liver transglutaminase.,» *Protein Expr Purif,* 33, 12, 256-264, **2004**.

[121] H. Eagle, «Amino acid metabolism in mammalian cell cultures.,» *Science*, 130, 13373, 432-437, **1959**.

[122] E. Kariya, S. Ohki, T. Hayano et M. Kainosho, «Backbone 1H, 13C, and 15N resonance assignments of an 18.2 kDa protein, E. coli peptidyl-prolyl cis-trans isomerase b (EPPIb).,» *J Biomol NMR*, 18, 11, 75-76, **2000**.

[123] K. Ozawa, M. J. Headlam, P. M. Schaeffer, B. R. Henderson, N. E. Dixon et G. Otting, «Optimization of an Escherichia coli system for cell-free synthesis of selectively N-labelled proteins for rapid analysis by NMR spectroscopy.,» *Eur J Biochem*, 271, 120, 4084-4093, **2004**.

[124] K. Ozawa, M. J. Headlam, D. Mouradov, S. J. Watt, J. L. Beck, K. J. Rodgers, R. T. Dean, T. Huber, G. Otting et N. E. Dixon, «Translational incorporation of L-3,4-dihydroxyphenylalanine into proteins.,» *FEBS J*, 272, 112, 3162-3171, **2005**.

[125] T. Kigawa, T. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, T. Shibata et S. Yokoyama, «Cell-free production and stable-isotope labeling of milligram quantities of proteins.,» *FEBS Lett*, 442, 11, 15-19, **1999**.

[126] T. Torizawa, M. Shimizu, M. Taoka, H. Miyano et M. Kainosho, «Efficient production of isotopically labeled proteins by cell-free synthesis: a practical protocol.,» *J Biomol NMR*, 30, 13, 311-325, **2004**.

[127] L. Banci, I. Bertini, S. Ciofi-Baffoni, N. G. Kandias, N. J. Robinson, G. A. Spyroulias, X.-C. Su, S. Tottey et M. Vanarotti, «The delivery of copper for thylakoid import observed by NMR.,» *Proc Natl Acad Sci U S A*, 103, 122, 8320-8325, **2006**.

[128] A. {De Simone}, A. Cavalli, S.-T. D. Hsu, W. Vranken et M. Vendruscolo, «Accurate random coil chemical shifts from an analysis of loop regions in native states of proteins.,» *J Am Chem Soc*, 131, 145, 16332-16333, **2009**.

[129] D. S. Wishart, B. D. Sykes et F. M. Richards, «Relationship between nuclear magnetic resonance chemical shift and protein secondary structure.,» *J Mol Biol*, 222, 12, 311-

333, **1991**.

[130] D. C. Dalgarno, B. A. Levine et R. J. Williams, «Structural information from NMR secondary chemical shifts of peptide alpha C-H protons in proteins.,» *Biosci Rep*, 3, 15, 443-452, **1983**.

[131] A. Pastore et V. Saudek, «The Relationship between Chemical Shift and Secondary Structure in Proteins,» *J Magn Reson*, 90, 165-176, **1990**.

[132] G. Cornilescu, F. Delaglio et A. Bax, «Protein backbone angle restraints from searching a database for chemical shift and sequence homology.,» *J Biomol NMR*, 13, 13, 289-302, **1999**.

[133] Y. Shen et A. Bax, «Identification of helix capping and b-turn motifs from NMR chemical shifts.,» *J Biomol NMR*, 52, 13, 211-232, **2012**.

[134] D. Roccatano, G. Colombo, M. Fioroni et A. E. Mark, «Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: a molecular dynamics study.,» *Proc Natl Acad Sci U S A*, 99, 119, 12179-12184, **2002**.

[135] J. F. Povey, C. M. Smales, S. J. Hassard et M. J. Howard, «Comparison of the effects of 2,2,2-trifluoroethanol on peptide and protein structure and function.,» *J Struct Biol*, 157, 12, 329-338, **2007**.

[136] X. Wei, S. Ding, Y. Jiang, X.-G. Zeng et H.-M. Zhou, «Conformational changes and inactivation of bovine carbonic anhydrase II in 2,2,2-trifluoroethanol solutions.,» *Biochemistry (Mosc)*, 71 Suppl 1, S77-S82, **2006**.

[137] K. Shiraki, K. Nishikawa et Y. Goto, «Trifluoroethanol-induced stabilization of the alpha-helical structure of beta-lactoglobulin: implication for non-hierarchical protein folding.,» *J Mol Biol*, 245, 12, 180-194, **1995**.

[138] R. A. Bednar, «Reactivity and pH dependence of thiol conjugation to N-ethylmaleimide: detection of a conformational change in chalcone isomerase.,»

Biochemistry, 29, 115, 3684-3690, **1990**.

[139] G. Bulaj, T. Kortemme et D. P. Goldenberg, «Ionization-reactivity relationships for cysteine thiols in polypeptides.,» *Biochemistry,* 37, 125, 8965-8972, **1998**.

[140] C. Tanford, «The Interpretation of Hydrogen Ion Titration Curves of Proteins,» *Advanced Protein Chemistry,* 17, 69-165, **1962**.

[141] Y. Nozaki et C. Tanford, «Examination of titration behaviour,» *Methods in Enzymology,* 11, 715-734, **1967**.

[142] S. R. Lehrman, J. L. Tuls et M. Lund, «Peptide alpha-helicity in aqueous trifluoroethanol: correlations with predicted alpha-helicity and the secondary structure of the corresponding regions of bovine growth hormone.,» *Biochemistry,* 29, 123, 5590-5596, **1990**.

[143] J. W. Nelson et N. R. Kallenbach, «Persistence of the alpha-helix stop signal in the S-peptide in trifluoroethanol solutions.,» *Biochemistry*, 28, 112, 5256-5261, **1989**.

[144] M. I. Page, «The energetics of intramolecular reactions and enzyme catalysis.,» *Philos Trans R Soc Lond B Biol Sci,* 332, 11263, 149-156, **1991**.

[145] M. I. Page et W. P. Jencks, «Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect.,» *Proc Natl Acad Sci U S A,* 68, 18, 1678-1683, **1971**.

[146] A. Tomas, C. E. Futter et E. R. Eden, «EGF receptor trafficking: consequences for signaling and cancer.,» *Trends Cell Biol*, 24, 11, 26-34, **2014**.

[147] R. D. Kornberg, «Structure of chromatin.,» *Annu Rev Biochem,* 46, 931-954, **1977**.

[148] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer et A. Bax, «NMRPipe: a multidimensional spectral processing system based on UNIX pipes.,» *J Biomol NMR,* 6, 13, 277-293, **1995**.

[149] T. D. Goddard et D. G. Kneller, «SPARKY 3,» San Francisco.

[150] K. Wüthrich, «NMR - this other method for protein and nucleic acid structure determination.,» *Acta Crystallogr D Biol Crystallogr,* 51, 249-270, **1995**.

[151] L. E. Kay et K. H. Gardner, «Solution NMR spectroscopy beyond 25 kDa.,» *Curr Opin Struct Biol,* 7, 15, 722-731, **1997**.

[152] T. Yamazaki, W. Lee, C. H. Arrowsmith, D. R. Muhandiram et L. E. Kay, «A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity,» *J Am Chem Soc,* 116, 11655-11666, **1994**.

[153] S. W. Muchmore, M. Sattler, H. Liang, R. P. Meadows, J. E. Harlan, H. S. Yoon, D. Nettesheim, B. S. Chang, C. B. Thompson, S. L. Wong, S. L. Ng et S. W. Fesik, «X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death.,» *Nature,* 381, 16580, 335-341, **1996**.

[154] R. A. Venters, C. C. Huang, 2. B. Farmer, R. Trolard, L. D. Spicer et C. A. Fierke, «High-level 2H/13C/15N labeling of proteins for NMR studies.,» *J Biomol NMR,* 5, 14, 339-344, **1995**.

[155] D. S. Garrett, Y. J. Seok, D. I. Liao, A. Peterkofsky, A. M. Gronenborn et G. M. Clore, «Solution structure of the 30 kDa N-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase system by multidimensional NMR.,» *Biochemistry,* 36, 19, 2517-2530, **1997**.

[156] K. H. Gardner et L. E. Kay, «The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins.,» *Annu Rev Biophys Biomol Struct,* 27, 357-406, **1998**.

[157] R. Kerfah, M. J. Plevin, R. Sounier, P. Gans et J. Boisbouvier, «Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins.,» *Curr Opin Struct Biol,* 32, 113-122, **2015**.

[158] R. Sprangers et L. E. Kay, «Quantitative dynamics and binding studies of the 20S proteasome by NMR.,» *Nature,* 445, 17128, 618-622, **2007**.

[159] N. K. Goto, K. H. Gardner, G. A. Mueller, R. C. Willis et L. E. Kay, «A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins.,» *J Biomol NMR,* 13, 14, 369-374, **1999**.

[160] N. K. Goto et L. E. Kay, «New developments in isotope labeling strategies for protein solution NMR spectroscopy.,» *Curr Opin Struct Biol,* 10, 15, 585-592, **2000**.

[161] X.-C. Su et G. Otting, «Paramagnetic labelling of proteins and oligonucleotides for NMR.,» *J Biomol NMR,* 46, 11, 101-112, **2010**.

[162] K. N. Allen et B. Imperiali, «Lanthanide-tagged proteins--an illuminating partnership.,» *Curr Opin Chem Biol,* 14, 12, 247-254, **2010**.

[163] G. Otting, «Prospects for lanthanides in structural biology by NMR.,» *J Biomol NMR,* 42, 11, 1-9, **2008**.

[164] R. R. Biekofsky, F. W. Muskett, J. M. Schmidt, S. R. Martin, J. P. Browne, P. M. Bayley et J. Feeney, «NMR approaches for monitoring domain orientations in calcium-binding proteins in solution using partial replacement of Ca2+ by Tb3+.,» *FEBS Lett,* 460, 13, 519-526, **1999**.

[165] M. a. B. I. a. J. M. B. L. a. L. Y.-M. a. L. G. a. L. C. Allegrozzi, «Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion,» *J. Am. Chem. Soc.,* 122, 4154-4161, **2000**.

[166] G. Pintacuda, M. John, X.-C. Su et G. Otting, «NMR structure determination of protein-ligand complexes by lanthanide labeling.,» *Acc Chem Res,* 40, 13, 206-212, **2007**.

[167] I. Bertini, M. B. Janik, Y. M. Lee, C. Luchinat et A. Rosato, «Magnetic susceptibility tensor anisotropies for a lanthanide ion series in a fixed protein matrix.,» *J Am Chem*

Soc, 123, 118, 4181-4188, **2001**.

[168] L. Banci, I. Bertini, K. L. Bren, M. A. Cremonini, H. B. Gray, C. Luchinat et P. Turano, «The use of pseudocontact shifts to refine solution structures of paramagnetic metalloproteins: Met80Ala cyano-cytochrome c as an example,» *J Biol Inorg Chem,* 1, 11-126, **1996**.

[169] I. Bertini, C. Luchinat et G. Parigi, «Paramagnetic Constraints: An Aid for Quick Solution Structure Determination of Paramagnetic Metalloproteins,» *Concepts in Magn Reson,* 14, 259-286, **2002**.

[170] H. Yagi, K. B. Pilla, A. Maleckis, B. Graham, T. Huber et G. Otting, «Three-dimensional protein fold determination from backbone amide pseudocontact shifts generated by lanthanide tags at multiple sites.,» *Structure,* 21, 16, 883-890, **2013**.

[171] G. Pintacuda, A. Y. Park, M. A. Keniry, N. E. Dixon et G. Otting, «Lanthanide labeling offers fast NMR approach to 3D structure determinations of protein-protein complexes.,» *J Am Chem Soc,* 128, 111, 3696-3702, **2006**.

[172] M. John, G. Pintacuda, A. Y. Park, N. E. Dixon et G. Otting, «Structure determination of protein-ligand complexes by transferred paramagnetic shifts.,» *J Am Chem Soc,* 128, 139, 12910-12916, **2006**.

[173] R. Barbieri, I. Bertini, G. Cavallaro, Y.-M. Lee, C. Luchinat et A. Rosato, «Paramagnetically induced residual dipolar couplings for solution structure determination of lanthanide binding proteins.,» *J Am Chem Soc,* 124, 119, 5581-5587, **2002**.

[174] J. Feeney, B. Birdsall, A. F. Bradbury, R. R. Biekofsky et P. M. Bayley, «Calmodulin tagging provides a general method of using lanthanide induced magnetic field orientation to observe residual dipolar couplings in proteins in solution.,» *J Biomol NMR,* 21, 11, 41-48, **2001**.

[175] C. Ma et S. J. Opella, «Lanthanide ions bind specifically to an added "EF-hand" and orient a membrane protein in micelles for solution NMR spectroscopy.,» *J Magn Reson*, 146, 12, 381-384, **2000**.

[176] M. Prudêncio, J. Rohovec, J. A. Peters, E. Tocheva, M. J. Boulanger, M. E. P. Murphy, H.-J. Hupkes, W. Kosters, A. Impagliazzo et M. Ubbink, «A caged lanthanide complex as a paramagnetic shift agent for protein NMR.,» *Chemistry*, 10, 113, 3252-3260, **2004**.

[177] B. Graham, C. T. Loh, J. D. Swarbrick, P. Ung, J. Shin, H. Yagi, X. Jia, S. Chhabra, N. Barlow, G. Pintacuda, T. Huber et G. Otting, «DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra.,» *Bioconjug Chem*, 22, 110, 2118-2125, **2011**.

[178] H. Yagi, A. Maleckis et G. Otting, «A systematic study of labelling an α-helix in a protein with a lanthanide using IDA-SH or NTA-SH tags.,» *J Biomol NMR*, 55, 12, 157-166, **2013**.

[179] X. Jia, A. Maleckis, T. Huber et G. Otting, «4,4'-dithiobisdipicolinic acid: a small and convenient lanthanide binding tag for protein NMR spectroscopy.,» *Chemistry*, 17, 124, 6830-6836, **2011**.

[180] Q.-F. Li, Y. Yang, A. Maleckis, G. Otting et X.-C. Su, «Thiol-ene reaction: a versatile tool in site-specific labelling of proteins with chemically inert tags for paramagnetic NMR.,» *Chem Commun (Camb)*, 48, 121, 2704-2706, **2012**.

[181] J. Evenäs, V. Tugarinov, N. R. Skrynnikov, N. K. Goto, R. Muhandiram et L. E. Kay, «Ligand-induced structural changes to maltodextrin-binding protein as studied by solution NMR spectroscopy.,» *J Mol Biol*, 309, 14, 961-974, **2001**.

[182] K. H. Gardner, X. Zhang, K. Gehring et L. E. Kay, «Solution NMR Studies of a 42 kDa Escherichia Coli Maltose Binding Protein/ beta-Cyclodextrin Complex: Chemical Shift Assignment and Analysis,» *J. Am. Chem. Soc.*, 120, 11738-11748, **1998**.

[183] A. Potapov, H. Yagi, T. Huber, S. Jergic, N. E. Dixon, G. Otting et D. Goldfarb, «Nanometer-scale distance measurements in proteins using Gd3+ spin labeling.,» *J Am Chem Soc*, 132, 126, 9040-9048, **2010**.

[184] D. L. Burk, N. Ghuman, L. E. Wybenga-Groot et A. M. Berghuis, «X-ray structure of the AAC(6')-Ii antibiotic resistance enzyme at 1.8 A resolution; examination of oligomeric arrangements in GNAT superfamily members.,» *Protein Sci*, 12, 13, 426-437, **2003**.

[185] L. A. Freiburger, O. M. Baettig, T. Sprules, A. M. Berghuis, K. Auclair et A. K. Mittermaier, «Competing allosteric mechanisms modulate substrate binding in a dimeric enzyme.,» *Nat Struct Mol Biol*, 18, 13, 288-294, **2011**.

[186] G. D. Wright et P. Ladak, «Overexpression and characterization of the chromosomal aminoglycoside 6'-N-acetyltransferase from Enterococcus faecium.,» *Antimicrob Agents Chemother*, 41, 15, 956-960, **1997**.

[187] J. W. Williams et D. B. Northrop, «Kinetic mechanisms of gentamicin acetyltransferase I. Antibiotic-dependent shift from rapid to nonrapid equilibrium random mechanisms.,» *J Biol Chem*, 253, 117, 5902-5907, **1978**.

[188] E. N. Rizkalla et G. R. Choppin, «Hydration of lanthanides and actinides in solution,» *J Alloys Compd*, 180, 325-336, **1992**.

[189] J. B. Lamture, Z. H. Zhou, A. S. Kuet T. G. Wensel, «Luminescence Properties of Terbium(III) Complexes with 4-Substituted Dipicolinic Acid Analogues,» *Inorg Chem*, 34, 864-869, **1995**.

[190] A.-S. Chauvin, F. Gumy, D. Imbert et J.-C. G. Bunzli, «Europium and Terbium tris9Dipicolinates) as Secondary Standards for Quantum Yield Determination,» *Spectrosc Lett*, 37, 517-532, **2006**.

[191] N. Candelon, N. D. H?dade, M. Matache, J.-L. Canet, F. Cisnetti, D. P. Funeriu, L. Nauton et A. Gautier, «Luminogenic "clickable" lanthanide complexes for protein

labeling.,» *Chem Commun (Camb)*, 49, 180, 9206-9208, **2013**.

[192] M. D. Purdy, P. Ge, J. Chen, P. R. Selvin et M. C. Wiener, «Thiol-reactive lanthanide chelates for phasing protein X-ray diffraction data.,» *Acta Crystallogr D Biol Crystallogr*, 58, 1111-1117, **2002**.

[193] T. Madl, W. Bermel et K. Zangger, «Use of relaxation enhancements in a paramagnetic environment for the structure determination of proteins using NMR spectroscopy.,» *Angew Chem Int Ed Engl*, 48, 144, 8259-8262, **2009**.

[194] O. Julien, Y.-B. Sun, A. C. Knowles, B. D. Brandmeier, R. E. Dale, D. R. Trentham, J. E. T. Corrie, B. D. Sykes et M. Irving, «Toward protein structure in situ: comparison of two bifunctional rhodamine adducts of troponin C.,» *Biophys J*, 93, 13, 1008-1020, **2007**.

[195] O. Julien, P. Mercier, L. Spyracopoulos, J. E. T. Corrie et B. D. Sykes, «NMR studies of the dynamics of a bifunctional rhodamine probe attached to troponin C.,» *J Am Chem Soc*, 130, 18, 2602-2609, **2008**.

[196] P. Gorostiza, M. Volgraf, R. Numano, S. Szobota, D. Trauner et E. Y. Isacoff, «Mechanisms of photoswitch conjugation and light activation of an ionotropic glutamate receptor.,» *Proc Natl Acad Sci U S A*, 104, 126, 10865-10870, **2007**.

[197] A. Martín-Quirós, L. Nevola, K. Eckelt, S. Madurga, P. Gorostiza et E. Giralt, «Absence of a stable secondary structure is not a limitation for photoswitchable inhibitors of β-arrestin/β-Adaptin 2 protein-protein interaction.,» *Chem Biol*, 22, 11, 31-37, **2015**.

[198] A. M. Ali et G. A. Woolley, «The effect of azobenzene cross-linker position on the degree of helical peptide photo-control.,» *Org Biomol Chem*, 11, 132, 5325-5331, **2013**.

[199] S. Samanta et G. A. Woolley, «Bis-azobenzene crosslinkers for photocontrol of

peptide structure.,» *Chembiochem,* 12, 111, 1712-1723, **2011**.

[200] J. H. Tomlinson, V. L. Green, P. J. Baker et M. P. Williamson, «Structural origins of pH-dependent chemical shifts in the B1 domain of protein G.,» *Proteins,* 78, 114, 3000-3016, **2010**.

[201] B. K. McMahon et T. Gunnlaugsson, «Selective detection of the reduced form of glutathione (GSH) over the oxidized (GSSG) form using a combination of glutathione reductase and a Tb(III)-cyclen maleimide based lanthanide luminescent 'switch on' assay.,» *J Am Chem Soc,* 134, 126, 10725-10728, **2012**.

[202] K. Pervushin, R. Riek, G. Wider et K. Wüthrich, «Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.,» *Proc Natl Acad Sci U S A,* 94, 123, 12366-12371, **1997**.

[203] K. Moffat, «Time-resolved macromolecular crystallography.,» *Annu Rev Biophys Biophys Chem,* 18, 309-332, **1989**.

[204] J. Hajdu, «Fast crystallography and time-resolved structures.,» *Annu Rev Biophys Biomol Struct,* 22, 467-498, **1993**.

[205] J. Hajdu, R. Neutze, T. Sjögren, K. Edman, A. Szöke, R. C. Wilmouth et C. M. Wilmot, «Analyzing protein functions in four dimensions.,» *Nat Struct Biol,* 7, 111, 1006-1012, **2000**.

[206] I. Schlichting et K. Chu, «Trapping intermediates in the crystal: ligand binding to myoglobin.,» *Curr Opin Struct Biol,* 10, 16, 744-752, **2000**.

[207] L. Ronda, S. Bruno, S. Bettati, P. Storici et A. Mozzarelli, «From protein structure to function via single crystal optical spectroscopy.,» *Front Mol Biosci,* 2, p. 12, **2015**.

[208] G. Taylor, «The phase problem.,» *Acta Crystallogr D Biol Crystallogr,* 59, 1881-1890, **2003**.

[209] G. L. Taylor, «Introduction to phasing.,» *Acta Crystallogr D Biol Crystallogr*, 66, 325-338, **2010**.

[210] K. Cowtan, «Phase Problem in X-ray Crystallography, and Its Solution,» *eLS*, **2003**.

[211] G. Scapin, «Molecular replacement then and now.,» *Acta Crystallogr D Biol Crystallogr*, 69, 2266-2275, **2013**.

[212] M. G. Rossmann, The Molecular Replacement Method, New York: Gordon & Breach, **1972**.

[213] M. G. Rossmann, «Molecular replacement--historical background.,» *Acta Crystallogr D Biol Crystallogr*, 57, 1360-1366, **2001**.

[214] M. G. Rossmann et D. M. Blow, «The Detection of Sub-Units Within the Crystallographic Asymmetric Unit,» *Acta Cryst*, 15, **1962**.

[215] M. Fujinaga et R. J. Read, «Experiences with a New Translation-Function Program,» *J Appl Cryst*, 20, 517-521, **1987**.

[216] R. A. Crowther, The Molecular Replacement Method, M. G. Rossmann, Éd., New York: Gordon & Breach, 173-178, **1972**.

[217] N. Ban, C. Escobar, R. Garcia, K. Hasel, J. Day, A. Greenwood et A. McPherson, «Crystal structure of an idiotype-anti-idiotype Fab complex.,» *Proc Natl Acad Sci U S A*, 91, 15, 1604-1608, **1994**.

[218] A. Vagin et A. Teplyakov, «MOLREP: an Automated Program fro Molecular Replacement,» *J Appl Cryst*, 30, 1022-1025, **1997**.

[219] A. Vagin et A. Teplyakov, «Molecular replacement with MOLREP.,» *Acta Crystallogr D Biol Crystallogr*, 66, 22-25, **2010**.

[220] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni et R.

J. Read, «Phaser crystallographic software.,» *J Appl Crystallogr,* 40, 658-674, **2007**.

[221] M. F. Perutz, «Isomorphous Replacement and Phase Determination in Non-centrosymmetric Space Groups,» *Acta Cryst,* 9, 867-873, **1956**.

[222] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff et D. C. Phillips, «A three-dimensional model of the myoglobin molecule obtained by x-ray analysis.,» *Nature,* 181, 14610, 662-666, **1958**.

[223] Z. Dauter, M. Dauter et K. R. Rajashankar, «Novel approach to phasing proteins: derivatization by short cryo-soaking with halides.,» *Acta Crystallogr D Biol Crystallogr,* 56, 232-237, **2000**.

[224] R. A. Nagem, Z. Dauter et I. Polikarpov, «Protein crystal structure solution by fast incorporation of negatively and positively charged anomalous scatterers.,» *Acta Crystallogr D Biol Crystallogr,* 57, 996-1002, **2001**.

[225] N. Moiseeva et M. Allaire, «Using barium ions for heavy-atom derivatization and phasing of xylanase II from Trichoderma longibrachiatum.,» *Acta Crystallogr D Biol Crystallogr,* 63, 1025-1028, **2007**.

[226] A. Cohen, P. Ellis, N. Kresge et S. M. Soltis, «MAD phasing with krypton.,» *Acta Crystallogr D Biol Crystallogr,* 57, 233-238, **2001**.

[227] P. D. Sun, S. Radaev et M. Kattah, «Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases.,» *Acta Crystallogr D Biol Crystallogr,* 58, 1092-1098, **2002**.

[228] P. D. Sun et S. Radaev, «Generating isomorphous heavy-atom derivatives by a quick-soak method. Part II: phasing of new structures.,» *Acta Crystallogr D Biol Crystallogr,* 58, 1099-1103, **2002**.

[229] T. Beck, A. Krasauskas, T. Gruene et G. M. Sheldrick, «A magic triangle for experimental phasing of macromolecules.,» *Acta Crystallogr D Biol Crystallogr,* 64,

1179-1182, **2008**.

[230] T. Beck, T. Gruene et G. M. Sheldrick, «The magic triangle goes MAD: experimental phasing with a bromine derivative.,» *Acta Crystallogr D Biol Crystallogr,* 66, 374-380, **2010**.

[231] J. Agniswamy, M. G. Joyce, C. H. Hammer et P. D. Sun, «Towards a rational approach for heavy-atom derivative screening in protein crystallography.,» *Acta Crystallogr D Biol Crystallogr,* 64, 354-367, **2008**.

[232] M. G. Joyce, S. Radaev et P. D. Sun, «A rational approach to heavy-atom derivative screening.,» *Acta Crystallogr D Biol Crystallogr,* 66, 358-365, **2010**.

[233] K. Fütterer, C. L. Murray, R. S. Bhatnagar, G. W. Gokel, J. I. Gordon et G. Waksman, «Crystallographic phasing of myristoyl-CoA-protein N-myristoyltransferase using an iodinated analog of myristoyl-CoA.,» *Acta Crystallogr D Biol Crystallogr,* 57, 393-400, **2001**.

[234] J. C. Spurlino, G. Y. Lu et F. A. Quiocho, «The 2.3-A resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis.,» *J Biol Chem,* 266, 18, 5202-5219, **1991**.

[235] R. Abdur, O. O. Gerlits, J. Gan, J. Jiang, J. Salon, A. Y. Kovalevsky, A. A. Chumanevich, I. T. Weber et Z. Huang, «Novel complex MAD phasing and RNase H structural insights using selenium oligonucleotides.,» *Acta Crystallogr D Biol Crystallogr,* 70, 354-361, **2014**.

[236] J. Xie, L. Wang, N. Wu, A. Brock, G. Spraggon et P. G. Schultz, «The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination.,» *Nat Biotechnol,* 22, 110, 1297-1301, **2004**.

[237] W. A. Hendrickson, J. R. Horton et D. M. LeMaster, «Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure.,» *EMBO J,* 9, 15, 1665-1672,

1990.

[238] L. Moroder, «Isosteric replacement of sulfur with other chalcogens in peptides and proteins.,» *J Pept Sci,* 11, 14, 187-214, **2005**.

[239] C. Evrard, J. Fastrez et J. P. Declercq, «The incorporation of a non-natural amino acid (aza-tryptophan) help to crystallize a protein and to solve its crystal structure. Application to bacteriophage lambda lysozyme.,» *Acta Crystallogr D Biol Crystallogr,* 55, 430-435, **1999**.

[240] C. A. McNamara, F. King et M. Bradley, «Experimental design and the optimization of a polymer supported palladium complex for use in the Heck reaction,» *Tetrahedron Lett,* 45, 8239-8243, **2004**.

[241] A. J. Sharff, L. E. Rodseth, J. C. Spurlino et F. A. Quiocho, «Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis.,» *Biochemistry,* 31, 144, 10657-10663, **1992**.

[242] A. J. Sharff, L. E. Rodseth et F. A. Quiocho, «Refined 1.8-A structure reveals the mode of binding of beta-cyclodextrin to the maltodextrin binding protein.,» *Biochemistry,* 32, 140, 10553-10559, **1993**.

[243] A. Abuchowski, J. R. McCoy, N. C. Palczuk, T. van Es et F. F. Davis, «Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase.,» *J Biol Chem,* 252, 111, 3582-3586, **1977**.

[244] J. E. Whitaker, R. P. Haugland, P. L. Moore, P. C. Hewitt, M. Reese et R. P. Haugland, «Cascade blue derivatives: water soluble, reactive, blue emission dyes evaluated as fluorescent labels and tracers.,» *Anal Biochem,* 198, 11, 119-130, **1991**.

[245] T. Terai et T. Nagano, «Small-molecule fluorophores and fluorescent probes for bioimaging.,» *Pflugers Arch,* 465, 13, 347-359, **2013**.

[246] M. P. D. Hatfield, R. F. Murphy et S. Lovas, «Molecular dynamics analysis of the conformations of a beta-hairpin miniprotein.,» *J Phys Chem B*, 114, 18, 3028-3037, **2010**.

[247] B. Maestro, C. M. Santiveri, M. A. Jiménez et J. M. Sanz, «Structural autonomy of a ?-hairpin peptide derived from the pneumococcal choline-binding protein LytA.,» *Protein Eng Des Sel*, 24, 11-2, 113-122, **2011**.

[248] M. S. Searle, R. Zerella, D. H. Williams et L. C. Packman, «Native-like beta-hairpin structure in an isolated fragment from ferredoxin: NMR and CD studies of solvent effects on the N-terminal 20 residues.,» *Protein Eng*, 9, 17, 559-565, **1996**.

[249] L. Wu, D. McElheny, R. Huang et T. A. Keiderling, «Role of tryptophan-tryptophan interactions in Trpzip beta-hairpin formation, structure, and stability.,» *Biochemistry*, 48, 143, 10362-10371, **2009**.

[250] L. Wu, D. McElheny, T. Takekiyo et T. A. Keiderling, «Geometry and efficacy of cross-strand Trp/Trp, Trp/Tyr, and Tyr/Tyr aromatic interaction in a beta-hairpin peptide.,» *Biochemistry*, 49, 122, 4705-4714, **2010**.

[251] Y. Yano et K. Matsuzaki, «Tag-probe labeling methods for live-cell imaging of membrane proteins.,» *Biochim Biophys Acta*, 1788, 110, 2124-2131, **2009**.

[252] S. Uchinomiya, A. Ojida et I. Hamachi, «Peptide tag/probe pairs based on the coordination chemistry for protein labeling.,» *Inorg Chem*, 53, 14, 1816-1823, **2014**.

[253] Y. Yano, N. Furukawa, S. Ono, Y. Takeda et K. Matsuzaki, «Selective amine labeling of cell surface proteins guided by coiled-coil assembly.,» *Biopolymers*, **2015**.

List of Publications

Strmiskova, M.; Keillor, J. W.; « Optimized peptide tags for fluorogenic protein labelling. » (manuscript in preparation to be submitted to *Chem & Biol*)

Strmiskova, M., Keillor, J. W.; «Peptide tags for fluorescent labelling of proteins. » US Provisional Patent Application 62/106,881

Chen, Y.; Clouthier, C. M.; Tsao, K.; **Strmiskova, M**.; Lachance, H.; Keillor, J. W.; «Coumarin-based fluorogenic probes for no-wash protein labeling. » *Angew Chem Int Ed Engl* **2014**, *53*, 13785–13788

Prchal, J., Junkova, P., **Strmiskova, M**., Lipov, J., Hynek, R., Ruml, T., Hrabal, R; «Expression and purification of myristoylated matrix protein of Mason-Pfizer monkey virus for NMR and MS measurements. » *Protein Expr Purif* **2011**, *79*, 122-127