Fakultät für Mathematik, Informatik und Naturwissenschaften | RWTH AACHEN UNIVERSITY

Selective dual-labeling of cell-free synthesized proteins for single-molecule FRET studies.

A case study of human calmodulin.

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften genehmigte Dissertation

vorgelegt von

M.Sc. in Biochemistry, Molecular and Cellular Biology

Mayuri Sadoine

aus Ixelles (Brussels), Belgium

Berichter: Universitätprofessor Dr.rer. nat. Jörg Fitter
Universitätprofessor Dr. rer. nat. Ulrich Schwaneberg

Tag der mündlichen Prüfung: 02 February 2018

Table of content

Abbreviation

µl	Microliter
µm	Micrometer
µs	Microsecond
3D	Three-dimensional
Å	Ångström
a.u.	Arbitrary unit
AA	Amino acid
AF488	Alexa Fluor 488
AF647	Alexa Fluor 647
AmF	p-Aminophenylalanine
APD	Avalanche photodiode
APTES	3-aminopropyltriethoxysilane
AV	Accessible Volume
AzF	p-Azido-phenylalanine
c	Concentration
Cα	Alpha carbon
CAMKII	CAM dependent kinase II
CD	Circular Dichroism
CECF	Continuous-exchange cell-free protein synthesis system
CF	Correction factor
CFPS	Cell-Free Protein Synthesis
Cys	Cysteine
D	Coefficient of diffusion
ddNTP	Didesoxy-nucleotide triphosphate
DM	Dichroic mirror
DNA	Deoxyribonucleic acid
DOL	Degree of labeling
E	FRET efficiency
E.coli	*Escherichia coli*
EDTA	Ethylendiethyltetraacetic acid
EF-G	Elongation factor G

EF-Tu	Elongation Factor Tu
EGTA	Ethylene glycol-bis (β-aminoethyl ether)-N,N,N',N'-tetraacetic acid
Emerald GFP	Emerald Green fluorescent protein
FCS	Fluorescence Correlation Spectroscopy
fMet-tRNA	formylmethionine tRNA
FRET	Förster Resonance Energy Transfer
h	Hour
hCaM	human calmodulin
ID	Identity
IF1	Initiation factor 1
IF2	Initiation factor 2
IF3	Initiation factor 3
IMAC	Immobilized metal affinity chromatography
IPD	Inter-photon time-distance
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kd	Dissociation constant
KDa	Kilodalton
l	Liter
Laser	Light amplification by stimulated emission of radiation
M	Molar
M13	Myosin light chain kinase peptide
MB	Molecular brightness
min	Minute
ml	Milliliter
MLCK	Myosin Light Chain Kinase
mM	Millimole
MOPS	3-N-Morpholino-propansulfonic acid
mRNA	Messenger RNA
ms	Millisecond
MW	Molecular weight
MWCO	Molecular weight cutoff
ng	Nanogram
NHS ester	Succinimidyl ester

Ni-NTA	Nickel-nitrilotriacetic acid
nm	Nanometer
nM	Nanomole
NMR	Nuclear magnetic resonance
ns	Nanosecond
NTA	Nitrilotriacetic
O.D.	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
pH	potential hydrogenii
PIE	Pulse interleaved excitation
PIFE	Protein Induced Fluorescence Enhancement
pM	Picomole
PTC	Peptidyl transferase center
PTM	Post-translational modification
QY	Quantum yield
R_{DA}	Donor-acceptor distance
RF1	Release factor 1
RF2	Release factor 2
RF3	Release factor 3
RNA	Ribonucleic acid
rpm	Rotation per minute
rRNA	Ribosomal RNA
s	Second
SD	Shine-Delgarno
SDS	Sodium dodecyl sulfate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
smFRET	Single molecule Förster Resonance Energy Transfer
SPAAC	Strain promoted azide-alkyne cycloadditions
SPAD	Single-photon avalanche diodes

TCCD	Two-Color Coincidence Detection
TCEP	Tris (2-chlorethyl)phosphine
TCSPC	Time-correlated single-photon counting module
TEMED	Tetramethylethylenediamine
TFP	Trifluoperazine
TM	Macro time
tm	Micro time
TRIS	Tris (hydroxymethyl) aminomethane
TTTR	Time-tagged time-resolved
UAA	Unnatural amino acid
UV	Ultraviolet
VMD	Visual molecular dynamics software
wt	Wild-type
yPGK	Yeast phosphoglycerate kinase

1. Introduction

1.1 Cell-free protein synthesis and expansion of the genetic code

1.1.1 Protein synthesis

Protein synthesis is one of the most fundamental processes in living organisms. It allows cells to produce their specific functional entities, the proteins. Together with other cellular processes, it permits information originally encoded in deoxyribonucleic acid (DNA) to be translated into functional polypeptide chains. Protein synthesis requires both the transcription into messenger ribonucleic acids (mRNA) of the DNA region encoding the specific information and the translation of the produced mRNA into proteins by ribosomes. Both transcriptional and translational processes ensure a proper sequential flow of the genetic information in biological systems. This constitutes the central dogma of molecular biology [1].

In prokaryotes, ribosomes are constituted of two subunits, the small 30S and the large 50S (**Fig.1**).

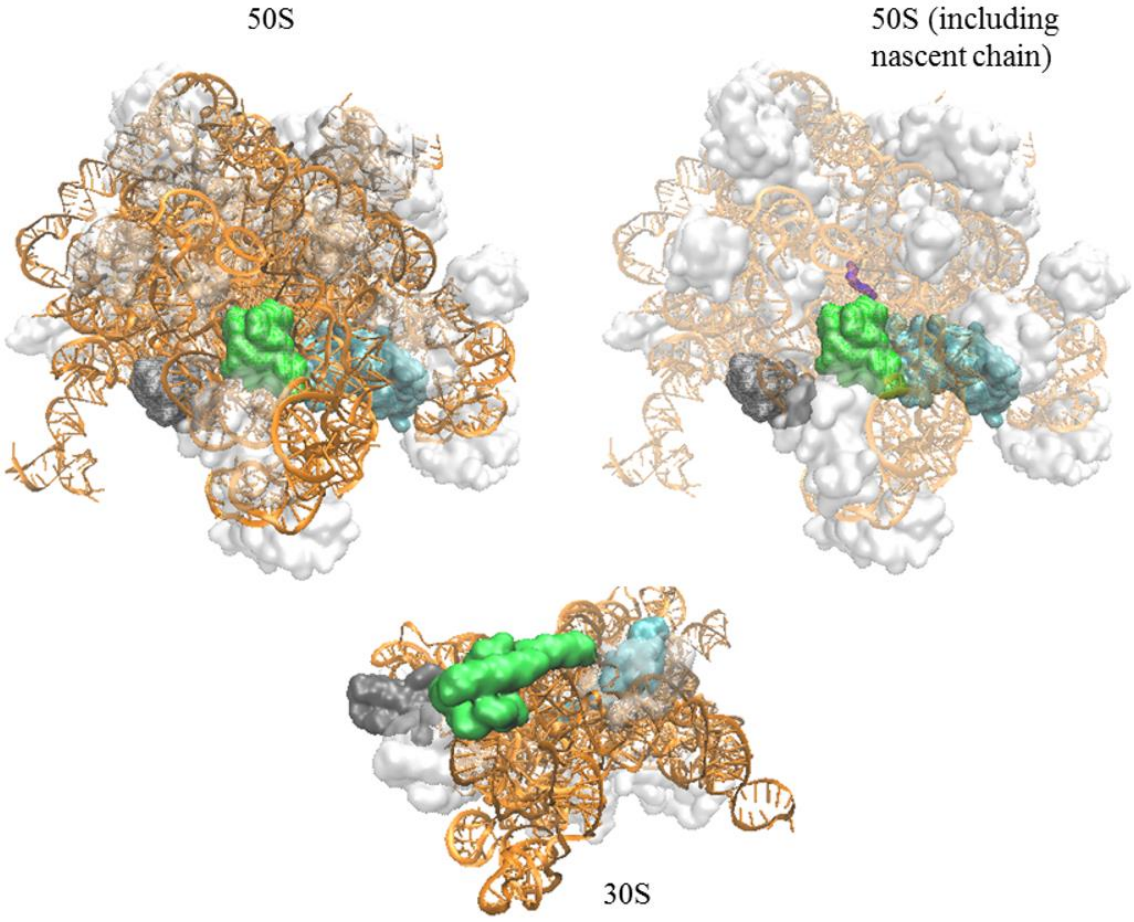


Figure 1. Overview of the small (30S) and large (50S) subunits of a prokaryotic ribosome. Representation of the large subunit (top) and the small subunit (bottom). The ribosomal RNA and proteins are displayed in orange and white, respectively. Proteins shown in cyan, green and gray are involved or are at close proximity of the A, P and E sites, respectively. The ribosome-bound nascent chain is represented in magenta.

Both subunits associate to form a 70S complex prior to translation. In *E. coli*, approximately one-third of the 70S complex is constituted of proteins and the remaining two-thirds of ribosomal RNA (rRNA). The small subunit includes 21 ribosomal proteins and a 16S rRNA whereas the large subunit is made of 33 proteins and two rRNAs, 23S and 5S [2].

Four main steps define the translation; the initiation, elongation cycle, termination and recycling which are supported sequentially by initiation, elongation, release and ribosome recycling factors (**Fig. 2**). During initiation, the 30S subunit recognizes the Shine-Dalgarno (SD) sequence of the mRNA via its 16S rRNA. Subsequently, the start codon (usually AUG) is positioned into the P site in a process that involves initiation factors 1, 2 and 3 (IF1, IF2 and IF3). Thus, the mRNA, the initiator aminoacyl-tRNA and the 70S ribosome form the initiation complex around the start codon [3]. At the beginning of the elongation cycle, the aminoacyl-tRNA occupies the P site while the A site is empty. Consequently, a new aminoacyl-tRNA can enter the A site in complex with elongation factor Tu-GTP (EF-Tu-GTP). The successful formation of the codon-anticodon complex leads to GTP hydrolysis and subsequent release of EF-Tu-GDP. At that stage, the ends of both amnioacyl-tRNAs are located in the peptidyl transferase center (PTC) of the 50S subunit and a new peptide bound is formed when the nascent chain from the P site is attached to the new amino acid. It is followed by a translocation driven by elongation factor G-GTP (EF-G-GTP) where the deacylated tRNA leaves the P site and enters the E site from which it will be later ejected. In the meanwhile, the peptidyl tRNA moves from the A site towards the P site allowing the reception of a new aminoacyl-tRNA [4]. All along the elongation, the nascent chain migrates through an 80-100 Ångströms (Å) exit tunnel [5-9] until a stop codon (UAG, UGA or UAA) is encountered and termination is triggered. Indeed, once the stop codon enters the A site, it is recognized by one of the release factors RF1 or RF2 resulting in the liberation of the polypeptide chain followed by the dissociation of RF1or RF2 from the ribosome, which is facilitated by the release factor 3 (RF3). Finally, thanks to the ribosome recycling factors together with EF-G the whole complex is dissociated and the mRNA and deacylated tRNA are released. Still, IF3 is required for replacing the deacylated tRNA thanks to its GTPase activity which leads *in fine* to the complete release of the mRNA [5].

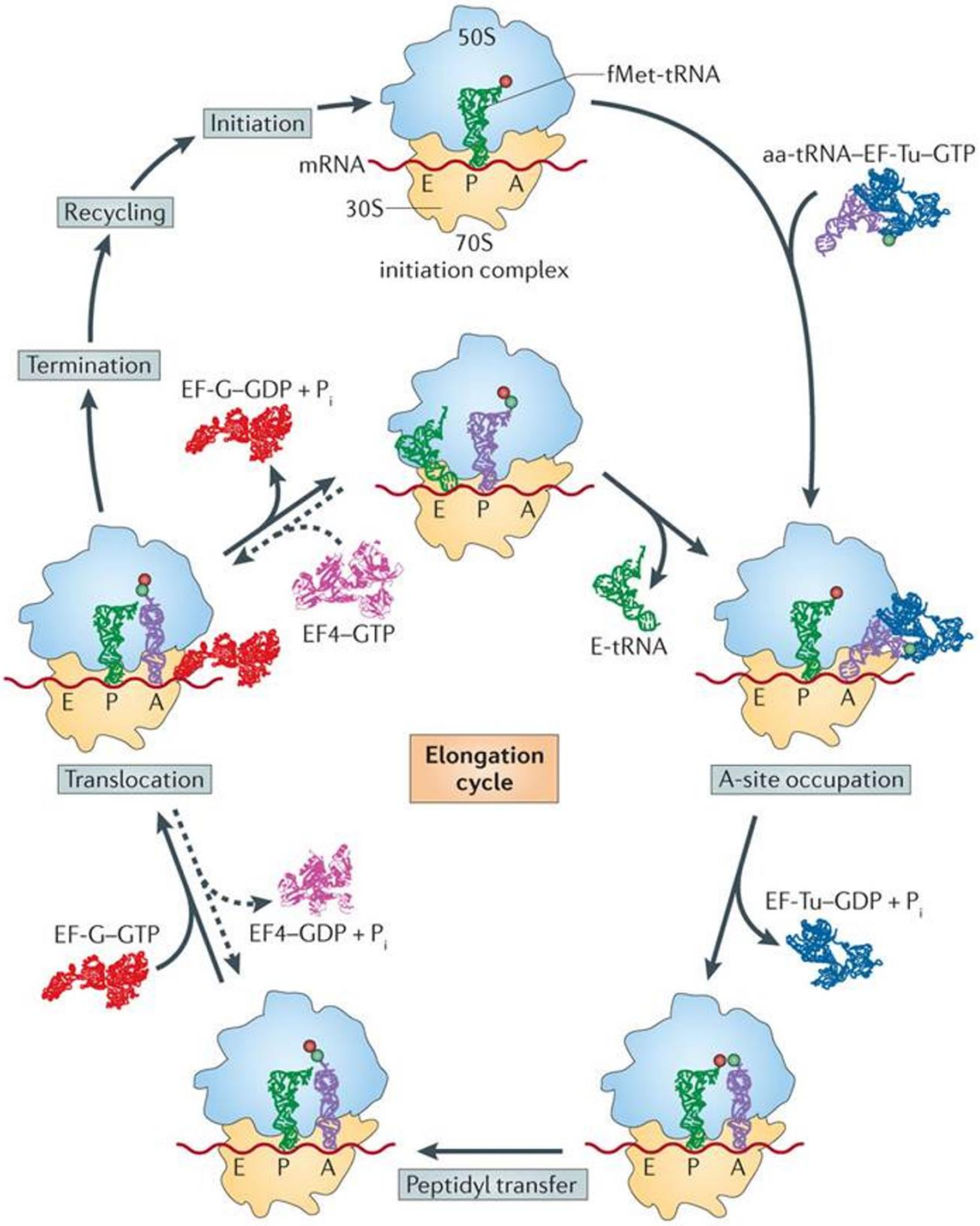


Figure 2. Translation of a mRNA by a ribosome in *E.coli*. The initiation complex is constituted of the 70S ribosome (associated 30S and 50S subunits), the mRNA and the initiator tRNA (formylmethionine tRNA (fMet-tRNA)) which is located at the P site. The elongation cycle starts when the complex aminoacyl-tRNA–elongation factor Tu–GTP (aa-tRNA–EF-Tu–GTP) enters the A site. After formation of the codon-anticodon complex, GTP is hydrolyzed and EF-Tu–GDP is released. During the peptidyl transfer, a new peptide bound is then formed when the peptide chain is attached to the new aa-tRNA. Then, both the deacylated and peptidyl tRNAs are translocated by a distance of one codon towards the E and P sites, respectively. In some case, EF4-GTP can reverse the reaction leading to the

mobilization of stalled ribosomes (dashed arrows). Once a stop codon enters the A site, the termination is triggered by release factors 1, 2 and 3 (RF1, RF2 and RF3). Finally, the whole complex is dissociated for recycling [10].

The primary sequence of the polypeptide chain is usually insufficient for getting a functional entity. Indeed, the function of a protein is highly associated with its three-dimensional (3D) structure. Therefore, the polypeptide chain has to fold in order to acquire its final native structure [11]. All the information required for protein folding is contained in the primary sequence of amino acids. In the cell, some proteins fold spontaneously whereas most of them require the assistance of specific proteins such as those commonly called chaperones. Chaperones bind and stabilize unfolded or partially folded intermediates preventing both misfolding and aggregation that are highly likely to occur in the very crowded environment. *In vivo*, the folding of protein is thought to start co-translationally and to occur in a vectorial manner [6, 9, 12, 13]. However, whether the nascent chain can significantly fold during biosynthesis is still debated.

1.1.2 Cell-free protein synthesis

Cell-free protein synthesis (CFPS) is the production of proteins outside the cell. The synthesis of proteins *in vitro* is performed by using systems which mimic the cell environment. *E. coli* CFPS systems were developed for the first time more than 50 years ago by Matthaei and Nirenberg [14], even though the concept of cell-free synthesis was much earlier stated by Buchner in 1897 [15]. Nowadays, CFPS is widely used for obvious reasons that will be described below. Firstly, commercial cell-free systems originating from Archea, prokaryotes and eukaryotes are currently available. Still, *E.coli* crude extracts are the most common source of cell-free systems even though it is mostly limiting for post-translational modifications. In coupled reaction scheme, CFPS is performed from either linear or circular DNA templates [16] in extracts that are supplemented with a set of necessary components (i.e. amino acids, nucleotides, energy substrates, cofactors and salts) that feed the transcription and translation machineries. The initial CFPS systems were associated with low yields and significant costs that were making them unattractive. In fact, the most challenging issue was linked to the short reaction duration. In the aim to overcome it, current modern cell-free systems are improved and often combined with different modes of reaction going from the classical batch-mode to the dialysis based continuous-exchange cell-free (CECF), which is able to reach pretty high yields (up to few mg/ml) [17].

Secondly, CFPS is often preferred to *in vivo* expression for its easy and fast manipulation. While it takes a several days to produce recombinant proteins in *E. coli*, CFPS can be achieved within a day [18]. Additionally, the open system allows a level of flexibility that cannot be reached *in vivo* permitting direct manipulation and real-time monitoring. The latter point is of interest for fundamental research, which attempts to understand processes such as transcription, translation and co-translational folding [19]. As a result of the major advancements made in CFPS, its current applications are various going from structural and high throughput proteomics to protein evolution and therapeutics [20].

Finally, cell-free systems are of particular interest for the synthesis of difficult-to-express proteins. Indeed, in CFPS, no viability or growth issue is encountered and therefore the production of toxic proteins can be easily handled. Toxic proteins include pharmaceutical, as well as plasma membrane proteins. The latter are of special concern in biomedical research since they account for three quarters of all potential drug targets [18, 20]. Altogether, CFPS extend considerably the possibilities in proteomics.

1.1.3 Expansion of the genetic code

The canonical genetic code is made of 64 triplets of nucleotides from which 61, called sense codons, encode for amino acids and 3, commonly named nonsense or stop codons, are used for terminating the translation. Despite this, they encode for only about 20 to 22 amino acids in most of the organisms. This is due to the fact that the code is degenerative, meaning that one amino acid is usually encoded by more than one codon [21]. The idea of adding new building blocks to the genetic code is therefore extremely attractive. Indeed, expansion of the genetic code for incorporation of unnatural amino acids (UAAs) into proteins constitutes a powerful tool for the scientific research community. Recent advancements in the manipulation of the genetic code and the incorporation of UAAs have opened new promising perspectives [22]. Up to now various UAAs carrying some particularly interesting properties (e.g. photoreactivity, fluorescence or post-translational modification) have been incorporated into proteins in *E. coli*, yeast and mammalian cells [23-25]. Expanding the genetic code often requires the use of an orthogonal pair of tRNA/aminoacyl-tRNA synthetase previously engineered in another organism (**Fig. 3**) [26]. The synthetase is used to charge the orthogonal tRNA with the desired UAA, whereas the tRNA is designed for decoding a unique codon. Different methodological approaches have been successfully used for incorporation of UAAs into protein [27, 28]. Among them, a very common strategy is to take advantage of the degeneracy of the genetic code by using one or two of the three nonsense codons; UAA

(ochre codon), UAG (amber codon) and UGA (opal codon) for site-specific incorporation [29]. Indeed, all the three stop codons are used to terminate the translation by recruiting release factors. Therefore, two of them can be used to perform unique UAA incorporation into proteins. So far, nonsense suppressor tRNAs have been used to incorporate different UAAs both *in vitro* and *in vivo* [30-33]. For instance, two distinct suppressor aminoacyl-tRNAs carrying two different fluorescent UAAs and decoding two distinct unique codons were reported for the dual-labeling of proteins [32, 33].

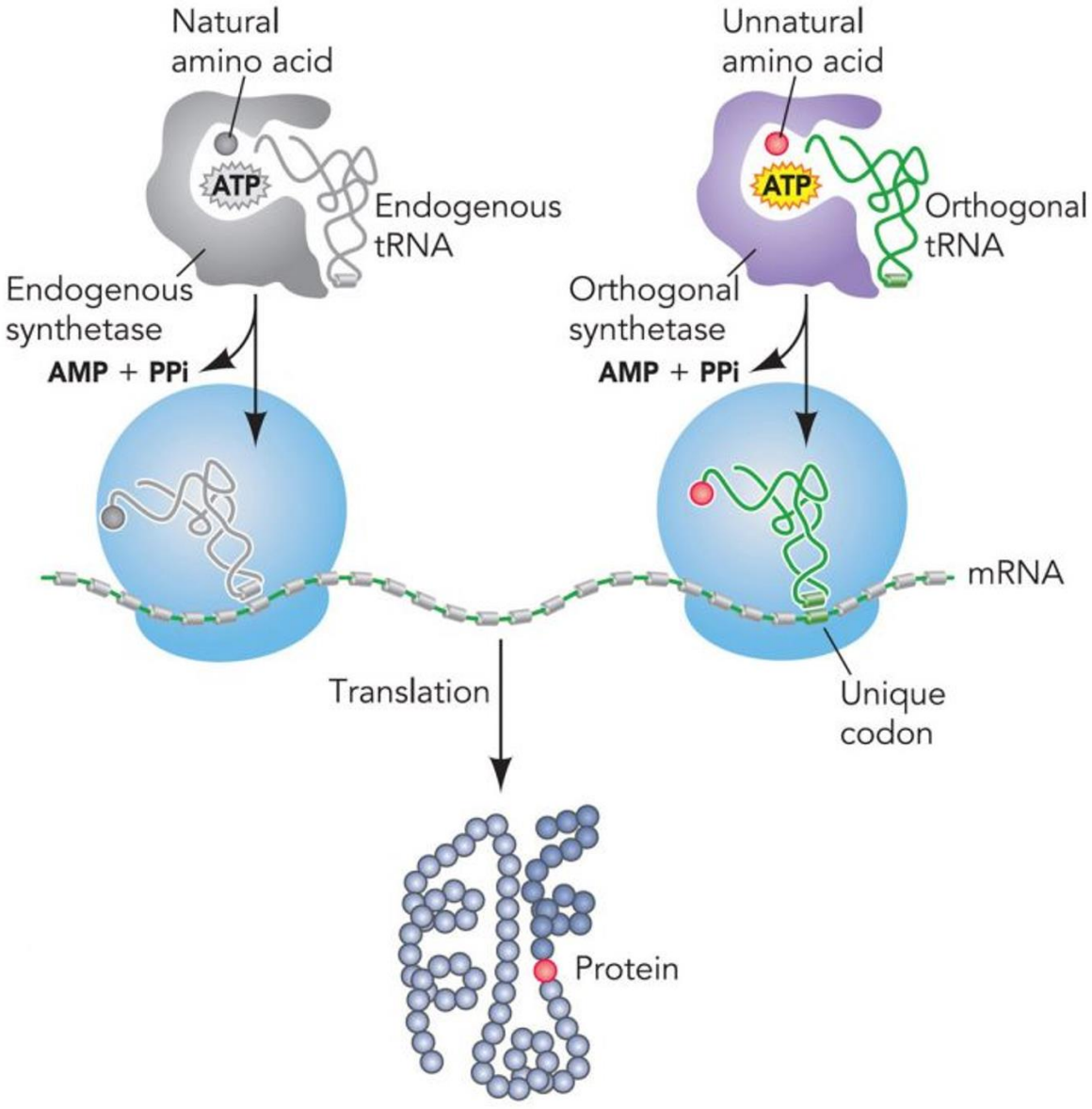


Figure 3. Incorporation of unnatural amino acids (UAAs) into proteins. A natural amino acid is incorporated thanks to an endogenous pair of tRNAs/aminoacyl-tRNA synthetase (left) while an orthogonal pair of tRNA/aminoacyl-tRNA synthetase is used to incorporate an UAA into proteins (right) [26]

Even though the combination of CFPS and UAAs incorporation leads to obvious issues concerning the protein yield, recent advancements have permitted to reach reasonable amounts of sample (i.e. µg/mL). For example, deletion or depletion of RF1 from cell extracts was used to improve the reading of UAG codons by suppressor tRNAs [34-40].

1.2 Single-molecule Förster resonance energy transfer (smFRET)

1.2.1 Single-molecule methods and FRET

Fluorescence is a physical process in which a specific molecule, so-called fluorophore, has the capability to emit photons upon proper excitation. For instance, a fluorophore can absorb photons of proper energy which bring it to an electronically excited state. Then, the return to the ground state can be done through several pathways including non-radiative processes and direct emission of photons.

In some specific conditions, the return to the lowest energy state can be performed via energy transfer to another molecule by making FRET (**Fig. 4,** top). The latter phenomenon was described by Förster in 1948 with his theory of non-radiative transfer of energy from one electronically excited molecule to a neighboring molecule [41]. In FRET, the energy from an excited donor fluorophore is transferred by dipole-dipole interaction to an acceptor fluorophore. This happens in very specific conditions that require, among others, the overlay of the emission spectrum of the donor with the absorption spectrum of the acceptor, close vicinity of the fluorophores (i.e. less than 10 nm) and at least partial alignment of their respective dipoles. Following FRET, the acceptor is found in an excited state and can subsequently return to the ground state by emitting photons. Interestingly, FRET is effective over distances (i.e. from few nm to 10 nm) which correspond to the size of biological macromolecules. Since FRET is a dipole-dipole interaction, its efficiency (E) depends strongly on the distance separating the two fluorophores (**Fig. 4,** bottom) [42]. The scientific community has taken advantages of fluorescence and FRET phenomena and nowadays fluorescent probes are commonly used. Among them, chemical fluorophores constitute a very large class of chemical compounds used to tag proteins and probe biological processes [43]. In particular, FRET occurring between two fluorophores is extensively used as a "spectroscopic ruler" to measure molecular distances and distance dynamics in and between proteins, nucleic acids or even big molecular complexes [44, 45].

Single-molecule FRET (smFRET) constitutes a peerless tool for studying protein structures, conformational dynamics and protein-protein interactions in heterogeneous systems both *in*

vivo and *in vitro* [46-49]. For example, the unique combination of FRET and single molecule techniques has revealed interesting but normally hidden subpopulations in ensembles [50, 51]. From its first use in 1996, smFRET has contributed to address various biological questions linked to important fundamental processes such as transcription, translation, protein folding and conformational dynamics [52, 53].

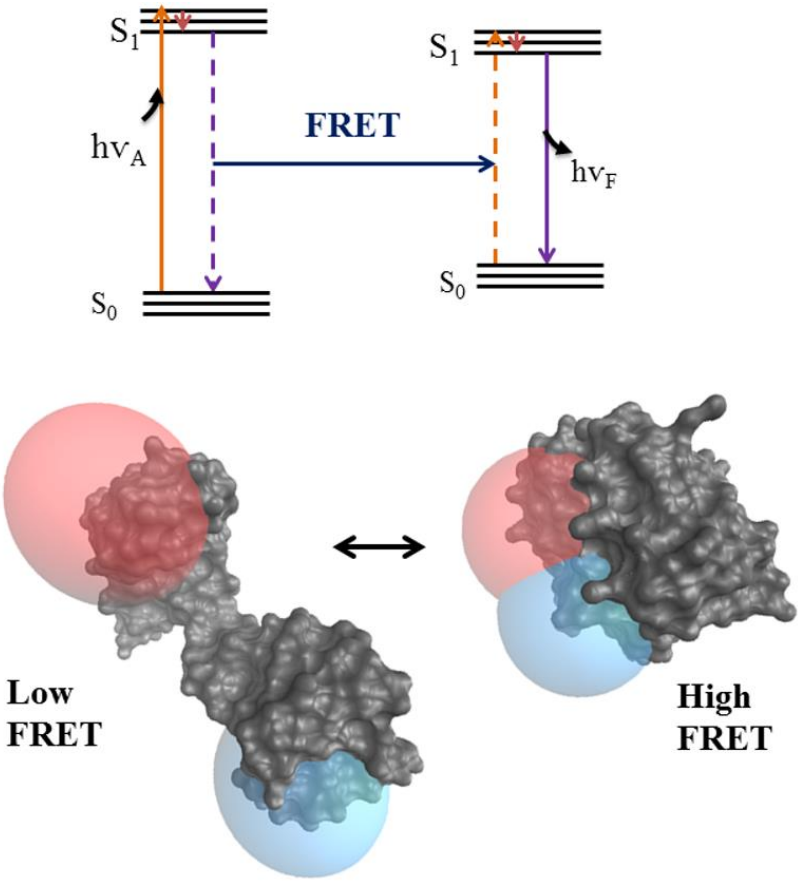


Figure 4. Principle of Förster resonance energy transfer (FRET). The basic principle of FRET is described by using a summarized Jablonski Diagram (top) and an example of its application in biophysical studies is shown (bottom). Absorption of photons ($h\nu_A$) of proper energy by a fluorophore 1 (top, left) leads to its transition from the ground state (S_0) towards an electronically excited state (S_1) (orange solid arrow). The return to S_0 can be performed through various processes including FRET (blue arrow). FRET allows the transfer of energy to a compatible neighborhood fluorophore 2 (top, right) leading to its transition towards S_1 (orange dashed arrow). The return to the ground state can be done by emission of photons ($h\nu_F$) (violet solid arrow). The distance dependency of FRET is very useful in biophysical studies for probing conformational changes (bottom). The protein is displayed in gray while the dyes are represented by blue and red spheres.

Combination of single molecule techniques (e.g. smFRET) and cell-free systems is highly advantageous. Firstly, CFPS takes advantage of the single molecule detection, since the latter

decreases significantly the amount of sample that is required for the analysis. Secondly, by analyzing single molecules individually, meaningful information that is hidden in an ensemble can be extracted. Moreover, single molecule techniques gain in turn from CFPS, which allows fast and easy sample preparation. Finally, that specific combination is of special interest for studying difficult-to-express proteins as well as for probing fundamental processes in real-time.

1.2.2 smFRET and selective dual-labeling

smFRET requires site-specific dual-labeling of the protein of interest with a relevant dye pair. This could be achieved by using site-specific modification chemistries. A common strategy is based on reacting two exposed cysteine residues, a lowly abundant natural amino acid, with maleimide-functionalized dyes [54, 55]. It can be conducted on exposed endogenous cysteine residues when they are located at relevant positions. On the contrary, if other positions are required, new cysteine residues need to be incorporated at the desired positions by site-directed mutagenesis after removal of the endogenous ones. Subsequently, specific functionalized (e.g. maleimide-functionalized) fluorophores react with the accessible thiol groups of the cysteine residues [56]. Even though the approach is site-specific over the cysteine residues, the two dye conjugations are performed via the same chemistry leading to non-selective labeling, since both cysteine residues can react with both the donor and the acceptor. This generates a subsequent highly heterogeneous sample that prevents fine data analysis [57].

The site-specific incorporation of UAAs is the most attractive alternative to the double cysteine mutagenesis [58, 59]. The incorporation of fluorescent UAAs seems to be the straightest forward approach for the selective dual-labeling of proteins. However, up to now, efficient co-translational incorporation of fluorophores into proteins is limited to small fluorophores like BODIPY dyes, which are not suitable for single-molecule detection due to their weak photostability [60, 61]. The incorporation of large fluorescent groups is thought to be limited by a likely weak affinity of the fluorophore for EF-Tu and obvious physical constrains related to the vicinity of the ribosome leading to consecutive low protein yields (i.e. ng/mL up to few μg/mL) [62-64]. As an alternative, the site-specific incorporation of UAAs carrying specific functional groups is more widely used. Actually, successful site-specific protein labeling through bioorthogonal chemical reactions performed on genetically encoded UAAs was already reported [55, 58, 65, 66]. This methodological approach has the key advantage to be available for a wide range of dyes, since the labeling is performed

subsequently to UAAs incorporation and does not depend at all on the degree of tolerance of the cellular translational machinery for the specific fluorophores. Such orthogonal systems were successfully used *in vivo* for selective dual-labeling of proteins by combining the site-specific incorporation of an UAA and a cysteine residue [30, 31] or two different UAAs [57, 67]. Furthermore, a potential gain in smFRET accuracy due to selective dual-labeling was suggested for protein expressed *in vivo* [31]. Nevertheless, despite the current availability of orthogonal CFPS systems [39, 68, 69] bearing acceptable protein yields (i.e. several hundreds µg/ml), smFRET performed on selectively dually labeled cell-free synthesized proteins remains limited to a couple of studies due to the lack of simple and robust labeling methodologies [70].

1.3 The Ca^{2+}-signaling protein calmodulin: a model protein

The Ca^{2+}-signaling protein calmodulin (CaM) is a highly conserved acidic protein (pI=4.5) of around 17 kDa and 148 amino acids (AAs) which is expressed in all eukaryotic cells. Indeed, it is a particularly important protein, since it is involved in the transduction of Ca^{2+} signals. A proper answer to Ca^{2+} is necessary for many cellular functions going from metabolism-related to apoptosis and muscle contraction [71]. Recently, mutations in the primary sequence of human calmodulin (hCaM) were linked to important diseases such as cardiac arrhythmias, long QT syndrome and cancers [72, 73]. Despite its particularities, CaM is relatively well known since it has been intensively studied for decades.

The three-dimensional structure of that two-domain protein displays 4 E-F hand motifs, which are constituted sequentially of a N-terminal helix, a Ca^{2+}-coordinating loop containing a 12-AA acidic motif and a C-terminal helix (**Fig. 5**). The two N-terminal E-F hand motifs form the N-lobe whereas the two C-terminal constitute the homologous C-lobe, which is separated from the previous one by an interdomain helix [74]. In cells, intracellular Ca^{2+} binds to CaM with an affinity that depends on the lobe (i.e. Kd of about 2 µM and 0.2 µM for the N- and C- lobes, respectively) [75]. Upon Ca^{2+} binding, major rearrangements occur in the three-dimensional structure of the protein, which exposes hydrophobic methionine-rich target-protein binding sites. Interestingly, the methionine side chain is flexible allowing a certain level of plasticity that is thought to be necessary for the binding of a broad range of targets [76]. Indeed, CaM-dependent proteins are continuously identified and account already for hundreds of different proteins [77]. They are various and include kinases, phosphatases, metabolic enzymes, ion channels, ion pumps, as well as transcription factors.

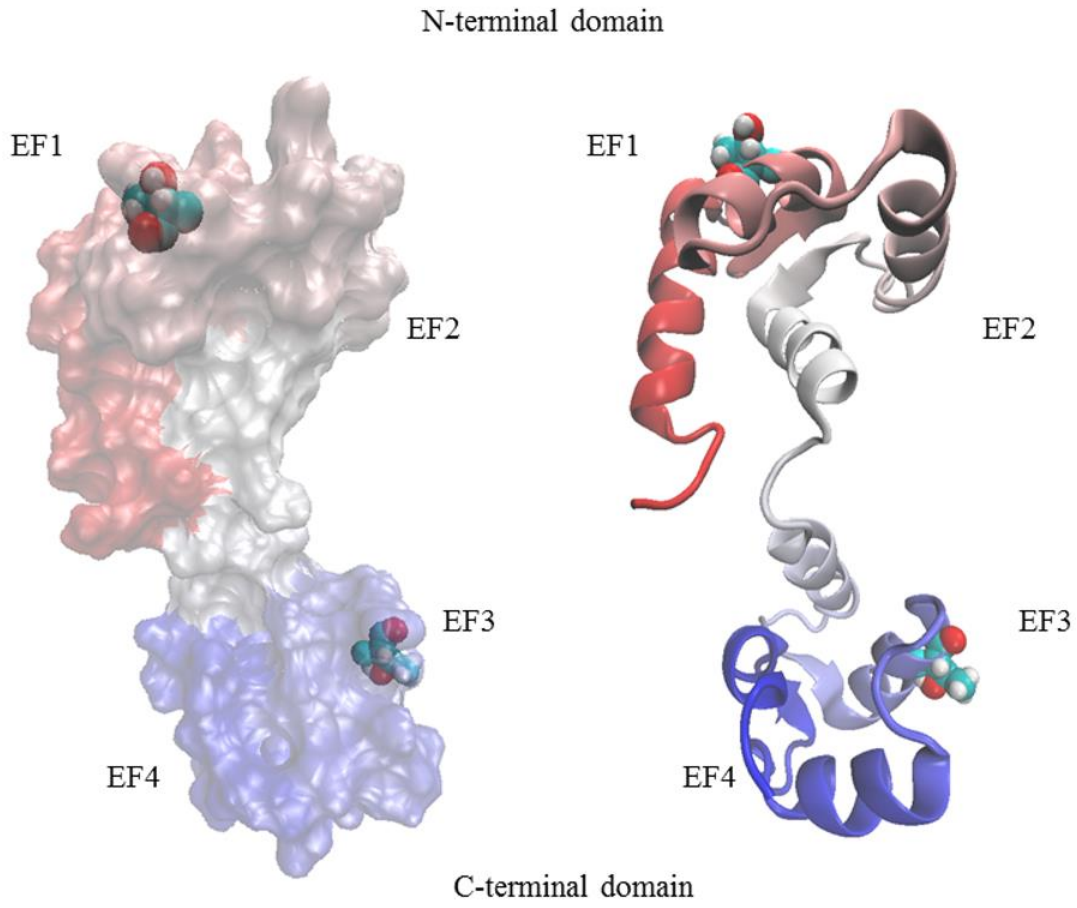


Figure 5. **Representation of the three-dimensional structure of human calmodulin.** The structure of CaM is shown according to its volume (left) or its secondary structure (right) and colored along the amino acid sequence from N-terminal (red) towards C-terminal (blue). The residues 34 and 110 used for selective dual-labeling are displayed in atomic view.

CaM has the particularity to bind its partners through multiple recognition modes either in presence or in absence of Ca^{2+}, while the former is more often reported, in a reversible or irreversible manner and by acting as an inhibitor or an activator. The binding of CaM to its targets is usually done with high affinity (i.e. Kd from 10^{-7} to 10^{-11} M) and leads to specific structural adaptation that depends on the partner [72]. The binding mechanism of CaM is still not fully understood. Interestingly, multiple conformations were reported for the protein in solution, which could explain at least partially its ability to interact with such a broad range of targets [78]. Nowadays, many different three-dimensional structures of CaM either in a free- or bound-form (i.e. bound either to Ca^{2+} or to a target or to both) are available in the Protein Data Base (PDB) (**Fig. 6**). Even though it is intensively studied, how CaM is able to bind its targets remains unclear and subject to discussion. The flexible apo-form of calmodulin is characterized by a relatively compact conformation whereas its more rigid

holo-form is associated with multiple reported three-dimensional structures, which go from highly extended to very compact, even more than the respective apo conformation (**Fig. 6**, bottom) [79-81].

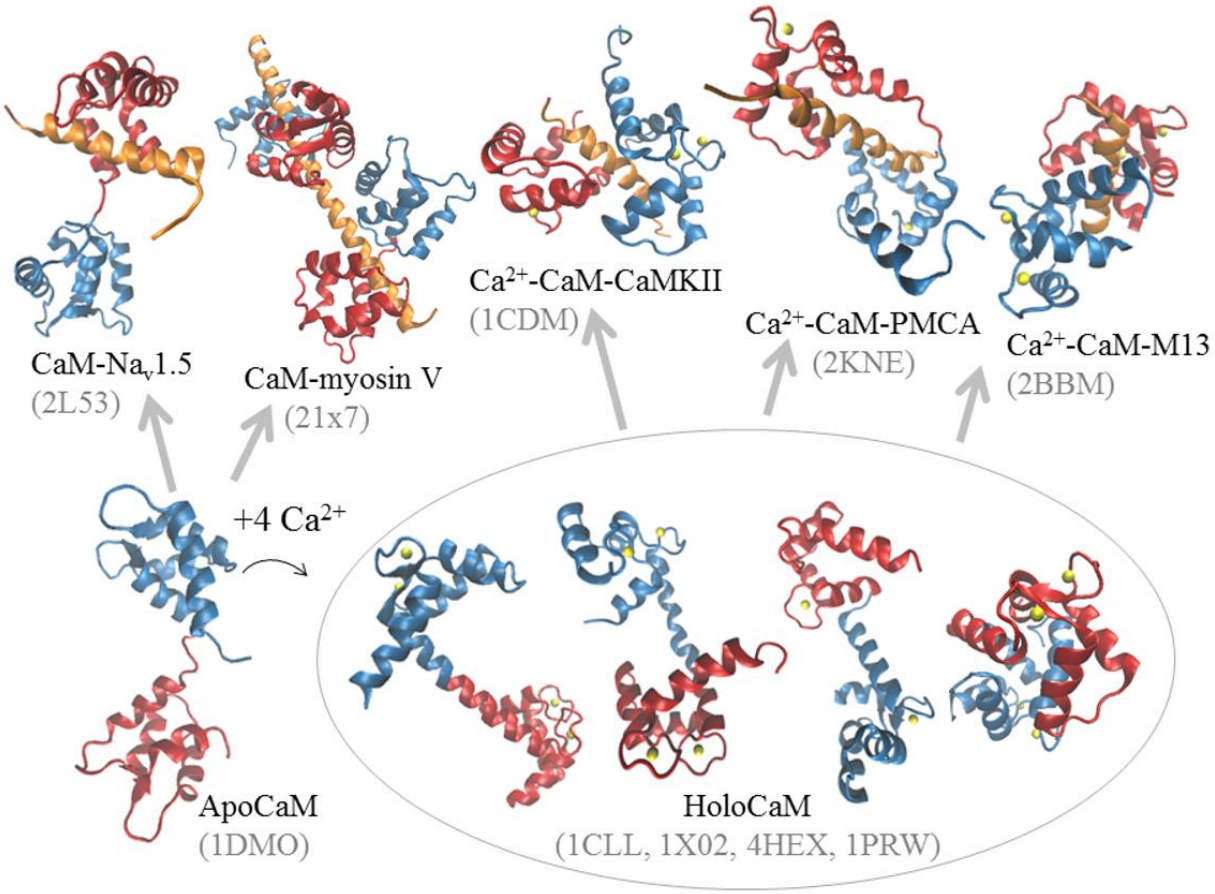


Figure 6. **Structural diversity of CaM.** ApoCaM (bottom, left) can bind up to 4 calcium ions. Upon binding, CaM undergoes conformational changes towards its holo-form (bottom, right). The holoCaM presents multiple conformational substates (gray circle) that are thought to be necessary for its functionality. Both apo- and holo-forms of CaM can bind partners (top). Interestingly, functional conformations of CaM show an exceptionally high structural diversity. The structures are displayed and the N-terminal and C-terminal domains are shown in red and blue, respectively. The calcium ions are represented by yellow balls and partners are colored in orange. PDB IDs are mentioned (gray).

The structural particularities associated to the holo-form are of strong interest. Indeed, CaM adopts very similar extended and compact conformations when bound to its targets (**Fig .6**). Therefore, the fact that distinct conformations were resolved for the holoCaM strongly suggests the actual co-existence of various conformations in solution that would be necessary for the binding of the multiple partners [81]. Furthermore, it was even suggested that CaM could sample a quasi-continuous spectrum of conformational substates displaying specific

statistical weights depending on the conditions [82, 83]. The statistical weights would be redistributed upon binding of a specific partner. Hence, it was suggested that the functional forms of CaM are most probably less distinct than previously thought allowing the broad range of activities of that particular protein [83]. Nevertheless, additional information about these conformational substates is still needed to fully understand CaM functionality. In particular, smFRET is singularly suitable for studying conformational diversity due to its high power of discrimination [84, 85]. For instance, Johnson *et al.* have used smFRET to demonstrate the presence in solution of multiple conformational substates for the holoCaM [78]. In this way, smFRET data have confirmed a previously stated hypothesis. Indeed, the fact that numerous conformational substates were observed in crystalline environment has previously suggested that CaM samples a large range of conformations [83]. In addition, the co-existence of various conformational substates is fully consistent with other NMR- and X-ray scattering-based studies which have suggested particular dynamics for CaM [86, 87]. The structure-function relationship, numerous conformational substates and dynamics of CaM make it an ideal candidate for smFRET studies [35].

1.4 Aim of the thesis

smFRET is a very powerful technique for studying protein conformations, dynamics and interactions in heterogeneous systems. To date, the technique is mainly limited to *in vivo* expressed proteins. However, the combination of smFRET and *in vivo* protein expression is disadvantageous in many ways (see above).

In the present study, we have aimed to combine smFRET and CFPS for allowing fast, simple and efficient protein dual-labeling and expanding the application of smFRET to the large family of difficult-to-express proteins. Therefore, the thesis was dedicated to the development of two methodological approaches for the selective dual-labeling of *in vitro* synthesized proteins for smFRET studies. Here, we have taken advantage of an expanded genetic code for the incorporation of UAAs into proteins. The biological relevance of the approaches was then demonstrated through the case-study of human calmodulin. Once available for the scientific community, the newly methodological approaches should offer new interesting possibilities for smFRET studies.

2. Materials

2.1 Instruments

Absorption spectrophotometer

NanoDrop 2000c (Thermo Fisher Scientific Life Technologies GmbH, Darmstadt, Germany)

UV 2401PC Spectrophotometer (Shimadzu, Kyoto, Japan) with 10 x 4 mm quartz cuvettes

Cell-Disruptor

Constant cell disruption system (Constant System, Northants, UK)

Emulsiflex-C35 (Avestin, Ottawa, Ontario, Canada)

Centrifuge

Centrifuge 5804/5804R (Eppendorf, Wesseling-Berzdorf, Germany)

Centrifuge Avanti J-20XP/ Rotor: JLA-8.1 (Beckman, Krefeld, Germany)

Ultracentrifuge Optima XPN-80/ Rotor: Ti 70 (Beckman, Krefeld, Germany)

Circular-Dichroism spectrometer

J-1100 CD Spectropolarimeter (Jasco GmbH, Gross Umstadt, Germany) with 1 mm quartz cuvettes

Confocal microscope

50/50-Beam Splitter (Linos, Goettingen, Germany)

Avalanche photodiodes (APD): 2x photon counting module, SPCM-CD3077-H and SPCM-AQR-14 (Perkin Elmer, Woofbridge, Ontario, Canada) and 2x single photon counting module, τ-SPAD, VLoK silicon avalanche photodiode (laser Components, Olching, Germany) with quenching electronics (PicoQuant, Berlin, Germany)

Emission filters: FF01 530/55 and LP665 (Semrock Inc., NY, USA)

FCU-II Fiber Coupling Unit (PicoQuant, Berlin, Germany)

Laser driver: computer controlled multichannel Picosecond Diode Laser Driver PDL828 sepia II (PicoQuant, Berlin, Germany)

Major dichroic: XF2401 (Omega Optical Inc., Brattleboro, VT, USA)

Microscope body: inverted microscope Olympus IX-81 (Olympus, Hamburg, Germany)

MicroTime 200 (PicoQuant, Berlin, Germany)

Minor dichroic: T600lpxr (Chroma Technology, Bellows Falls, VT, USA)

Objective: UPLanSApo 60x, N. A. 1.2, W.D 0.28, F.N. 26.5, water immersion, cover glass thickness 0.13-0-21, (400-900) nm (Olympus, Hamburg, Germany)

Picosecond Laser Heads: LDH-D-C-485B, LDH-D-C-640B (PicoQuant, Berlin, Germany)
Pinhole 75 µm (Chroma Technology, Bellows Falls, VT, USA)
Polarizing Beam Splitter Cube (Linos, Goettingen, Germany)
TCSPC-Acquisition unit: HydraHarp-400, Picosecond histogram accumulating processor, time-correlated single photon counting system with USB interface (PicoQuant, Berlin, Germany)

DNA-Sequencer

Long Redair 4200 (MWG Biotech, Ebersberg, Germany)

Fluorescence spectrometer

Spectrofluorometer QuantaMaster-7 (Photon-Technology-International, Birmingham, NJ, USA)

Gel imaging systems

ChemiDoc MP Imaging System (Bio-Rad, Hercules, CA, USA)
Typhoon 9500 (GE Healthcare Life science, Buckinghamshire, UK)

Lyophilizer

Christ Alpha 1-2 LD Plus Freeze Dryer (Shrewsbury, UK)

PCR thermo Cycler

FlexCycler (Analityk Jena AG, Jena, Germany)

pH-meter

S20-SevenEasyTM mit InLab® Micro Electrode (Mettler Toledo, Giessen, Germany)

Vacuum pump

WP6222050 (Millipore, Billerica, MA, USA)

Western blot chamber

semi-dry blotter V20DB (Scie-Plas Ltd., Cambridge, UK)

2.2 Softwares

MATLAB R2012a (The MarhWorks Inc.)
NanoDrop™ software (Thermo Fisher Scientific Life Technologies GmbH, Darmstadt, Germany)
OriginPro (OriginaLab Corporation, Northampton, MA, USA)
PyMol 1.8.6.0 (Python Software Foundation, Dalware, USA)
SymPhoTime64 (PicoQuant, Berlin, Germany)
The Molecular Modeling Toolkit (MMTK) [88]

Visual Molecular Dynamics (University of Illinois, Champaign, Illinois, USA)
wxDev-C++ 5.11

2.3 Competent cells

E. coli BL21 CodonPlus (DE3)-RP (Agilent Technologies Inc., Santa Clara, CA, USA)
E. coli Top10 (Thermo Fisher Scientific Inc., Waltham, MA, USA)

2.4 Vectors, nucleotides and oligonucleotides

Nucleotides

dNTP-Mix (Thermo Fisher Scientific Inc., Waltham, MA, USA)

Oligonucleotides

Sequences of wild-type and mutant proteins used in the present work are presented in Supplementary information.

Vectors

pCR® Blunt Vector (Thermo Fisher Scientific Inc., Waltham, MA, USA)
pRSET (Thermo Fisher Scientific Inc., Waltham, MA, USA)

2.5 Kits and enzymes

Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare, Chalfont St. Giles, UK)
Antartic phosphatase (New England Bioloabs, Ipswich, MA, USA)
Buffer QX1, Solubilization Buffer (Quiagen, Hilden, Germany)
Coupled transcription/translation cell-free system from E. Coli (Biotechrabbit, Hennigsdorf, Germany)
Cysteine-depleted coupled transcription/translation cell-free system from *E. Coli* (Biotechrabbit, Hennigsdorf, Germany)
NucleoBond®Xtra Midi (Macherey-Nagel GmbH, Düren, Germany)
Phusion Hot Start DNA-Polymerase (New Englands Biolabs, Ipswich, MA, USA)
PURE *in vitro* transcription-translation system (New England Biolabs, Ipswich, MA, USA)
QIAprepSpin Miniprep Kit (Qiagen, Hilden, Germany)
Rapid DNA Ligation Kit (Fermentas, St. Leon-Rot, Germany)
Rapid Ligation Kit containing T4-Ligase (Fermentas, St. Leon-Rot, Germany)
Restriction Enzymes Fast Digest (Fermentas, St. Leon-Rot, Germany)

RF1/cysteine-depleted coupled transcription/translation cell-free system from E. Coli (Biotechrabbit, Hennigsdorf, Germany)
Site-directed protein labeling (Biotechrabbit, Hennigsdorf, Germany)
SUPERase In Rnase inhibitor (Thermo Fischer Scientific, Waltham, MA, USA)
Zero Blunt PCR Cloning Kit (Thermo Fisher Scientific, Waltham, MA, USA)

2.6 Chemicals and other consumables

1.5 DNA-LoBind tubes, PCR clean (Eppendorf AG, Hamburg, Germany)
1.5 Protein-LoBind tubes, PCR clean (Eppendorf AG, Hamburg, Germany)
2-Amino-2-hydrxymethyl-propane-1,3-diol (TRIS), buffer grade (Applichem, Darmstadt, Germany)
3-N[Morpholino]-propansulfonic acid (MOPS), buffer grade (Applichem, Darmstadt, Germany)
Acetone, CHROMASOLV® for HPLC ≥ 99.9 % (Sigma-Aldrich, St Louis, MO, USA)
Acrylamide/Bisacrylamide: Rotiphorese Gel 30 (37.5:1) (ROTH GmbH, Karlsruhe, Germany)
Agar for Molecular Biology (Applichem, Darmstadt, Germany)
Agarose UltraPure for gel electrophoresis (Thermo Fischer Scientific, Waltham, MA, USA)
Ampicillin (Roth, Karlsruhe, Germany)
Calcium chloride (Sigma-Aldrich, St Louis, MO, USA)
Calmodulin (human), (recombinant) (Enzo, New York, NY, USA)
Calmodulin-dependent Protein Kinase II fragment 290-309 (Sigma-Aldrich, St Louis, MO, USA)
Complete EDTAfree Protease inhibitor cocktail (Roche, Basel, Switzerland)
Coomassie brilliant blue G250 (Applichem, Darmstadt, Germany)
Coomassie Brilliant Blue R (SERVA Electrophoresis GmbH, Heidelberg, Germany)
Dimethyl sulfoxide (DMSO), anhydrous, ≥ 99 % (Sigma-Aldrich, St Louis, MO, USA)
EGTA, disodium salt (Sigma-Aldrich, St Louis, MO, USA)
Ethanol, HPLC grade (Applichem, Darmstadt, Germany)
Ethylenediaminetetraacetic acid (EDTA), disodium salt (MP Biomedicals, South Chillicothe Rd., Aurora, USA)
Gelred (Biotium Inc., Hayward, CA, USA)
Glucose (Sigma-Aldrich, St Louis, MO, USA)
Guanidine hydrochloride (Applichem, Darmstadt, Germany)
High precision glass slides (170±5µm, No. 1.5H) (Menzel-Gläser GmbH, Braunschweig,

Germany)
Hydrogen chloride (Merck, Darmstadt, Germany)
Hydrogen peroxide solution, 35wt % in H2O2 (Sigma-Aldrich, St Louis, MO, USA)
Imidazole (Applichem, Darmstadt, Germany)
Isopropyl β-D-1-thiogalactopyranoside (IPTG) (VWR International LLC, Radnor, PA, USA)
Kanamycin (Applichem, Darmstadt, Germany)
M13 Skeletal Muscle Myosin Light Chain Kinase Peptide (AnaSpec, Fremont, CA, USA)
Methanol (for HPLC) (VWR International LLC, Radnor, PA, USA)
Monoclonal Anti-polyHistidine-Peroxidase antibody (Sigma-Aldrich, St. Louis, MO, USA)
Ni-NTA Agarose resin (Quiagen, Hilden, Germany)
Ni-NTA Magnetic Agarose Beads (Quiagen, Hilden, Germany)
PEG-NHS (Rapp polymere GmbH, Tuebingen, Germany)
Potassium chloride (Sigma-Aldrich, St Louis, MO, USA)
RNAse-free pipette tips (Biozym Scientific GmbH, Hessisch Oldendorf, Germany)
Roti®-PVDF membrane 0,45µm (Carl Roth GmbH & Co.KG, Karlsruhe, Germany)
Sigmacote® (Sigma-Aldrich, St Louis, MO, USA)
Sodium chloride (Fluka Chemie GmbH, Oberhaching, Germany)
Sodium dodecyl sulfate (SDS) (Applichem, Darmstadt, Germany)
Sodium hydroxide ≥ 99 % p.a. ISO in Plaetzchen (ROTH GmbH, Karlsruhe, Germany)
Sulfuric acid, (95.0-98.0) % (Sigma-Aldrich, St Louis, MO, USA)
Tetramethylethylendiamin (TEMED) (Sigma-Aldrich, St Louis, MO, USA)
Trifluoperazine dihydrochloride ≥ 99 % (Sigma-Aldrich, St Louis, MO, USA)
Triptone (Applichem, Darmstadt, Germany)
Tris hydrocholride, buffer grade (Applichem, Darmstadt, Germany)
Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) (Sigma-Aldrich, St Louis, MO, USA)
Tween20 (ROTH GmbH, Karlsruhe, Germany)
Yeast cell extract (Applichem, Darmstadt, Germany)
ZebaSpin Desalting Columns, 7K MWCO (Thermo Fisher Scientific Life Technologies GmbH, Darmstadt, Germany)

2.7 Fluorescence dyes and unnatural amino acids

Alexa Fluor 488 C5-maleimide (Invitrogen, Carlsbad, CA, USA)
Alexa Fluor 488 N-Hydroxysuccinimide-ester (NHS) (Invitrogen, Carlsbad, CA, USA)

Alexa Fluor 647 C2-maleimide (Invitrogen, Carlsbad, CA, USA)

Alexa Fluor 647 N-Hydroxysuccinimide-ester (NHS) (Invitrogen, Carlsbad, CA, USA)

ATTO488 N-Hydroxysuccinimide-ester (NHS) (ATTO-TEC, Siegen, Germany)

ATTO488-C-tRNA, cysteine (Biotechrabbit, Hennigsdorf, Germany)

ATTO488-K-tRNA, amber (Biotechrabbit, Hennigsdorf, Germany)

ATTO633 alkyne (Sigma-Aldrich, St Louis, MO, USA)

ATTO633-AmF-tRNA, amber (ProteinExpress, Chibo, Japan)

ATTO633-C-tRNA, cysteine (Biotechrabbit, Hennigsdorf, Germany)

ATTO633-K-tRNA, amber (Biotechrabbit, Hennigsdorf, Germany)

ATTO655 N-Hydroxysuccinimide-ester (NHS) (ATTO-TEC, Siegen, Germany)

ATTO655 N-Hydroxysuccinimide-ester (NHS) (ATTO-TEC, Siegen, Germany)

BODIPY 576/589-C-tRNA, cysteine (Biotechrabbit, Hennigsdorf, Germany)

BODIPY 576/589-K-tRNA, amber (Biotechrabbit, Hennigsdorf, Germany)

BODIPY FL C3-maleimide (Thermo Fisher Scientific Life Technologies GmbH, Darmstadt, Germany)

BODIPY FL N-Hydroxysuccinimide-ester (NHS)

BODIPY FL-C-tRNA, cysteine (Biotechrabbit, Hennigsdorf, Germany)

BODIPY FL-K-tRNA, amber (Biotechrabbit, Hennigsdorf, Germany)

Click-IT Alexa Fluor 488 DIBO alkyne (Thermo Fisher Scientific Life Technologies GmbH, Darmstadt, Germany)

Click-IT Alexa Fluor 647 DIBO alkyne (Thermo Fisher Scientific Life Technologies GmbH, Darmstadt, Germany)

CY5-C-tRNA, cysteine (Biotechrabbit, Hennigsdorf, Germany)

2.8 Buffer, solutions and media

Circular Dichroism

25 mM phosphate buffer

Cell culture

DYT: 16 g Trypton, 10 g yeast extract, 5 g NaCl in 1 L dH_20

DNA sequencing

10xTBE: 162 g Tris Base, 27.5 g Boric acid, 9.3 g EDTA pH 8.3-8.7

Licor Loading dye: 95 % Formamide, 20mM EDTA, 0.1 % Xylen cyanol FF

sequencing gel: 21g Urea, 32 ml water, 5 ml 10xTBE, 500µl DMSO, 4.3 ml Rapide Gel XL,

50 µl TEMED, 350 µl APS

FRET experiments

Apo-binding-buffer: 50mM MOPS, 10mM EGTA, 150 mM KCl, pH 7.5

Apo-buffer: 50mM MOPS, 10mM EGTA, pH 7.5

Holo-binding-buffer: 50mM MOPS, 10mM CaCl2, 150mM KCl pH 7.5

Holo-buffer: 50mM MOPS, 10mM CaCl2, pH 7.5

Glass slides passivation

Functionalizing buffer: 2 % solution of 3-aminopropyltriethoxysilane (APTES) (Sigma-Aldrich) in acetone

PEG solution: 200 mg/ml PEG-NHS in 1M NaHCO3, pH 8.3

Pirhanha solution: H_2SO_4 and 30 % H_2O_2, 2:1 v/v

Labeling buffers

Labeling buffer azide: PBS 1x, pH 7.5

Labeling buffer maleimide: PBS 1x, TCEP 10x concentration of cysteines, pH 7.5

Labeling buffer succinimidyl ester: 100mM $NaHCO_3$, pH 8.3

Phosphate buffered saline

PBS-Phosphate-Buffered saline (10x) pH 7.4

Protein purification buffers

Elution buffer: 50mM MOPS, 300 mM NaCl, 1mM TCEP, 500mM imidazole, pH 7.5

Guanidine hydrochloride buffer: 50 mM Tris, 150mM KCl, 0/2.5/4 M Guanidium hydrochloride

Lysis buffer: 50mM MOPS, 300 mM NaCl, 1mM TCEP, 5mM imidazole, pH 7.5

Wash buffer: 50mM MOPS, 300 mM NaCl, 1mM TCEP, 10mM imidazole, pH 7.5

SDS Polyacrylamide gel electrophoresis

Concentration gel (5 %): 15.2 ml water, 6.2 ml concentration gel buffer, 3.2 ml 30 % acrylamide, 225µl 10 % APS, 25µl TEMED

Concentration gel buffer: 0.5M TRIS, 4 % (w/v) SDS, pH 6. 8

Coomassie staining solution: 0.12 % (w/v) Coomassie blue R250, 10 % (v/v) ammonium sulfate, 10 % (v/v) acetic acid, 20 % (v/v) methanol

Running buffer: 25mM TRIS, 192mM Glycine, 0.1 % (w/v) SDS

SDS-Sample buffer: 60mM TRIS/HCl, 2 % (w/v) SDS, 14mM β-Mercaptoethanol, 25 % (w/v) glycerin, 0.1 % (w/v) bromophenol blue

Separation gel (15 %): 7.2 ml water, 7.5 ml separation gel buffer, 15 ml 30 % acrylamide,

300µl 10 % APS, 30µl TEMED

Separation gel buffer: 1.5 M TRIS, 0.4 % (w/v) SDS, pH 8.8

Western blot

Blocking buffer: 5 g milk, 2.5 g BSA in 50 ml TTBS

Blotting buffer: 39mM Glycin, 48 mM Tris (Base), 20 % (v/v) methanol

Wash buffer1: TBS: 20mM Tris, 137mM NaCl, pH 7.6

Wash buffer2: TTBS: 20mM Tris, 137mM NaCl, 0.1 % (v/v) Tween 20 pH 7.6

3. Method

3.1 Generation of the plasmids

Plasmids encoding either wild-type (wt) or mutant calmodulin (i.e. *CaM wt*, *CaM T34amber*, *CaM T110C*, *CaM T34C*, *CaM T34amber T110C* and *CaM T34C T110amber*) used in the present studies were kindly provided by Dr. Alexandros Katranidis.
A brief description of the method used for the generation of the mutants is given below.

3.1.1 Polymerase chain reaction and site-directed mutagenesis

All the plasmids containing mutant calmodulin were generated via site-directed mutagenesis by using the polymerase chain reaction (PCR). PCR is a very common technique used in molecular biology for amplifying DNA. It is constituted of a repeated cycle of three steps: denaturation, annealing and elongation (**Fig. 7**).

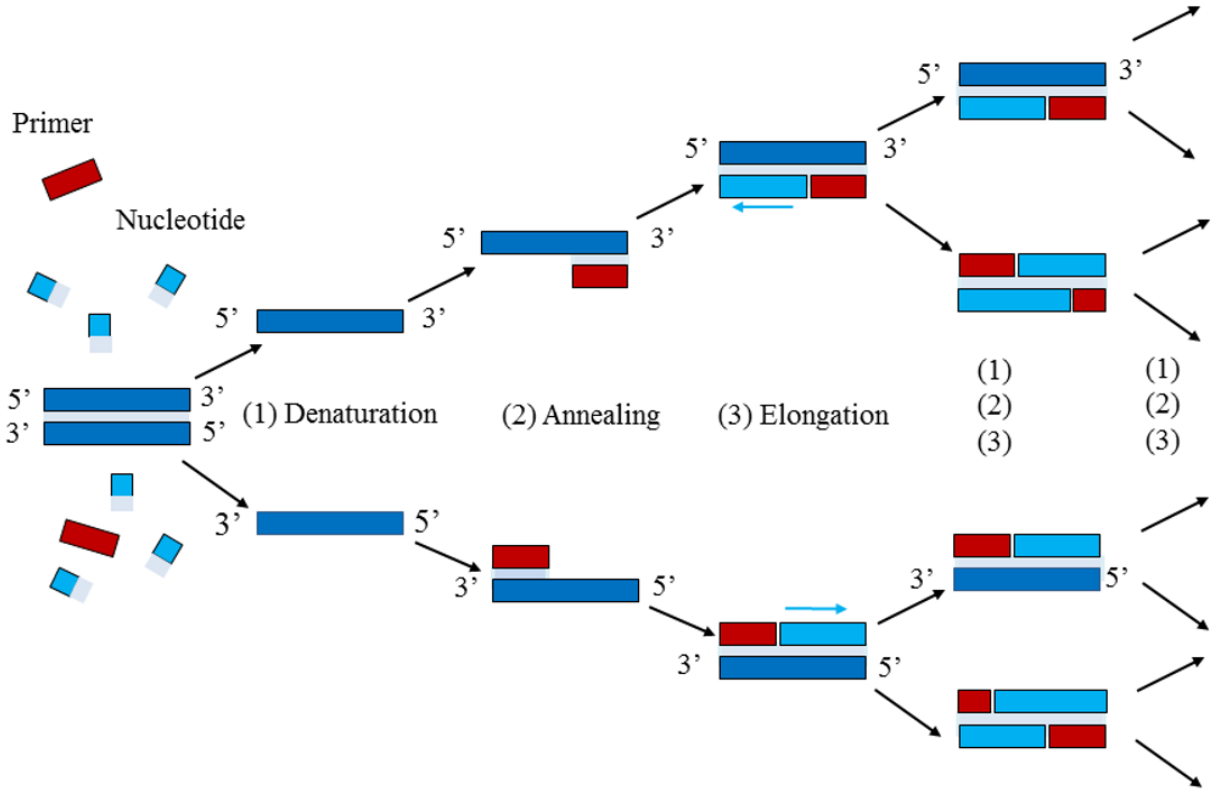


Figure 7. **Polymerase chain reaction (PCR).** Schematic representation of the PCR technique. PCR is performed by a repetition of cycles of (1) denaturation, (2) annealing and (3) elongation. The original DNA template is shown in dark blue whereas the newly generated DNA is shown in light blue and the DNA primers in red [89].

An initial denaturation step starts the reaction. It is followed by a repetition of around 20-40 cycles of denaturation-annealing-elongation. Each of the three steps is performed in a short timescale (i.e. from tens of seconds up to few minutes). The denaturation is typically done at high temperature (i.e. 90-100 °C). It is during this step that the DNA template is fully denatured. Later, the temperature significantly decreases (i.e. down to 30-65 °C) to allow the annealing and hybridization of a pair of primers to the complementary sequence. The primers allow the initiation of the elongation by a specific DNA polymerase [89, 90]. Depending on the type of polymerase used the temperature associated with the latter step varies. For the present study, the extension was performed by a high-fidelity DNA polymerase that is commonly used at 72 °C.

Firstly, the plasmids encoding the single mutants *CaM T34C*, *CaM T110C* and *CaM T34amber* were generated. To this aim, two PCR steps were performed. The first one was used to generate a fragment of the region of interest containing the desired mutation. To this aim, one of the primers was carrying the specific mutation. Later, the obtained product was used as primer to initiate a second PCR. Here, the aim was to get the entire sequence of the gene encoding the single mutant calmodulin.

Secondly, the double mutants *CAM T34C T110amber* and *CaM T34amber T110C* were generated via a similar two-step procedure. The first PCR served to insert the amber codon either at the position 34 or 110. Next, the PCR fragments containing the amber codon were used as primers for another PCR performed on the plasmid containing either *CaM T34C* or *CaM T110C*. As a consequence, we generated the plasmids containing the sequence of the gene encoding the double mutant calmodulins.

3.1.2 Cloning, sequencing and amplification

CaM wt, CaM T34amber, *CaM T110C, CaM T34amber T110C* and *CaM T34C T110amber* were cloned into pRSET for amplification in Top10 *E.coli* cells. The cloning is typically constituted of four steps: fragmentation, ligation, transfection and selection. Firstly, the DNA fragments containing, on one hand, the insert and, on the other hand, the linearized pRSET vector were obtained by digestion with proper restriction enzymes. Later, the two fragments were separated by agarose gel electrophoresis and isolated via ion-exchange chromatography. The vector was additionally dephosphorylated to remain linear and the two fragments were mixed together and treated with a T4 phage DNA ligase. Once the plasmid of interest containing the mutant gene was obtained, competent Top10 *E. coli* cells were transformed for plasmid amplification. Finally, the selection of the bacteria having incorporated the proper

plasmid was done by growing them in the presence of ampicillin for which a resistance gene is encoded by pRSET. In order to verify the correct sequence of the constructs, the plasmids were subsequently purified from single colonies by using a commercial miniprep kit according to the manufacturer's protocol. DNA sequencing was either performed by an external company or in our lab by using an automated infrared DNA-sequencer. In the latter case, the Sanger sequencing method [91] was used. In brief, PCR reactions were performed with a fluorescent labeled primer in the presence of dideoxynucleotide triphosphates (ddNTPs) lacking the 3'-hydroxyl group, which stop the replication after being inserted. For a complete sequencing, 4 different PCRs were performed for the 4 nucleotides A, T, C and G. After electrophoresis on a 4.8 % polyacrylamide gel, the sequence was checked and correct constructs were transformed in Top10 *E.coli* cells for isolating a larger amount by using a commercial midiprep kit according to the manufacturer's instructions.

3.2 Protein sample preparation

3.2.1 *In vivo* expression in BL21

CaM wt, *CaM T110C* and *CaM T34C* were previously cloned in pRSET vector (see 3.1.2). This vector contains a T7 promoter which binds specifically the T7 RNA-polymerase. In addition, the T7 promoter is associated to a *Lac* O operator that is, in absence of an inducer, bound to the *lac* repressor (*lac I*). The binding of the repressor prevents the expression of the target gene even in the presence of the T7 RNA-polymerase (**Fig. 8**, right).

His-tagged CaM wt, CaM T110C and CaM T34C were expressed in transformed *E. coli* BL21 (DE3) competent cells. These cells contain a chromosomal copy of the T7 RNA-polymerase necessary for the expression of the gene cloned into pRSET (see above). The RNA-polymerase is under the control of the lac promoter. Additionally, the promoter is associated to the *Lac* O operator, which prevents the expression of the T7 RNA-polymerase in absence of an inducer by binding the *lac* repressor (**Fig. 8**, left). Commonly, the isopropyl β-D-1-thiogalactopyranoside (IPTG), a lactose analogue, is used to induce the expression of the target gene by suppressing the *lac I* inhibition.

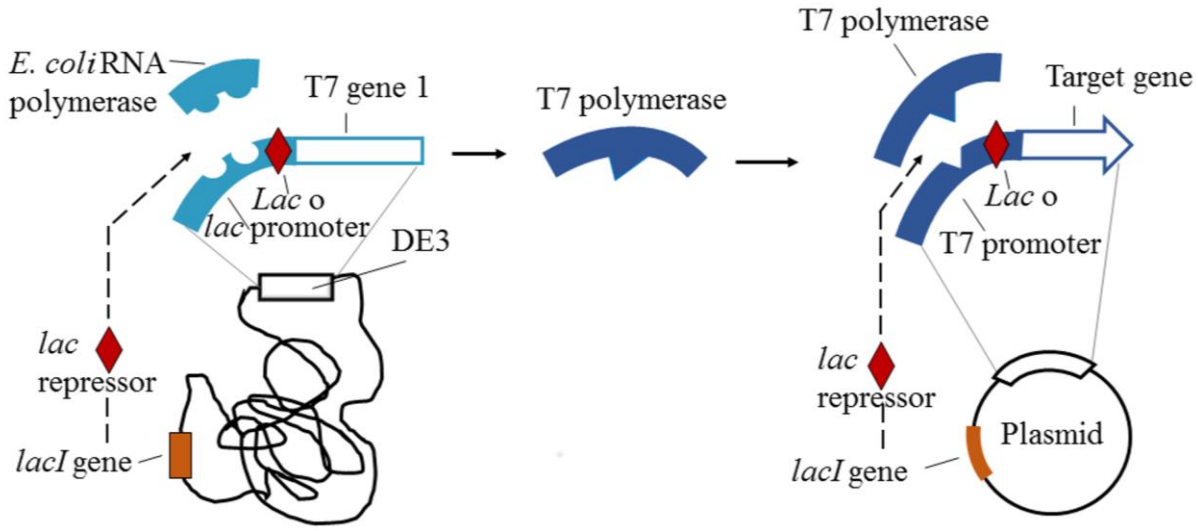


Figure 8. **Gene expression in BL21 (D3).** Schematic representation of the gene expression of a vector similar to pRSET. In absence of an inducer, the *lac* repressor (*lac I*) encoded by BL21 (D3) binds to the *Lac* O operator associated with the promoter of the genomic T7 RNA polymerase gene preventing its expression (left). Additionally, the *lac* repressor encoded by pRSET binds to the *lac* O operator of the target gene switching off its expression even when the basal expressed T7 RNA polymerase is present (right). IPTG leads to the removal of the inhibitions by *lac I*.

In vivo expression is constituted of 4 main steps: transformation, induction, lysis and purification. In the present work, the transformation was performed by heat shock. Firstly, cells were incubated with the DNA plasmids for 20 min at 4 °C. Subsequently, the heat shock was performed for 2 min at 42 °C and directly followed by a final incubation of 1 min at 4 °C. Then, cells were recovered for 1 h at 37 °C by incubating them with 800 µl of DYT medium. Later, cells were streaked on DYT agar plates containing ampicillin for the selection and incubated for 15 h at 37 °C. Afterwards, colonies were harvested and grown in liquid medium containing ampicillin. When the optical density (O.D.) reached 0.6, the induction was performed for 3 h at 37 °C with 1 mM of IPTG. After induction, cells were ultracentrifuged for 15 min at 5000 rpm. Next, the lysis of the cells was performed on resuspended pellets in lysis buffer by using a cell disruptor. Before purification, the lysate was ultracentrifuged an additional time. The purification of the proteins was performed by immobilized metal affinity chromatography (IMAC). IMAC is based on the specific interaction of few amino acids such as histidine residues with metal [92]. In the present work, we have used the nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography, which is based on the specific interaction of nickel with histidine residues. A column containing Ni-NTA agarose beads was first equilibrated with 10 times bed volumes of lysis buffer. Later,

the agarose beads were incubated with the lysate for 1 hour at 4 °C. Then, the column was washed 3 times with 10 times bed volumes of washing buffer. Finally, elution of his-tagged proteins was performed with 500 mM imidazole, a histidine analogue which competes for the binding to the resin.

3.2.2 CFPS and incorporation of unnatural amino acids (UAAs)

CaM wt, CaM T110C, CaM T34C, CaM T34AzF, CaM T34AzF T110C, CaM T34C T110AzF were synthesized by using *E. coli* derived *in vitro* coupled transcription-translation systems. Protein synthesis was performed in 50 µl at 32 °C under shaking either for 120 min in a batch-mode in vials or for 24 h in continuous-exchange in RTS 100 dialysis devices according to the manufacturer's instructions. The latter contains a dialysis membrane, which allows only small molecules to pass through and prevents the passage of larger ones. Therefore, it is permeable to substrates required for the transcription and translation reactions but not to proteins. The aim of the continuous-exchange cell-free system (CECF) is to increase the protein yield obtained from *in vitro* synthesis by feeding continuously the reaction with substrates that would be otherwise gradually depleted in a batch-mode. In addition, it prevents the accumulation of by-products such as inorganic phosphate that inhibit the reactions. Indeed, it is thought that the limited duration of the reaction in batch-mode is due to both the depletion of the substrates and the inactivation of protein synthesis by inhibitory by-products [93]. In the continuous-exchange cell-free system device, a first chamber containing the reaction solution is separated by a dialysis membrane from a second chamber filled with a feeding solution. In the present work, the device contained the feeding solution in a chamber that was below the reaction chamber (**Fig. 9**, right). Even though the cost of CFPS performed in continuous-exchange is significantly increased in respect to batch-mode, it could be necessary for some specific applications to boost the yield of proteins, which is one of the major issues encountered in cell-free synthesis.

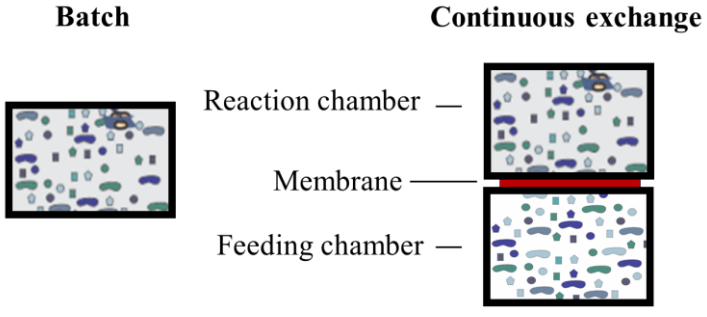


Figure 9. Comparison between a batch-mode and a CECF. In the batch-mode (left), the reaction compartment is depleted in substrates over time and inhibitory by-products accumulate into it whereas in the continuous-exchange (right) thanks to a dialysis membrane the substrates could be renewed and by-products removed.

The incorporation of the unnatural amino acid, p-azido-phenylalanine (AzF), into CaM was achieved with the constructs *CaM T34amber*, *CaM T34amber T110C* and *CaM T34C T110amber* via an amber codon by using an *E. coli* derived orthogonal coupled cell-free transcription-translation system depleted from RF1 [39] that would otherwise normally compete for the incorporation at the amber codon. In the presence of RF1, truncated CaM products tend to deplete the system from substrates decreasing significantly the amount of full-length proteins obtained. In the aim to increase the yield, modified *in vitro* coupled transcription-translation systems lacking RF1 were therefore developed. Additionally, the modified cell-free system used was supplemented with enriched fractions of orthogonal amber suppressor tRNA , p-azido-phenylalanyl-tRNA synthetase [94] and AzF (**Table 1**).

Table 1. Cell-free system for incorporation of AzF into CaM (provided by M. Gerrits).

***E. coli* extract**	**Reaction buffer**	**Feeding buffer**
70 A260/ml RF1-depleted E. coli extract	50 mM HEPES pH 7.6	35 mM HEPES pH7.6
1 mg/ml purified AzF tRNA synthetase	100 mM KOAc	70 mM KOAc
0.4 mg/ml total tRNA preparation with overexpressed amber suppressor tRNATyrCUA	50 mM NH4OAc	35 mM NH4OAc
500 U/ml T7 RNA polymerase	13 mM MgCl2	12 mM MgCl2
0.008 mg/ml pyruvate kinase	0.1 mM EDTA	0.07 mM EDTA
0.1 U/ml RNasin and proteinase inhibitor	0.65 mM DTT	3.15 mM DTT
	4 % PEG3000	4 % PEG3000
	100 µM Folic acid	100 µM Folic acid
	1 mM ATP and GTP	1 mM ATP and GTP
	0.5 mM CTP and UTP	0.5 mM CTP and UTP
	30 mM Phosphoenolpyruvate	30 mM Phosphoenolpyruvate
	10 mM acetyl phosphate	10 mM acetyl phosphate
	0.02 % sodium azide	0.02 % sodium azide
	1.2 mM of each amino acid including AzF	1.2 mM of each amino acid including AzF

The p-azido-phenylalanyl-tRNA synthetase attaches the AzF to the amber suppressor tRNA, which is subsequently used for the site-specific incorporation of the UAA at an amber codon. The incorporation of the fluorescent unnatural amino acids (ATTO633-AmF, ATTO633-Lys, ATTO488-Lys, BODIPY-FL-Lys, ATTO633-Cys, ATTO488-Cys BODIPY-FL-Cys, Cy5-Cys, BODIPY 576/589-Lys and BODIPY 576/589-Cys) into CaM was achieved with the constructs *CaM T34amber*, *CaM T34amber T110C* and *CaM T110C* by using an *E. coli* derived *in vitro* coupled transcription-translation system depleted from both RF1 and Cys. Cysteine residues were provided separately when the incorporation via cysteine codon was not required. Both RF1 and cysteine were depleted to prevent competition with the specific precharged tRNAs carrying the fluorescent UAAs. The incorporation of fluorophores by using a cysteine codon was performed by supplementing the reaction solution with either ATTO633-C-tRNA-GCA, ATTO488-C-tRNA-GCA, BODIPY-FL-C-tRNA-GCA, Cy5-C-tRNA-GCA and BODIPY 576/589-C-tRNA-GCA at a final concentration of 10 µM. Similarly, ATTO633-K-tRNA-CUA, ATTO488-K-tRNA-CUA, ATTO633-AmF-tRNA-CUA, BODIPY-FL-K-tRNA-CUA or BODIPY 576/589-K-tRNA-CUA was used for the incorporation made via an amber codon. When, ATTO633-AmF-tRNA-CUA was preferred to ATTO633-K-tRNA-CUA, it was used according to the manufacturer's instructions. The dual incorporation of BODIPY-FL-Lys and ATTO633-AmF was done by adding both BODIPY-FL-C-tRNA-GCA and ATTO633-AmF-tRNA-CUA to the reaction solution. Following the reaction, the remaining fluorescent UAA-charged tRNAs were deacylated for 1 h at 37 °C, pH 8.3.

CFPS products were purified by using either Ni-NTA agarose beads (i.e. continuous-exchange reactions) or Ni-NTA magnetic agarose beads (i.e. batch-mode reaction). First, 50 µl of beads were equilibrated with 4 times bed volume of lysis buffer. Later, the beads were mixed with 50 µl of sample and 150 µl of lysis buffer for 2 h at 4 °C. Then, the beads were washed with 3 times bed volume of wash buffer. Subsequently, the his-tagged products were eluted with 500 mM imidazole in 2 times 50 µl of elution buffer. Finally, by using size exclusion Zeba™ desalting columns, the product was desalted with either PBS (pH 7.5) or PBS-Tris(2-carboxyethyl) phosphine (TCEP) (pH 7.2). A special treatment was applied to CFPS products that were obtained after incorporation of fluorescent UAAs and for which the yield is drastically decreased. To avoid too high competition of the his-tagged products with imidazole, purification was made as previously stated but with lysis and wash buffers containing 10 times diluted imidazole concentrations.

3.2.3 Labeling by post-translational chemical modifications

Nonspecific labeling of commercial recombinant human calmodulin and purified CaM wt was made on exposed lysine residues and N-terminal amino groups by using 6-fold excess of BODIPY-FL-NHS Ester (succinimidyl Ester) in the presence of 100 mM $NaHCO_3$ (pH 8.3) for 5 h at 25 °C. In fact, amine residues react efficiently with NHS-functionalized dyes in a pH range of 8.0-9.0.

Specific labeling of purified CaM T110C was done in anaerobic conditions for 15 h at 4 °C on exposed cysteine residues by using a 20–fold excess of Alexa Fluor 488 (AF488)-maleimide and in PBS in presence of TCEP (pH 7.2), a reducing agent. Anaerobic conditions were achieved by degassing the labeling buffer for at least 120 min. In fact, in reducing conditions, thiol groups react efficiently with maleimide-functionalized dyes at neutral pH.

Specific labeling of CaM T34AzF was performed on AzF residues for 15 h at 25 °C with a 25-fold excess of Alexa Fluor 647 (AF647)-DIBO alkyne in PBS (pH 7.5). Azide groups usually react with alkyne-functionalized groups through a copper-dependent reaction. However, DIBO dyes are designed for more easy to handle copper free reactions so-called strain promoted azide-alkyne cycloadditions (SPAAC) and therefore were chosen for the present study [95].

Finally, selective dual-labeling of CaM T34AzF T110C and CaM T34C T110AzF was achieved in similar ways as previously stated on both azide and thiol groups. Firstly, CaM was reacted with AF647-DIBO alkyne. After being purified the single labeled calmodulin was dually labeled by using AF488-maleimide. After dual-labeling, the purification of the his-tagged CaM by using Ni-NTA magnetic agarose beads followed by size-exclusion chromatography was used for dye removal. In brief, size-exclusion chromatography is a chromatographic method in which the molecules in solution are separated according to their hydrodynamic volume which strongly depends on the size of the molecule. The separation is done by using pores of different sizes in the stationary phase and is therefore based on the specific permeation of the molecules in solution **(Fig. 10)** [96].

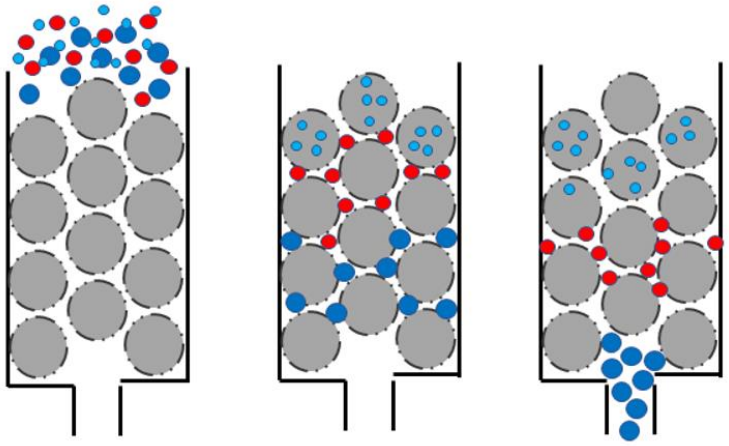


Figure 10. **Size-exclusion chromatography.** Schematic representation of size-exclusion chromatography. The different molecules in solution (red, light blue and dark blue balls) are running through pores of different sizes and are separated according to their hydrodynamic volume because of their specific permeation.

In the present work, the aim is to separate the low molecular weight dyes from the protein of interest. Good separation is achieved since proteins display much lower permeation rates compared to the dyes and are therefore moving faster through the column.

3.2.4 SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

15 % SDS PAGE was used to verify products from CFPS and *in vivo* expression. In short, SDS PAGE is a molecular biology technique used to separate proteins according to their size. The separation is achieved by submitting to an electrical field a mixture of proteins that were previously denatured and uniformly charged by using SDS. In such context, the electrophoretic mobility of each protein will depend only on the size. During their migration, proteins will sequentially encounter the so-called stacking and separation gels. The former contains a low percentage of acrylamide which allows fast migration of the sample. The aim of the stacking gel is to concentrate the sample at its border with the separation gel that contains a much higher percentage of acrylamide. This ensures a common starting point for all the proteins when entering the separation gel. Then, it is during migration in the separation gel that the proteins are separated according to their size [97].

In the present study, 20 µl of the sample were incubated under shaking for 30 min at 25 °C with 5 µl of loading buffer in the presence or absence of EGTA. Before loading, the mixture was shortly heated for 3 min at 90 °C to fully denature the proteins. For non-fluorescent products, the gel was stained for at least 15 h with 0.1 % Coomassie Brilliant Blue R-250. Later, the gel was visualized with a ChemiDoc MP gel imaging system after being washed

twice with MilliQ water. Alternatively, for fluorescent products, the gel was first read with a Typhoon 9500 laser scanner after being washed once with MilliQ water. After fluorescence imaging, it was stained with 0.1 % Coomassie Brilliant Blue R-250, as previously stated. Exceptionally, when western blots were performed for the detection of his-tagged products, the gel was not stained. Instead, proteins from the gel were transferred to a PVDF membrane by using the semidry method [98]. To this aim, the gel and the membrane are placed between two layers of filter papers that were impregnated with blotting buffer and an electrical field was applied. As a consequence, the proteins migrated out from the gel and bound to the membrane. Prior to the transfer, the membrane was incubating with methanol for allowing proper binding of the proteins. The transfer was performed in a western blot chamber at a constant voltage of 0.8 mA / cm^2 of gel area according to the manufacturer's instructions. Later, the PVDF membrane was incubating for 1 h at 25 °C with blocking buffer to prevent free binding to the surface. After being washed, it was incubated for 1 h at 25 °C under gentle shaking with monoclonal anti-His antibody HRP conjugate. Finally, the PVDF membrane was washed with wash buffer1 and wash buffer2 and the visualization was performed by using the western blotting detection kit according to the manufacturer's instructions.

3.3 Absorption spectroscopy measurements

Absorbing molecules can absorb light of proper wavelength. This property is used in absorption spectroscopy for the measurement of absorption spectra and the determination of sample concentration. In particular, phenylalanine, tyrosine and tryptophan residues are three naturally fluorescent aromatic amino acids that are often encountered in proteins. Their absorption at 280 nm is widely used for quantifying proteins [99]. Additionally, their near UV absorption spectra can also be used to identify purified recombinant proteins since the wavelength of maximum absorption is specific to each aromatic amino acid and their relative content is pretty specific to a protein. For these reasons, the shape and the maxima of the near UV absorption spectrum should vary from one protein to another and can be used for identifying purified proteins [100]. However, since tryptophan residues are absorbing much more than tyrosine and phenylalanine residues, the near UV spectrum is often mostly shaped by their presence and the application of that methodological approach for protein identification is therefore very restricted. Nevertheless, calmodulin is a particular case since its primary sequence does not contain any tryptophan residue making the near UV spectrum highly specific.

In the present study, absorption spectra were needed for determining protein concentrations and identifying CaM in purified samples. In addition, there were also required for quantum yield and Förster radius determinations of the FRET dye pairs. An absorption spectrophotometer was used to measure the absorption of unpolarized light by a specific sample by recording the intensity of the transmitted light.

For protein concentration determination, the absorption was measured at the specific wavelength λ of 280 nm and the following Beer-Lambert law was used [101]:

$$I_d = I_0 \cdot 10^{-\varepsilon \cdot c \cdot d} \qquad \text{(Eq. 1)}$$

Where I_d, I_0, ε, c and d are the intensity of the transmitted and the incident light, the molar absorption coefficient in $M^{-1}cm^{-1}$, the molar concentration and the optical pathway length in cm, respectively.

For the measurements of absorption spectra, the transmitted intensity was measured over a large range of wavelengths which covers the entire spectrum.

For near UV absorption spectroscopy used for identifying CaM, a commercial calmodulin was used as a reference and the samples were incubated under shaking for at least 30 min at 25 °C before all measurements.

Measurements made with the UV 2401PC Spectrophotometer were done with a sampling interval of 1 nm, an accumulation time of 0.1 s, a slit width of 0.5 nm and a path length of 10 mm. Measurements made with the NanoDrop 2000c were performed with the A280 application of the NanoDrop™ software with a sampling interval of 1 nm and a path length of 1 mm. For all the experiments, the background from the specific buffer was subtracted and data were analyzed by using OriginPro.

3.4 Circular dichroism spectroscopy measurements

Far UV circular dichroism (CD) spectroscopy was used to calculate the relative content of secondary structures of wild-type and mutant calmodulins. CD is based on the property of chiral and asymmetric molecules to absorb differently left-handed and right-handed circularly polarized light. As a consequence, the circularly polarized light becomes elliptically polarized after absorption. Secondary structures from proteins have such properties of dichroic absorption of circularly polarized light [102]. In fact, each type of secondary element can be characterized by a specific CD spectrum (**Fig. 11)**.

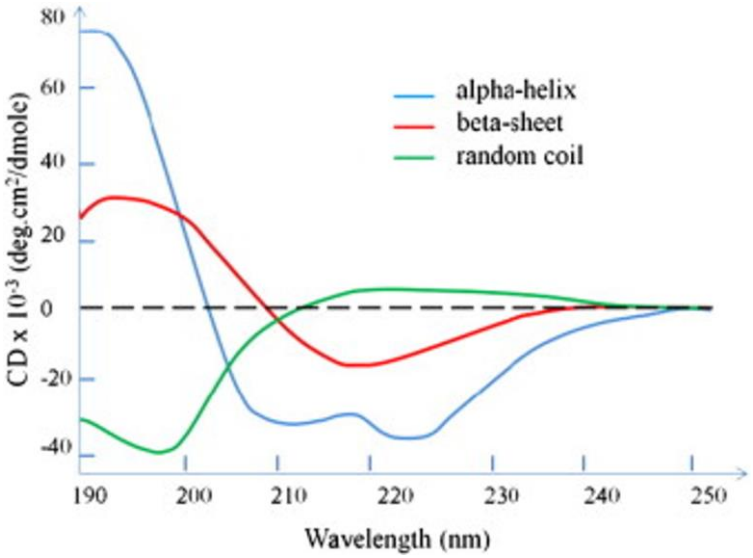


Figure 11. Far UV spectra of secondary structures. Far UV spectra of α-helices (bleu), β-sheets (red) and random coils (green) are displayed [103].

The spectrum of α-helical components has one positive peak at 191-193 nm and two negative peaks at 208 nm and 222 nm, whereas β-sheet-related structures display a spectrum with one positive and one negative peak at 198 nm and 215 nm, respectively. On the contrary, a random coil is characterized by negative and positive peaks that are roughly opposite from those previously stated [103, 104].

CD spectra are recorded in degrees of ellipticity according to the following equation [102]:

$$\theta = 32.98\Delta A \qquad \text{(Eq. 2)}$$

with

$$\Delta A = A_L - A_R \qquad \text{(Eq. 3)}$$

Where θ is the ellipticity in degrees and ΔA represent the difference in absorption of left-handed and right-handed light. Usually, for comparison purposes, the mean residue ellipticity θ_{MRW} in $[\frac{\text{mdeg cm}^2}{\text{dmol}}]$ is preferred to the degree of ellipticity. It can be calculated by using the following equation [104]:

$$\theta_{MRW} = \frac{MRW \cdot \theta(\lambda)}{10 \cdot d \cdot c} \qquad \text{(Eq. 4)}$$

Where θ (λ), MRW, d and c are the ellipticity at the wavelength λ, the mean residue weight, the optical path length in cm and the concentration in g/ml, respectively. The MRW represents the molecular weight in Dalton divided by the number of peptide bonds.

For circular dichroism spectroscopy measurements, non-labeled proteins were used to prevent saturation of the signal by absorption of the dye. In the aim to reach an optimal concentration (i.e. few hundreds of μg/ml), CFPS was performed in continuous-exchange. CD measurements were performed in 25 mM phosphate buffer (pH 7.5) at 25 °C in a 1 mm quartz cell by using a Jasco J-1100-Spectropolarimeter over a large range of wavelengths going from 190 nm to 260 nm and the buffer background was subtracted. In the present study, the results are shown in degree of ellipticity. In fact, the conversion in molar residual ellipticity required a precise measurement of the sample concentration that is practically not possible in the present situation due to the low amount of CaM obtained by CFPS combined to its particularly low absorption coefficient related to the absence of tryptophan residue in the amino acid sequence. Therefore, the ellipticity was kept in degrees of ellipticity and the relative amount of secondary structural elements was evaluated by using the Secondary Structure Estimation software included in the Jasco package.

3.5 Fluorescence spectroscopy measurements

3.5.1 Basic principle of fluorescence

The principle of fluorescence is well defined by the energy diagram of Jablonski. As shown in **Fig. 12**, a fluorophore has the particularity to emit light after being excited. Excitation occurs when it absorbs photons of proper energy according to the different electronic states. The absorbance is therefore only possible for some wavelengths leading to a specific distinct absorption spectrum. After photon absorption, electrons undergo a transition from the electronic ground state S_0 to an excited state (e.g. S_1 or S_2). Fluorescence occurs when they return to the ground state. Indeed, once excited the electron has to go back to the lowest electronic state, S_0. It is achieved by dissipation of energy via numerous processes that are either non-radiative such as vibrational relaxation and internal conversion or radiative through emission of photons in the so-called fluorescence process [43].

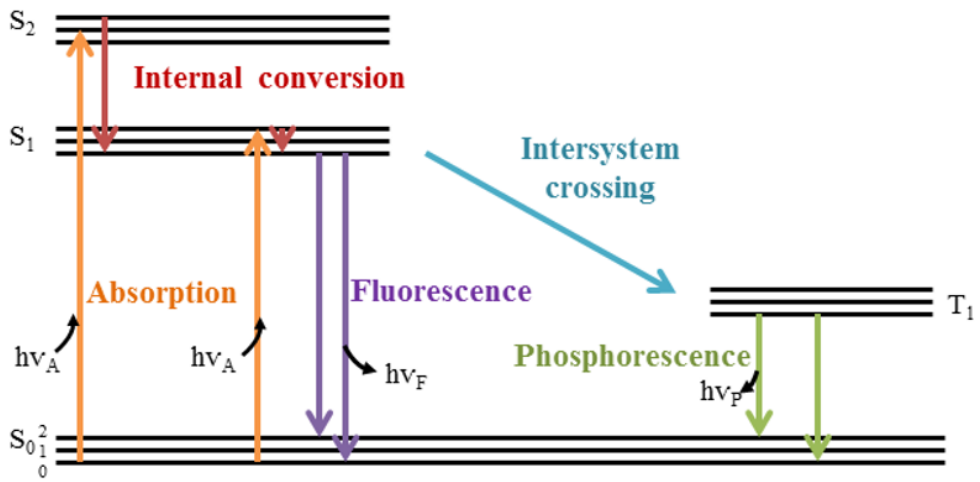


Figure 12. Jablonski diagram and principle of fluorescence. The fluorescence physical process is governed by transitions between different electronic states. The absorption of photons of proper energy by a fluorophore leads to its transition towards an electronically excited state (e.g. S_1 or S_2) (orange arrows). The return to the ground states (S_0) can be achieved through various processes including fluorescence (violet arrows) or intersystem crossing (cyan arrow) which leads to the transition towards triplet states and which can be followed by phosphorescence (green arrows).

Since vibrational relaxation and internal conversion are much faster processes (i.e. 10^{-14}- 10^{-11} seconds) than fluorescence (i.e. 10^{-9}- 10^{-7} seconds), they are more likely to occur immediately after absorbance. On the contrary, the lifetime of the fluorescence decay T_{fluo} is typically of few nanoseconds. As for the absorption, emission of photon occurs only for certain wavelengths leading to a distinct fluorescence spectrum red-shifted in respect to the absorption spectrum due to the loss of energy resulting from non-radiative processes. This specific phenomenon is known as the Stokes shift. Additionally, electrons in excited states can also proceed to dissipation of energy through intersystem crossing. In that case they undergo a transition towards the triplet state. This process is pretty slow and therefore less probable but has to be carefully considered in fluorescence spectroscopy. Indeed, the lifetime of the triplet state is much higher compared to the one of a singlet state leading to a long period in which the electron cannot be excited anymore. This results in unwanted so-called dark states in fluorescence spectroscopy. The return to the ground state takes place via different radiative and non-radiative ways. In that specific case, the transition taking place by emission of photons is called phosphorescence [43]. The phosphorescence lifetime is typically from milliseconds up to few minutes and therefore much longer than the fluorescence lifetime. In fact, phosphorescence is a very slow process that is in turn less probable than non-radiative processes and quenching. The latter has to be carefully

considered in fluorescence spectroscopy since it impacts on the fluorescence properties of a molecule. Dynamic quenching happens by collision with a molecule called the quencher whereas static quenching occurs by formation of complexes with other molecules. The former affects both the lifetime and the quantum yield of the fluorophore while the latter affects only the quantum yield [43, 105].

3.5.2 Fluorescence spectrophotometry

In the present work, a fluorescence spectrophotometer was used to measure the emission spectra of the various free- and calmodulin-bound-fluorophores and to quantify the protein yields obtained with different cell-free systems by measuring fluorescence intensities from *in vitro* synthesized Emerald green fluorescence protein (Emerald GFP). In addition, the same device was used to detect FRET from selectively dually labeled calmodulins. On one hand, fluorescence spectra were required for quantum yield, Förster radius, detection efficiency and GFP concentration determination. On the other hand, ensemble FRET measurements were performed for verifying the selective dual-labeling of our protein of interest. A schematic representation of the setup is shown in **Fig. 13**.

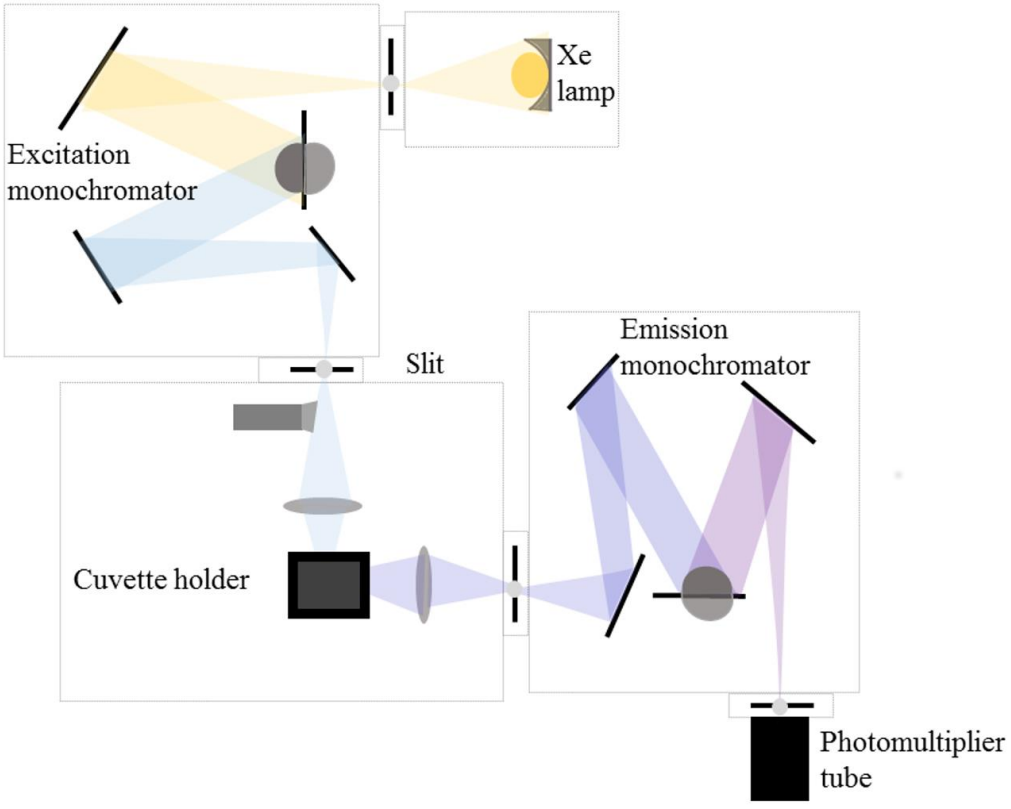


Figure 13. Schematic representation of a fluorescence spectrophotometer. The light path is shown in colors going from yellow to violet [106-108].

As shown, the excitation light of a xenon lamp which covers a broad range of wavelengths is used. Firstly, a fixed wavelength for exciting specifically the fluorophore is selected in the excitation monochromator. Afterwards, the light leaves the monochromator and reaches the sample. After excitation, the fluorophores present in the sample emit light. The fluorescence signal is detected by a photomultiplier tube after having gone through the emission monochromator, which selects the wavelengths to be recorded [43, 106, 107]. In the present study, all the samples were measured in a 10 mm quartz cuvette that was previously washed with MiliQ water and ethanol. For all the measurements, the adjustable slits were used with widths of 0.5 and 1 mm for excitation and emission, respectively. The data were recorded using the FelixGX 4.1.2 software. In all case, the data recorded were corrected for the wavelength dependency of the detection efficiency. The typical integration time used was 1 s. For ensemble FRET experiments, the excitation light was set at 470 nm and the emission signals was recorded over a range of wavelengths that covers the emission spectra of both the donor and acceptor fluorophores. The idea was to excite the donor only but record the fluorescence signals from both the donor and the acceptor. In all the other measurements, the excitation light used was set in such a way that the emission signal can be recorded over a range of wavelengths that covers the entire emission spectrum of the fluorophore of interest (i.e. 470 nm for BODIPY-FL and AF488 and 540 nm for ATTO633 and AF 647). Finally, data were analyzed by using OriginPro after having corrected all the spectra for the background contribution of the buffer. For the determination of Emerald GFP concentration, a relevant calibration curve was used during data analysis.

3.5.3 Confocal fluorescence measurements

3.5.3.1 Slide preparation

The passivation of a cover slide is often a necessary step in fluorescence microscopy. The idea is to treat the material in such a way that it becomes passive towards its environment. When working with proteins, the risk is that they bind to the cover slide. In the present study, the cover slides used for all the experiments were treated in such a way that sticking of proteins to the surface was prevented. To this aim, high precision glass cover slides were PEG-passivated (**Fig. 14**) [109, 110]. Passivation was performed in three steps: cleaning, functionalization and PEGylation. Firstly, the cleaning was performed at 25 °C for 15 h with piranha solution. Later, the surface was functionalized with amino-groups by incubating with a 2 % acetone solution of 3-aminopropyltriethoxysilane (APTES) for 10 min at 25 °C.

Finally, the passivation was made at 4 °C by incubating for at least 12 h with a solution of 1 M $NaHCO_3$ (pH 8.3) containing 200 mg/ml PEG-NHS ester. Before the measurement, all cover slides were washed with MilliQ water, cleaned with acetone and methanol and dried with nitrogen. In addition, a closed chamber was built for preventing evaporation by gluing a half-cut protein LoBind Eppendorf tube to the slide (**Fig. 14**).

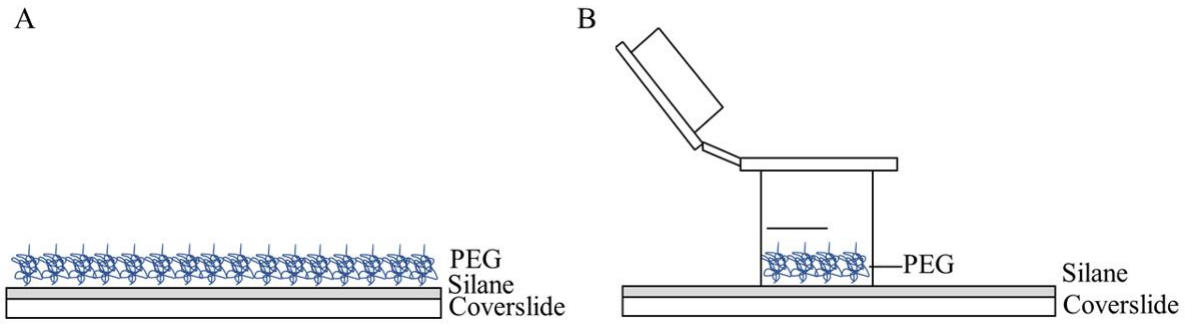


Figure 14. Passivation of cover slides with polyethylene glycol (PEG). (A) The passivation is made in three steps: cleaning, functionalization and PEGylation. Functionalization is done by covering the surface with silane (gray) which allows further PEGylation (blue). (B) In the present studies glued half-cut protein LoBind Eppendorf tube were used to prevent evaporation.

3.5.3.2 Fluorescence confocal microscope

In opposition to classical light microscopy, confocal microscopy allows the illumination and light-collection from a defined focal spot excluding out-of-focus glare, which contaminates the detected signal. That particular feature improves greatly the resolution of the technique [111]. In the present work, fluorescence correlation spectroscopy (FCS), single-molecule FRET (smFRET) and single-molecule two-color coincidence detection (TCCD) measurements were performed on freely diffusing fluorescent calmodulin with a MicroTime 200 confocal inverted microscope [112]. A schematic representation of the setup is shown in **Fig. 15**. In short, pulsed excitation light generated by 485 nm and 640 nm lasers is reflected towards a high numerical aperture water immersion objective for reaching the sample. The focus is set up in such a way that the focal spot is located deeply inside the solution (i.e. about 30 μm). Therefore, freely diffusing labeled molecules are efficiently excited. Subsequently, the fluorescence emission light is collected with the same objective and then sent through a major dichroic mirror (DM) before reaching a 75 μm pinhole for spatial filtering. Later, depending on the type of measurements, the emission light is either split between both donor and acceptor channels by passing through a dichroic mirror or sent or reflected towards one of the two channels only. In each detection channel, the emission light

is then either split into two equivalent beams by a 50:50 beam splitter or sent towards one of the two single-photon avalanche diodes (SPAD). Later, the light is filtered by using a 535 nm band pass filter for the donor channel and a 635 nm long-pass filter for the acceptor channel. Finally, the photons are detected by SPADs and a time-correlated single-photon counting module (TCSPC) is used to record the arrival time of each photon. Additionally, for smFRET and TCCD measurements, a pulsed interleaved excitation (PIE) scheme is applied, in which excitation of the donor and acceptor is alternated (See below for details).

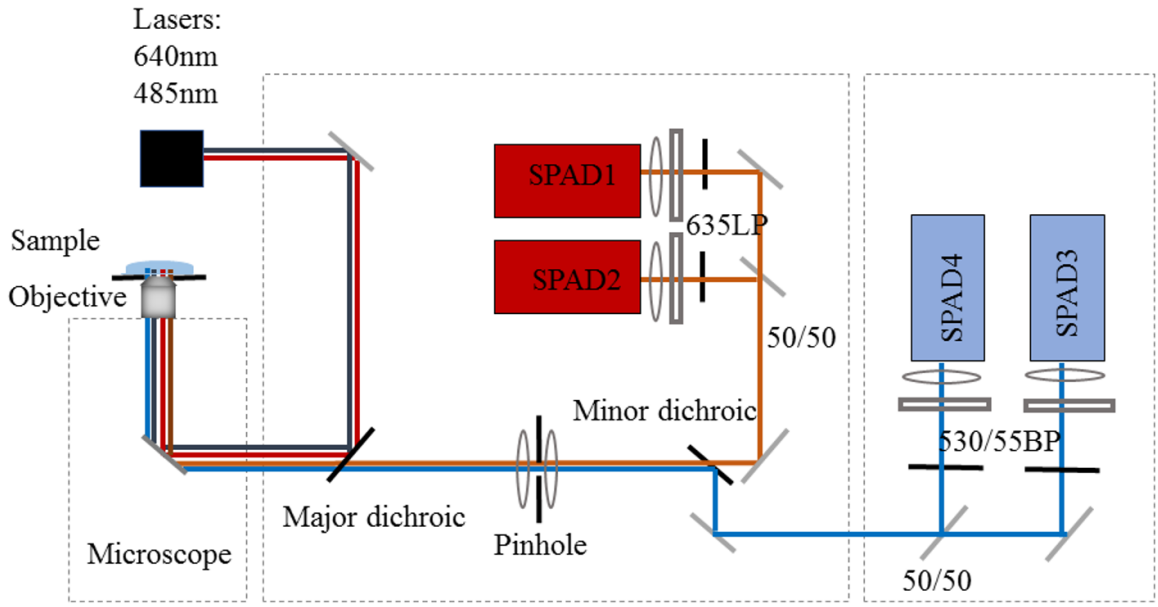


Figure 15. Set up of the confocal microscope used in the present work. Both excitation and emission light paths are shown [113].

In the present study, data were recorded by using a dedicated computer and saved in a time-tagged time-resolved (TTTR)-mode. This specific format contains three types of information: the specific SPAD that has detected the photon, the macro time (TM) which is the time interval between the start of the measurement and the detection of the photon and the micro time (tm) that is the time interval between the detection of the photon and the last laser pulse.

3.5.3.3 Fluorescence correlation spectroscopy (FCS) measurements

In the present study, FCS was used to determine the diffusion coefficient (D) and the concentration of the protein of interest. FCS takes advantages of fluorescence intensity fluctuations recorded over time. Indeed, molecules that are submitted to Brownian motion will randomly enter and leave the confocal volume (**Fig. 16A**). When the concentration of the sample is low enough (i.e. pM up to few nM with about 0.1-10 molecules inside the confocal

volume), significant fluorescence intensity fluctuations can be recorded on the respective time-trace of the experiment (**Fig. 16B**) [114]. From its autocorrelation function, both the diffusion time and the number of molecule can be estimated by fitting the real data with a specific model (**Fig. 16C**).

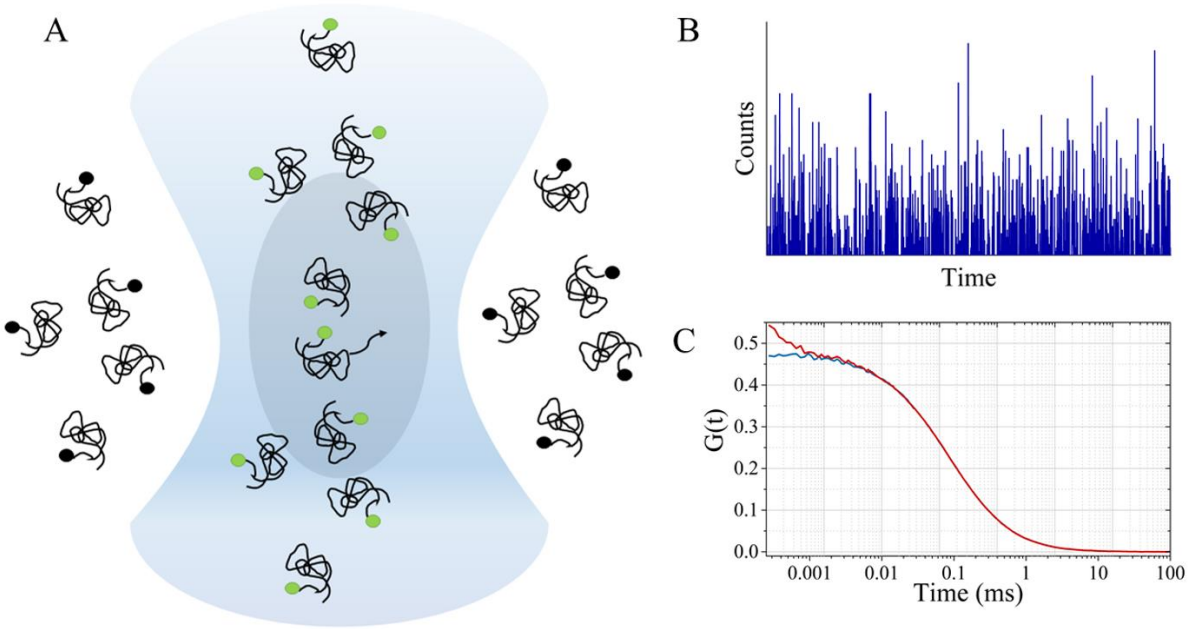


Figure 16. Typical time-trace and corresponding auto-correlation curves from FCS. The labeled molecules go through a defined illumination volume (A) and generate intensity fluctuations over time (B). From the time-trace auto-correlation curves are calculated (C).

In our lab, for removal of possible artifact, the auto-correlation curve is generated by cross-correlating the two auto-correlation curves obtained from two SPADs. To this aim, a 50:50 beam splitter is used to divide the signal into two equivalent beams (see 3.5.3.2 for details). The normalized auto-correlation function is calculated as follows [114]:

$$G(\Delta t) = \frac{< \delta I_i(t) \cdot \delta I_j(t + \Delta t) >}{< I_i\,(t) > \cdot < I_j\,(t) >} \qquad \text{(Eq. 5)}$$

with

$$\delta I_{i,j}(t) = I_{i,j}(t) - < I_{i,j}(t) > \qquad \text{(Eq. 6)}$$

$$< I_{i,j}(t) > = \frac{1}{T} \cdot \int_0^T I_{i,j}(t)\, dt \qquad \text{(Eq. 7)}$$

where $I_{i,j}\,(t)$ is the fluorescence intensity detected by SPADi and SPADj at the time t, $< I_{i,j}\,(t) >$ is the mean of the fluorescence intensity detected by SPADi and SPADj, $\delta I_{i,j}$ is the fluctuations in fluorescence intensity detected by SPADi and SPADj and Δt is the time lag.

Interestingly, the correlation function can also be written as a function of both the diffusion time and the concentration as seen in the following equation [114]:

$$G(\Delta t) = \frac{1}{V_{eff} \cdot < c >} \cdot \frac{1}{1 + \frac{\Delta t}{\tau_d}} \cdot \frac{1}{\sqrt{1 + \frac{\Delta t}{\tau_d \cdot \kappa^2}}} \quad \text{(Eq. 8)}$$

with

$$\kappa = \frac{z_0}{r_0} \quad \text{(Eq. 9)}$$

where V_{eff} , $< c >$, Δt, τ_d , κ are the effective detection volume, the average sample concentration, the time lag, the lateral diffusion time and the elongation of the detection volume, respectively. The terms r_0 and z_0 correspond to the axial and radial radii of the detection volume.

Additionally, D can be easily calculated from the diffusion time by using the following equation [114]:

$$\tau_d = \frac{r_0^2}{4D} \quad \text{(Eq. 10)}$$

In our lab, a calibration measurement is performed before each experiment. The aim of the calibration is to determine both V_{eff} and κ, a prerequisite for the determination of both the concentration and the diffusion coefficient of the protein of interest (see above). To this aim, a FCS measurement is performed with a dye having relevant spectral characteristic in respect to the fluorophore used in the study and for which the diffusion coefficient is known. That diffusion coefficient is usually obtained from the literature and is always corrected for the temperature dependency as follows [115]:

$$D(T) = \frac{D\,(T_0) \cdot T \cdot 8.9.10^{-4} \mathrm{Pa.s}}{T_0 \cdot \eta\,(T)} \quad \text{(Eq. 11)}$$

where $D\,(T_0)$, $\eta\,(T)$, T and T_0 are the coefficient of diffusion calculated at the temperature T_0, the viscosity at the current temperature T and the temperature in Kelvin, respectively.
Then, from the correlation function, both V_{eff} and κ can be estimated according to the known coefficient of diffusion by fitting the data with a specific model. Once V_{eff} and κ calculated, they are kept fixed for the further FCS measurements.

In the present work, FCS measurements of 10 min were made on calmodulin prepared by using CFPS in batch-mode. Later, the proteins were labeled by chemical modification as previously stated in 3.2.3 or by incorporation of fluorophores as described in 3.2.2. Before all measurements, the samples were incubated for at least 1 hour at 25 °C in the final buffer. In addition, FCS experiments were always preceded with calibration measurements performed for 5 min with either ATTO488-NHS ester or ATTO655-NHS ester. FCS data analyses were performed by using the SymPhoTime64 and OriginPro.

3.5.3.4 smFRET measurements

In the present work, single-molecule Förster resonance energy transfer (smFRET) experiments were used to investigate the structure and the functionality of selectively dually labeled calmodulins. FRET was performed on single molecules and used as a spectroscopic ruler to probe intramolecular distances distributions in calmodulin conformational (sub)states. FRET is a physical phenomenon based on dipole-dipole interactions and which relies on the distance-dependent radiationless transfer of energy between two fluorophores, a donor and an acceptor. The distance dependency of FRET is summarized in the following equation [116]:

$$E = \frac{R_0^6}{R_0^6 + R_{DA}^6} \quad \text{(Eq. 12)}$$

where E, R_0 and R_{DA} correspond to the FRET efficiency, the Förster radius and the distance separating the two fluorophores in Å, respectively. The range of distances that can be relevantly studied depends therefore on the Förster radius of the specific dye pair used. R_0 corresponds to the distance for which E is 0.5 and can be calculated as follows [116]:

$$R_0 = (8.79 \cdot 10^{-5} \cdot \kappa^2 \cdot n^{-4} \cdot \phi_{FD} \cdot J(\lambda))^{1/6} \quad \text{(Eq. 13)}$$

with

$$J(\lambda) = \int_0^\infty F_D(\lambda) \cdot \varepsilon_A(\lambda) \cdot \lambda^4 d\lambda \quad \text{(Eq. 14)}$$

where κ^2, n, ϕ_{FD} and J(λ) are the relative orientation of the dipoles, the refractive index of the solvent, the quantum yield of the donor and the spectral overlap of the donor emission and the acceptor absorption in M^{-1} cm^{-1} nm^4, respectively and F_D, ε_A are the normalized fluorescence intensity of the donor and the molar absorption coefficient of the acceptor, respectively. In the present study, the dyes were free to move around the anchoring position.

Therefore, it was considered that they are randomly oriented according to each other. For such situation, the value of κ^2 is estimated at 2/3 [46]. J(λ) is an important parameter, which determines the quality of a pair of fluorophores (**Fig. 17A**).

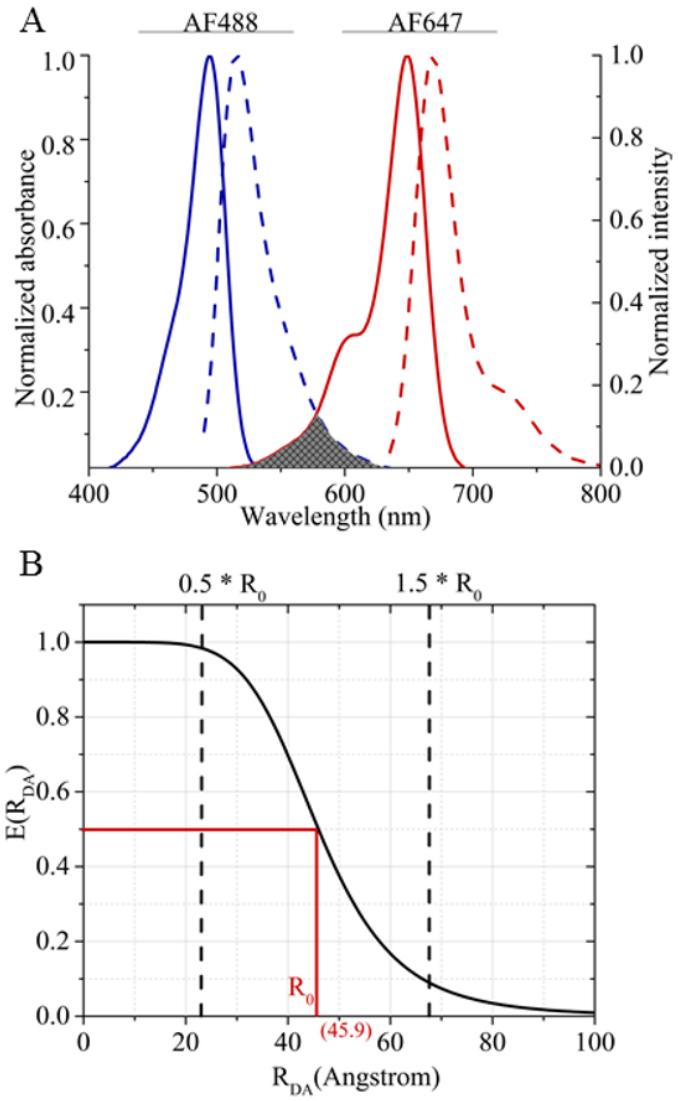


Figure 17. **Förster radius determination of AF488-AF647 pair.** (A) Absorption (solid lines) and emission (dashed lines) spectra of Alexa Fluor 488 and Alexa Fluor 647 bound to CaM. The gray area represents the overlap integral (λ). (B) The corresponding transfer efficiency (E) in function of the distance.

On one hand, it should be high enough to ensure efficient energy transfer and, on the other hand, the emission spectra of both fluorophores have still to be separated enough for preventing detection of the donor in the acceptor channel a phenomenon known as spectral cross-over. Due to the relation of E and R_{DA}, the sensitivity of FRET to small distance changes is the highest around the Förster radius, whereas it is not anymore sensitive at very

low or very high distances (**Fig. 17B**). For that reason, the pair is usually chosen depending on the proximity of R_0 with the specific molecular distance to be studied.

In the present work, the following dye pairs were used: AF488-AF647, AF488-ATTO633 and BODIPY-FL-ATTO633. The R_0 of the latter was determined in the present study whereas the Förster radii of the two others were previously determined in our group or taken from the literature.

In the present study, smFRET were performed on CaM produced by CFPS in batch-mode. The protein was selectively dually labeled either by chemical modification (see 3.2.3 for details) or by incorporation of fluorophores (see 3.2.2 for details) or by a combination of both. All samples were then incubated for at least 1 h at 25 °C in the final buffer. smFRET experiments made with apoCaM were done in apo-buffer whereas binding experiments with calcium ions were made in holo-buffer after having incubated overnight to reach equilibrium. Binding experiments made in presence of functional binding partners were done in holo-binding-buffer supplemented with a high excess of calcium/calmodulin-dependent protein kinase II (CaMKII) (150 μM) or skeletal muscle myosin light chain kinase (M13) (100 μM) binding peptide to be in saturated conditions (Kd from sub-picomolar to nanomolar). Addition of an equivalent volume of apo-binding-buffer was used to inhibit the calcium-dependent binding of CaMKII and M13 by chelating calcium ions. smFRET measurements were performed on 50 pM samples for at least 60 min using a pulsed-interleaved-excitation (PIE) scheme (**Fig.18**) [117]. This specific scheme was used to verify the presence of the acceptor by direct excitation and to minimize artifacts due to the presence of donor-only molecules. It was achieved by the sequential excitations of the donor and the acceptor. The time interval between the two pulses was 25 ns. The first pulse allowed to record fluorescent bursts coming from directly excited donor and FRET excited acceptor, whereas the second pulse excited directly the acceptor (**Fig. 18**). Therefore, events coming from donor-only molecules were sorted out preventing the generation of an associated zero peak in the FRET histogram. In the aim to check the quality of the sample, all experiments were preceded by FCS measurements. Later, smFRET measurements were made in presence of 0.001 % of Tween. Additionally, a trolox- cysteamine cocktail was used for photoprotection purposes [118].

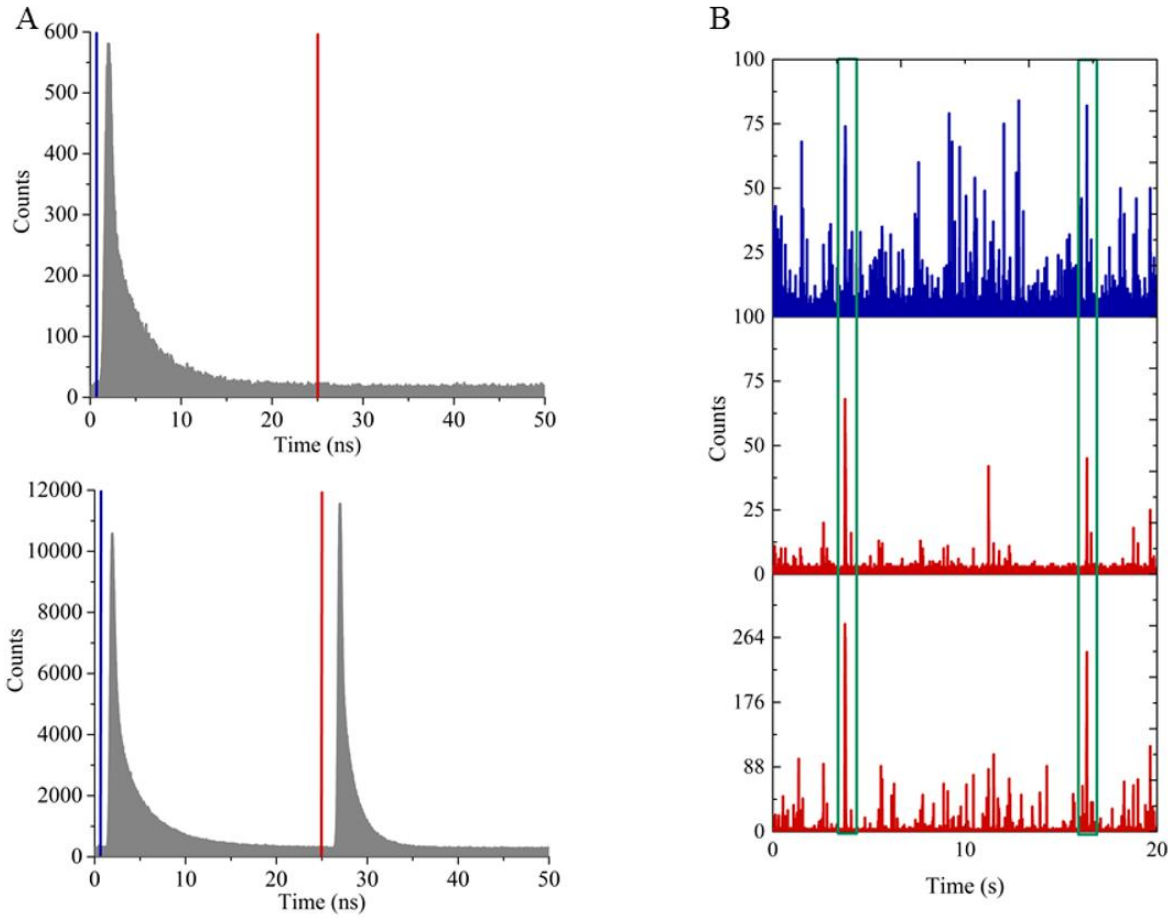


Figure 18. Time-correlated single-photon counting (TCSPC) histograms of fluorescence decays and time-traces of a typical sample. (A) TCSPC histogram of fluorescence decays of a donor-only population (top) compared to a dually labeled population (bottom). (B) Typical time-traces obtained from a dually labeled population from both donor and acceptor channel using PIE. The three time-traces that are obtained represent the donor emission (top), the acceptor emission after donor excitation (middle) and the acceptor emission after direct excitation (bottom). Green boxes represent bursts recorded from dually labeled molecules.

smFRET data analyses were done by using self-written Matlab routines, as previously described by Gabba *et al*. [119]. In brief, we used the inter-photon time-distance (IPD) to select events over the background [120]. The idea is to take advantage of the time-distance separating two subsequent photons which is much shorter for the transit of a molecule than for the background. For each experiment, IPD values were calculated as follows:

$$IPD_{i+1} = TM_{i+1} - TM_i \qquad \text{(Eq. 15)}$$

where the macro time TM of a photon i is subtracted from the macro time of the photon i+1. Later, the time-trace of IPD values was used to discriminate events from background by defining a specific threshold (**Fig. 19**).

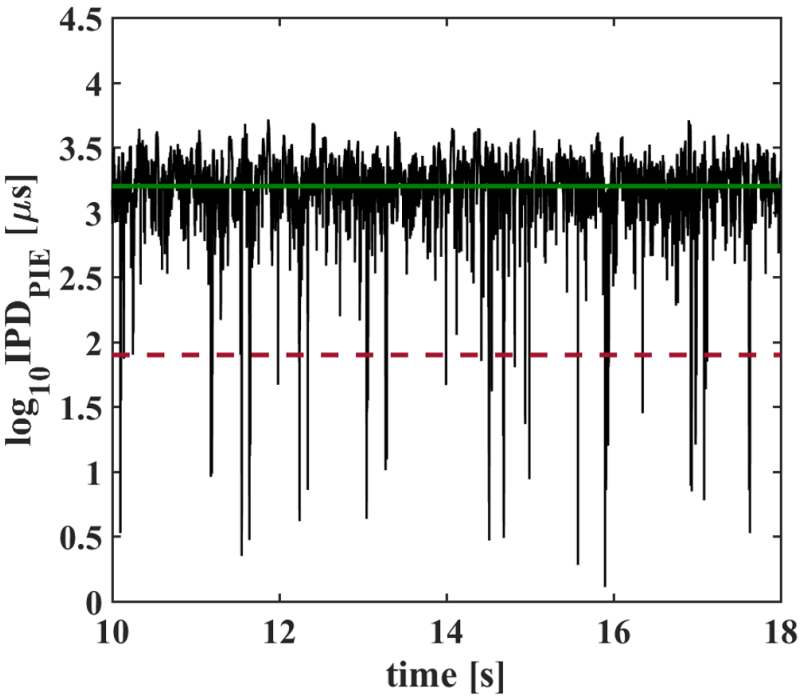


Figure 19. Inter-photon distance (IPD) time-trace. The average signals originated from the background (solid green line) and the threshold of selection (dashed red line) for relevant burst are both shown.

Subsequently, E was calculated for each FRET event by using the following equation:

$$E = \frac{I_A}{I_D + \gamma \cdot I_A} \quad \text{(Eq. 16)}$$

with

$$\gamma = \frac{\Phi_A}{\Phi_D} \cdot \frac{g_A}{g_D} \quad \text{(Eq. 17)}$$

where I_A and I_D are the intensities of the donor and the acceptor and γ is a correction factor for the difference in the detection efficiencies (g) and quantum yields (ϕ) of both fluorophores [121].

3.5.3.5 Quantum yield determination

In the present work, quantum yields (QYs) of the AF488, AF647, BODIPY-FL and ATTO633 bound to CaM were determined for the calculation of the correction factor γ that is included in the FRET efficiency calculation (Eq. 16) and for the Förster radius calculation (Eq. 13). The quantum yield of a fluorophore defines how many photons are emitted in respect to how many photons were previously absorbed. QY can be calculated as follow:

$$\phi = \frac{N_{em}}{N_{abs}} \quad \text{(Eq. 18)}$$

where N_{em} and N_{abs} correspond to the number of emitted and absorbed photons, respectively.

In the present work, QYs were determined by using a comparative method based on the molecular brightness (MB), previously described by Kempe *et al.* [122]. MB corresponds to the number of photons emitted per molecule per second [116]. Here, FCS measurements were performed to determine MB of the specific fluorophore as a function of the intensity [123]. The MB can be calculated from FCS by using the following equation [123]:

$$MB = \frac{< f >}{< N >} \quad \text{(Eq. 19)}$$

where <f> and <N> are the fluorescence count rate without any background contribution and the average number of molecule in the detection volume, respectively.

In short, the method is based on the comparison of the fluorophore of interest with a reference dye of known QY. The following equation is used to determine the unknown QY of a specific sample [122]:

$$\phi_S = \frac{g_{Ref} \cdot \varepsilon_{Ref}(\lambda) \cdot m_S}{g_S \cdot \varepsilon_S(\lambda) \cdot m_{Ref}} \cdot \phi_{Ref} \quad \text{(Eq. 20)}$$

where g, ε and m correspond to the optical detection efficiency, the absorption coefficient at the excitation wavelength (λ) and the slopes of the MB as a function of the excitation intensity, respectively. Here, $_S$ and $_{Ref}$ are used for the sample and the reference, respectively. In the present study, <f> was calculated from the fluorescence decay histogram and corrected for the background contribution. <N> was determined from the correlation function (see 3.5.3.3 for details). FCS measurements of 20 min were performed for both the fluorophore of interest and the reference at 5, 10, 15 and 20 arbitrary units (a.u) to ensure linearity of MB as a function of the intensity. FCS measurements of 2 min were performed with the respective buffer at similar intensities for background subtraction. AF488-NHS ester and AF647-NHS ester were used as reference for the donor and acceptor calculations, respectively. FCS data analyses were made by using SymphoTime64 and self-written Matlabs routines.

Additionally, absorption coefficients at the excitation wavelength were calculated by multiplying normalized absorption spectrum with the specific absorption coefficient of the dye. Measurements of the absorption spectra were performed by using an absorption

spectrophotometer (see 3.3 for details) for BODIPY-FL-tRNA, ATTO633-tRNA, AF488-NHS ester and AF647-NHS ester instead of the CaM-bound-dyes since the amount of labeled CaM obtained in CFPS is too low for obtaining reliable measurements. Finally, the optical detection efficiency was calculated by using the transmission function (see 3.5.3.7 for details).

3.5.3.6 Donor-acceptor distance determination and accessible volume simulations

3.5.3.6.1 Donor-acceptor distances based on alpha-carbon distances

The donor-acceptor distance (R_{DA}) is often approximate to the distance separating the α-carbons of the anchor positions of the dyes. In the present work, such approximations were performed for comparing the calculated R_{DA}s with those determined by using accessible volume simulations (see 3.5.3.6.2). The three-dimensional structures of CaM were extracted from the PDB and R_{DA}s were determined by using the Visual molecular dynamics software (VMD). The expected FRET efficiencies were then calculated by using Eq.12.

3.5.3.6.2 Donor-acceptor distances based on accessible volumes

In the present work, R_{DA} determinations were based on the accessible volumes (AVs) of each fluorophore. AVs were simulated as previously described by Höfig *et al.* [124]. To this aim, a WxDevC++ algorithm was used for the generation of an AV cloud for each fluorophore bound to the protein. The cloud is constituted of all the spatial points which are sterically accessible to the dye. It takes into account the shape of the dye, the length of the linker and the fact that the dye is free to move around the anchoring position. Then, by averaging all the possible positions for both dyes, mean weighted and unweighted donor-acceptor distances are calculated. The R_{DA}s were determined for different three-dimensional structures extracted from the PDB. According to the mutant used, point mutations were inserted into them at the relevant positions by using PyMol. The three-dimensional structures, as well as the accessible volumes were visualized in PyMol. Finally, the expected FRET efficiencies were calculated by using Eq.12.

3.5.3.7 Detection efficiency determination

The detection efficiency of the MicroTime200 confocal microscope has to be estimated since it is required for quantum yield and FRET efficiency calculations (Eq. 16 and Eq. 20). The detection efficiency g comprises two components as shown in the following equation:

$$g = g_{optical} \cdot g_{electrical} \quad \text{(Eq. 21)}$$

For the optical efficiency $g_{optical}$, properties of the major dichroic mirror, the minor dichroic mirror and the filters have to be considered for each measurement condition. To determine the optical efficiency, transmission functions of the components of the detection pathway have to be determined both in blue and red (**Fig. 20**). The optical efficiency is then calculated by using the convolution of the transmission efficiency with the normalized emission spectrum.

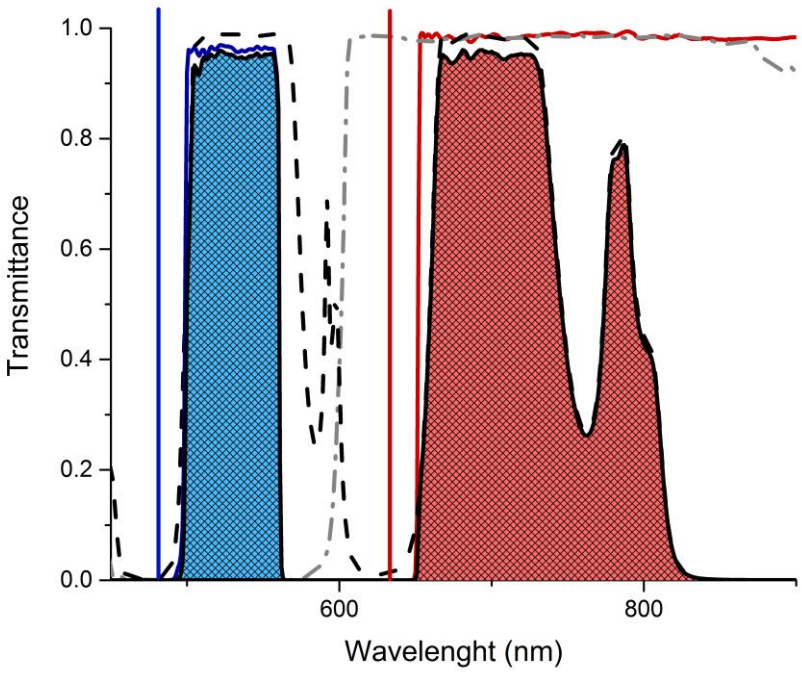


Figure 20. **Transmission functions of the optics.** Blue and red emission filters (solid blue and red lines), minor dichroic (dashed light gray line), and major dichroic (dashed black line) are shown. The blue area represents the transmission function in blue and the red area the transmission function in red. Lasers pulses in both colors are also included.

The electrical efficiency $g_{electrical}$ takes into account, among others, the electrical processing and wavelength dependency of the detection efficiency of the SPAD. To estimate its contribution, measurements are made with two populations of dually labeled DNA for which the values that should be obtained after data analysis are already known. Here, its contribution to the total detection efficiency was negligible and therefore the detection efficiency was approximated to the $g_{optical}$. Detection efficiency calculations were made for BODIPY-FL-CaM, ATTO633-CaM, AF488-CaM and AF647-CaM.

3.5.3.8 TCCD measurements

Single-molecule two-color coincidence detection (TCCD) measurements were performed on 10 pM samples in similar conditions previously stated in 3.5.3.4. TCCD analyses were done by using self-written Matlab routines in a similar way to Li *et al.* [125].

4. Results

4.1 Cell-free synthesis of calmodulin and incorporation of UAAs

4.1.1 Screening of the literature for calmodulin mutants

In the present work, a case-study was performed on a model protein to prove the biological relevance of newly designed methodological approaches for selective dual-labeling in smFRET. Here, we have used the human calmodulin (hCaM) for multiple reasons (see Introduction for details). Specific double and single mutant calmodulins were necessary for selective dual-labeling of CaM. Therefore, the wild-type sequence of CaM (**Fig, 21**) has to be modified for the incorporation of specific functional groups and fluorophores via cysteine and amber stop codons.

```
MADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQ      50
DMINEVDADGNGTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGY     100
ISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQMMTAK      149
```

Figure 21. Amino acid sequence of human calmodulin. The primary sequence of CaM encoded by the human CALM1 gene contains 148 amino acids in addition to the initiator methionine.

In this aim, literature was screened and double mutant calmodulins suitable for labeling were found and listed. The list was restricted to double mutants that were used for probing conformational changes that are similar to those we wanted to detect. It was found that the specific functional mutant CaM T34C T110C was the most widely reported in the literature (**Table 2**) [35, 126-129]. Here, for the validation of our methodological approaches, we have chosen to keep the same two mutated positions for labeling since they were previously selected by others for smFRET studies performed with non-selectively labeled CaM [35, 78].

Table 2. Non-exhaustive list of studies using CaM T34C T110C.

Study	Dye1	Dye2	REF
Slaugther *et al*. 2007	Alexa Fluor 488	Alexa Fluor 488	[126]
Allen *et al.* 2003	Alexa 488	Texas red	[35]
Torok *et al.* 2001	AEDANS	DDP	[127]
Putkey *et al.* 2007	IAEDANS	DDPM	[128]
Xiong *et al.* 2005	IAEDANS	DDPM	[129]

Mutations at position 34 (P34) and 110 (P110) were relevant for the present work for various reasons. Firstly, P34 and P110 are located in the N and C termini, respectively. Secondly, the various three-dimensional structures of CaM which were deposited in the protein data bank (PDB) range molecular distances going from less than 15 Å to more than 50 Å for the mentioned positions. Finally, the positions 34 and 110 were previously used in smFRET studies for detecting conformational changes in CaM [35] and therefore allow direct comparison of the labeling procedures [78].

4.1.2 *In vivo* expression of human calmodulin in *E.coli*

In the present work, our objective was to selectively label CaM at the position 34 and 110. For that purpose, *CaM T34C T110amber* and *CaM T34amber T110C* constructs were used instead of the classical double mutant *CaM T34C T110C*. Firstly, single mutants *CaM T34C* and *CaM T110C* were generated by site-directed mutagenesis. Both mutants were subsequently used for generating *CaM T34C T110amber* and *CaM T34amber T110C*. Additionally, all the constructs were fused to a His_6-tag at the C terminus (**Fig. 22**) for purification purposes.

Figure 22. His_6-tagged single mutants CaM T34C and CaM T110C. The respective point mutations are marked and shown in cyan.

In the aim to verify the two original constructs, expressions of *CaM T34C* and *CaM T110C* were performed in BL21 (DE3) competent *E. coli* cells. As shown in **Fig. 23**, the expected His-tag protein was eluted in fractions 1-5 (**Fig. 23 and Supplementary Fig. 2**). The apparent size (i.e. between 10 and 15 KDa) is consistent with the Ca^{2+}-loaded form of CaM which migrates significantly faster than its free form (i.e. about 17 KDa) [130]. In addition, His-tag detection by western blot has revealed the presence of tagged proteins in the elution fractions 1, 2 and 3. All in all, our results show that His_6-tagged hCaM carrying mutation either at P34 or P110 is properly expressed in BL21 (DE3) competent *E. coli* cells.

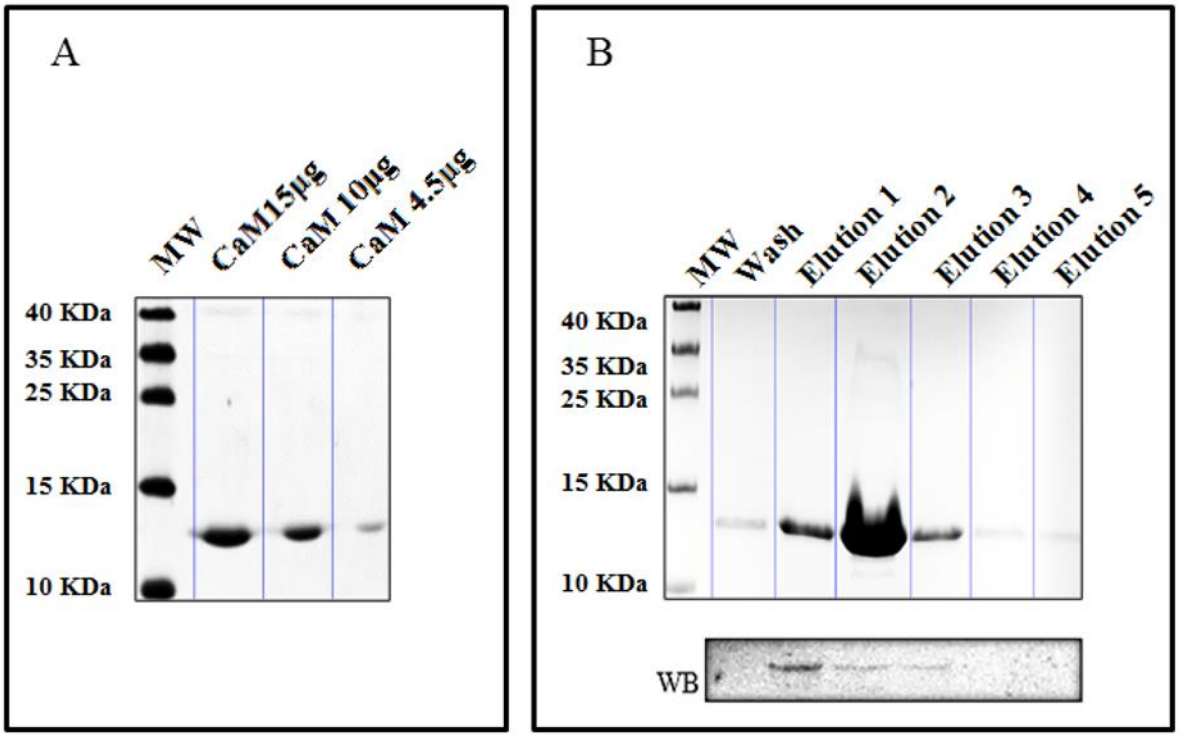


Figure 23. Overexpression of CaM T34C in BL21 *E.coli* cells. (A) SDS PAGE of different amounts of commercial hCaM. (B) SDS PAGE of different fractions collected during the purification of CaM T34C overexpressed into BL21 cells. Western blots (WB) immunodetection of His_6-tagged CaM are displayed.

4.1.3 Cell-free synthesis and incorporation of UAAs into calmodulin

In the present study, our objective was to selectively label cell-free synthesized proteins by taking the advantage of UAA incorporation. UAAs were incorporated via sense and non-sense codons. It was limited for the former category to cysteine codons that present a relatively low frequency and for the latter to amber stop codons, which have been successfully used by others. The incorporation of UAAs was performed thanks to a modified coupled transcription-translation cell-free system, which was either enriched or depleted in some specific components (see Materials and Method for details). *CaM T34C*, *CaM T110C*, *CaM T34C T110amber* and *CaM T34amber T110C* were used for the incorporation of UAAs into calmodulin.

Currently, two categories of cell-free systems are available. In the first one, the cell-free system is fully reconstituted from recombinants components of *E.coli*. In the second one, the *in vitro* system is produced from either eukaryotic or prokaryotic cell extracts. Here, we have used coupled transcription-translation cell-free systems originated from E.*coli* extracts for three reasons: (i) studies have suggested that they are more efficient [131, 132], (ii) the

folding of CaM required physiological concentration of Ca^{2+} [133-135] and (iii) modified extract-based systems are available for optimal incorporation of UAAs.

Firstly, to verify the proper synthesis of CaM in a cell-free system, *CaM wt*, *CaM T110C* and *CaM T34C* were expressed in *vitro* by using a non-modified coupled transcription-translation cell-free system (**Fig. 24A**). As shown in **Fig. 24**, products of expected sizes were purified for *CaM wt*, *CaM T110C* and *CaM T34C* constructs. Similar experiments were not performed with *CaM T34C T110amber* and *CaM T34amber T110C* since the presence of the amber stop codon will lead to the termination of translation before full-length synthesis of the His_6-tagged protein. Our results strongly suggest that expression of human CaM *in vitro* does not encounter major problems.

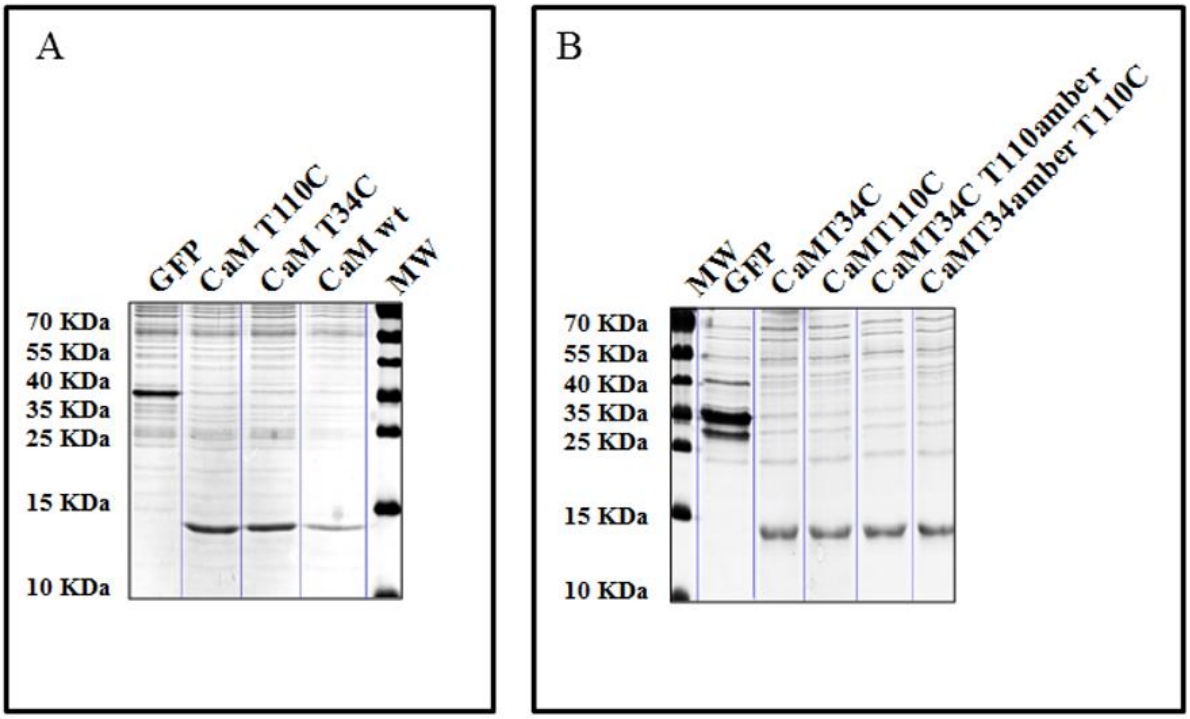


Figure 24. Cell-free synthesis and incorporation of UAAs into CaM. (A) synthesis of CaM by using a coupled transcription-translation cell-free system in batch-mode. (B) incorporation of UAAs into CaM by using a RF1-depleted coupled transcription-translation cell-free system in continuous-exchange. The 26.9 KDa Emerald green fluorescent protein (Emerald GFP) was used as a control.

To characterize further the product, the specificities of the near UV absorption spectrum of calmodulin were used. The near UV absorption spectrum of a protein is mainly shaped by its specific content in aromatic amino acids (AAs) (tryptophan (W), tyrosine (T) and phenylalanine (F)). Therefore, in theory, the near UV absorption spectrum can be used to identify or at least confirm the identity of a purified product. In facts, UV spectra-based

methods for the identification of recombinant proteins were previously reported [100]. However, in most cases the discrimination power is low due to the huge difference in the near UV absorption properties of the three aromatic AAs. Indeed, tryptophan residues absorb 5 and 50 times more than the tyrosine and phenylalanine residues, respectively. Since often a protein contains at least one tryptophan, it means that its near UV absorption spectrum is pretty similar to those of others. Nevertheless, CaM is particular in that way that its primary sequence does not contain any tryptophan residue making its near UV absorption spectrum highly specific (**Table 3**).

Table 3. Aromatic amino acid content of hCAM, Emerald GFP and yPGK

	CaM	Emerald GFP	yPGK
Phenylalanine (F)	8	13	19
Tryptophan (W)	0	1	2
Tyrosine (Y)	2	11	7

Actually, the spectrum is only shaped by 2 tyrosine and 8 phenylalanine residues. As shown in **Fig. 25**, the eluted product displays the expected specific near UV spectrum of calmodulin indicating that we have properly expressed our protein of interest. All in all, our results show that *in vitro* synthesis of CaM was successful and did not encounter major obstacles.

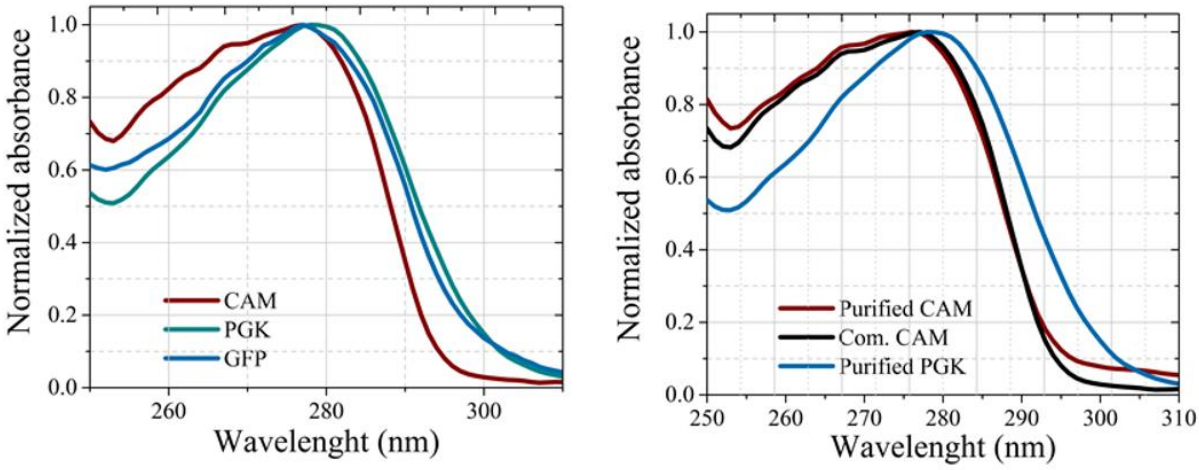


Figure 25. Identification of cell-free synthesized calmodulin (CaM). Comparison of the near UV absorption spectra of purified calmodulin, commercial calmodulin, purified yPGK and purified Emerald GFP. The specific near UV spectra of CaM is shown in comparison to those of the Emerald green fluorescent protein (Emerald GFP), the phosphoglycerate kinase (yPGK) (left). The near UV spectrum was compared to the one of commercial calmodulin (right).

Secondly, to investigate the incorporation into calmodulin of a specific UAA, the azido-phenylalanine (AzF), *CaM T34C T110amber* and *CaM T34amber T110C* were expressed *in vitro* by using a modified RF1- depleted coupled transcription-translation cell-free system in either batch-mode or continuous-exchange (see Materials and Method for details) (**Fig, 24B** and **Supplementary Fig. 3**). Full length CaM T34C T110AzF and CaM T34AzF T110C were successfully eluted from IMAC resins. Our results indicate that recognition by suppressor tRNAs and incorporation of UAAs are efficient in the tested modified cell-free system.

Finally, in the aim to investigate the efficiencies of the cell-free systems used in the present study, we have quantified the synthesis yield. Since the yield is pretty low in CFPS, we have used the Emerald green fluorescent protein (Emerald GFP), a 26.9KDa polypeptide of 238 amino acids, as a reference and taken advantage of its fluorescent properties. From fluorescence spectroscopy measurements, it appears that the yields of proteins is consistent with those reported in the literature with about few hundreds of µg per ml in batch-mode and around few mg per ml in continuous-exchange [132] (**Table 4**).

Table 4. Fluorescence intensity and equivalent concentration of cell-free synthesized Emerald GFP

Coupled transcription-translation cell-free system	**Wild-type**	**ΔRF1 in batch-mode**	**ΔRF1 in continuous-exchange**
Concentration nM	3457	9405	52156
Concentration µg/ml	92.9	253	1403

For CaM, the quantification by UV absorption at 280 nm was only possible in continuous-exchange due to the low yield, the low coefficient of absorption of CaM at 280 nm and imidazole contamination of the samples. The results with and without UAA incorporation were similar to those obtained with Emerald GFP with about 1 mg per ml (**Supplementary Table 1).** This is consistent with the fact that efficient incorporation of AzF is reported in the literature [27]. Altogether, our results indicate that both cell-free synthesis of CaM and incorporation of AzF at either P34 or P110 are efficient.

4.2 Selective dual-labeling by using a bioorthogonal chemistry

Here, we have investigated the possibility to use a bioorthogonal chemistry for the selective dual-labeling of cell-free synthesized proteins for smFRET. A similar chemistry was reported for selective dual-labeling of *in vivo* expressed proteins for smFRET studies [31]. However, it is the first time, to our knowledge, that such labeling with photostable dyes suitable for smFRET experiments is combined to CFPS.

4.2.1 Incorporation of UAAs and labeling of CaM by post-translational chemical modifications

In the present work, our objective was to selectively dually label cell-free synthesized CaM for smFRET studies. One of the two methodological approaches tested was based on the combined site-specific incorporation of AzF and cysteine residues into CaM for subsequent selective dual-labeling by using a bioorthogonal chemistry via maleimide-cysteine and azide-alkyne ligations [94]. To demonstrate the relevance of this approach, AzF and cysteine residues were site-specifically incorporated into CaM by using the constructs *CaM T34C T110amber* and *CaM T34amber T110C.* AzF residues were incorporated at the amber stop codon, whereas cysteine residues were incorporated at their respective codon. The incorporations of both AzF and cysteine residues into CaM were done by using a modified RF1-depleted coupled transcription-translation cell-free system (see Materials and Method for details). Firstly, CFPS was performed with either *CaM T34C T110amber* or *CaM T34amber T110C* for the production of CaM T34C T110AzF and CaM T34AzF T110C, respectively (**Fig. 26A**). In the order to selectively label CaM, maleimide-functionalized Alexa Fluor 488 (AF488) and alkyne-functionalized Alexa Fluor 647 (AF647) dyes were used to react with the incorporated azide and sulfhydryl functional groups, respectively (**Fig. 26B**). Later, the specificity of labeling was checked by gel imaging through the detection of fluorescence signals originated from the donor and the acceptor (**Fig. 26C**). As shown in **Fig. 26**, after dual-labeling, two fluorescent signals corresponding to the double mutant CaM T34AzF T110C were detected for AF488 and AF647 whereas CaM wt displayed no fluorescence at all. Furthermore, as expected, a single fluorescent signal corresponding to either AF647 or AF488 was detected for similarly treated CaM T34AzF and CaM T110C, respectively demonstrating the absence of cross-reactivity.

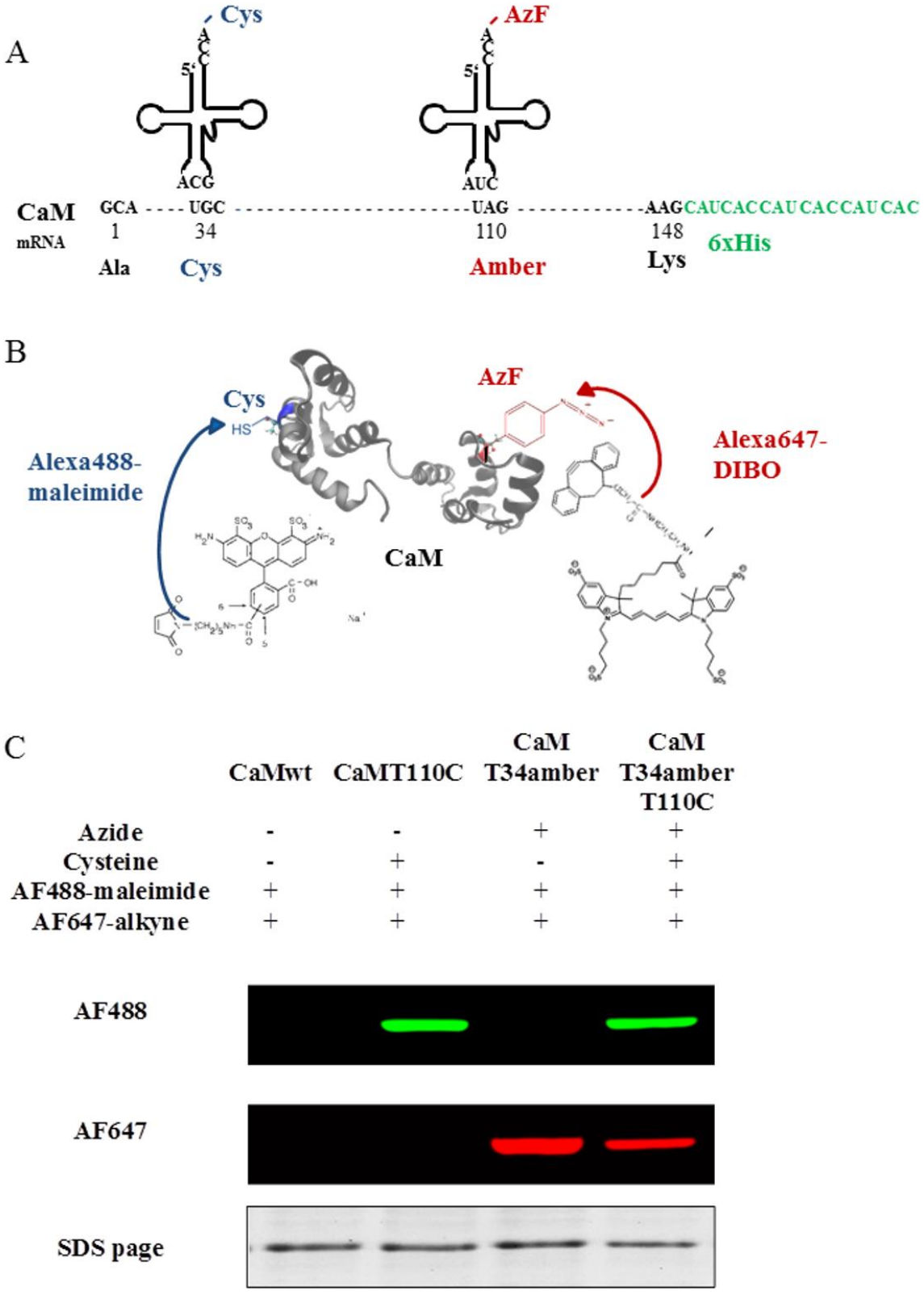


Figure 26. Selective dual-labeling of CaM by using a bioorthogonal chemistry. (A) Site-specific incorporation of cysteine and AzF residues into CaM. (B) Selective dual-labeling by using maleimide-cysteine and azide-alkyne reactions. (C) Fluorescent signals from differently labeled CaM on SDS PAGE.

Additionally, the presence of both dyes on the same molecule was investigated by using ensemble FRET measurements. After donor excitation, the presence of CaM T34AzF^{AF647} T110C^{AF488} was revealed by emission of both AF488 and AF647 while a mixed population constituted of single labeled CaM T34 AzF^{AF647} and CaM T110C^{AF488} displayed emission light from the donor only (**Fig. 27**, top). This strongly indicates that emission light recorded from AF647 was generated by FRET. All in all, our results show proper dual-labeling of the double mutant calmodulin.

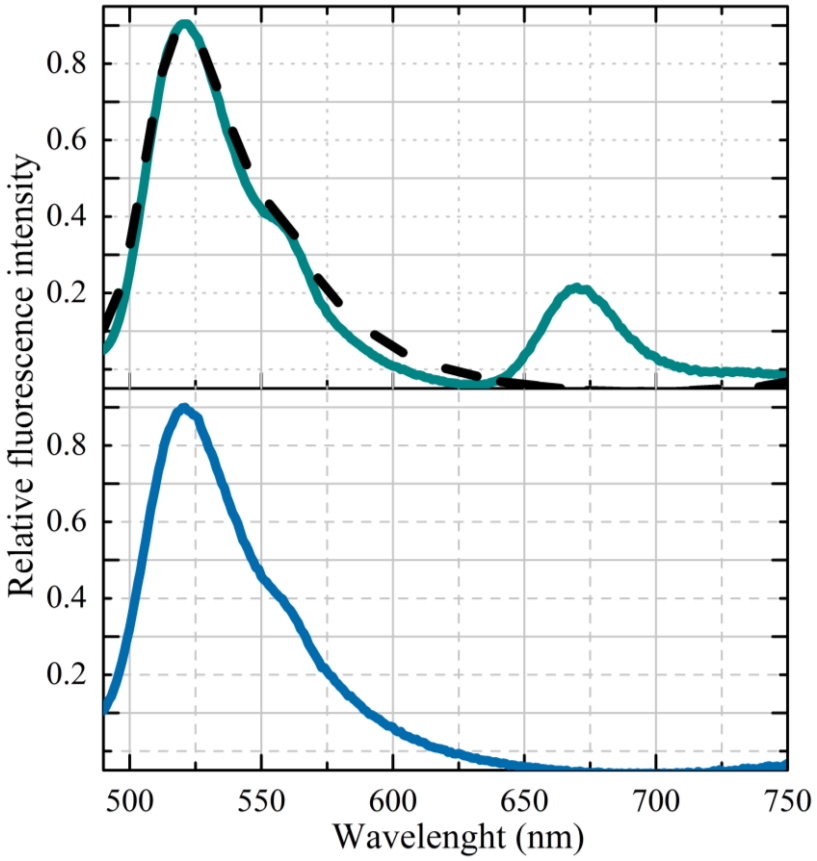


Figure 27. Dual-labeling of CaM by using a bioorthogonal chemistry. The emission spectrum of dually labeled CaM T34AzF^{AF647} T110C^{AF488} after excitation of the donor (top, solid green line) is shown in respect to the spectra of a mixed population constituted of single labeled CaM T34 AzF^{AF647} and CaM T110C^{AF488} (top, dashed black line) and of CaM T110C^{AF488} (bottom, blue solid line).

Finally, in the aim to quantify the different fluorescent populations in presence, single molecule two-color coincidence detection (TCCD) measurements were performed (see Materials and Method for details). Here, the percentages of molecules labeled with the acceptor (P_A), the donor (P_D) and both dyes (P_{A+D}) were determined (**Table 5**).

Table 5. TCCD measurements made on both constructs

CaM T34AzF^{AF647} T110C^{AF488}		
P_A (%)	P_D (%)	P_{A+D} (%)
22.0 (+/- 3.7)	49.2 (+/- 5.5)	28.8 (+/- 2.5)
CaM T34C^{AF488} T110AzF^{AF647}		
P_A (%)	P_D (%)	P_{A+D} (%)
29.4 (+/- 5.2)	55.3 (+/- 5.3)	15.3 (+/- 3.0)

As shown in **Table 5,** optimization of the procedure has led to 30 % and 15 % of dually labeled individuals for the constructs *CAM T34amber T110C* and *CAM T34C T110amber* (**Table 5**, column 3**)**, respectively. Altogether, our results have shown the successful selective dual-labeling of cell-free synthesized CaM with a degree of labeling (DOL) that is largely sufficient for further smFRET investigations.

4.2.2 Characterization of selectively dually labeled CaM

Our objective was to assess the biological relevance of the present methodological approach. To this aim, we have investigated the impact of the point mutations, the UAA incorporations and the labeling procedure on the structure and the functionality of CaM.

4.2.2.1 Three-dimensional structure of double mutant calmodulins

To analyze the secondary structure elements of CaM T34C T110AzF and CaM T34AzF T110C, circular dichroism (CD) measurements were performed [104]. Our aim was to study the effect on the three-dimensional structure of both mutations inserted at positions 34 and 110 and the incorporation of cysteine and AzF residues at the cited positions.

Here, circular dichroism was used for estimating the content in secondary structure of CaM wt, CaM T34C T110AzF and CaM T34AzF T110C. Such investigations were done in a concentration-independent manner in a similar way to what was already reported by using secondary structure databases that have served for the fitting of the measured curves (see Materials and Method for details) [136-138]. Here, the fitting was performed by using a specific software which aims to estimate the content in secondary structures thanks to two reference databases. As shown in **Fig. 28**, similar contents were found for CaM wt, CaM T34C T110AzF and CaM T34AzF T110C with less than 10 % of variability in the α-helical content.

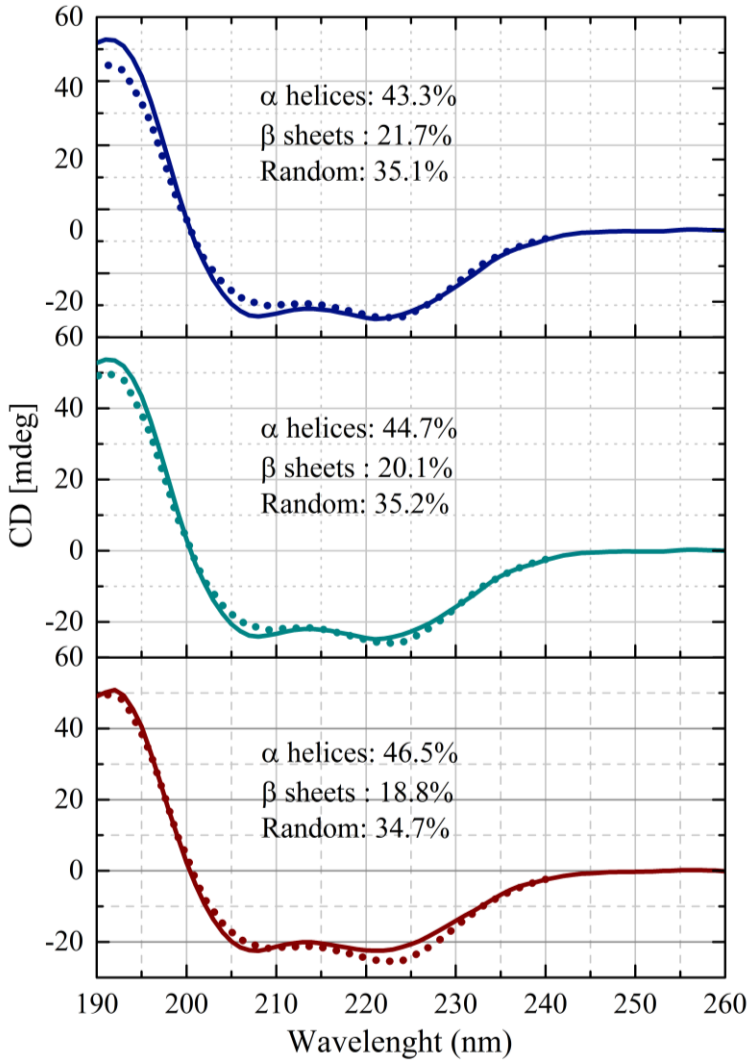


Figure 28. Structure of mutant calmodulins. CD spectra of CaM wt (top, blue solid line), CaM T34AzF T110C (middle, green solid line) and CaM T34C T110AzF (bottom, red solid line) measured in 25 mM phosphate buffer (pH 7.5). Fits to the experimental data are represented by dotted lines.

Our results were obtained for an heterogeneous population constituted of both free and Ca^{2+}-loaded-calmodulin and are consistent with those previously reported by Gao *et al.* for the apoCaM and holoCaM (**Table 6**) [73].

Table 6. Content in secondary structures of various conformations of CaM [73].

CaM	α-Helical (%)	β-Sheet (%)	Random (%)
Without Ca^{2+}	37.3	27.1	35.4
With Ca^{2+}	48.8	22.4	28.8

This clearly indicates that the three-dimensional structures of cell-free synthesized CaM T34C T110AzF and CaM T34AzF T110C are not significantly affected even considering that CD spectroscopy is a low-resolution technique. Thus, we can confidently state that the insertion of the specific mutations at P34 and P110 as well as the incorporation of a UAA into cell-free synthesized CaM do not disturb significantly its three-dimensional structure.

4.2.2.2 Diffusion properties of selectively dually labeled CaM

To verify the proper behavior of cell-free synthesized calmodulins in solution, fluorescence correlation spectroscopy (FCS) was used. Auto-correlation curves were analyzed and the diffusion coefficient was determined (see Materials and Method for details). Then, values from a labeled reference commercial calmodulin were compared to those of selectively dually labeled cell-free synthesized CaM.

The coefficient of diffusion measured for a non-specifically labeled commercial calmodulin was 91.66 ± 1.09 μm²/s. This result is consistent with what was previously determined for that specific protein in our lab (Lamprou, unpublished data) and with what is expected for a small protein. As expected**,** the diffusion coefficients measured for CaM T34C^{AF488} T110Az^{AF647} were similar to the one of the reference. Indeed, the values obtained from CaM-bound AF488 and CaM-bound AF647 display a pretty low deviation in respect to the value determined for the non-specifically labeled commercial calmodulin with 85.17 ± 4.40 μm²/s and 90.33 ± 4.37 μm²/s, respectively. Therefore, our results indicate that cell-free synthesis of CaM, as well as incorporation of UAA into it and subsequent selective dual-labeling does not affect the diffusion properties of the protein.

4.2.2.3 Structure and functionality of selectively dually labeled CaM

4.2.2.3.1 Accessible volume measurements of CaM-bound fluorophores

In the present work, selectively dually labeled calmodulins were used for smFRET. For proper data analysis, the expected FRET efficiency (E) has to be precisely determined for all the situations that could be encountered. To this aim, diverse three-dimensional structures of CaM were extracted from the PDB. Later, the molecular distance separating the donor from the acceptor (R_{DA}) was calculated for each conformation. This was done for apoCaM (ID:1DMO) [80], holoCaM (IDs: 1PRW, 1CLL, 4HEX and 1X02) [81], as well as for the two complexes Ca^{2+}-CaM-CaMKII (ID:1CDM) [139] and Ca^{2+}-CaM-M13 (ID:2BBM) [140].

R_{DA}s were firstly calculated based on the distance separating the α carbons (Cα) of the amino acid anchors (**Table 7**, Column 3 **Fig. 29A**).

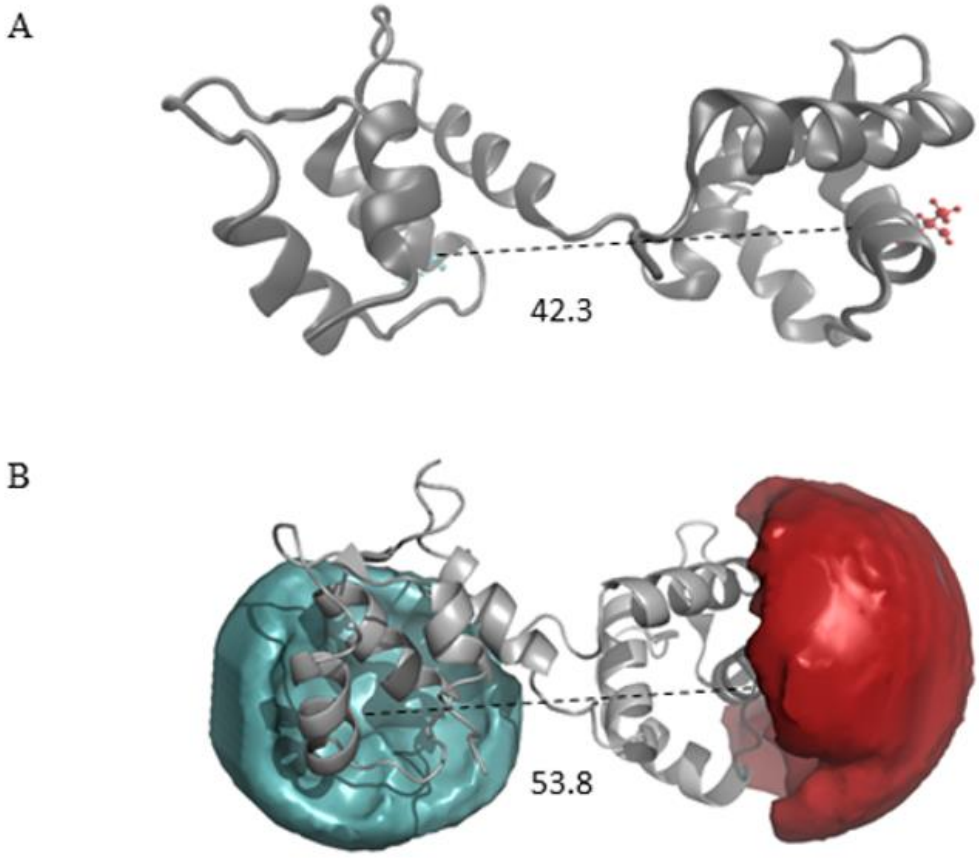


Figure 29. R_{DA} calculation. (A) The R_{DA} calculated from the Cα distance separating P34 (red) and P110 (blue) in apoCaM. (B) The R_{DA} calculated from accessible volumes of the donor and acceptor fluorophores (blue and red, respectively) of apoCaM T34Az^{AF647} T110C ^{AF488}.

However, such a way to calculate R_{DA} was reported to be very inaccurate [141]. Indeed, a significant error is introduced when the location of the dye is approximated to the unique position of the Cα of the amino acid anchor. Actually, the length of the linker used for anchoring, the shape of the fluorophore and the fact that it can freely move around the Cα position are not taken into account. R_{DA} can be estimated in another way that takes into account all these parameters by using accessible volume (AV) simulations. There, a specific AV is calculated for each dye, which represents all the sterically accessible positions that can be sampled by the freely moving dye [124, 141] (**Fig. 29B**). R_{DA} values were then re-calculated based on AVs for both CaM T34AzF^{AF647} T110C^{AF488} and CaM T34C^{AF488} T110AzF^{AF647}. As shown in **Table 7** and **Supplementary Table 2**, the calculated distance is

significantly longer compared to those previously obtained (**Table 7**, columns 3 and 5 and **Supplementary Table 2**). Finally, the expected FRET efficiencies (see 3.5.3.4, Eq. 12) were calculated by using the specific Förster radius of the dye pair AF488/AF647 (53.4 Å) [142] (see Materials and Method for details).

Table 7. Expected FRET efficiencies calculated for CaM T34AzF^{AF647} T110C^{AF488}.

		calculated from Cα		calculated from AV	
PDB ID	**Description**	**R_{DA} (Å)**	**E**	**R_{DA} (Å)**	**E**
1DMO	ApoCaM	42.3	0.80	53.8	0.49
1CLL	HoloCaM	52.4	0.53	64.4	0.25
1PRW	HoloCaM	16.3	1.00	23.5	1.00
4HEX	HoloCaM	49.6	0.61	59.9	0.34
1X02	HoloCaM	40.4	0.84	52.6	0.53
1CDM	Ca^{2+}-CaM-CaMKII	14.8	1.00	20.5	1.00
2BBM	Ca^{2+}-CaM-M13	15.4	1.00	23.2	0.99

As shown in **Table 7**, R_{DA} values go from about 20 Å to more than 60 Å. As a consequence, a very broad range of E is expected according to the different conformations of CaM, which goes from around 0.25 to nearly 1. In the present study, the dye pair AF488/AF647 was used to verify the integrity of the structure of apoCaM, as well as for detecting with high sensitivity the conformational changes that CaM undergoes after ligand and partner binding.

4.2.2.3.2 Structure of selectively dually labeled CaM

To verify the structural integrity of CaM T34AzF^{AF647} T110C^{AF488} and CaM T34C^{AF488} T110AzF^{AF647}, smFRET measurements were performed. All experiments were done in the presence of 10 mM EGTA, a calcium chelating agent. In the presence of EGTA CaM is in its apo-form. For CaM T34AzF^{AF647} T110C^{AF488}, it means that the donor and acceptor fluorophores are separated by 53.8 Å (**Table 7**). As displayed in **Fig. 30**, a well-defined single FRET population was observed around E = 0.5 (**Fig. 30**, bottom). This is consistent with R_{DA} calculations (**Table 7**, columns 5 and 6) based on the structure reported by Zang *et al*. [80] and with the fluorescence lifetime obtained for the donor in presence of the acceptor τ_{DA} (**Fig. 30,** top left).

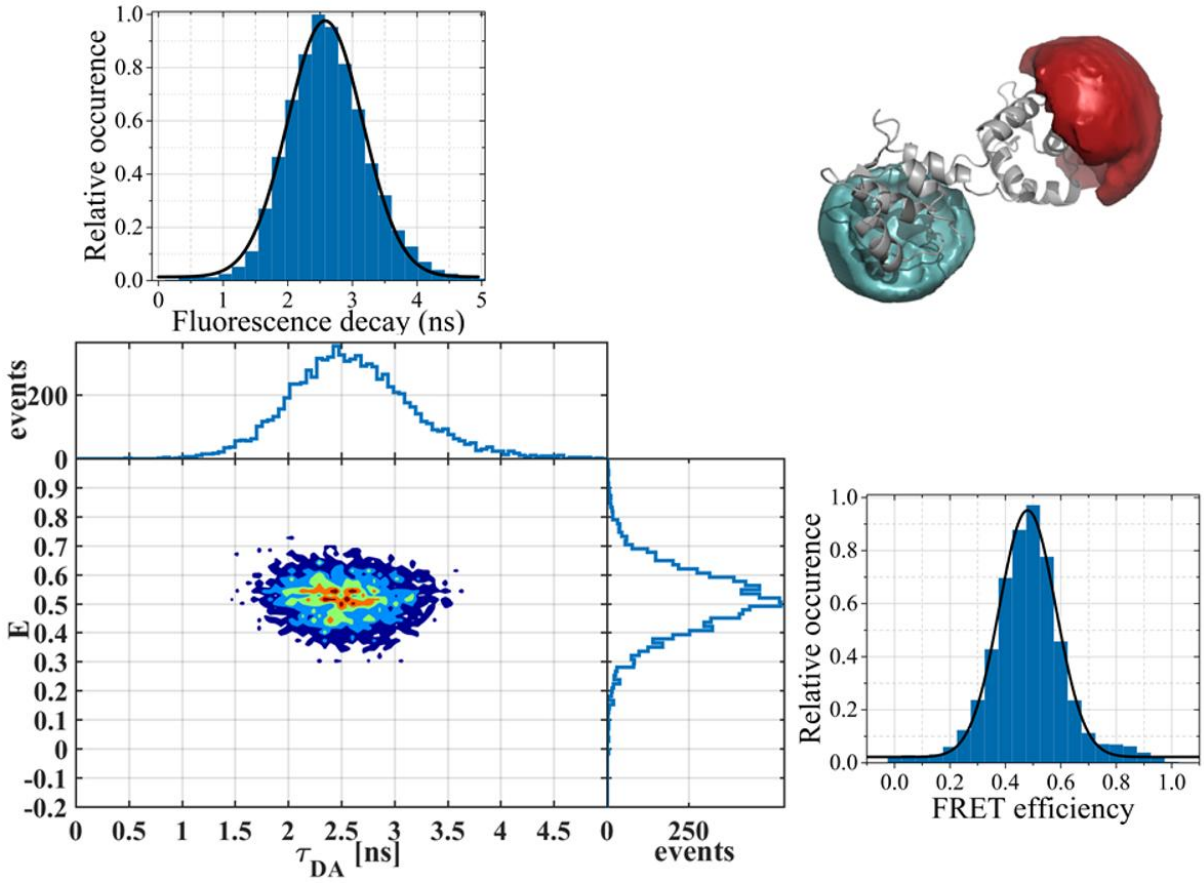


Figure 30. Structure of apoCaM T34AzF^{AF647} T110C^{AF488} verified by Förster Energy Transfer. 2D plot of FRET efficiency (E) and lifetime of the donor (τ_{DA}) of a dually labeled population of CaM (bottom, left), the FRET efficiency histogram (bottom, right) and the lifetime histogram (top, left) as well as the three-dimensional structure of apoCaM with the accessible volumes of the dyes (top, right) are shown. Gaussian fits were applied to the histograms and a mean FRET efficiency value was calculated for every peak.

Similar results were obtained for CaM T34C^{AF488} T110AzF^{AF647} (**Supplementary Fig. 5** and **Supplementary Table 2**). Our results indicate that the three-dimensional structure of CaM remains mostly unaffected by our labeling procedure.

4.2.2.3.3 Functionality of selectively dually labeled CaM

To investigate the functionality of CaM T34AzF^{AF647} T110C^{AF488} and CaM T34C^{AF488} T110AzF^{AF647}, smFRET measurements were performed in the presence of ligands and functional binding partners. Firstly, we have performed smFRET experiments in the presence of 10 mM $CaCl_2$. In such calcium-saturated conditions, CaM is bound to 4 calcium ions. Much more three-dimensional structures were reported in the literature for the holoCaM than for the previously mentioned apoCaM (**Fig 31**). It was even suggested that the holoCaM could sample a quasi-continuous spectrum of conformations [83]. For CaM T34AzF^{AF647}

T110C^{AF488}, the R_{DA} was found to vary significantly depending on the considered structure going from nearly 20 Å to around 65 Å (**Table 7**, column 5). As a consequence, a very broad range of FRET efficiencies is expected going from about 0.25 to nearly 1 (**Table 7**, column 6). Such broad range of molecular distances cannot be precisely covered by any currently available FRET dye pair. However, all holoCaM-related structures are distinct from the respective apoCaM allowing comparative analyses and detection of conformational changes upon calcium binding.

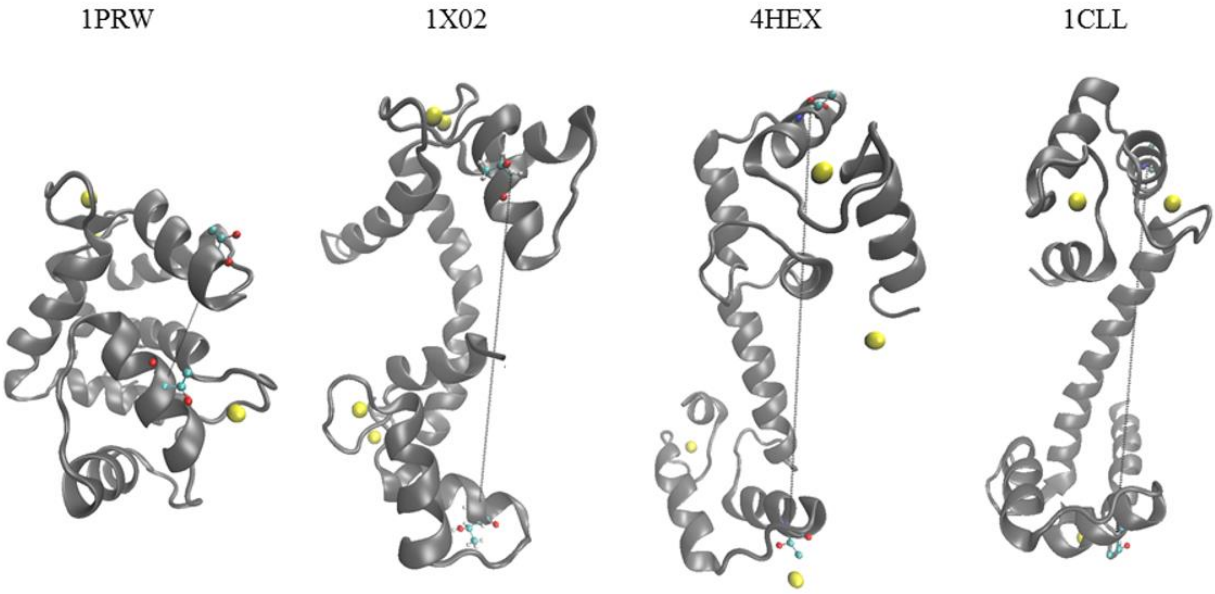


Figure 31. Various conformations of calmodulin bound to Ca^{2+}. The different three-dimensional structures of holoCaM are shown in gray. The two labeling positions P34 and P110 chosen for the present work are represented in atomic view and the molecular distance that separates them is represented by a dark line. Calcium ions are represented by yellow balls. PDB IDs are mentioned.

From data analysis, three distinct conformational subpopulations were visible for the holoCaM (**Fig. 32**, right) which is consistent with the fact that multiple three-dimensional structures were reported in the literature (**Fig. 31**). Moreover, the low- and high-FRET populations are fully compatible with the reported extended and compact forms of holoCaM, respectively [79, 81]. R_{DA} associated with the latter is about 20 Å. Due to the specific Förster radius of the dye pair AF488/AF647 (53.4 Å), this extremely short distance could not be properly resolved (**Supplementary Fig. 4**) leading therefore to a pretty narrow peak in high FRET efficiencies in the histogram. Additionally, we have observed a highly populated FRET population displaying a FRET efficiency of around 0.6 which was also previously reported by others [78, 143]. Nevertheless, this subpopulation cannot be reliably associated to

any known available structure. However, if a quasi-continuous spectrum of conformations is sampled by holoCaM as it was suggested [83], it would fall in this range of E values. In addition, our results are also in line with the particular conformational dynamics of holoCaM that was reported by others [144-146]. Similar results were got for CaM T34C^{AF488} T110AzF^{AF647} (**Supplementary Fig. 6** and **Supplementary Table 2**). Altogether, our results demonstrate that CaM T34AzF^{AF647} T110C^{AF488} and CaM T34C^{AF488} T110AzF^{AF647} behave as expected and undergo proper conformational changes upon calcium binding.

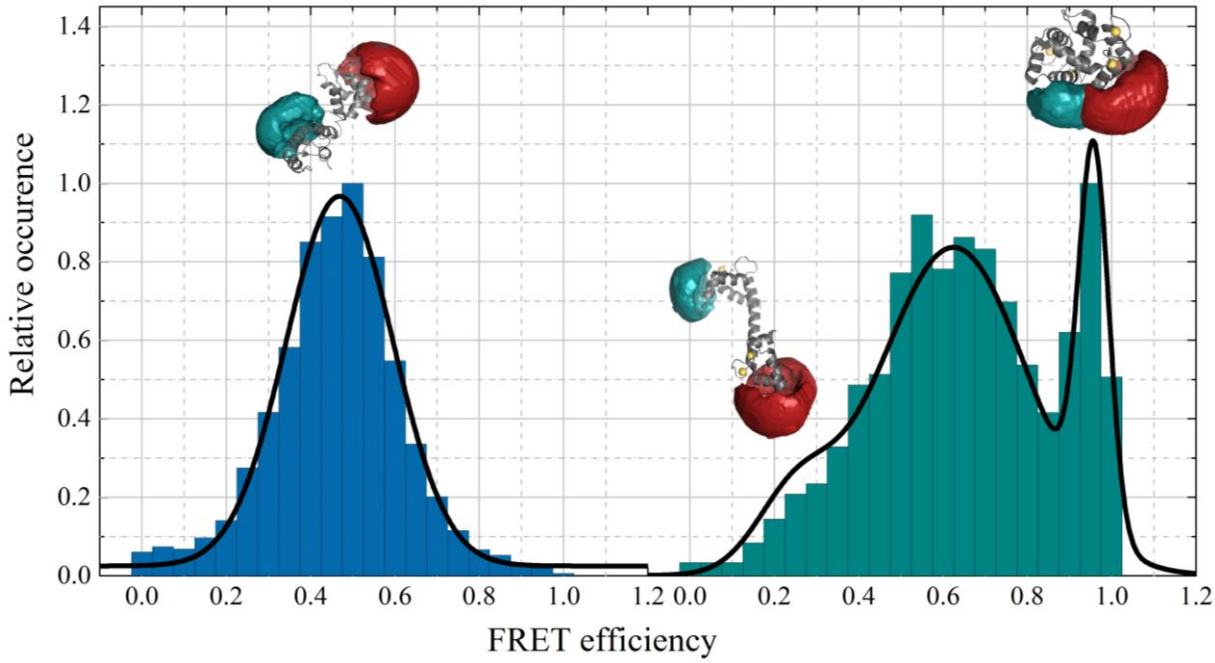


Figure 32. **Conformational changes that CaM T34AzF^{AF647} T110C^{AF488} undergoes upon Ca^{2+} binding.** FRET histograms of apoCaM (left) and holoCaM (right). The apoCaM structure with the accessible volumes for the two dyes is illustrated, as well as two extreme structures of holoCaM. Gaussian fits were applied to the histograms and a mean FRET efficiency value was calculated for every peak.

Secondly, in the aim to investigate further the functionality of CaM T34AzF^{AF647} T110C^{AF488} and CaM T34C^{AF488} T110AzF^{AF647}, similar smFRET measurements were performed in the presence of either CaMKII or M13 binding peptide (**Fig. 34** and **Fig. 35**). Even though the concentration regime is low, successful "binding experiments" performed at a single molecule level were already reported in the literature [57, 147]. As shown in **Fig. 33**, the structures of Ca^{2+}-CaM-CaMKII and Ca^{2+}-CaM-M13 complexes were reported [139, 140]. In both complexes, CaM is bound to 4 calcium ions and displays a compact form for which the calculated R_{DA} is about 20 Å (**Table 7**, column 5).

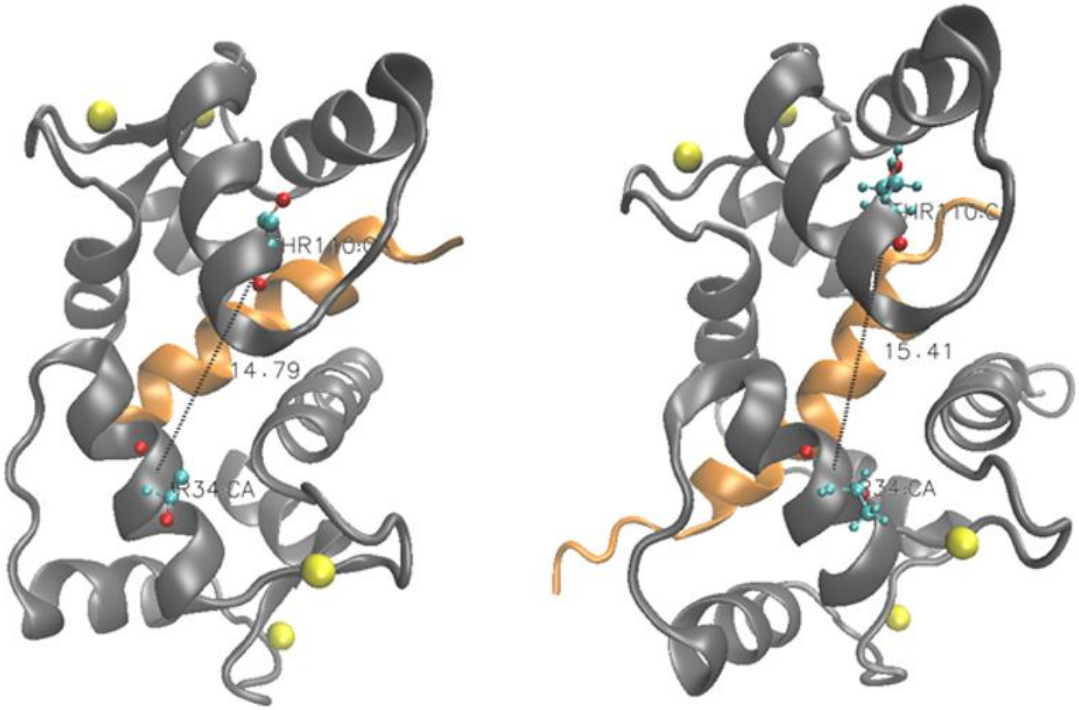


Figure 33. **Conformations of CaM bound to a partner.** Representation of CaM (gray) bound to CaMKII (orange) (left) and to M13 binding peptide (orange) (right) (PDB IDs: 1CDM and 2BBM). The calcium ions are represented by yellow balls.

smFRET measurements performed either with M13 or with CaMKII binding peptides were done in the presence of 10 mM $CaCl_2$ and 150 mM KCl. Here, physiological concentration of KCl were used since it was suggested that salts favored the interactions of CaM with its targets [148, 149]. As expected, when the partner is present, calculated FRET histograms display a narrow high-FRET population corresponding to the complexes (**Fig. 34** and **35**, middle). Furthermore, consistent with the fact that the interaction is calcium-dependent, the complex was dissociated by addition of EGTA (**Fig. 34** and **35**, right) leading to a related FRET histogram similar to the one obtained for apoCaM (**Fig. 30**, left**)**. Similar results were acquired for CaM T34C^{AF488} T110AzF^{AF647} (**Supplementary Fig. 7**, **Supplementary Fig. 8** and **Supplementary Table 2).**

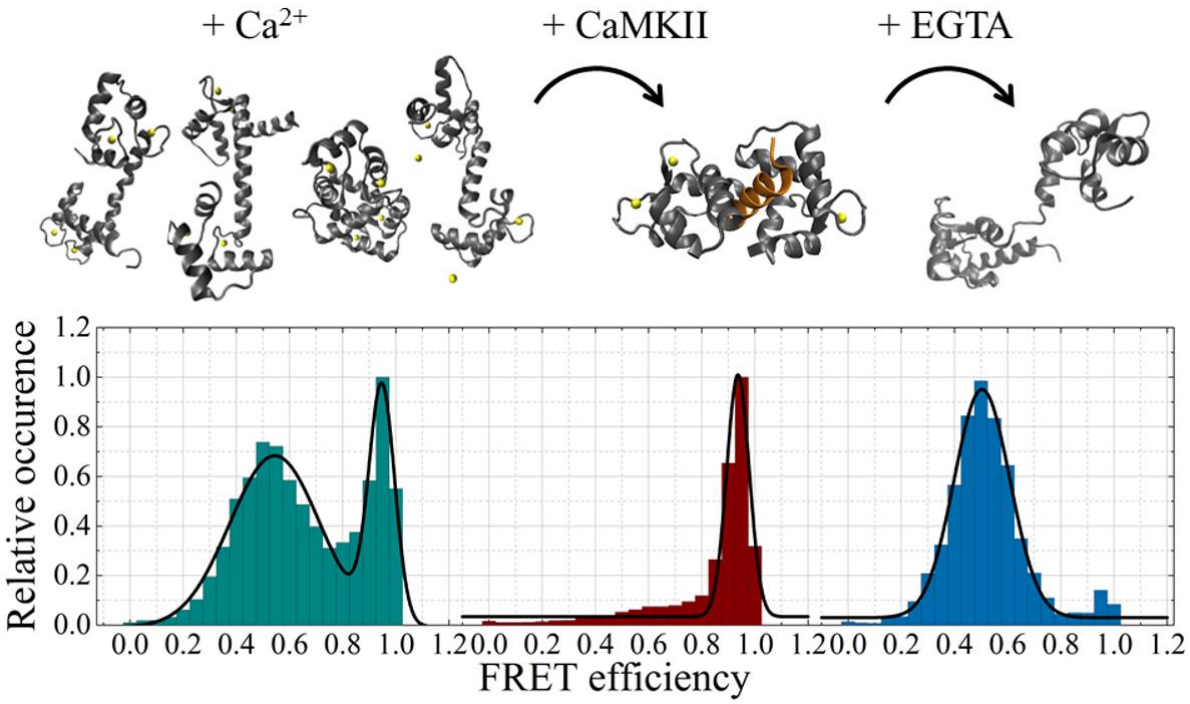


Figure 34: **Binding of CaMKII to CaM T34AzF^{AF647} T110C^{AF488}.** FRET histograms of holoCaM in the absence (left) or presence of CaMKII binding peptides (middle) and after addition of EGTA (right). Gaussian fits were applied to the histograms and a mean FRET efficiency value was calculated for every peak. Representations of holoCaM, Ca^{2+}-CaM-CaMKII and apoCaM are shown. The partner is displayed in orange.

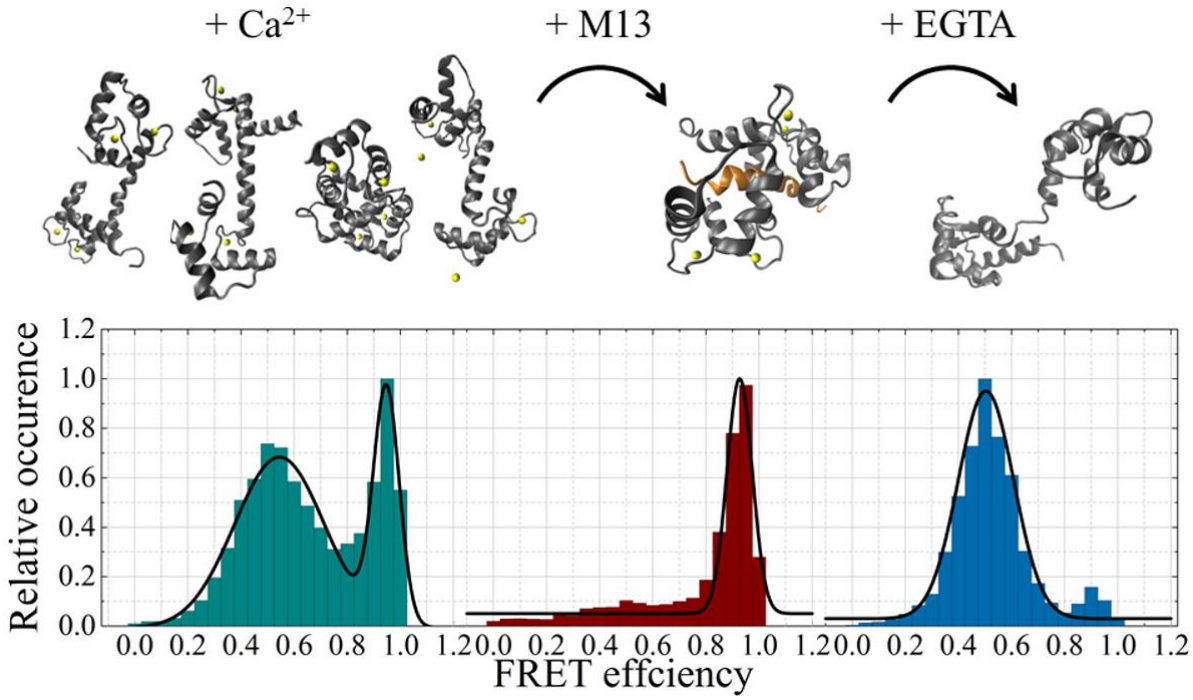


Figure 35. **Binding of M13 to CaM T34AzF^{AF647} T110C^{AF488}.** FRET histograms of holoCaM in the absence (left) or presence of M13 binding peptides (middle) and after addition of EGTA (right). Gaussian fits were applied to the histograms and a mean FRET efficiency value was calculated for every peak. Representations of holoCaM, Ca^{2+}-CaM-M13 and apoCaM are shown. The partner is displayed in orange.

The reversible binding of CaMKII and M13 partners strongly indicates that the functionality of the selectively dually labeled CaM remains unaffected. This highly suggests that the present methodological approach of selective dual-labeling of cell-free synthesized proteins does not disturb the structure and the functionality of CaM.

4.3 Selective dual-labeling by using fluorescent UAAs

Here, we have investigated the possibility to co-translationally incorporate into proteins fluorophores that are suitable for smFRET. Small fluorophores belonging to the BODIPY family were already shown to incorporate efficiently into proteins [27, 33]. However, BODIPY dyes are not suitable for single-molecule studies due to their weak photostability [60, 61]. Recently, larger fluorophores with better properties were also shown to incorporate into proteins but, only at the N-terminal region and when linked to p-aminophenylalanine (AmF) [150]. Nevertheless, the incorporation of large fluorophores remains limited to a couple of studies mainly due to the low yield it generates. In the present study, we have demonstrated for the first time, to our knowledge, that the site-specific co-translational incorporation of ATTO dyes, which are suitable for single-molecule detection, is compatible with smFRET studies.

4.3.1 Co-translational incorporation of a single fluorophore into CaM

4.3.1.1 Co-translational incorporation of fluorescent UAAs into CaM

In the present work, various dyes were site-specifically incorporated into CaM (**Table 8**).

Table 8. Characteristics of incorporated fluorophores.

	BODIPY-FL	BODIPY 576/589	ATTO633	Cy5	ATTO488
MW(g/mol)*	386	426	749	616	981
λabs (nm)	502	576	629	649	501
λfl (nm)	510	590	657	670	523
Charge	No net charge	No net charge	+1	+1	-1

*NHS ester, λabs = λ of max. absorption, λfl = λ of max. fluorescence emission

As presented in **Table 8**, these dyes display very different chemical and physical properties. For instance, ATTO dyes are significantly larger than BODIPY dyes. In addition, when the former are usually charged, the latter are generally hydrophobic.

Precharged tRNAs were used in combination to a RF1- and cysteine-depleted coupled transcription-translation cell-free system for the incorporation of a single fluorophore either into the N- or the C-terminal domain of CaM by using an amber stop codon or a cysteine codon, respectively (see Materials and Method for details) (**Fig. 36**).

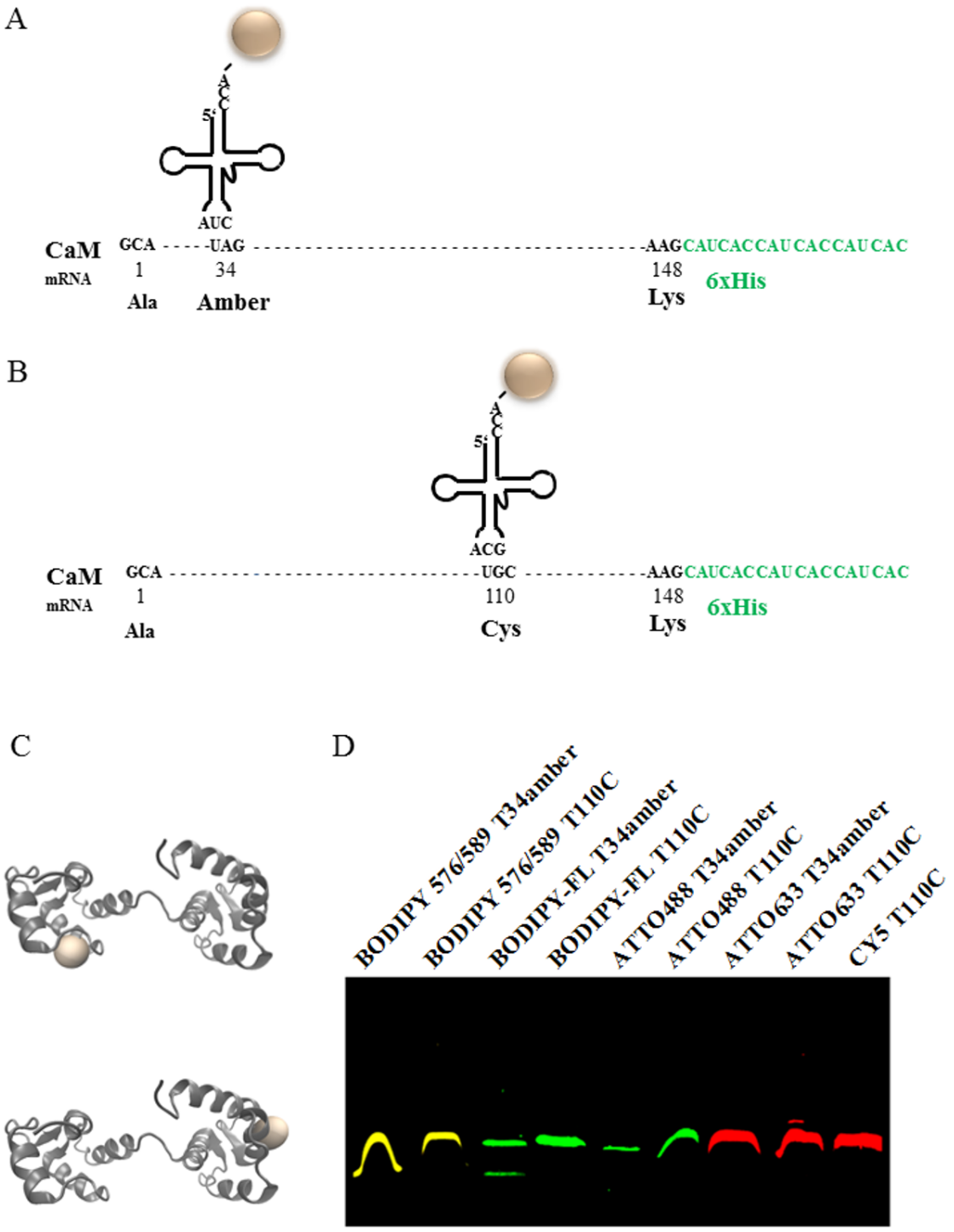


Figure 36. Incorporation of a single fluorophore into CaM. Co-translational incorporation of a fluorescent UAA into CaM via (A) an amber stop codon at P34 or (B) cysteine codon at P110. (C) Representation of CaM (gray) labeled with the incorporated fluorophore (beige sphere) either at the N- or at the C-terminal domain. (D) Fluorescent signals from different site-specifically labeled CaM on SDS PAGE.

The single incorporation of BODIPY-FL, ATTO633, ATTO488, Cy5 and BODIPY 576/589 were performed at either P34 or P110 (**Fig. 36A** and **Fig. 36B**) to generate site-specifically single labeled CaM (**Fig. 36**C). This was done by using fluorescently-modified cysteine, lysine or p-aminophenylalanine (AmF) residues. The successful incorporation of the UAAs was verified by gel imaging with the detection of fluorescent signals originated from the dyes. As shown, fluorescent signals corresponding to the dye of interest were clearly detected (**Fig. 36D**) demonstrating the successful incorporation of the various tested fluorophores. CaM was sometimes displaying a typical bell-shaped band which is thought to be due to the presence of both apo- and holo-forms that are characterized by different migration profiles (see Discussion for details)

Our results indicate that independently to their size and their charge, the incorporation of fluorophores into calmodulin N- and C- termini occurs successfully. This suggests that incorporations of large fluorophores such as ATTO dyes that display high quantum yield and photostability could be used for smFRET even though the incorporation efficiency would be much lower than for BODIPY dyes [27, 33, 150] .

4.3.1.2 Structure and functionality of selectively dually labeled CaM

Our objective was to assess the biological relevance of the methodological approach. To this aim, we have investigated the impact of the single incorporation of an ATTO dye on the structure and the functionality of CaM. Since in the present approach the incorporation of the fluorophore is made before the protein has reached its native structure, there is a non-negligible risk that the latter is affected. The risk is higher when the constraints associated with the dye are more significant as it is the case for the large and charged ATTO dyes. As a proof-of-concept, we have investigated the effect of the co-translational incorporation of ATTO633 on CaM by using smFRET.

4.3.1.2.1 Structure of selectively dually labeled CaM

To investigate the three-dimensional structure of CaM having incorporated ATTO633 at P34, smFRET measurements were performed with both CaM T34K^{ATTO633} T110C^{AF488} and CaM

T34AmF^{ATTO633} T110C^{AF488}. *CaM T34amber T110C* was used for incorporating ATTO633 via an amber stop codon at P34 and for the selective labeling of the cysteine residue at P110 with maleimide-functionalized AF488. Depending on the situation, ATTO633 was linked to either p-aminophenylalanine or lysine residues. smFRET measurements were performed in the presence of 10 mM EGTA. This means that the donor and acceptor fluorophores are separated by around 54 Å. Since the structure of ATTO633 is currently not available, the R_{DA} values for the different conformations were approximated those previously calculated for CaM T34 AzF^{AF647} T110C^{AF488} for which the size of the dyes and the linker lengths are quite similar. The expected FRET efficiencies (see 3.5.3.4, Eq. 12) were calculated by using the specific Förster radius of the dye pair ATTO633/AF488 (55.5 Å) [151] (see Materials and Method for details) (**Table 9**).

Table 9. Expected FRET efficiencies calculated for CaM T34AmF^{ATTO633} T110C^{AF488}

		calculated from Cα		**calculated from AV**	
PDB ID	**Description**	**R_{DA} (Å)**	**E**	**R_{DA} (Å)**	**E**
1DMO	ApoCaM	42.3	0.84	53.8	0.55
1CLL	HoloCaM	52.4	0.59	64.4	0.29
1PRW	HoloCaM	16.3	1.00	23.5	0.99
4HEX	HoloCaM	49.6	0.66	59.9	0.39
1X02	HoloCaM	40.4	0.87	52.6	0.58
1CDM	Ca^{2+}-CaM-CaMKII	14.8	1.00	20.5	1.00
2BBM	Ca^{2+}-CaM-M13	15.4	1.00	23.2	0.99

As shown in **Fig. 37**, a well-defined FRET population with a FRET efficiency of about 0.55 was observed, which is in line with the structure reported by Zang *et al*. [80] and consistent with the expected FRET efficiency of 0.55. Furthermore, the E values were similar for CaM T34K^{ATTO633} T110C^{AF488} and CaM T34AmF^{ATTO633} T110C^{AF488} with 0.56 and 0.58, respectively (**Fig.37**, left and right).

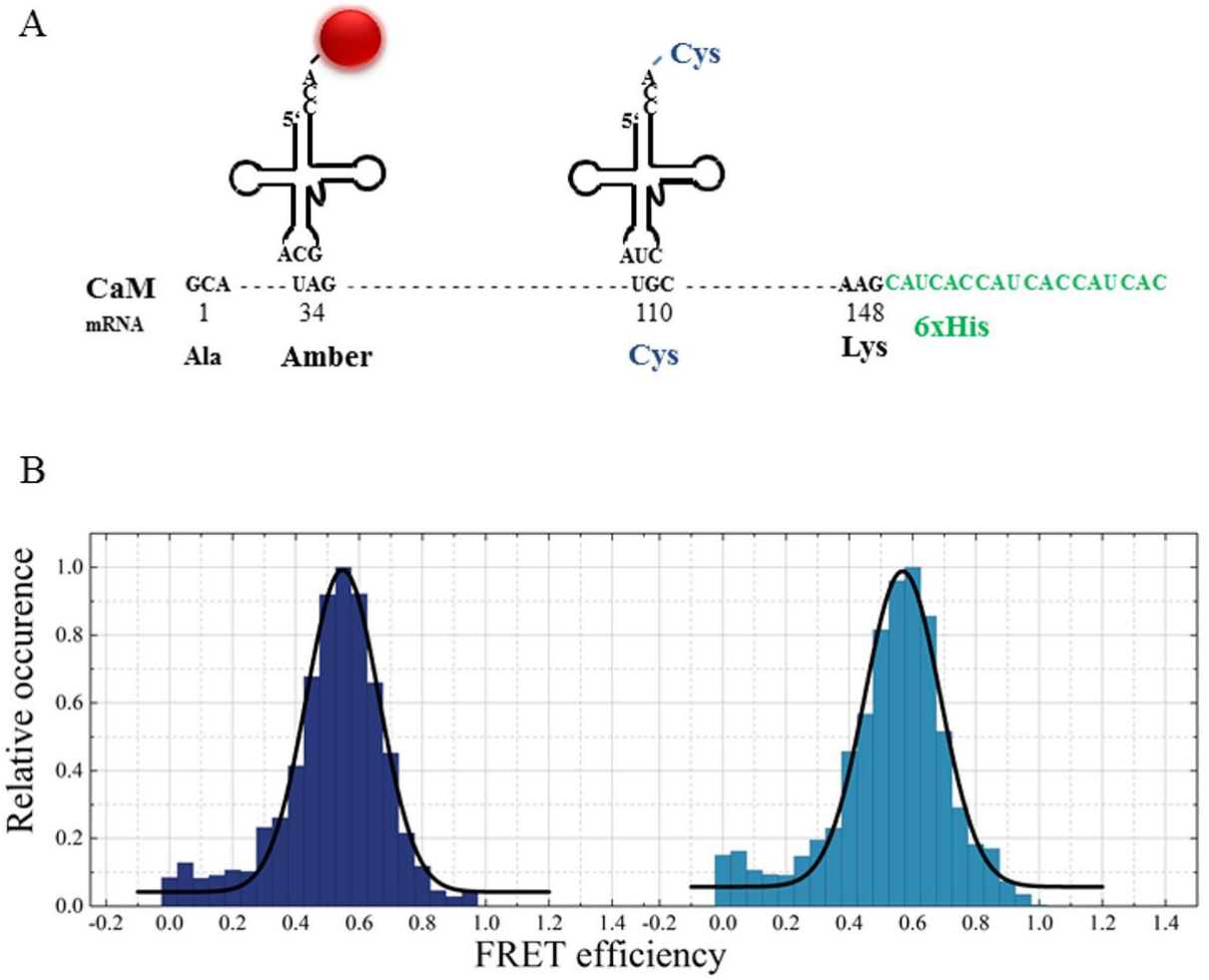


Figure 37. Structure of CaM T34AmF^{ATTO633} T110C^{AF488} and CaM T34K^{ATTO633} T110C^{AF488}. (A) Co-translational incorporation of ATTO633 into CaM via an amber stop codon at P34. (B) FRET histograms of apoCaM having incorporated at P34 either ATTO633-K (left, dark blue) or ATTO633-AmF (right, light blue). Gaussian fits were applied to the histograms and a mean FRET efficiency value was calculated for every peak.

Our results strongly indicate that the single incorporation of ATTO633 into CaM does not affect the three-dimensional structure of the protein.

4.3.1.2.2 Functionality of selectively dually labeled CaM

To investigate the functionality of CaM T34AmF^{ATTO633} T110C^{AF488}, smFRET experiments were performed in calcium-saturating conditions in the presence of CaMKII binding peptides [139]. In these conditions, the complex Ca^{2+}-CaM-CaMKII is allowed to form. When bound to CaMKII, CaM displays a compact form for which the calculated R_{DA} is about 20 Å [139, 140] (**Fig. 33**) (**Table 9**, column 5). Initially, in the absence of CaMKII, CaM responds to calcium by adopting multiple conformational substates (**Fig. 38**, left).

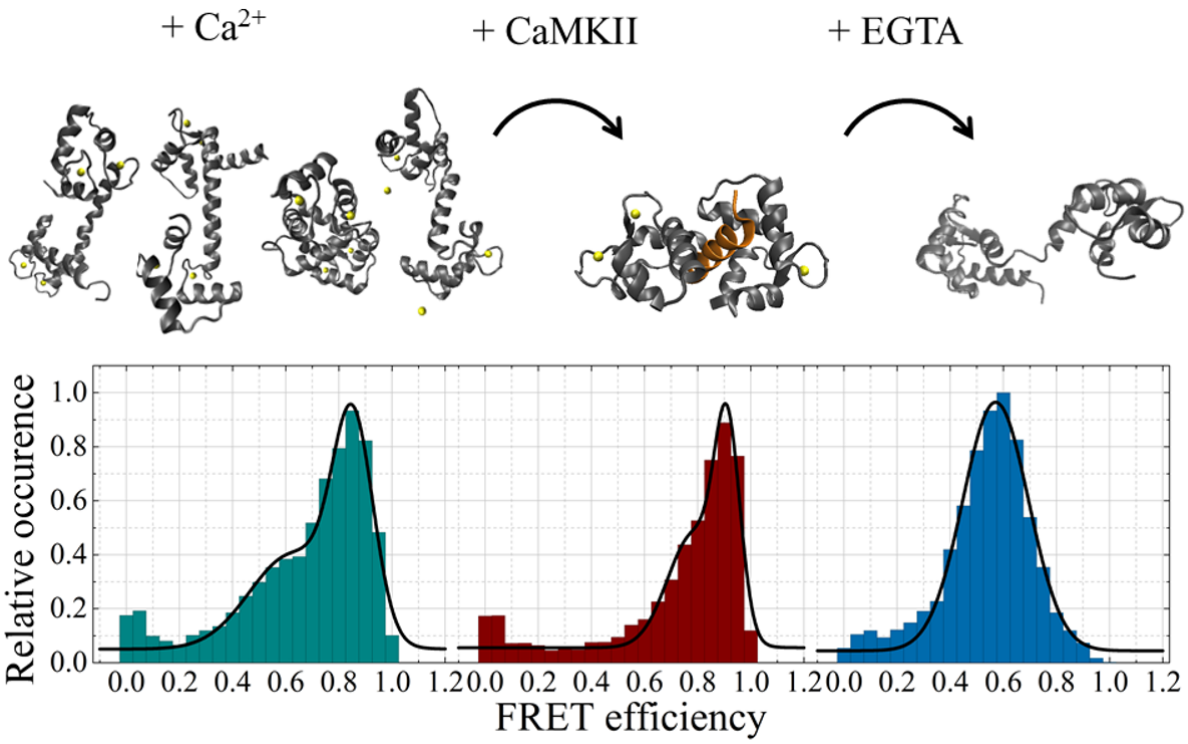


Figure 38. Functionality of CaM T34AmF^{ATTO633} T110C^{AF488}. FRET histograms of holoCaM in the absence (left) or presence of CaMKII binding peptides (middle) and after addition of EGTA (right). Gaussian fits were applied to the histograms and a mean FRET efficiency value was calculated for every peak. Representations of holoCaM, Ca^{2+}-CaM-CaMKII and apoCaM are shown. The partner is displayed in orange.

It is consistent with what we and others have previously reported (see 4.2.2.3.3) [78, 143]. In addition, in presence of CAMKII, the related FRET histogram displays the expected narrow high-FRET population corresponding to Ca^{2+}-CaM-CaMKII complexes (**Fig. 38**, middle). Finally, these complexes are properly dissociated by addition of EGTA, resulting in a FRET histogram which is similar to the one previously obtained for the apoCaM, in line with the fact that the specific interaction is calcium-dependent (**Fig. 38**, right and **Fig. 37**, left**)**. Our results indicate that CaM T34AmF^{ATTO633} T110C^{AF488} undergoes proper conformational changes upon both calcium and CaMKII binding. This strongly suggests that the co-translational incorporation of ATTO633 does not affect the structure and the functionality of CaM.

All in all, our results show the successful incorporation into CaM of a commercially available ATTO dye displaying high fluorescence quantum yield and photostability and which is therefore suitable for single-molecule detection.

4.3.2 Co-translational incorporation of two fluorophores into CaM

4.3.2.1 Co-translational incorporation of two fluorescent UAAs into CaM

Here, we have investigated whether the co-translational incorporation of an ATTO dye in combination with the co-translational incorporation of another fluorophore is appropriate for smFRET. In the present work, we have evaluated the relevance of incorporating two different fluorescent UAAs into CaM. As a proof-of-principle, smFRET measurements were performed with CaM T34AmF^{ATTO633} T110C$^{BODIPY\text{-}FL}$. *CaM T34 amber T110C* was used to incorporate ATTO633 at P34 and BODIPY-FL at P110 via an amber stop codon and a cysteine codon, respectively (**Fig. 39A** and **Fig. 39B**). As shown in **Fig. 39**, after dual-labeling, fluorescent signals corresponding to labeled CaM were detected for both ATTO633 and BODIPY-FL (**Fig. 39C**).

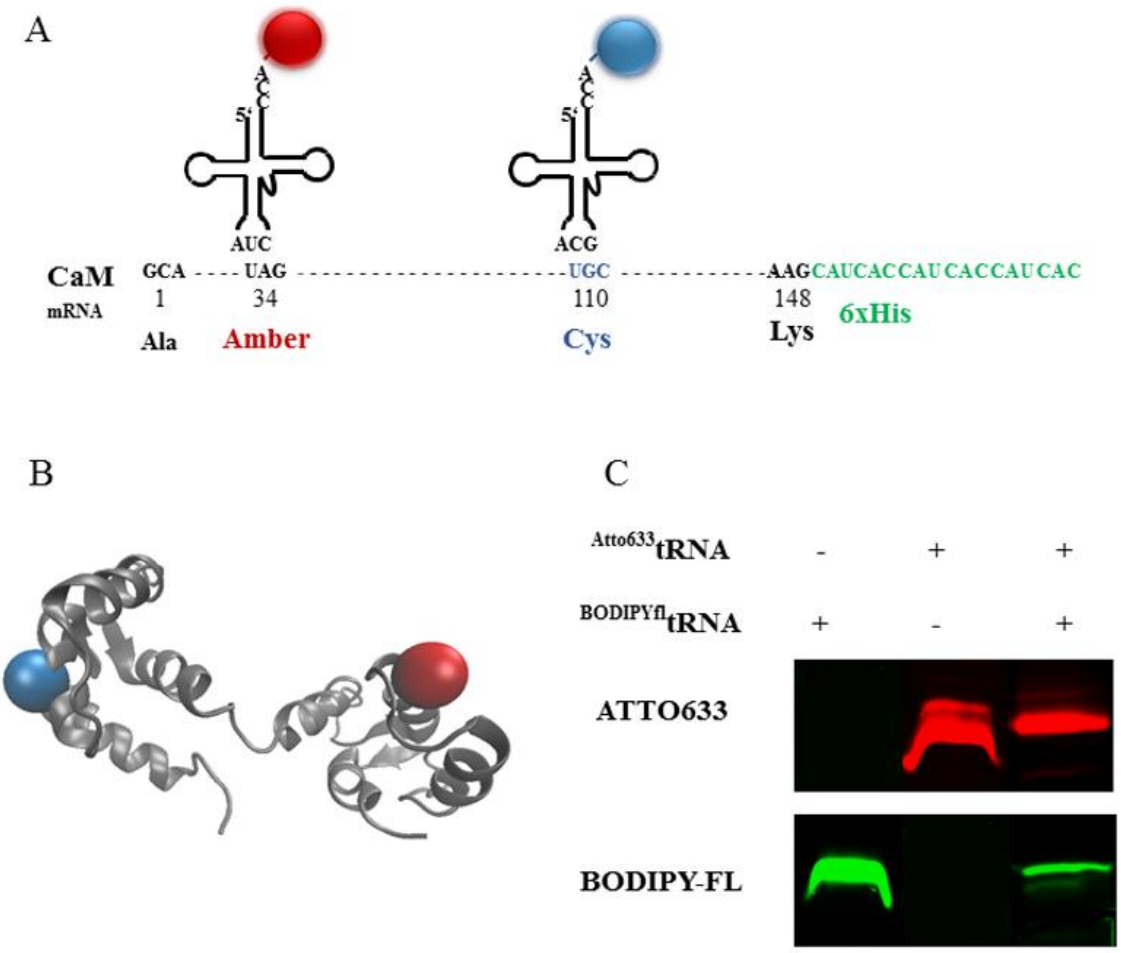


Figure 39. Incorporation of two fluorophores into CaM. (A) Co-translational incorporation of two fluorescent UAAs into CaM via amber stop and cysteine codons at P34 and P110, respectively. (B) Representation of CaM (gray) labeled with the two incorporated

fluorophores (blue and red spheres). (C) fluorescent signals from differently labeled CaM on SDS PAGE.

In addition, the presence of both dyes on the same molecule was investigated by using ensemble FRET measurements. After donor excitation, the presence of CaM T34AmF^{ATTO633} T110C$^{BODIPY-FL}$ was revealed by emission of both dyes (**Fig. 40**, top). When assuming that the co-translational incorporation of BODIPY-FL at P110 could only occurs if ATTO633 was previously properly incorporated at P34, it strongly indicates that emission light recorded from ATTO633 was generated by dually labeled CaM. Altogether, these results demonstrate the proper dual-labeling of the double mutant calmodulins (**Fig. 40**).

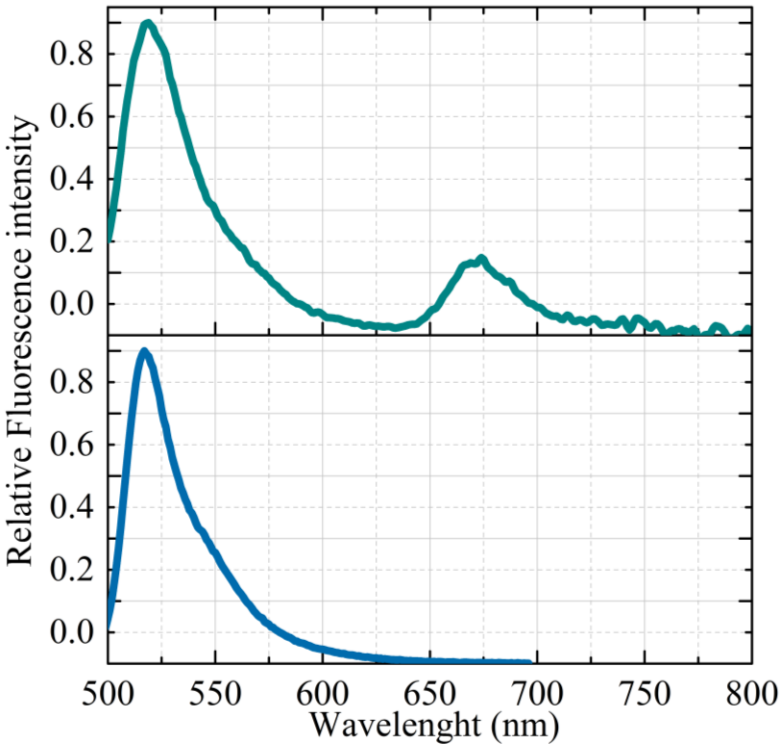


Figure 40. Dual-labeling of CaM by incorporation of fluorescent UAAs. Emission spectrum of dually labeled CaM T34AmF^{ATTO633} T110C$^{BODIPY-FL}$ after excitation of the donor (top, green solid line) is shown in respect to the spectrum of CaM T110C$^{BODIPY\ FL}$ (bottom, blue solid line).

All in all, our results strongly indicate that the incorporation of two fluorescent UAAs is relevant for the selective dual-labeling of cell-free synthesized proteins for smFRET. Even though the protein yield is low, it does not constitute significant issues for smFRET, which requires only a few picomoles of sample.

4.3.2.2 smFRET measurements with selectively dually labeled CaM

In the aim to investigate the relevance of the incorporation of ATTO dyes for single molecule detection, smFRET measurements were performed with CaM T34AmF^{ATTO633} T110C$^{BODIPY-FL}$ in the presence of 10 mM $CaCl_2$. In such conditions CaM is in its holo-form (**Fig. 41A**). For CaM T34AmF^{ATTO633} T110C$^{BODIPY-FL}$, R_{DA} values could not be precisely calculated due to the lack of available structure for ATTO633 and the precharged tRNAs used. Even though it is not optimal, they were approximated to those found for CaM T34AzF^{AF647} T110C^{AF488}. Since our objective, here, is only to assess the feasibility of incorporating two UAAs for smFRET, this should be relevant enough. The expected FRET efficiencies (see 3.5.3.4, Eq. 12) were calculated by using the calculated Förster radius of the specific dye pair BODIPY-FL/ATTO633 (45.9 Å) (see Materials and Method for details) (**Table 10** and **Supplementary Fig. 10**).

Table 10. Expected FRET efficiencies for holoCaM T34AmF^{ATTO633} T110C$^{BODIPY-FL}$

		calculated from Cα		calculated from AV	
PDB ID	**Description**	**R_{DA} (Å)**	**E**	**R_{DA} (Å)**	**E**
1CLL	HoloCaM	52.4	0.311	64.4	0.116
1PRW	HoloCaM	16.3	0.997	23.5	0.982
4HEX	HoloCaM	49.6	0.386	59.9	0.168
1X02	HoloCaM	40.4	0.683	52.6	0.306

As shown in **Fig. 41B**, the related FRET histogram displays, in addition to a zero peak thought to be due to low statistic and low percentage of dually labeled proteins, a narrow high FRET peak found around E = 1 associated to the expected compact form of the holoCaM and another highly populated FRET population of about E = 0.5 which does not correspond to any known structures (**Table 10**, column 6), consistent with what was previously reported by us (see 4.2.2.3.3) and others [78, 143].

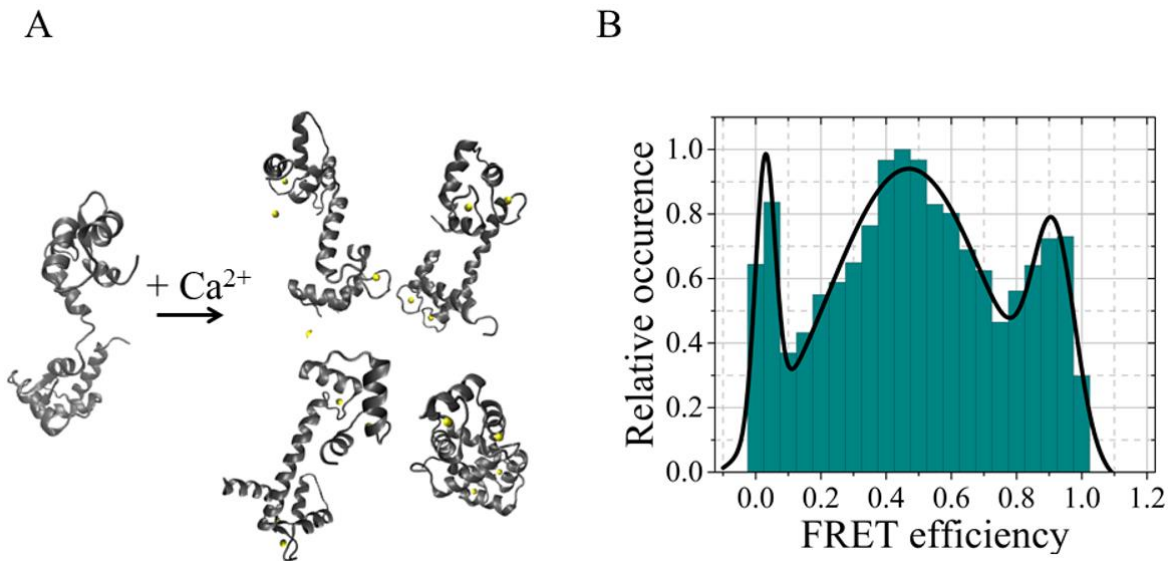


Figure 41. smFRET histogram of holoCaM T34AmF^{ATTO633} T110C$^{BODIPY-FL}$. (A) Conformational changes of CaM upon Ca^{2+} binding (B) smFRET histogram of holoCaM labeled with co-translationally incorporated ATTO633 at P34 and BODIPY-FL at the P110. Gaussian fits were applied to the histograms and a mean FRET efficiency value was calculated for every peak.

Our results demonstrate that the co-translational incorporation of an ATTO dye is relevant for smFRET studies. Further, it suggests that the co-translational incorporation of two fluorophores does not affect significantly the three-dimensional structure of CaM. This is consistent with the fact that both the structure and the functionality of CaM remain unaffected by single incorporated ATTO633 and BODIPY-FL (see 4.3.1.2.2) [33].

Altogether, our results have demonstrated the possibility to conduct smFRET measurements of cell-free synthesized proteins by co-translationally incorporating suitable dyes (i.e. ATTO633) for optimal single-molecule detection. Furthermore, even though the protein yield is decreased, we have proved that it is still largely sufficient for single molecule studies. For these reasons, co-translational incorporation of ATTO dyes appears as a very promising alternative approach for the selective dual-labeling of cell-free synthesized proteins for smFRET studies, in particular those requiring real-time measurements which prevent the use of the previously described approach based on post-translational chemical modifications.

4.4 Selective dual-labeling for more accurate smFRET

Here, we have investigated the effect of selective dual-labeling on smFRET. To this aim we have compared the situation encountered when the classical double cysteine mutagenesis is used with the one encountered when our methodological approach of selective labeling by

using a bioorthogonal chemistry is applied. Here, we demonstrate that selective dual-labeling of proteins allows more accurate smFRET studies.

4.4.1 QY measurements

In the present work, selectively dually labeled calmodulins were used for smFRET. Data analysis required to calculate the FRET efficiency as precisely as possible. Since, the quantum yields (ϕ) of the dyes enter into the FRET efficiency calculation (Eq. 16), they have to be determined with high accuracy. Additionally, ϕ of the donor is also included in the Förster radius calculation (Eq. 13).

The quantum yields for the free-forms of AF647 and AF488 were reported in the literature (**Table 11**, columns 1 and 4). However, the quantum yield of the dye when bound to a protein is usually different since the photophysical properties of a fluorophore can be highly dependent on the specific environment encountered [152, 153]. For instance, protein induced fluorescence enhancement (PIFE) was recently reported for various dyes displaying cis-trans isomerization [154]. Thanks to our selective methodological approach, the determination of QY values for each dye when bound to CaM at a specific position was achievable.

Table 11. Quantum yields of Alexa488 and Alexa647

Alexa Fluor 488			**Alexa Fluor 647**		
Free*	P34	P110	Free*	P34	P110
0.92	0.65	0.66	0.33	0.27	0.36

*NHS ester

In the present study, quantum yields were determined for each fluorophore when bound to CaM at a specific position. The method was based on the comparative analysis of the molecular brightness (MB) of the sample and a reference (Eq. 20) (see Method, for details) [122]. MB values were measured for AF488 and AF647 when bound to the CaM at positions 34 and 110 (Eq. 19). As shown in **Table 11**, the QY of the donor did not change significantly in respect to the position and the values are rather similar. This value is nevertheless drastically different to the value of the free dye with a decrease of about 30 %. On the contrary, the QY of the acceptor is much more weakly affected by the protein binding but varies significantly according to the position with nearly 10 % of decrease for P34 as compared to P110.

4.4.2 smFRET experiments with selectively dually labeled CaM

It was previously suggested by others that selective dual-labeling would significantly improve smFRET experiments [31]. Here, our objective was to directly assess whether a gain in accuracy due to selective labeling is appreciable. To this aim, smFRET measurements were performed with either CaM T34AzF^{AF647} T110C^{AF488} or CaM T34C^{AF488} T110AzF^{AF647}. The two constructs differ compared to each other for the position of the dyes; for the former AF647 is located at position 34 and AF488 at position 110, while for the latter the dyes are present in the opposite way (**Fig. 42A**, left and middle).

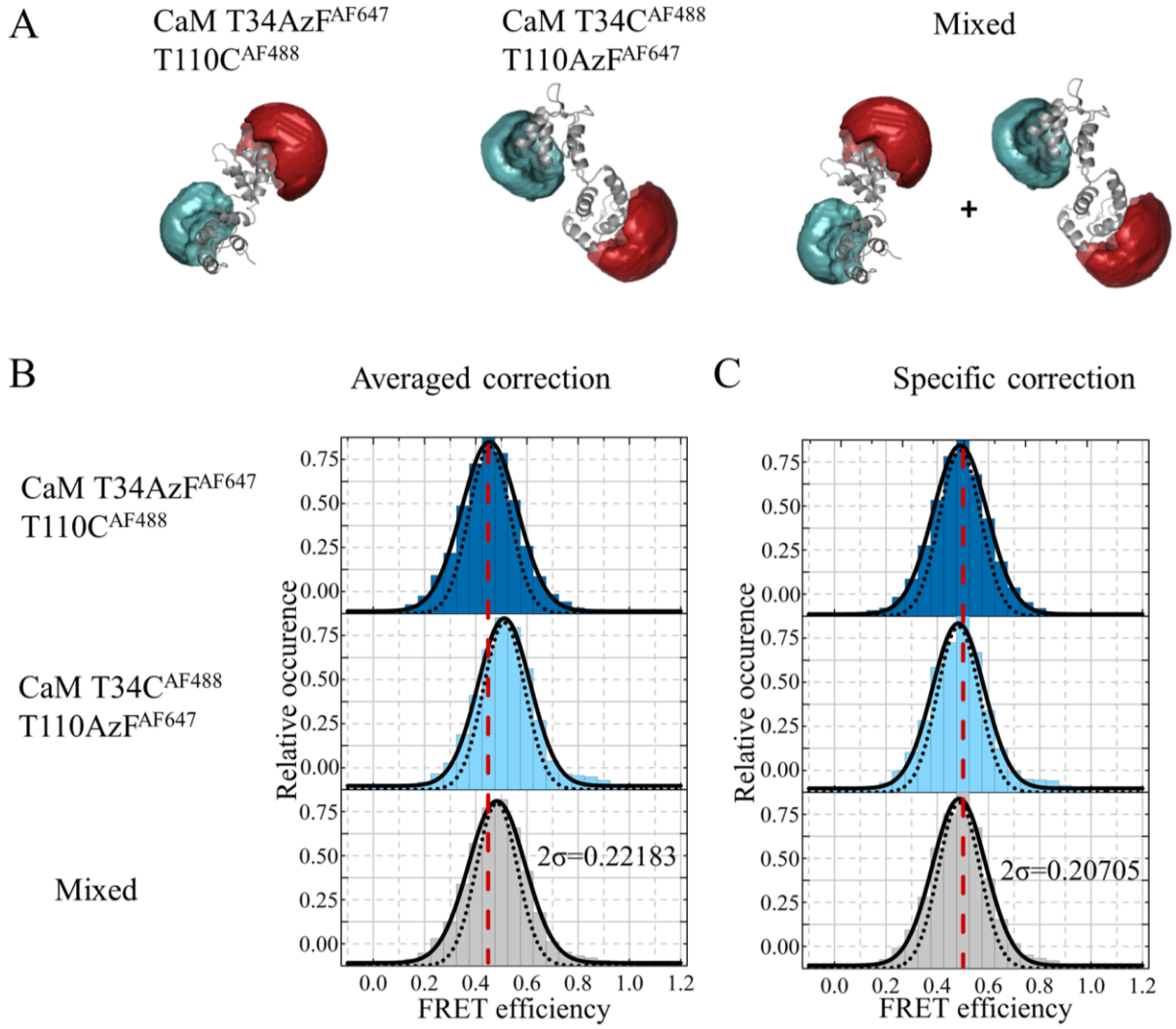


Figure 42. Gain in accuracy of the FRET histograms due to selective dual-labeling. (A) The constructs with the AVs are illustrated for each case given in (B) and (C). (B) FRET efficiency histograms of CaM T34AzF^{AF647} T110C^{AF488} (top), CaM T34C^{AF488} T110AzF^{AF647} (middle) and a mixed of both constructs (bottom) using an averaged correction for the QYs of the fluorophores in both positions. (C) FRET efficiency histograms from the same constructs using specific corrections for the QYs of the fluorophores at their specific positions. A

vertical dashed red line which is centered on the mean efficiency of CaM T34AzF^{AF647} T110C^{AF488} (top) is drawn for easier visualization of the relative shifts. Shot-noise are represented in dotted lines.

By using our methodological approach of selective dual-labeling by chemical modifications, we were able to distinguish both products (i.e. CaM T34AzF^{AF647} T110C^{AF488} and CaM T34C^{AF488} T110AzF^{AF647}) and measure them separately in smFRET. However, this is not possible when a non-selective approach is used for the dual-labeling of proteins, such as the classical double cysteine mutagenesis approach. Here, FRET efficiencies from measurements made with either CaM T34AzF^{AF647} T110C^{AF488} or CaM T34C^{AF488} T110AzF^{AF647} were then calculated by using two different approaches (**Fig. 42B** and **Fig. 42C)**. These approaches differ for the way in which they take into account the QYs of the fluorophores when bound to CaM (**Table 11**). Firstly, for each dye, we have applied the classical "averaged correction" (**Fig. 42B)**. In that case, we have calculated for each fluorophore the arithmetical average of the QYs of every position (i.e. 34 and 110). The two obtained averaged values (i.e. for the donor and acceptor) were subsequently used to calculate the FRET histograms of each construct (**Fig. 42B**, top and middle). This approach is the one classically used when a sample contains a mixed population of both constructs (**Figure 42A**, bottom) like those obtained with the very common non-selective labeling procedure based on double cysteine mutagenesis [56]. As shown in **Fig. 42**, we observed that the FRET histograms of the two constructs (i.e. CaM T34AzF^{AF647} T110C^{AF488} and CaM T34C^{AF488} T110AzF^{AF647}) were shifted according to each other (**Figure 42B**, top and middle). In order to reconstitute the situation encountered in non-selective dual-labeling (e.g. for the double cysteine mutagenesis approach) where the fluorophore does not favor any of the two positions, we have averaged the FRET histograms of both constructs (**Figure 42B**, bottom). From that operation, it appeared that a further broadening resulting from the average of two shifted populations becomes apparent in the newly derived FRET histogram.

Secondly, in another approach to calculate the FRET efficiencies, we have used the QY of each fluorophore at its exact position (i.e. 34 or 110) within the respective construct instead of averaging the values of both positions (**Table 11**). Here, we have applied what we have called the "specific correction" since it takes into account the real value of the QY of the dye at its specific position. This is only possible when selective dual-labeling is applied and would not be applicable in non-selective dual-labeling (e.g. for the double cysteine mutagenesis approach). As shown in **Fig. 42**, the FRET histograms are not shifted anymore according to each other, consistent with the similar R_{DA} previously calculated for both

constructs (**Figure 42C**, top and middle**, Table 7** and **Supplementary Table 2**). Furthermore, averaging of both histograms did not lead to any further broadening (**Figure 42B** and **42C,** bottom), clearly demonstrating that selective dual-labeling leads to more accurate smFRET.

The broadening in this particular case is not so big, due to the optimal use of labeling positions, which were previously reported in the literature [35, 126-129]. Nevertheless, even small, the effect is present. There are naturally cases, in which the QY of a fluorophore is much more affected (e.g. due to strong local quenching effects), resulting in larger broadening of the FRET histogram. In extreme cases, the corresponding labeling positions are discarded as not suitable for performing single molecule data analysis. In these situations, the selective dual-labeling that allows the "specific correction" of the QYs of the fluorophores will offers a considerable gain in accuracy and the possibility to analyze all the data without the need to discard any of them.

5. Discussion and outlook

Here, we have presented two methodological approaches for the selective dual-labeling of cell-free synthesized proteins for smFRET studies. Firstly, in the aim to selectively dually label proteins, we have used an expanded genetic code to incorporate UAAs that carry either fluorophores or specific functional groups, which can be used for further selective chemical modifications. Secondly, we have demonstrated the biological relevance of the methodological approaches by performing a case-study of the well-studied human calmodulin. It is the first time, to our knowledge, that selective dual-labeling of cell-free synthesized proteins with suitable dyes for smFRET is reported. In the following sections, we will discuss the main strengths and weaknesses of combining CFPS and smFRET, its potential applications, the importance of selective dual-labeling for smFRET and the choice of CaM as a model protein.

5.1 Combination of CFPS and smFRET

The association of FRET with single molecule detection combines high sensitivity and single molecule discrimination [155]. Consequently, smFRET constitutes an incomparable tool for accessing quantitative and real-time information about biomolecule interactions and conformational dynamics that would not be approachable in ensemble measurements. In particular, smFRET allows sensitive and accurate molecular distance quantification and FRET-based structural modeling of co-existing conformational substates [156, 157]. For the latter, smFRET measurements have the strong advantage over classical X-ray crystallography and NMR that it does not required highly ordered scheme. This becomes an immense advantage when the protein of interest displays some high levels of flexibility like it is the case for intrinsically disordered proteins (IDPs). Additionally, smFRET appeared to be very precise in probing complex systems and biological processes in real-time [158]. Here, we have described two methodological approaches, which allow its combination with CFPS. Since, smFRET requires very small amounts of sample (i.e. few picomoles of sample), it is suitable for CFPS which major downside is linked to the limited protein yield (hundreds µg/ml up to few mg/ml) in respect to the classical *in vivo* expression (several mg/ml) [18]. Despite this, the application of smFRET is still often limited to *in vivo* expressed proteins. However, such a combination can be disadvantageous in many ways. Firstly, *in vivo* expression is time-consuming and implicates complex multi-step procedures which are often

associated with important organism-related issues (e.g. those associated with cell growth, cell transformation and protein degradation by endogenous phosphatases). Secondly, smFRET remains limited by the host strain's tolerance and is hardly accessible to difficult-to-express candidates (i.e. plasma membrane and toxic proteins), which are reluctant to *in vivo* expression [159]. Since smFRET is particularly powerful in probing protein folding and interactions, its near restriction to *in vivo* expressed proteins greatly limits the field of structure-function studies by excluding particularly interesting proteins. Finally, *in vivo* expression implies the necessity of having fully equipped working place dedicated to that purpose and linked to rigorous safety regulations [160]. Here, the combination of smFRET and CFPS (i) simplifies and shortens the sample preparation procedure, (ii) allows a level of flexibility and monitoring that would never be reached *in vivo* and (iii) expands smFRET to the study of difficult-to-express proteins.

5.2 Cell-free synthesis of calmodulin and incorporation of UAAs

In the present study, modified coupled transcription–translation cell-free systems were used in the aim to incorporate UAAs into calmodulin. In general, the amounts of protein obtained with CFPS are lower than those resulting from *in vivo* expression [18]. Moreover, the incorporation of UAAs tends to reduce further the yield (i.e. ng/ml up to hundreds μg/ml) [39, 161]. The decrease is strongly related to the incorporation efficiency of the specific UAA which depends on the positions, orthogonal tRNA/tRNA synthetase pair, expanded genetic code and modified cell-free system used. In the present work, both amber and cysteine codons were used for incorporating UAAs. In the aim to increase significantly the efficiencies of UAA incorporation, either RF1- or RF1 and Cys-depleted modified coupled transcription-translation cell-free systems were used [162]. In the following sections, we will specifically discuss the choice of the mutants and the main issues associated with cell-free synthesis of the human calmodulin and incorporation of UAAs into it.

5.2.1 Choice of the mutant calmodulins

In the present study, double mutant calmodulins were generated from the constructs *CaM T34amber T110C* and *CaM T34C T110amber*. Among the various double mutant calmodulins reported in the literature, CaM T34C T110C was found to have been widely used by others [35, 127, 128]. Whatever the dye pair used, the molecular distance separating the two mutated positions was optimal for studying CaM structure and functionality.

Moreover, the latter was shown to remain unaffected by the two inserted mutations at P34 and P110. Additionally, CaM T34C T110C was also specifically used for smFRET studies [35, 78] allowing a direct comparison with the present work. For instance, our results for the apo-form and holo-form of CaM were similar to those obtained by Johnson's group demonstrating the viability of our methodological approaches [78].

5.2.2 Cell-free synthesis of calmodulin

Here, we have shown the successful cell-free expression of human calmodulin by using a prokaryotic coupled transcription-translation system. In the present study, the product has presented some particularities, which have prevented its identification by using classical molecular biology methods and the specific near UV absorption spectrum of calmodulin was used instead [100]. Indeed, classical his-tag detection by western blot was unsuccessful (data not shown). This is thought to be due to the combination of the low amount of cell-free synthesized CaM with the low accessibility of the His_6-tag. The latter suggests that CaM would be only partially denatured by SDS. This is consistent with the exceptional high stability of its holo-form, which most probably allows only partial unfolding of the protein by denaturants and goes in line with the fact that we and others have observed faster SDS PAGE migration of the holoCaM in respect to the apoCaM which could reflect the presence of different conformations [130]. Furthermore, low accessibility of the His_6-tag is fully consistent with the fact that purification of his-tagged calmodulins was inefficient when cobalt resins were used whereas it did not encounter any problem by using nickel resins (**Supplementary Fig. 11**). This observation can easily be explained by the fact that the latter displays a higher affinity towards polyhistidine tag in respect to the former [163].

5.2.3 Incorporation of UAAs into calmodulin

In the aim to achieve selective dual-labeling, one or two UAAs were incorporated into CaM. The UAA carried either a fluorophore or a specific functional group that was further used for chemical modifications. On one hand, we have observed that incorporation of AzF into CaM by using a RF1-depleted orthogonal coupled cell-free transcription-translation system did not significantly decrease the protein yield, which is consistent with what was previously reported by others [162]. This suggests that the incorporation is rather efficient even though we cannot exclude some misreading events. However, such misreading events should be low considering the acceptable degree of labeling of about 40 to 80 % for the two chemistries used in the present study. On the other hand, we have seen that the co-translational

incorporation of one or two fluorophores strongly reduced the protein yield (i.e. hundreds of nM to few nM) as it is expected for incorporation of such a large group [27]. Consistent with previous studies, the decrease was highly dependent on the position of incorporation and the type of incorporated fluorophore for which the size and the charge were found to be crucial [33, 150]. It was suggested that the chemical and physical properties of the fluorophores affect its interaction with key factors of the translational machinery such as the EF-Tu elongation factor and imply obvious physical constrains related to the vicinity of neighborhood amino acid side-chains of the nascent chain and ribosomal subunits [27]. These important challenges associated with the co-translational incorporation of fluorophores explain why it is so far restricted to only a couple of studies. However, the decrease in yield should not constitute an issue for single molecule methods that require only limited amounts of sample.

In the present work, we have demonstrated through a case-study that incorporation of UAAs into cell-free synthesized proteins is relevant for single molecule applications. Furthermore, we have shown that the incorporation of functional groups and fluorophores is suitable for smFRET by achieving proper selective dual-labeling of the protein of interest.

5.3 Selective labeling of cell-free synthesized CaM

Dual-labeling of the protein of interest is a necessity for smFRET and a critical step in the sample preparation process. The classical approach based on double cysteine mutagenesis required relatively large amounts of proteins and complex procedures that are hardly compatible with CFPS [31, 164]. The problem resides in the fact that an identical chemistry (i.e. via maleimide-cysteine reaction) is used for the dual-labeling, which obviously prevent selectivity since both donor and acceptor fluorophores can react with each thiol group of the two mutated positions. Therefore, this leads to evident heterogeneous samples containing four types of dually labeled products (i.e. donor-donor, acceptor-acceptor, donor-acceptor and acceptor-donor labeled proteins) [165]. In the aim to deal with this unwanted heterogeneity different strategies were developed which are however sample-consuming and therefore non-suitable for CFPS [165, 166]. For that reason, the double cysteine mutagenesis approach is currently strictly restricted to *in vivo* expressed proteins. With this study, we achieved to replace the classical *in vivo* method by more convenient cell-free approaches. Here, we have described two simple and efficient methodological approaches for the selective dual-labeling of cell-free synthesized proteins. Both approaches lead to a unique homogeneous selectively

dually labeled population, rendering the labeling procedure suitable for CFPS. Both approaches are biologically relevant as demonstrated by the proper folding of the selectively dually labeled CaM.

5.3.1 Selective dual-labeling by using a bioorthogonal chemistry

Here, we have site-specifically incorporated two distinct chemoselective reactive groups for the subsequent selective dual-labeling of cell-free synthesized proteins by post-translational chemical modifications. An RF1-depleted orthogonal coupled cell-free system was used for incorporating both unique cysteine and azido-phenylalanine residues into CaM [34-40]. The latter is an UAA carrying a specific bioorthogonal functional group [31, 94]. Then, selective dual-labeling was performed by reacting sulfhydryl- and azide-groups with functionalized fluorophores. The relevance of this fast and simple approach for the selective dual-labeling of cell-free synthesized proteins for smFRET was demonstrated by the case-study of the human calmodulin. In the present work, we have shown (i) that the combination of azide- and sulfhydryl- reactive reagents is easy to handle and allows efficient selective dual-labeling, (ii) that incorporation of the functional groups and chemical-modification-based dual-labeling with photostable fluorophores do not disturb the structure and the functionality of the protein and (iii) that our approach produces enough proteins for smFRET studies.

5.3.1.1 Selective dual-labeling by post-translational chemical modifications

In the present study, we have assessed the occurrence of the selective dual-labeling of cell-free synthesized calmodulin. To this aim, we have used both gel imaging and TCCD measurements. We have observed a degree of labeling of about 40-80 % for both chemistries and around 30 % of dually labeled proteins. The latter percentage was calculated over the whole labeled population and therefore implies that about 70 % of the labeled molecules passing through the confocal volume carry only one dye. Even though 70 % of the detected individuals are not of interest, 30 % of dually labeled proteins is more than enough for achieving statistical significance in smFRET. In addition, the remaining single labeled population should not disturb the measurements since they are sorted out with the PIE excitation scheme (see Method for details) [117]. Consistent with TCCD measurements, gel imaging has revealed labeling occurrence for both functionalized fluorophores. However, an unexpected faint red band corresponding to the single mutant CaM T110C was sometimes visible while it should show no reactivity with AF647-alkyne and react exclusively with AF488-maleimide. This could be due to either optical or chemical crosstalk. In fact, it was

previously reported that strain promoted alkyne–azide cycloadditions (SPAAC) can cross-react with thiol groups but, the reaction was shown to be rather inefficient for the specific DIBO chemistry used here, which is pretty selective for azide [167]. Even though a marginal chemical crosstalk cannot be fully rejected, the probability of having dually labeled protein carrying the fluorophores in the opposite positions would be very low.

5.3.1.2 Structure and functionality of selectively dually labeled CaM

Here, we have investigated the effect of our methodological approaches on the three-dimensional structure of CaM by using CD and smFRET measurements. Even though CD gives pretty low resolution information and cannot detect small conformational disarrangement, it can reliably reveal strong structural disturbances [168] while smFRET can detect much more tiny effects. Here, consistently with other studies, we have shown that the structure of the protein remains unaffected by the inserted point mutations, the incorporated functional groups and the subsequent labeling procedure [31, 35, 57]. Even though very small disturbances cannot be excluded, they should be minimal since selective dually labeled proteins were fully functional. Indeed, we have demonstrated the ability of selectively dually labeled CaM to undergo proper conformational changes upon binding of both Ca^{2+} and partners. As expected, smFRET measurements performed with selectively dually labeled CaM at low ionic strength in the presence of 10 mM $CaCl_2$ have revealed multiple conformational substates consistent with what was previously reported by others [78, 143, 169, 170]. Furthermore, the functionality of CaM was demonstrated by performing similar measurements in the presence of partners like CaMKII or M13 binding peptides. This set of experiments were instead performed at physiological ionic strength since it was previously shown that the binding affinity of calmodulin is enhanced by increasing salt concentration [148]. In our results, we have observed that the relative statistical weight of the different holoCaM conformational substates was strongly sensitive to changes in the experimental conditions. In particular, the statistical weight of the low FRET population decreases significantly at physiological ionic strength (i.e. 50 mM MOPS, 10 mM $CaCl_2$, 150 mM KCl) compared to low ionic strength (i.e. 50 mM MOPS, 10 mM $CaCl_2$, 0 mM KCl) and is not visible anymore in the related FRET histograms. This is most probably explained by the effect of the ionic strength on the stability of CaM conformational substates, since it is known that salts affect the stability of protein conformation [171]. In particular, it was reported that CaM stability is strongly influenced by the experimental conditions such as molecular crowding and ionic strength [172].

5.3.2 Co-translational selective dual-labeling by using fluorescent UAAs

Here, by using commercially available precharged tRNAs carrying fluorescently-modified amino acids [32, 33], we have shown the successful incorporation of a fluorophore displaying high photostability characteristics, which is suitable for single molecule detection. Furthermore, we have shown that, in combination to it, a second fluorophore can be co-translationally incorporated leading to sufficient amounts of dually labeled proteins for smFRET. This constitutes a major progress, since until now co-translational incorporated fluorophores used for the selective dual-labeling of proteins were limited to a couple of dyes with inferior photostabilily properties forbidding the use of this elegant technique in smFRET studies. In the present study, we have demonstrated (i) that the co-translational incorporation of dyes is suitable for single molecule detection, (ii) that incorporation of large and charged ATTO fluorophores does not disturb the three-dimensional structure of the protein and (iii) that the combined incorporation of two fluorophores is pertinent for smFRET. The present methodological approach opens therefore new exciting possibilities for smFRET by allowing, for instance, direct real-time measurements of biological processes at single molecule resolution (e.g. co-translational folding).

5.3.2.1 Incorporation of a single fluorophore into CaM

In the present work, we have shown the successful co-translational incorporation of BODIPY-FL, BODIPY 576/589, ATTO633, ATTO488 and Cy5. These dyes differ from each other in their physical and chemical properties (e.g. ATTO633 is large and charged while BODIPY-FL is small and neutral). Beside this, all the dyes were successfully incorporated showing that the translational machinery can accommodate the incorporation of a large range of fluorophores. However, we and others have observed that the incorporation efficiency varies significantly according to the dye and the position chosen [150]. For example, we have observed that the yield of CaM having incorporated ATTO633 at P34 was about 50 nM, whereas it was increased to few hundreds nM for BODIPY-FL at P110. Actually, the fact that BODIPY-FL incorporated into CaM much more efficiently is fully consistent with what was reported in the literature [27, 33]. Even though the incorporation efficiency depends on the fluorophore and the position chosen, the related protein yield should not be a major issue in smFRET, where only few microliters of picomolar protein concentration is required.

5.3.2.1.1 Structure and functionality of selectively dually labeled CaM

Here, we have proved that the co-translational incorporation of ATTO633 affects neither the structure nor the functionality of calmodulin. Firstly, smFRET measurements made with apoCaM T34AmF^{ATTO633} T110C^{AF488} and apoCaM T34K^{ATTO633} T110C^{AF488} were consistent with the reported structure of Zhang *et al*. [80], strongly suggesting that the three-dimensional structure of calmodulin remains unaffected by the incorporated fluorophore. Secondly, we have proved the functionality of CaM T34AmF^{ATTO633} T110C^{AF488} by performing smFRET experiments in the presence of calcium and CaMKII binding peptide. For the holoCaM, the related FRET histogram has revealed, as expected, multiple conformational substates [78, 143, 169, 170]. However, we have also observed that the middle- and high-FRET populations were slightly merging. This is most probably caused by specific electrostatic interactions and favored spatial occupancy of ATTO633, which are encountered in that extremely compact conformation and are not considered by our AV simulations. Despite this, the functionality of CaM T34AmF^{ATTO633} T110C^{AF488} was clearly demonstrated through the reversible Ca^{2+}-CaM-CaMKII complex formation, strongly indicating that the structure of CaM is not affected.

5.3.2.2 Incorporation of two fluorophores into CaM

In the present study, we have further demonstrated that incorporation of an ATTO dye can be performed in combination with the co-translational incorporation of another fluorophore. Here, we have shown, as proof-of-principle, the successful incorporation of two fluorophores, ATTO633 and BODIPY-FL, into CaM. BODIPY-FL was chosen since its incorporation into functional CaM was previously reported [32]. In the present work, ATTO633 was incorporated via an amber stop codon at P34, whereas BODIPY-FL was incorporated by suppressing a cysteine codon at P110. In the aim of increasing both rates of incorporation, a RF1- and Cys-depleted coupled transcription-translation cell-free system was used. However, it appeared that such a system was only partially depleted in both components leading to unwanted basal competitions with remaining RF1s and cysteine residues. Even though it does not significantly disturb the incorporation of UAAs if the latter is efficient, these endogenous competitors can be problematic when fluorescent UAAs need to be incorporated. In fact, both cysteine residue and release factor 1 can generate undesired products. The former will lead to unwanted full length his-tagged proteins whereas the latter can give rise to undesirable truncated proteins that will be, in turn, discarded during Ni-NTA purification. In the specific context of dual incorporation of fluorophores, the competition with endogenous remaining cysteine residues will lead to contaminating single-labeled products while, on the contrary,

the termination of translation by RF1 will not contaminate the sample. For that reason, the small and more competitive BODIPY fluorophore was chosen for being incorporated via a cysteine codon while the incorporation of the large ATTO dye was performed through amber codon suppression. Furthermore, in order to limit the depletion of the cell-free system by truncated products, the construct *CaM T34amber T110C* was preferred to *CaM T34C T110amber*. By using this scheme, we should be able to limit the cell-free production of undesired products and favor, as much as possible, the synthesis of selectively dually labeled CaM.

5.3.2.2.1 Structure of selectively dually labeled CaM

Here, we have further demonstrated that the co-translational incorporation of two fluorophores is relevant for smFRET. To this aim, we have performed smFRET measurements with CaM T34AmF^{ATTO633} T110C$^{BODIPY\text{-}FL}$ for a proof-of-principle. Measurements performed with the holoCaM T34AmF^{ATTO633} T110C$^{BODIPY\text{-}FL}$ were conclusive and our results strongly indicate that the three-dimensional structure of CaM remains unaffected by the co-translational incorporation of ATTO633 and BODIPY-FL at P34 and P110, respectively. On the contrary, the apoCaM T34AmF^{ATTO633} T110C$^{BODIPY\text{-}FL}$ could not be properly measured in smFRET due to strong local quenching of the donor fluorophore in that specific conformation leading to a significant drop in the quantum yield of the dye (**Supplementary Table 3** and **Supplementary Table 4)**. The consequent decrease in FRET efficiencies has prevent proper detection of FRET events already affected by the low transmission efficiency associated with the acceptor (**Supplementary Table 5**). For these reasons, we were not able to verify the functionality of the selectively dually labeled CaM. However, the functionality should not be drastically affected since, on the one hand, we have demonstrated that the unique incorporation of large ATTO dyes does not significantly disturb neither the structure nor the functionality (see point 5.3.2.1.1) and, on the other hand, incorporation of small BODIPY fluorophores into CaM was already reported by others [32]. Moreover, the successful incorporation of two distinct BODIPY dyes into CaM has been previously demonstrated suggesting that multiple co-translational incorporations of fluorophores into proteins can be achieved without disturbing their functionality [33].

5.4 Selective labeling for more accurate smFRET

In the present study, we have demonstrated that selective dual-labeling leads to more accurate FRET histograms allowing more precise data analysis. Actually, it was already previously suggested that dual-labeling via double cysteine mutagenesis could cause further broadening of FRET histograms [31]. In our results, the related broadening is not very pronounced. Indeed, here, we have analyzed two labeling positions that were previously selected by others for smFRET studies performed with non-selectively labeled CaM [35, 127-129]. Therefore, it did not show a large difference between the respective quantum yields of the fluorophores at the two positions (**Table 11**). Yet, we can see that a slight difference of the quantum yields of the acceptor (i.e. 0.27 and 0.36) between the two labeled positions (i.e. 34 and 110) have led to a significant broadening. If other positions associated with higher differences in quantum yield are chosen, the present broadening will be much more obvious. In such situations, our selective approaches would be even more necessary. Additionally, when a non-selective dual-labeling scheme is used, positions displaying very strong local quenching for the attached fluorophore are usually discarded as they result in strongly broadened FRET histograms. However, for some specific applications such as FRET-based structural modeling, they are required for a complete mapping of molecular distances [156, 157]. In such case, our selective approach could be a very interesting alternative.

5.5 CaM: a model protein

CaM is a very important protein which is intensively studied. For that reason, it was an ideal model protein for the present work. Additionally, CaM is a particularly interesting candidate for smFRET studies due to its multiple recognition modes, conformational substates and particular dynamics [72, 76, 78]. Nowedays, the binding of CaM to its very broad range of partners is still not fully understood. It is thought that the co-existence of multiple conformational substates is necessary for CaM functionality [81]. Furthermore, it was suggested that the protein samples a quasi-continuous spectrum of conformational substates which explains its multiple recognition modes [82, 83]. Years ago, multiple conformational substates of CaM were reported in smFRET by the group of Johnson partially demonstrating the stated hypothesis [78, 143]. In the present work, our results further confirmed the co-existence of multiple conformational substates for the holoCaM in solution. Interestingly, we have also shown that the binding of CaM to CaMKII and M13 binding peptides leads to the

redistribution of the whole population towards a unique compact form in line with the sampling of a large spectrum of conformational substates by CaM.

5.6 Conclusions

With the present work, we report two fast, robust and simple methodological approaches for the selective dual-labeling of cell-free synthesized proteins for smFRET. Here, we have shown with the case-study of human calmodulin (i) that the incorporation of UAAs into proteins by expansion of the genetic code is easy to handle and allows efficient selective dual-labeling (ii) that the approach is biologically relevant and leads to functional proteins and (iii) that selective dual-labeling allows more accurate data analysis of smFRET measurements, a requirement for precise molecular distance determination. By combining smFRET and CFPS, our methodological approaches open new interesting possibilities by allowing fast, simple and efficient selective dual-labeling of cell-free synthesized proteins for more precise measurements and expanding the application of smFRET to the large family of difficult-to-express proteins.

5.7 Outlook

In the present work, the two methodological approaches for the selective dual-labeling of cell-free synthesized proteins present a couple of limitations. Firstly, both approaches require the use of cysteine residues for labeling. This greatly limits the application to proteins with no or a limited number of non-essential cysteine residues. For the latter, it requires, in addition, site-directed mutagenesis for the removal of endogenous cysteine residues as well as careful verifications of its effect on the three-dimensional structure and functionality of the protein of interest. In the future, our approaches could be extended to a larger range of proteins by performing the incorporation of two different UAAs via two distinct unique codons either non-sense (e.g. ochre and amber) or multi-base artificial codons (e.g. 4-base and 5-base codons) or a combination of both. Secondly, the modified coupled transcription-translation cell-free system is originated from *E.coli* where post-translational modifications (PTMs) are largely restricted. This strongly restricts our approaches to proteins which require no or limited post-translational modifications, non-required for the functionality. Our approaches could therefore gain in interest from the development of similarly modified eukaryotic cell-free systems. However, the currently commercially available eukaryotic systems are associated with pretty low yields, which could be problematic when the incorporation of

UAAs is needed [17]. Thirdly, while dual-labeling by post-translational chemical modifications is accessible to a broad range of dyes, the co-translational incorporation of fluorophores remains limited to few of them. Here, by proving the successful incorporation of ATTO633 into CaM, we have opened new interesting possibilities for smFRET. Still, for the generalization of the approach, it has to be proved that it is applicable to other dyes displaying high photostability characteristics. Here, we have demonstrated that the incorporation of ATTO633 can take place in combination to the incorporation of the lowly photostable BODIPY-FL which is a non-suitable partner for smFRET. Yet, selective dual-labeling of proteins by incorporation of two fluorescent UAAs has to be demonstrated for a suitable dye pair (e.g. ATTO633- ATTO488). This was not possible in the present work since we did not have enough amount of suitable donor for it (the commercial production of precharged tRNAs carrying ATTO488 has stopped during the course of the study). Lastly, our methodological approaches were designed for the small field of smFRET studies and lack generality otherwise. This prevents greatly their applications in multimethodology research which would use smFRET methods among others. Here, the main disadvantage is linked to the limited protein yield of CFPS. However, major advancements are continuously reported in the literature and let us hope that CFPS could rapidly achieve protein yields much more suitable for its general application.

Summary

Single-molecule Förster Resonance Energy Transfer (smFRET) is a peerless tool for studying molecular structures, dynamics, functions and interactions of biological macromolecules in heterogeneous systems. Nowadays, the technique is mainly limited to *in vivo* expressed proteins. However, its combination with cell-free protein synthesis (CFPS) appears to be much more advantageous and even, to be necessary in certain situations. For instance, it could be indispensable for studying candidates that required cell-free expression such as difficult-to-express proteins that are reluctant to *in vivo* expression.

In general, smFRET studies necessitate protein samples that are site-specifically dually labeled with a pair of donor and acceptor fluorophores. The common strategy based on double cysteine mutagenesis requires relatively large amounts of proteins and complex procedures that are hardly compatible with CFPS. Therefore, this approach is restricted to *in vivo* expressed proteins.

Here, we present two methodological approaches for the selective dual-labeling of cell-free synthesized proteins which take advantage of an expanded genetic code for the incorporation of unnatural amino acids (UAAs) into proteins via cysteine and amber codon suppression. The first approach provides proteins with two unique functional groups, a sulfhydryl-group and an azido-group, carried by a site-specifically incorporated cysteine residue and a p-azido-L-phenylalanine (AzF), a specific UAA, which allows selective dual-labeling by post-translational chemical modifications thanks to a bioorthogonal chemistry. The second approach takes advantage of the site-specific incorporation of fluorescent UAAs into proteins for selective dual-labeling by the co-translationally inserted fluorophores.

By using a case-study of human calmodulin (hCaM), we demonstrate (i) the viability of both strategies for producing suitable samples for smFRET, (ii) the biologically relevance of the approaches by proving the structural integrity and full functionality of hCaM and (iii) that selective dual-labeling permits the highest accuracy in data analysis.

In summary, with the present thesis, we report two fast, robust and simple methodological approaches that achieve optimal selective dual-labeling of cell-free synthesized proteins for smFRET. Once available for the scientific community, they will offer new interesting possibilities for smFRET studies, for instance, by expanding the application of the technique to the large family of difficult-to-express proteins.

Zusammenfassung

Einzelmolekül Förster Resonanz Energie Transfer (smFRET) ist ein einzigartiges Werkzeug zur Untersuchung molekularer Strukturen, Dynamik, Funktion und Interaktionen biologischer Makromoleküle in heterogenen Systemen. Heutzutage ist diese Technik meistens auf *in vivo* synthetisierte Proteine beschränkt. Jedoch erscheint die Kombination dieser Technik mit zellfreier Protein Synthese (CFPS) sehr vorteilhaft und sogar nötig in bestimmten Situationen. Zum Beispiel kann zellfreie Expression unabdingbar sein, wenn Proteine untersucht werden sollen, die nur durch zellfreie Synthese hergestellt werden können, weil sie *in vivo* nur schlecht synthetisierbar sind.

Im Allgemeinen sind für smFRET Untersuchungen Proteinproben nötig, die mit einem Paar aus Donor- und Akzeptor-Fluoreszenzfarbstoffen ortsspezifisch doppelt markiert sind. Die gebräuchlichste Strategie zum Markieren der Proteine basiert auf zweifachen Cystein-Varianten. Sie benötigt große Proteinmengen und beinhaltet komplexe Reinigungsstrategien, die mit CFPS nicht gut vereinbar sind. Daher ist diese Markierungsstrategie auf *in vivo* synthetisierte Proteine beschränkt.

Hier präsentieren wir zwei methodische Ansätze zur selektiven Doppelmarkierung von zellfrei synthetisierten Proteinen. Diese Methoden nutzen einen erweiterten genetischen Kode zur Einführung von unnatürlichen Aminosäuren (UAA) in Proteine durch die Unterdrückung des Cystein- und Amber-Stop-Kodons. Der erste Ansatz liefert Proteine mit zwei einzigartigen funktionellen Gruppen: der Sulfhydryl-Gruppe eines ortsspezifisch eingeführten Cysteins und eine Azido-Gruppe an p-Azido-L-Phenylalanin (AzF), einer speziellen UAA. Diese funktionellen Gruppen erlauben die selektive Doppelmarkierung durch post-translationale Modifizierung mittels bioorthogonaler Chemie. Beim zweiten Ansatz werden direkt kotranslational ortsspezifisch fluoreszente UAAs in das Protein zur Doppelmarkierung eingeführt.

Anhand des humanen Calmodulins (hCaM) als Modelprotein zeigen wir hier (i) die Realisierbarkeit beider Strategien zur Produktion geeigneter Proben für smFRET, (ii) die biologische Relevanz der Ansätze, indem wir die strukturelle Integrität und volle Funktionsfähigkeit von hCaM zeigen und (iii) das die selektive Doppelmarkierung die höchste Genauigkeit der Datenanalyse ermöglicht.

Zusammengefasst präsentieren wir mit dieser Arbeit zwei schnelle, robuste und einfache methodische Ansätze zur selektiven Doppelmarkierung von zellfrei synthetisierten Proteinen für smFRET. Sobald diese Methoden der Wissenschaftsgemeinschaft zur Verfügung stehen,

werden sie neue interessante Möglichkeiten für smFRET Studien, beispielsweise die Ausweitung dieser Technik auf die große Gruppe der schwer in vivo zu synthetisierenden Proteine, eröffnen.

Bibliography

1. Crick, F., *Central dogma of molecular biology.* Nature, 1970. **227**(5258): p. 561-563.
2. Kaczanowska, M. and M. Rydén-Aulin, *Ribosome biogenesis and the translation process in Escherichia coli.* Microbiology and Molecular Biology Reviews, 2007. **71**(3): p. 477-494.
3. Marintchev, A. and G. Wagner, *Translation initiation: structures, mechanisms and evolution.* Quarterly reviews of biophysics, 2004. **37**(3-4): p. 197-284.
4. Wintermeyer, W., et al., *Mechanisms of elongation on the ribosome: dynamics of a macromolecular machine.* Biochemical Society Transactions 2004. **32** (5): p. 733-737.
5. Ramakrishnan, V., *Ribosome structure and the mechanism of translation.* Cell, 2002. **108**(4): p. 557-572.
6. Fedorov, A.N. and T.O. Baldwin, *Cotranslational protein folding.* Journal of Biological Chemistry, 1997. **272**(52): p. 32715-32718.
7. Kim, S.J., et al., *Translational tuning optimizes nascent protein folding in cells.* Science, 2015. **348**(6233): p. 444-448.
8. Kramer, G., et al., *The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins.* Nature structural & molecular biology, 2009. **16**(6): p. 589-597.
9. Kramer, G., V. Ramachandiran, and B. Hardesty, *Cotranslational folding—omnia mea mecum porto?* The international journal of biochemistry & cell biology, 2001. **33**(6): p. 541-553.
10. Yamamoto, H., et al., *EF-G and EF4: translocation and back-translocation on the bacterial ribosome.* Nature Reviews Microbiology, 2014. **12**(2): p. 89-100.
11. Cooper, G.M. and R.E. Hausman, *The cell*. Vol. 85. 2000: Sinauer Associates Sunderland.
12. Holtkamp, W., et al., *Cotranslational protein folding on the ribosome monitored in real time.* Science, 2015. **350**(6264): p. 1104-1107.
13. Zhang, G. and Z. Ignatova, *Folding at the birth of the nascent chain: coordinating translation with co-translational folding.* Current opinion in structural biology, 2011. **21**(1): p. 25-31.
14. Nirenberg, M.W. and J.H. Matthaei, *The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides.* Proceedings of the National Academy of Sciences, 1961. **47**(10): p. 1588-1602.
15. Buchner, E., *Alkoholische Gärung ohne Hefezellen.* European Journal of Inorganic Chemistry, 1897. **30**(1): p. 1110-1113.
16. Rosenblum, G. and B.S. Cooperman, *Engine out of the chassis: Cell-free protein synthesis and its uses.* FEBS letters, 2014. **588**(2): p. 261-268.
17. Zemella, A., et al., *Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems.* ChemBioChem, 2015. **16**(17): p. 2420-2431.
18. Carlson, E.D., et al., *Cell-free protein synthesis: applications come of age.* Biotechnology advances, 2012. **30**(5): p. 1185-1194.
19. Soni, R.K. and M.-F.B. Tsou, *A Cell-Free System for Real-Time Analyses of Centriole Disengagement and Centriole-to-Centrosome Conversion.* Methods in molecular biology (Clifton, NJ) 2016. **1413**: p. 197-206.
20. Katzen, F., G. Chang, and W. Kudlicki, *The past, present and future of cell-free protein synthesis.* Trends in biotechnology, 2005. **23**(3): p. 150-156.

21. de Pouplana, L.R., *The genetic code and the origin of life*. 2007: Springer Science & Business Media.
22. Wang, L., J. Xie, and P.G. Schultz, *Expanding the genetic code.* Annu. Rev. Biophys. Biomol. Struct., 2006. **35**: p. 225-249.
23. Liu, W., et al., *Genetic incorporation of unnatural amino acids into proteins in mammalian cells.* Nature methods, 2007. **4**(3): p. 239.
24. Hancock, S.M., et al., *Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair.* Journal of the American Chemical Society, 2010. **132**(42): p. 14819-14824.
25. Ryu, Y. and P.G. Schultz, *Efficient incorporation of unnatural amino acids into proteins in Escherichia coli.* Nature methods, 2006. **3**(4): p. 263.
26. Lin, M.Z. and L. Wang, *Selective labeling of proteins with chemical probes in living cells.* Physiology, 2008. **23**(3): p. 131-141.
27. Quast, R.B., et al., *Cotranslational incorporation of non-standard amino acids using cell-free protein synthesis.* FEBS letters, 2015. **589**(15): p. 1703-1712.
28. Davis, L. and J.W. Chin, *Designer proteins: applications of genetic code expansion in cell biology.* Nature reviews. Molecular cell biology, 2012. **13**(3): p. 168.
29. Noren, C.J., et al., *A general method for site-specific incorporation of unnatural amino acids into proteins.* Science, 1989. **244**(4901): p. 182.
30. Brustad, E.M., et al., *A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer.* Journal of the American Chemical Society, 2008. **130**(52): p. 17664-17665.
31. Seo, M.-H., et al., *Efficient single-molecule fluorescence resonance energy transfer analysis by site-specific dual-labeling of protein using an unnatural amino acid.* Analytical chemistry, 2011. **83**(23): p. 8849-8854.
32. Kajihara, D., et al. , *FRET analysis of proteinconformational change through position-specific incorporation of fluorescent aminoacids.* Nature Methods 2006. **3**(11): p. 923-929.
33. Iijima, I. and T. Hohsaka, *Position-Specific Incorporation of Fluorescent Non-natural Amino Acids into Maltose-Binding Protein for Detection of Ligand Binding by FRET and Fluorescence Quenching.* ChemBioChem, 2009. **10**(6): p. 999-1006.
34. Aerni H. R. , et al., *Revealing the amino acid composition of proteins within an expanded genetic code.* . Nucleic acids research, 2015. **43**(2).
35. Allen, M.W., et al., *Fluorescence labeling, purification, and immobilization of a double cysteine mutant calmodulin fusion protein for single-molecule experiments.* Analytical biochemistry 2004. **325**(2): p. 273-284.
36. Bartlett, A.I. and S.E. Radford, *An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms.* Nature structural & molecular biology, 2009. **16**(6): p. 582-588.
37. Bernhard, F. and Y. Tozawa, *Cell-free expression—making a mark.* Current opinion in structural biology, 2013. **23**(3): p. 374-380.
38. Short, G.F., S.Y. Golovine, and S.M. Hecht, *Effects of release factor 1 on in vitro protein translation and the elaboration of proteins containing unnatural amino acids.* Biochemistry, 1999. **38**(27): p. 8808-8819.
39. Hong, S.H., et al., *Cell-free protein synthesis from a release factor 1 deficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation.* ACS synthetic biology, 2014. **3**(6): p. 398-409.
40. Gerrits, M., et al., *Cell-free synthesis of defined protein conjugates by site-directed cotranslational labeling*, in *Madame Curie Bioscience Database* 2013, Landes Bioscience.

41. Medintz, I.L. and N. Hildebrandt, *FRET-Förster resonance energy transfer: from theory to applications*. 2013: John Wiley & Sons.
42. Johnson-Buck, A.E., M.R. Blanco, and N.G. Walter, *Single-molecule fluorescence resonance energy transfer*, in *Encyclopedia of Biophysics*. 2013, Springer. p. 2329-2335.
43. Lakowicz, J.R. and B.R. Masters, *Principles of fluorescence spectroscopy.* Journal of Biomedical Optics, 2008. **13**(2): p. 029901.
44. Schuler, B. and W.A. Eaton, *Protein folding studied by single-molecule FRET.* Current opinion in structural biology, 2008. **18**(1): p. 16-26.
45. Hussain, S.A., *An introduction to fluorescence resonance energy transfer (FRET).* arXiv preprint arXiv:0908.1815, 2009.
46. Michalet, X., S. Weiss, and M. Jäger, *Single-molecule fluorescence studies of protein folding and conformational dynamics.* Chemical reviews, 2006. **106**(5): p. 1785-1813.
47. Deniz, A.A., et al., *Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Förster distance dependence and subpopulations.* Proceedings of the National Academy of Sciences, 1999. **96**(7): p. 3670-3675.
48. Ha, T., *Single-molecule fluorescence resonance energy transfer*. Methods, 2001. **25**(1): p. 78-86.
49. Sustarsic, M. and A.N. Kapanidis, *Taking the ruler to the jungle: single-molecule FRET for understanding biomolecular structure and dynamics in live cells.* Current opinion in structural biology, 2015. **34**: p. 52-59.
50. Ha, T., *Single-molecule methods leap ahead.* Nature methods, 2014. **11**(10): p. 1015-1018.
51. Kapanidis, A.N. and T. Strick, *Biology, one molecule at a time.* Trends in biochemical sciences, 2009. **34**(5): p. 234-243.
52. Weiss, S., *Fluorescence spectroscopy of single biomolecules.* Science, 1999. **283**(5408): p. 1676-1683.
53. Kapanidis, A.N., et al., *Alternating-laser excitation of single molecules.* Accounts of chemical research, 2005. **38**(7): p. 523-533.
54. Obermaier, C., A. Griebel, and R. Westermeier, *Principles of protein labeling techniques.* Methods Mol. Biol., 2015. **1295**: p. 153-165.
55. Krall, N., et al., *Site-selective protein-modification chemistry for basic biology and drug development.* Nature chemistry, 2016. **8**(2): p. 103.
56. Kim, Y., et al., *Efficient site-specific labeling of proteins via cysteines.* Bioconjugate chemistry, 2008. **19**(3): p. 786-791.
57. Kim, J., et al. , *Simple and efficient strategy for site-specific dual labeling of proteins for single-molecule fluorescence resonance energy transfer analysis.* Analytical chemistry 2013. **85**(3): p. 1468-1474.
58. Gautier, A. and M.J. Hinner, *Site-Specific Protein Labeling: Methods and protocols*. Vol. 1266. 2015: Springer Science+Business Media.
59. Hohsaka, T. and M. Sisido, *Incorporation of non-natural amino acids into proteins.* Current opinion in chemical biology, 2002. **6**(6): p. 809-815.
60. Perronet, K., et al., *Single molecule fluorescence detection of BODIPY-FL molecules for monitoring protein synthesis.* J. Lumin., 2007. **127**(1): p. 264-268.
61. Dulin, D., et al., *Reduced photobleaching of BODIPY-FL.* Phys. Procedia, 2010. **3**(4): p. 1563-1567.
62. Hohsaka, T., et al., *Efficient incorporation of nonnatural amino acids with large aromatic groups into streptavidin in in vitro protein synthesizing systems.* J. Am. Chem. Soc., 1999. **121**(1): p. 34-40.

63. Doi, Y., et al., *Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system.* J. Am. Chem. Soc., 2007. **129**(46): p. 14458-14462.
64. Mittelstaet, J., A.L. Konevega, and M.V. Rodnina, *A kinetic safety gate controlling the delivery of unnatural amino acids to the ribosome.* J. Am. Chem. Soc., 2013. **135**(45): p. 17031-17038.
65. Gong, Y. and L. Pan, *Recent advances in bioorthogonal reactions for site-specific protein labeling and engineering.* Tetrahedron Letters, 2015. **56**(17): p. 2123-2132.
66. Lang, K. and J.W. Chin, *Bioorthogonal reactions for labeling proteins.* ACS Chem. Biol., 2014. **9**(1): p. 16-20.
67. Wu, B., et al., *Catalyst-Free and Site-Specific One-Pot Dual-Labeling of a Protein Directed by Two Genetically Incorporated Noncanonical Amino Acids.* ChemBioChem, 2012. **13**(10): p. 1405-1408.
68. Goerke, A.R. and J.R. Swartz, *High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids.* Biotechnology and bioengineering, 2009. **102**(2): p. 400-416.
69. Uzagare, M.C., et al. , *A Novel Method for the Labelling of Peptides and Proteins through a Bioorthogonal Staudinger Reaction by Using 2-Cyanoethyl Phosphoramidites.* ChemBioChem 2012. **13**(15): p. 2204-2208.
70. Cui, Z., et al., *Combining sense and nonsense codon reassignment for site-selective protein modification with unnatural amino acids.* ACS synthetic biology, 2016. **6**(3): p. 535-544.
71. Van Eldik, L.J. and D.M. Watterson, *Calmodulin and signal transduction*. 1998: Gulf Professional Publishing.
72. Tidow, H., and Poul Nissen., *Structural diversity of calmodulin binding to its target sites.* Febs Journal 2013. **280**(21): p. 5551-5565.
73. Gao, J., et al., *Size-dependent impact of CNTs on dynamic properties of calmodulin.* Nanoscale, 2014. **6**(21): p. 12828-12837.
74. Cheung, W.Y., *Calcium and cell function*. Vol. 3. 2013: Academic Press.
75. Potter, J.D., et al., *[12] Ca 2+ binding to calmodulin.* Methods in enzymology, 1983. **102**: p. 135-143.
76. Ådén, J.r. and P. Wittung-Stafshede, *Folding of an unfolded protein by macromolecular crowding in vitro.* Biochemistry, 2014. **53**(14): p. 2271-2277.
77. Nussinov, R. and G. Schreiber, *Computational protein-protein interactions*. 2009: CRC Press.
78. Johnson, C.K., et al., *Single-molecule fluorescence spectroscopy: new probes of protein function and dynamics.* Physiology, 2005. **20**(1): p. 10-14.
79. Chattopadhyaya, R., et al., *Calmodulin structure refined at 1.7 Å resolution.* Journal of molecular biology, 1992. **228**(4): p. 1177-1192.
80. Zhang, M., T. Tanaka, and M. Ikura, *Calcium-induced conformational transition revealed by the solution structure of apo calmodulin.* Nature Structural & Molecular Biology, 1995. **2**(9): p. 758-767.
81. Fallon, J.L. and F.A. Quiocho, *A closed compact structure of native Ca 2+-calmodulin.* Structure, 2003. **11**(10): p. 1303-1307.
82. Gsponer, J., et al., *A coupled equilibrium shift mechanism in calmodulin-mediated signal transduction.* Structure, 2008. **16**(5): p. 736-746.
83. Wilson, M.A. and A.T. Brunger, *The 1.0 Å crystal structure of Ca 2+-bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity.* Journal of molecular biology, 2000. **301**(5): p. 1237-1256.

84. Price, E.S., M.S. DeVore, and C.K. Johnson, *Detecting intramolecular dynamics and multiple Forster resonance energy transfer states by fluorescence correlation spectroscopy.* The Journal of Physical Chemistry B, 2010. **114**(17): p. 5895-5902.
85. Ramanathan, A., et al., *Protein conformational populations and functionally relevant substates.* Accounts of chemical research, 2013. **47**(1): p. 149-156.
86. Barbato, G., et al., *Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible.* Biochemistry, 1992. **31**(23): p. 5269-5278.
87. Seaton, B., et al., *Calcium-induced increase in the radius of gyration and maximum dimension of calmodulin measured by small-angle X-ray scattering.* Biochemistry, 1985. **24**(24): p. 6740-6743.
88. Hinsen, K. and R.C. Sadron, *MMTK user's guide.* Centre de Biophysique Moleculaire, CNR Orleans, France, 2002.
89. Pray, L., *The biotechnology revolution: PCR and the use of reverse transcriptase to clone expressed genes.* Nature Education, 2008. **1**(1): p. 94.
90. Old, R.W. and S.B. Primrose, *Principles of gene manipulation: an introduction to genetic engineering*. Vol. 2. 1981: Univ of California Press.
91. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proceedings of the national academy of sciences, 1977. **74**(12): p. 5463-5467.
92. Block, H., et al., *Immobilized-metal affinity chromatography (IMAC): a review.* Methods in enzymology, 2009. **463**: p. 439-473.
93. Kim, D.M. and C.Y. Choi, *A semicontinuous prokaryotic coupled transcription/translation system using a dialysis membrane.* Biotechnology progress, 1996. **12**(5): p. 645-649.
94. Chin, J.W., et al., *Addition of p-Azido-l-phenylalanine to the Genetic Code of Escherichia c oli.* Journal of the American Chemical Society, 2002. **124**(31): p. 9026-9027.
95. Mbua, N.E., et al., *Strain-Promoted Alkyne–Azide Cycloadditions (SPAAC) Reveal New Features of Glycoconjugate Biosynthesis.* ChemBioChem, 2011. **12**(12): p. 1912-1921.
96. Mori, S. and H.G. Barth, *Size exclusion chromatography*. 2013: Springer Science & Business Media.
97. Fraga, D., T. Meulia, and S. Fenster, *Current protocols: essential laboratory techniques*. 2008, Hoboken: John Wiley & Sons.
98. Gravel, P. and O. Golaz, *Protein blotting by the semidry method*, in *The protein protocols handbook*. 2009, Humana Press. p. 249-260.
99. Grimsley, G.R. and C.N. Pace, *Spectrophotometric determination of protein concentration.* Current protocols in protein science, 2004: p. 3.1. 1-3.1. 9.
100. Mach, H., C.R. Middaugh, and N. Denslow, *Determining the identity and purity of recombinant proteins by UV absorption spectroscopy.* Current Protocols in Protein Science, 2001: p. 7.2. 1-7.2. 21.
101. Schmidt, W., *Optical spectroscopy in chemistry and life sciences*. 2005: Wiley-VCH.
102. Kelly, S.M. and N.C. Price, *Circular Dichroism: Studies of Proteins*, in *Encyclopedia of life science*. 2009, John Wiley & Sons. p. 1-6.
103. Wei, Y., A.A. Thyparambil, and R.A. Latour, *Protein helical structure determination using CD spectroscopy for solutions with strong background absorbance from 190 to 230nm.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2014. **1844**(12): p. 2331-2337.

104. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism.* Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 2005. **1751**(2): p. 119-139.
105. Lakowicz, J.R., *Quenching of fluorescence*, in *Principles of fluorescence spectroscopy*. 1983, Springer. p. 257-301.
106. Würth, C., et al., *Relative and absolute determination of fluorescence quantum yields of transparent samples.* Nature protocols, 2013. **8**(8): p. 1535-1550.
107. Lakowicz, J.R., *Principles of fluorescence spectroscopy*. 2013: Springer Science & Business Media.
108. Norda, C., *Single Molecule FRET Studies on Folding States of the Phosphoglycerate Kinase. Matser's Thesis.* 2014, RWTH Aachen Universtity.
109. Groll, J. and M. Moeller, *Surface Passivation for Single Molecule Detection*, in *Encyclopedia of Biophysics*. 2013, Springer. p. 2531-2536.
110. Schlapak, R., et al., *Selective protein and DNA adsorption on PLL-PEG films modulated by ionic strength.* Soft Matter, 2009. **5**(3): p. 613-621.
111. Dean, P.N., *Confocal Microscopy: Principles and Practices*, in *Current Protocols in Cytometry*. 2001, John Wiley & Sons, Inc.
112. Sisamakis, E., et al., *Accurate single-molecule FRET studies using multiparameter fluorescence detection.* Methods in enzymology, 2010. **475**: p. 455-514.
113. Schoene, A., *Etablierung eines Systems aus Cysteinmutanten der Phosphoglycerat-Kinase für Entfaltungsstudien mit Einzelmolekül-FRET. PhD Thesis.* 2017, RWTH Aachen Universtity.
114. Haustein, E. and P. Schwille, *Fluorescence correlation spectroscopy: novel variations of an established technique.* Annu. Rev. Biophys. Biomol. Struct., 2007. **36**: p. 151-169.
115. Kapusta, P., *Absolute diffusion coefficients: compilation of reference data for FCS calibration.* Application note (PicoQuant GmbH, Berlin), 2010.
116. Lakowicz, J.R., *Principles of Fluorescence Spectroscopy, 3rd edition* 2006: Springer Science+Business.
117. Müller, B.K., et al., *Pulsed interleaved excitation.* Biophysical journal, 2005. **89**(5): p. 3508-3522.
118. Campos, L.A., et al., *A photoprotection strategy for microsecond-resolution single-molecule fluorescence spectroscopy.* Nature methods, 2011. **8**(2): p. 143-146.
119. Gabba, M., et al., *Conformational state distributions and catalytically relevant dynamics of a hinge-bending enzyme studied by single-molecule FRET and a coarse-grained simulation.* Biophysical journal, 2014. **107**(8): p. 1913-1923.
120. Fries, J.R., et al., *Quantitative identification of different single molecules by selective time-resolved confocal fluorescence spectroscopy.* The Journal of Physical Chemistry A, 1998. **102**(33): p. 6601-6613.
121. Selvin, P. and T. Ha, *Single-Molecule Techniques: A Laboratory Manual*. 2008: Cold Spring Harbor Laboratory Press.
122. Kempe, D., et al., *Accurate Fluorescence Quantum Yield Determination by Fluorescence Correlation Spectroscopy.* The Journal of Physical Chemistry B, 2015. **119**(13): p. 4668-4672.
123. Buschmann, V., et al., *Quantitative FCS: determination of the confocal volume by FCS and bead scanning with the microtime 200.* Application note (PicoQuant GmbH, Berlin), 2009.
124. Höfig, H., et al., *Inter-dye distance distributions studied by a combination of single-molecule FRET-filtered lifetime measurements and a weighted accessible volume (wAV) algorithm.* Molecules, 2014. **19**(12): p. 19269-19291.

125. Li, H., et al., *Ultrasensitive coincidence fluorescence detection of single DNA molecules.* Analytical chemistry, 2003. **75**(7): p. 1664-1670.
126. Slaughter, B.D., et al., *Mechanism of calmodulin recognition of the binding domain of isoform 1b of the plasma membrane Ca2+-ATPase: kinetic pathway and effects of methionine oxidation.* Biochemistry, 2007. **46**(13): p. 4045-4054.
127. Török, K., et al., *Dual effect of ATP in the activation mechanism of brain Ca2+/calmodulin-dependent protein kinase II by Ca2+/calmodulin.* Biochemistry, 2001. **40**(49): p. 14878-14890.
128. Putkey, J.A., et al., *Acidic/IQ motif regulator of calmodulin.* Journal of Biological Chemistry, 2008. **283**(3): p. 1401-1410.
129. Xiong, L., et al., *Sites on calmodulin that interact with the C-terminal tail of Cav1. 2 channel.* Journal of Biological Chemistry, 2005. **280**(8): p. 7070-7079.
130. Bahloul, A., et al., *The unique insert in myosin VI is a structural calcium–calmodulin binding site.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(14): p. 4787-4792.
131. Hillebrecht, J.R. and S. Chong, *A comparative study of protein synthesis in in vitro systems: from the prokaryotic reconstituted to the eukaryotic extract-based.* BMC biotechnology, 2008. **8**(1): p. 58.
132. Li, J., et al., *Improved cell-free RNA and protein synthesis system.* PloS one, 2014. **9**(9): p. e106232.
133. Stigler, J. and M. Rief, *Calcium-dependent folding of single calmodulin molecules.* Proceedings of the National Academy of Sciences, 2012. **109**(44): p. 17814-17819.
134. Junker, J.P., F. Ziegler, and M. Rief, *Ligand-dependent equilibrium fluctuations of single calmodulin molecules.* Science, 2009. **323**(5914): p. 633-637.
135. Lakowski, T.M., et al., *Calcium-induced folding of a fragment of calmodulin composed of EF-hands 2 and 3.* Protein science, 2007. **16**(6): p. 1119-1132.
136. Raussens, V., J.-M. Ruysschaert, and E. Goormaghtigh, *Protein concentration is not an absolute prerequisite for the determination of secondary structure from circular dichroism spectra: a new scaling method.* Analytical biochemistry, 2003. **319**(1): p. 114-121.
137. Reed, J. and T.A. Reed, *A set of constructed type spectra for the practical estimation of peptide secondary structure from circular dichroism.* Analytical biochemistry, 1997. **254**(1): p. 36-40.
138. Yang, J.T., C.-S.C. Wu, and H.M. Martinez, *[11] Calculation of protein conformation from circular dichroism.* Methods in enzymology, 1986. **130**: p. 208-269.
139. Meador, W.E., A.R. Means, and F.A. Quiocho, *Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures.* Science-AAAS-Weekly Paper Edition-including Guide to Scientific Information, 1993. **262**(5140): p. 1718-1720.
140. Ikura, M. and M. Clore, *Solution structure of a calmodulin-target peptide complex by multidimensional NMR.* Science, 1992. **256**(5057): p. 632.
141. Sindbert, S., et al., *Accurate distance determination of nucleic acids via Forster resonance energy transfer: implications of dye linker length and rigidity.* Journal of the American Chemical Society, 2011. **133**(8): p. 2463-2480.
142. Rosenkranz, T., *Time-Resolved Single Molecule FRET Studies on Folding/Unfolding Transitions and on Functional Conformational Changes of Phosphoglycerate Kinase. PhD Thesis*. 2011, Heinrich-Heine-Universität Düsseldorf.
143. Priddy, T.S., et al., *Single molecule analyses of the conformational substates of calmodulin bound to the phosphorylase kinase complex.* Protein science, 2007. **16**(6): p. 1017-1023.

144. Price, E.S., M. Aleksiejew, and C.K. Johnson, *FRET-FCS detection of intralobe dynamics in calmodulin.* The Journal of Physical Chemistry B, 2011. **115**(29): p. 9320-9326.
145. Shepherd, C.M. and H.J. Vogel, *A molecular dynamics study of Ca 2+-calmodulin: evidence of interdomain coupling and structural collapse on the nanosecond timescale.* Biophysical journal, 2004. **87**(2): p. 780-791.
146. Yang, C., G.S. Jas, and K. Kuczera, *Structure and dynamics of calcium-activated calmodulin in solution.* Journal of Biomolecular Structure and Dynamics, 2001. **19**(2): p. 247-271.
147. Allen M. W. , Urbauer R. J. , and J.C. K., *Single-molecule assays of calmodulin target binding detected with a calmodulin energy-transfer construct.* Analytical chemistry, 2004. **76**(13): p. 3630-3637.
148. André, I., et al., *Salt enhances calmodulin-target interaction.* Biophysical journal, 2006. **90**(8): p. 2903-2910.
149. Ogawa, Y. and M. TANOKURA, *Calcium binding to calmodulin: effects of ionic strength, Mg2+, pH and temperature.* The Journal of Biochemistry, 1984. **95**(1): p. 19-28.
150. Abe, R., et al., *Incorporation of fluorescent non-natural amino acids into N-terminal tag of proteins in cell-free translation and its dependence on position and neighboring codons.* Journal of bioscience and bioengineering, 2010. **110**(1): p. 32-38.
151. Berney, C. and G. Danuser, *FRET or no FRET: a quantitative comparison.* Biophysical journal, 2003. **84**(6): p. 3992-4010.
152. Moerner, W. and M. Orrit, *Illuminating single molecules in condensed matter.* Science, 1999. **283**(5408): p. 1670-1676.
153. Brasselet, S. and W. Moerner, *Fluorescence Behavior of Single-Molecule pH-Sensors.* Single Molecules, 2000. **1**(1): p. 17-23.
154. Hwang, H. and S. Myong, *Protein induced fluorescence enhancement (PIFE) for probing protein–nucleic acid interactions.* Chemical Society Reviews, 2014. **43**(4): p. 1221-1229.
155. Yang, K., Y. Yang, and C.-y. Zhang, *Single-molecule FRET for ultrasensitive detection of biomolecules.* Nanobioimaging, 2014. **1**(1).
156. Okamoto, K., *Introduction of FRET application to biological single-molecule experiments.* International Journal of Biophysics, 2013. **3**(1A): p. 9-17.
157. Dimura, M., et al., *Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems.* Current opinion in structural biology, 2016. **40**: p. 163-185.
158. Deniz, A.A., S. Mukhopadhyay, and E.A. Lemke, *Single-molecule biophysics: at the interface of biology, physics and chemistry.* Journal of the Royal Society Interface, 2008. **5**(18): p. 15-45.
159. Saccardo, P., J.L. Corchero, and N. Ferrer-Miralles, *Tools to cope with difficult-to-express proteins.* Applied microbiology and biotechnology, 2016. **100**(10): p. 4347-4355.
160. Rosano, G.L. and E.A. Ceccarelli, *Recombinant protein expression in Escherichia coli: advances and challenges.* Frontiers in microbiology, 2014. **5**: p. 172.
161. Tuckey, C., et al., *Protein Synthesis Using a Reconstituted Cell-Free System.* Current protocols in molecular biology, 2014: p. 16.31. 1-16.31. 22.
162. Johnson, D.B., et al., *RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites.* Nature chemical biology, 2011. **7**(11): p. 779-786.
163. Bornhorst, J.A. and J.J. Falke, *[16] Purification of proteins using polyhistidine affinity tags.* Methods in enzymology, 2000. **326**: p. 245-254.

164. Kuiper, J.M., et al., *A method for site-specific labeling of multiple protein thiols.* Protein Science, 2009. **18**(5): p. 1033-1041.
165. Ratner, V., et al., *A general strategy for site-specific double labeling of globular proteins for kinetic FRET studies.* Bioconjugate chemistry, 2002. **13**(5): p. 1163-1170.
166. Jäger, M., E. Nir, and S. Weiss, *Site-specific labeling of proteins for single-molecule FRET by combining chemical and enzymatic modification.* Protein Science, 2006. **15**(3): p. 640-646.
167. Tian, H., et al., *Bioorthogonal Fluorescent Labeling of Functional G-Protein-Coupled Receptors.* ChemBioChem, 2014. **15**(12): p. 1820-1829.
168. Kelly, S.M. and N.C. Price, *The use of circular dichroism in the investigation of protein structure and function.* Current protein and peptide science, 2000. **1**(4): p. 349-384.
169. Johnson, C.K., *Calmodulin, conformational states, and calcium signaling. A single-molecule perspective.* Biochemistry, 2006. **45**(48): p. 14233-14246.
170. Slaughter, B.D., et al., *Conformational substates of calmodulin revealed by single-pair fluorescence resonance energy transfer: influence of solution conditions and oxidative modification.* Biochemistry, 2005. **44**(10): p. 3694-3707.
171. Ugwu, S.O. and S.P. Apte, *The effect of buffers on protein conformational stability.* Pharmaceutical Technology, 2004. **28**(3): p. 86-109.
172. Wang, Q., et al., *The effect of macromolecular crowding, ionic strength and calcium binding on calmodulin dynamics.* PLoS Comput Biol, 2011. **7**(7): p. e1002114.

Acknowledgement

I owe many thanks to Prof. Dr. Jörg Fitter for his supervision of the thesis and fruitful discussions and to Prof. Dr. Schwaneberg for having kindly accepted to serve as co-supervisor.

I would like to thank Dr. Alexandros Katranidis and Dr. Michele Cerminara for their supervision concerning molecular biology techniques and smFRET; Dr. Noemie Kempf and Dr. Nadine Jordan for their precious help in TCCD measurements and fluorescence gel imaging; Dr. Matteo Gabba for developing and establishing the smFRET data analysis programs in our institute; Ilona Ritter and Ramona Choy for their technical assistance; Dr. Iris Von der Hocht for the translation of the summary of the present thesis and Dr. Antonie Schöne for her precious advices concerning the PhD thesis.

I am also grateful to Guido, Clara, Viviane, Elsa, Kassem, Nicolas, Philippe, Massimo, Catherine, Jean-Christophe and Mohamadou. You are outstanding scientists, often friends and sometimes more. The accomplishment of the present thesis would not have been possible without you. Thank you.

Thank you to my parents who were supporting me over years. That is very difficult to imagine that a little half-blood girl from the school number 17 in Schaerbeek could go so far in her education path. Things were not always easy. You were part of the adventure even if not physically at least in mind. Thank you.

Finally, I would like to give a special thanks to Prof. Dr. Wolf B. Frommer. An incredible scientist, an exceptional Alexander von Humboldt-Professor, a wonderful PI and most importantly, a real mentor. Years ago, you gave me the reasons for dedicating my career to scientific research. Thank you. Despite the difficulties, I have never regretted that choice. For years, you have supported me even when things were tricky. I would never thank you enough for that.

Supplementary information

(A)

```
ATGGCAGATCAGCTGACTGAGGAGCAGATTGCAGAGTTCAAAGAGGCCTT     50
CTCCCTCTTTGACAAGGATGGAGATGGCACTATCACCACCAAGGAGTTGG    100
GGACAGTGATGCGTTCCCTGGGTCAGAACCCGACTGAAGCAGAGCTGCAG    150
GATATGATCAATGAGGTGGATGCAGATGGTAACGGTACCATTGACTTCCC    200
GGAGTTCCTGACCATGATGGCCCGTAAGATGAAGGACACAGACAGCGAGG    250
AGGAGATCCGTGAGGCGTTCCGTGTCTTTGACAAGGATGGTAATGGCTAC    300
ATCAGCGCCGCAGAGCTGCGTCACGTAATGACGAACCTGGGTGAGAAGCT    350
GACCGATGAGGAGGTGGATGAGATGATCCGTGAGGCTGACATCGATGGTG    400
ATGGCCAGGTCAATTATGAAGAGTTTGTACAGATGATGACTGCAAAGCAT    450
CACCATCACCATCACTAA    468
```

(B)

```
ATGGCAGATCAGCTGACTGAGGAGCAGATTGCAGAGTTCAAAGAGGCCTT     50
CTCCCTCTTTGACAAGGATGGAGATGGCACTATCACCACCAAGGAGTTGG    100
GGTAGGTGATGCGTTCCCTGGGTCAGAACCCGACTGAAGCAGAGCTGCAG    150
GATATGATCAATGAGGTGGATGCAGATGGTAACGGTACCATTGACTTCCC    200
GGAGTTCCTGACCATGATGGCCCGTAAGATGAAGGACACAGACAGCGAGG    250
AGGAGATCCGTGAGGCGTTCCGTGTCTTTGACAAGGATGGTAATGGCTAC    300
ATCAGCGCCGCAGAGCTGCGTCACGTAATGACGAACCTGGGTGAGAAGCT    350
GACCGATGAGGAGGTGGATGAGATGATCCGTGAGGCTGACATCGATGGTG    400
ATGGCCAGGTCAATTATGAAGAGTTTGTACAGATGATGACTGCAAAGCAT    450
CACCATCACCATCACTAA    468
```

(C)

```
ATGGCAGATCAGCTGACTGAGGAGCAGATTGCAGAGTTCAAAGAGGCCTT     50
CTCCCTCTTTGACAAGGATGGAGATGGCACTATCACCACCAAGGAGTTGG    100
GGUGCGTGATGCGTTCCCTGGGTCAGAACCCGACTGAAGCAGAGCTGCAG    150
GATATGATCAATGAGGTGGATGCAGATGGTAACGGTACCATTGACTTCCC    200
GGAGTTCCTGACCATGATGGCCCGTAAGATGAAGGACACAGACAGCGAGG    250
AGGAGATCCGTGAGGCGTTCCGTGTCTTTGACAAGGATGGTAATGGCTAC    300
ATCAGCGCCGCAGAGCTGCGTCACGTAATGACGAACCTGGGTGAGAAGCT    350
GACCGATGAGGAGGTGGATGAGATGATCCGTGAGGCTGACATCGATGGTG    400
ATGGCCAGGTCAATTATGAAGAGTTTGTACAGATGATGACTGCAAAGCAT    450
CACCATCACCATCACTAA    468
```

(D)

```
ATGGCAGATCAGCTGACTGAGGAGCAGATTGCAGAGTTCAAAGAGGCCTT     50
CTCCCTCTTTGACAAGGATGGAGATGGCACTATCACCACCAAGGAGTTGG    100
GGACAGTGATGCGTTCCCTGGGTCAGAACCCGACTGAAGCAGAGCTGCAG    150
GATATGATCAATGAGGTGGATGCAGATGGTAACGGTACCATTGACTTCCC    200
GGAGTTCCTGACCATGATGGCCCGTAAGATGAAGGACACAGACAGCGAGG    250
AGGAGATCCGTGAGGCGTTCCGTGTCTTTGACAAGGATGGTAATGGCTAC    300
ATCAGCGCCGCAGAGCTGCGTCACGTAATGUGCAACCTGGGTGAGAAGCT    350
GACCGATGAGGAGGTGGATGAGATGATCCGTGAGGCTGACATCGATGGTG    400
ATGGCCAGGTCAATTATGAAGAGTTTGTACAGATGATGACTGCAAAGCAT    450
CACCATCACCATCACTAA    468
```

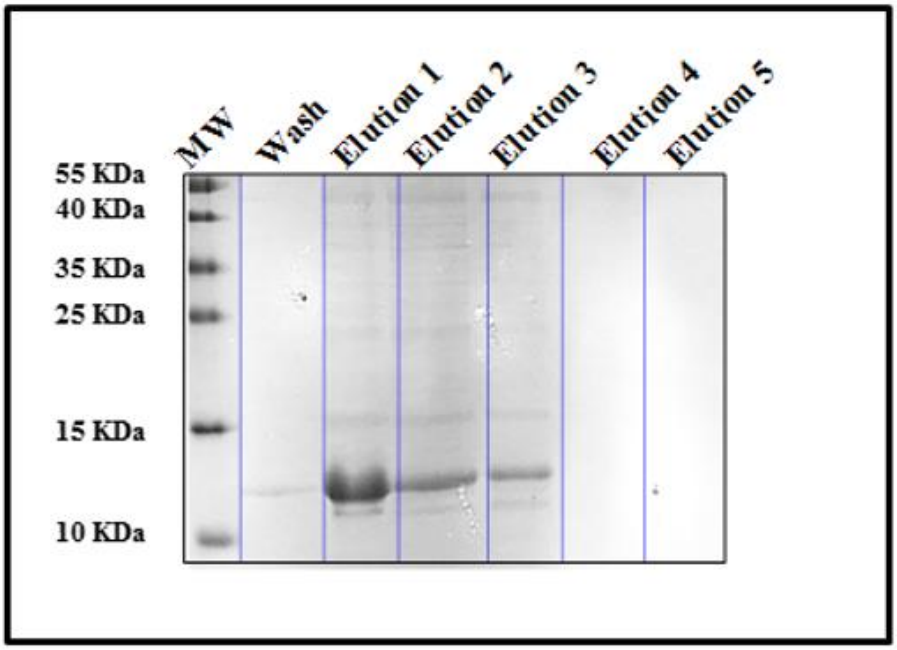
(E)

```
ATGGCAGATCAGCTGACTGAGGAGCAGATTGCAGAGTTCAAAGAGGCCTT      50
CTCCCTCTTTGACAAGGATGGAGATGGCACTATCACCACCAAGGAGTTGG     100
GGUGCGTGATGCGTTCCCTGGGTCAGAACCCGACTGAAGCAGAGCTGCAG     150
GATATGATCAATGAGGTGGATGCAGATGGTAACGGTACCATTGACTTCCC     200
GGAGTTCCTGACCATGATGGCCCGTAAGATGAAGGACACAGACAGCGAGG     250
AGGAGATCCGTGAGGCGTTCCGTGTCTTTGACAAGGATGGTAATGGCTAC     300
ATCAGCGCCGCAGAGCTGCGTCACGTAATGTAGAACCTGGGTGAGAAGCT     350
GACCGATGAGGAGGTGGATGAGATGATCCGTGAGGCTGACATCGATGGTG     400
ATGGCCAGGTCAATTATGAAGAGTTTGTACAGATGATGACTGCAAAGCAT     450
CACCATCACCATCACTAA     468
```

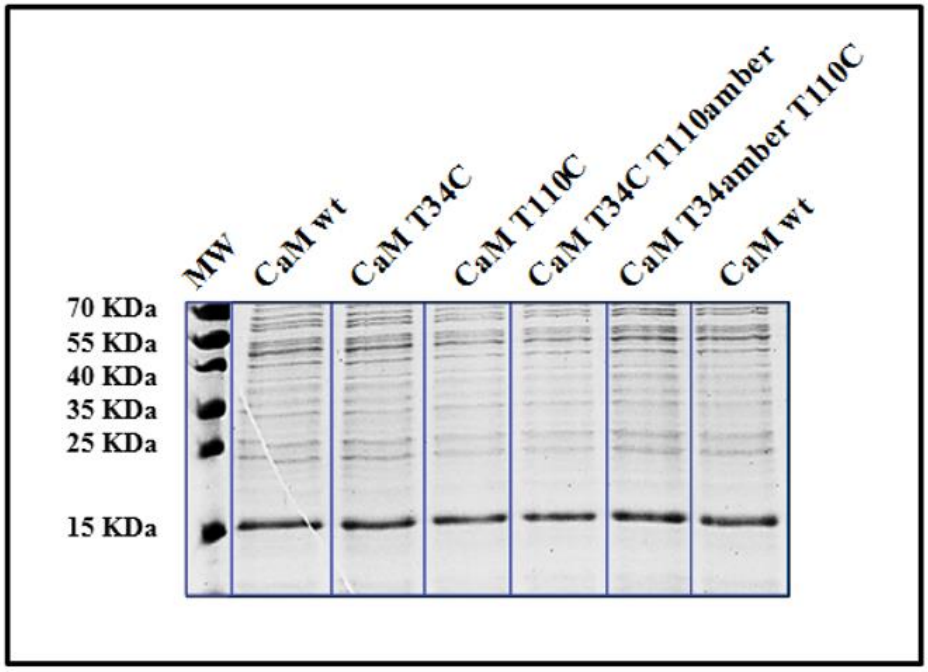
(F)

```
ATGGCAGATCAGCTGACTGAGGAGCAGATTGCAGAGTTCAAAGAGGCCTT      50
CTCCCTCTTTGACAAGGATGGAGATGGCACTATCACCACCAAGGAGTTGG     100
GGTAGGTGATGCGTTCCCTGGGTCAGAACCCGACTGAAGCAGAGCTGCAG     150
GATATGATCAATGAGGTGGATGCAGATGGTAACGGTACCATTGACTTCCC     200
GGAGTTCCTGACCATGATGGCCCGTAAGATGAAGGACACAGACAGCGAGG     250
AGGAGATCCGTGAGGCGTTCCGTGTCTTTGACAAGGATGGTAATGGCTAC     300
ATCAGCGCCGCAGAGCTGCGTCACGTAATGUGCAACCTGGGTGAGAAGCT     350
GACCGATGAGGAGGTGGATGAGATGATCCGTGAGGCTGACATCGATGGTG     400
ATGGCCAGGTCAATTATGAAGAGTTTGTACAGATGATGACTGCAAAGCAT     450
CACCATCACCATCACTAA     468
```

Supplementary Figure 1. Sequences of the plasmids used in the present study: (A) CaM wt, (B) CaM T34amber, (C) CaM T34C, (D) CaM T110C, (E) CaM T34C T110amber and (F) CaM T34amber T110C. The amber and cysteine codons used for labeling and the His_6-tag are shown in cyan, magenta and yellow, respectively.



Supplementary Figure 2. Overexpression of CaM T110C in BL21 *E.coli* cells. SDS PAGE of different fractions collected during the purification of CaM T110C overexpressed into BL21 cells.



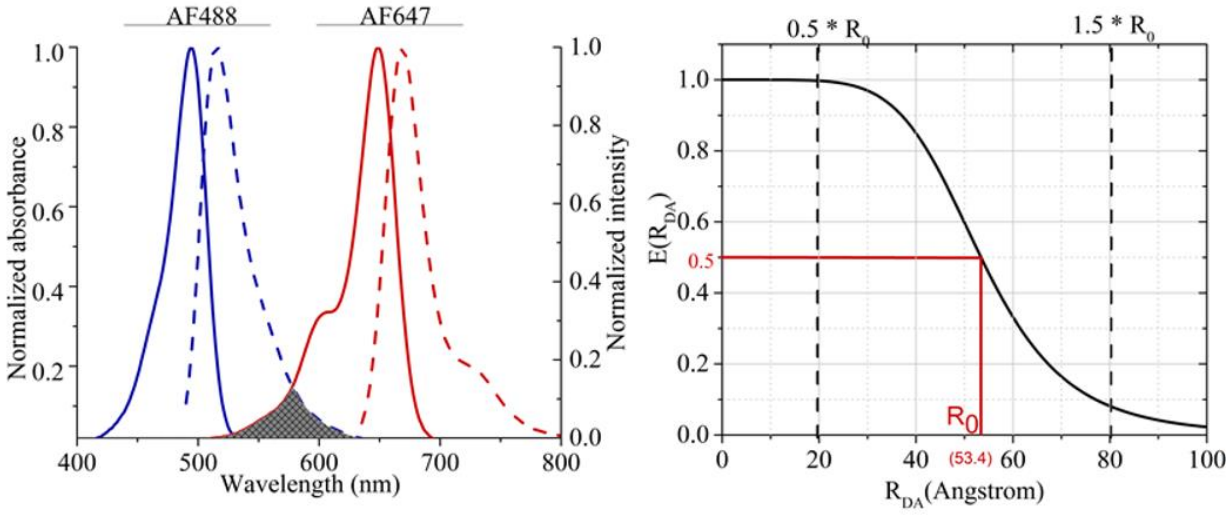
Supplementary Figure 3. **Cell-free synthesis and incorporation of UAA into CaM**. Incorporation of UAA into CaM by using a RF1-depleted coupled transcription-translation cell-free system in continuous-exchange.

Supplementary Table 1. Concentration of cell-free synthesize CaM

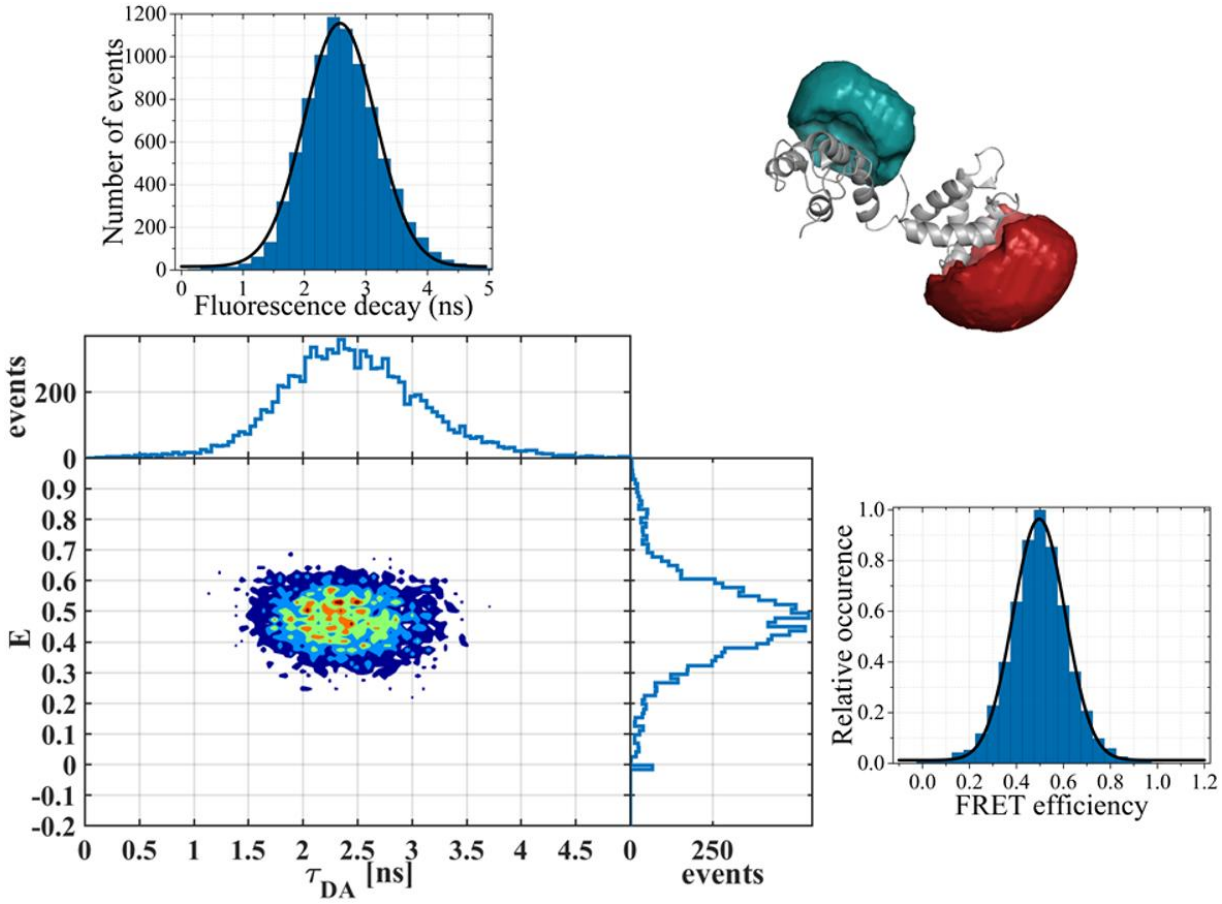
	Cam wt	**CaM T34amber T110C**	**CaM T34C T110amber**
Concentration µg/ml	1013	1083	1178

Supplementary Table 2. Expected FRET efficiencies calculated for CaM T34C^{AF488} T110AzF^{AF647}

		calculated from Cα		**calculated from AV**	
PDB ID	**Description**	**R_{DA} (Å)**	**E**	**R_{DA} (Å)**	**E**
1DMO	ApoCaM	42.3	0.79	53.3	0.50
1CLL	HoloCaM	52.4	0.53	64.7	0.24
1PRW	HoloCaM	16.3	1.00	23.9	0.99
4HEX	HoloCaM	49.6	0.61	59.4	0.35
1X02	HoloCaM	40.4	0.84	52.2	0.53
1CDM	Ca^{2+}-CaM-CaMKII	14.8	1.00	20.9	1.00
2BBM	Ca^{2+}-CaM-M13	15.4	1.00	23.1	0.99

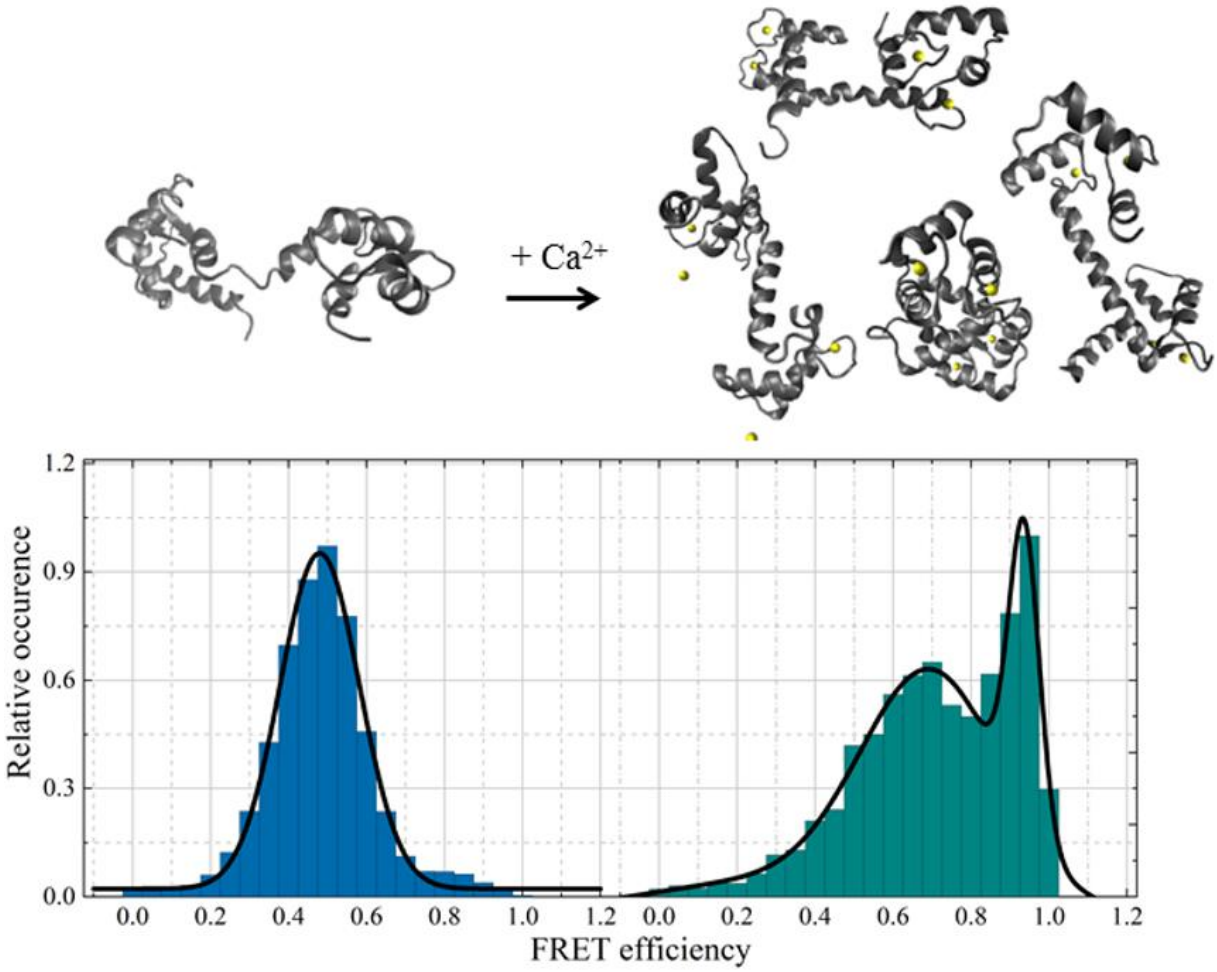


Supplementary Figure 4. Förster radius determination of AF488-AF647 pair. Absorption (solid lines) and emission (dashed lines) spectra of Alexa Fluor 488 and Alexa Fluor 647 bound to CaM (left). The gray area represents the overlap integral (λ). The corresponding transfer efficiency (E) in function of the distance (right).

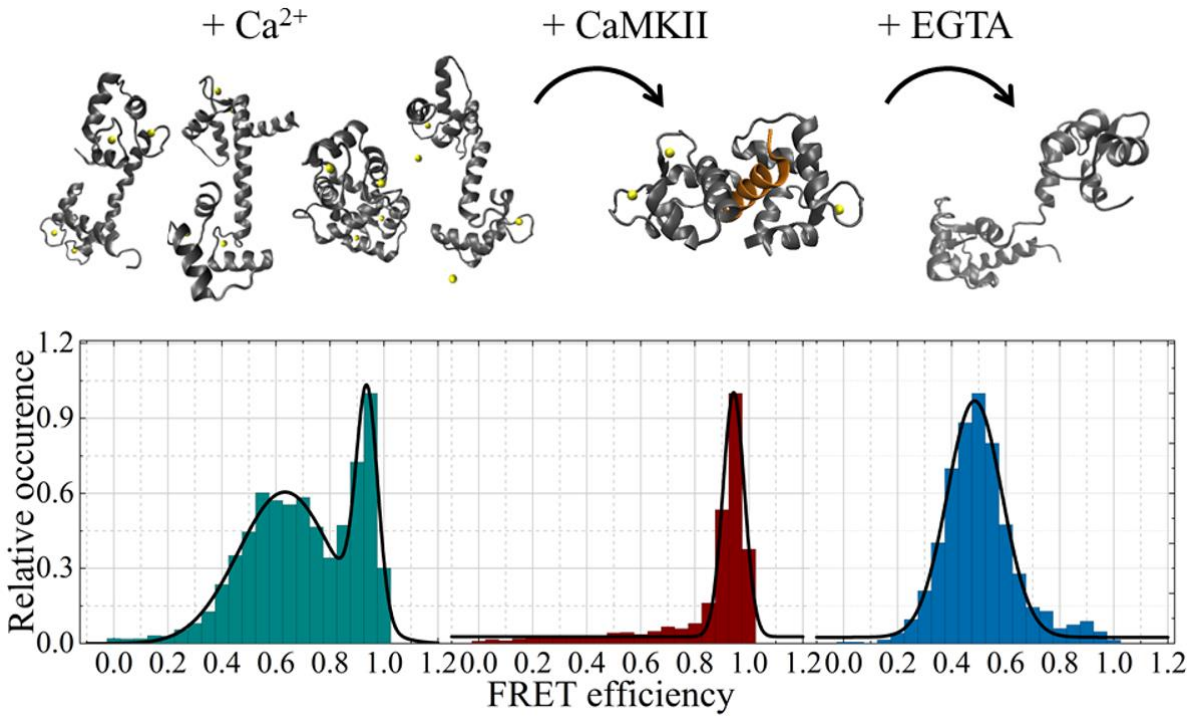


Supplementary Figure 5. Structure of apoCaM verified by Förster Energy Transfer. 2D plots of FRET efficiency (E) and lifetime of the donor (τ_{DA}) of a dually labeled population of CaM (left, bottom), the FRET efficiency histogram (right, bottom) and the lifetime histogram

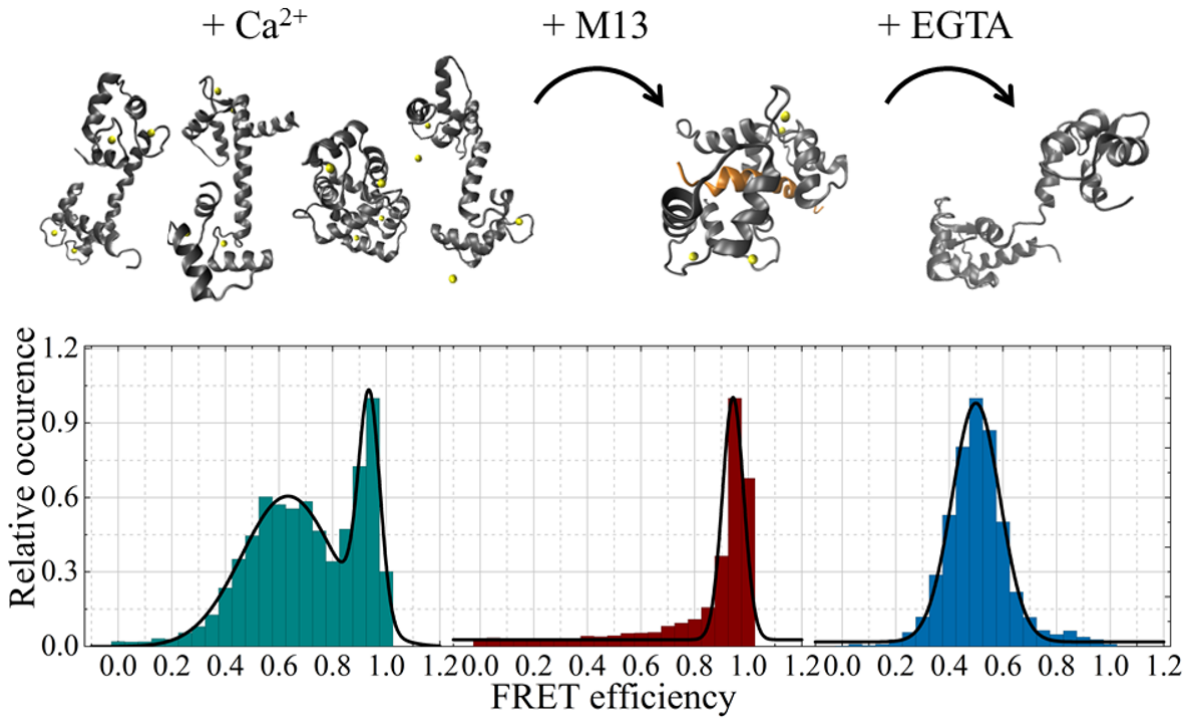
(right, top) as well as the three-dimensional structure of apoCaM with the accessible volumes of the dye (right, top) are shown.



Supplementary Figure 6. Conformational changes undergo by CaM T34C^{AF488} T110AzF^{AF647} upon Ca^{2+} binding. FRET histograms of apoCaM (left) and holoCaM (right). The apoCaM structure with the accessible volumes for the two dyes is illustrated, as well as two extreme structures of holoCaM. Gaussian fits were applied to the histograms and a mean FRET efficiency was calculated for every peak. Conformational changes of CaM upon Ca^{2+} binding are represented.

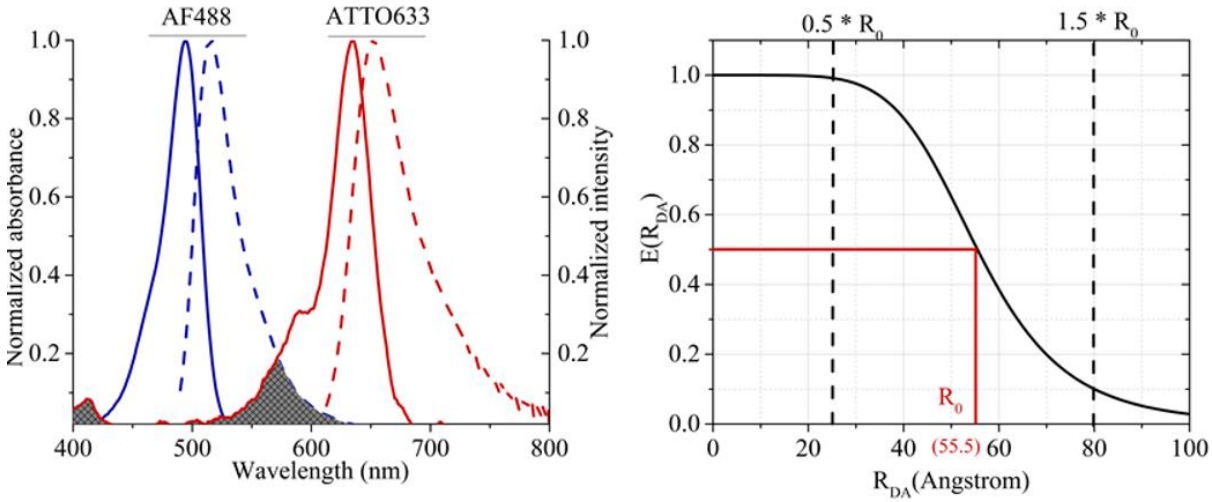


Supplementary Figure 7. **Binding of CaMKII to CaM T34C^{AF488} T110AzF^{AF647}.** FRET histograms of holoCaM in the absence (left) or presence of CaMKII (middle) and after addition of EGTA (right). Gaussian fits were applied to the histograms and a mean FRET efficiency was calculated for every peak. Representations of holoCaM, Ca^{2+}-CaM-CaMKII and apoCaM are shown. The partner is displayed in orange.

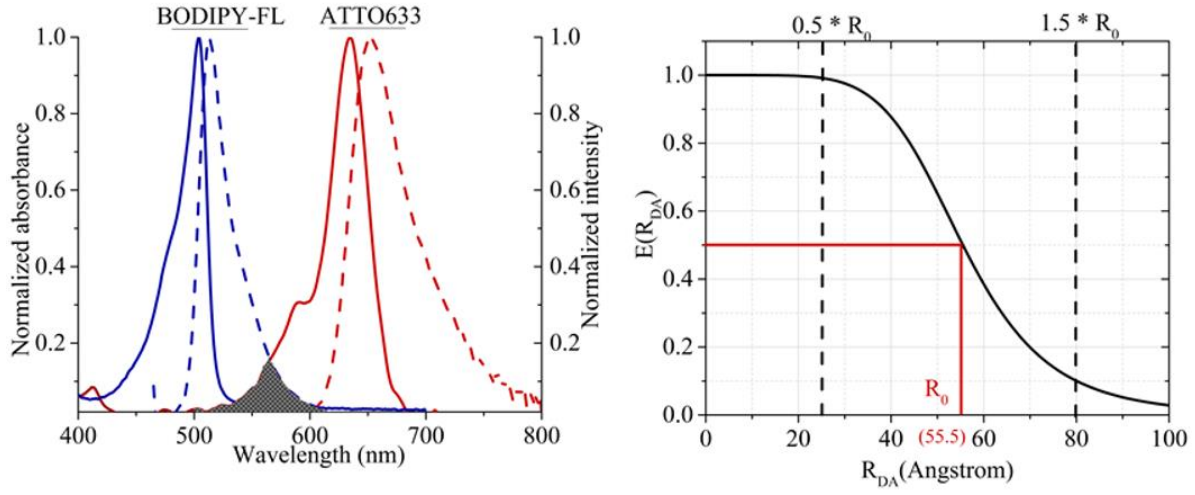


Supplementary Figure 8. **Binding of M13 to CaM T34C^{AF488} T110AzF^{AF647}.** FRET histograms of holoCaM in the absence (left) or presence of M13 (middle) and after addition

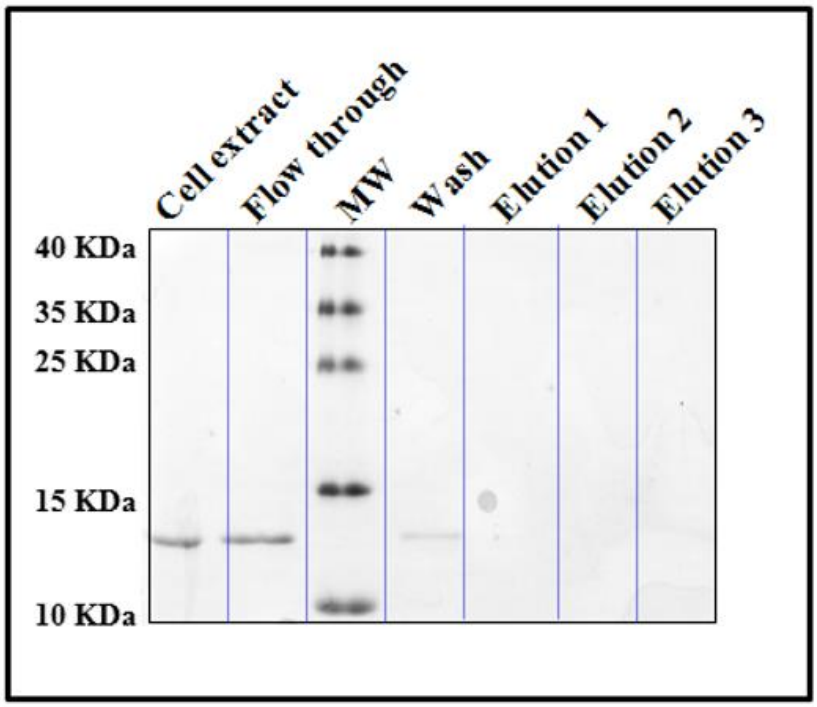
of EGTA (right). Gaussian fits were applied to the histograms and a mean FRET efficiency was calculated for every peak. Representations of holoCaM, Ca^{2+}-CaM-M13 and apoCaM are shown. The partner is displayed in orange.



Supplementary Figure 9. Förster radius determination of AF488-ATTO633 pair. Absorption (solid lines) and emission (dashed lines) spectra of Alexa Fluor 488 and ATTO633 bound to CaM and the tRNA, respectively (left). The gray area represents the overlap integral (λ). The corresponding transfer efficiency (E) in function of the distance (right).



Supplementary Figure 10. Förster radius determination of BODIPY FL-ATTO633 pair. Absorption (solid lines) and emission (dashed lines) spectra of BODIPY-FL and ATTO633 bound to the tRNAs (left). The gray area represents the overlap integral (λ). The corresponding transfer efficiency (E) in function of the distance (right).



Supplementary Figure 11. **Overexpression of hCaM in BL21 *E.coli* cells.** SDS PAGE of different fractions collected during the purification of overexpressed CaM T110C with cobalt resins.

Supplementary Table 3. Quantum yield of the dyes when bound to apoCaM

CaM T34AzF T110C		**CaM T34AmF T110C**	
AF647	AF488	ATTO633	BODIPY-FL
P34	**P110**	**P34**	**P110**
0.27	0.66	0.55	0.38

Supplementary Table 4. Quantum yield of the dyes when bound to holoCaM

CaM T34AzF T110C		**CaM T34AmF T110C**	
AF647	AF488	ATTO633	BODIPY FL
P34	**P110**	**P34**	**P110**
0.22	0.38	0.98	0.79

Supplementary Table 5. Transmission efficiency for the dyes used in the present study

	ATTO633	**BODIPY FL**	**AF488 CaM**	**AF647 CaM**	**AF647 NHS**	**AF488 NHS**
Transmission efficiency (%)	52.9	79.6	75.2	77.9	77.4	74.9