PAPER • **OPEN ACCESS**

Advantages and limitations of a supernegative GFP in facilitating MyoD intracellular tracking

To cite this article: Lucia Boeri *et al* 2020 *Methods Appl. Fluoresc.* **8** 025007

Methods and Applications in Fluorescence

PAPER

Advantages and limitations of a supernegative GFP in facilitating MyoD intracellular tracking

Lucia Boeri[1], **Emanuela Jacchetti**[1], **Monica Soncini**[2], **Alessandro Negro**[3], **Diego Albani**[4] and **Manuela Teresa Raimondi**[1]

[1] Department of Chemistry, Materials and Chemical Engineering 'Giulio Natta', Politecnico di Milano, Milan, Italy
[2] Department of Electronics Information and Bioengineering, Politecnico di Milano, Milan, Italy
[3] Department of Biomedical Sciences, University of Padua, Padua, Italy
[4] Department of Neuroscience, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy

E-mail: lucia.boeri@polimi.it

OPEN ACCESS

RECEIVED
20 September 2019

REVISED
22 January 2020

ACCEPTED FOR PUBLICATION
24 February 2020

PUBLISHED
13 March 2020

Keywords: fluorescence microscopy, molecular dynamics, protein transduction, supernegative GFP, MyoD

Supplementary material for this article is available online

Abstract

Despite intracellular molecular dynamics being fundamental to understand pathological, biomechanical or biochemical events, several processes are still not clear because of the difficulty of monitoring and measuring these phenomena. To engineer an effective fluorescent tool useful to improve protein intracellular tracking studies, we fused a supernegative green fluorescent protein, (−30)GFP, to a myogenic transcription factor, MyoD. The (−30)GFP-MyoD was able to pass the plasma membrane when complexed with cationic lipids. Fluorescence confocal microscopy showed the protein delivery in just 3 hours with high levels of protein transduction efficiency. Confocal acquisitions also confirmed the maintenance of the MyoD nuclear localization. To examine how the supernegative GFP influenced MyoD activity, we did gene expression analyses, which showed an inhibitory effect of (−30)GFP on transcription factor function. This negative effect was possibly due to a charge-driven interference mechanism, as suggested by further investigations by molecular dynamics simulations. Summarizing these results, despite the functional limitations related to the charge structural characteristics that specifically affected MyoD function, we found (−30)GFP is a suitable fluorescent label for improving protein intracellular tracking studies, such as nucleocytoplasmic transport in mechanotransduction.

1. Introduction

Optical microscopy has become an integral part of modern cell biology, not only thanks to huge advances in spatial resolution but above all for the fact that we can now study complex systems such as living cells and their dynamics. Over the last decades, from the discovery of the green fluorescent protein (GFP), extracted from the jellyfish *Aequorea victoria* [1, 2], remarkable advances have been made in molecular engineering to generate fluorescent tools for studying cell biology and physiology by optical and imaging techniques [3]. Beyond the GFP-derived labels, fluorophores and nanoparticles serve as alternative labeling strategies in cell and molecular visualization approaches [4].

Organic fluorophores such as Cy5 and Cy3 offer high brightness and small size. Since labeling the protein of interest (POI) is mediated by two complementary reactive groups—such as the maleimide group of the fluorophore and the cysteine residue of the POI—the main problem with this labeling strategy is low specificity. The fusion of protein tags (e.g. Halo, SNAP) to bind the fluorophores can increase the specificity but it alters the size of the labeling complex, nullifying the small dimension-related advantage [5, 6].

Quantum dots (Qdot) are nanoparticle labels exploited for their brightness and photostability. Their main disadvantage is that they have to be functionalized with reactive peptides (i.e. biotin, antibodies) making the whole label larger and changing the function and dynamics of the POI [7].

The most used labeling strategy to monitor protein dynamics is still the fusion of a fluorescent protein to the POI. This involves a specific labeling approach by

cloning the fluorescent tag at the N- or C-terminus of the POI. Most of this type of labels come from the original GFP, which is a very small protein of about 27 kDa with a 30 Å diameter. Today we can use improved GFP variants with different physicochemical properties, such as brightness, photostability, oligomerization tendency, sensitivity to pH variation and maturation rates [8]. For example, certain genetic modifications enable the GFP to be photoactivated, photoconverted or photoswitched [9–11].

The wide range of GFP variants allows us to investigate intracellular dynamics such as single protein movement or protein interaction kinetics. The fusion of a GFP variant to a POI and its detection could unveil intracellular mechanisms such as the ability to bind DNA, to migrate through lipid cell membranes or the mechanotransduction response [12–14].

The classic procedure to internalize the labeled POI into cells requires DNA transfection. This involves the internalization of a recombinant plasmid coding for the labeled POI and detection of the fluorescence after at least 24 h. This provides continuous high levels of protein expression that could potentially alter the interpretation of physiological protein dynamics and kinetics. Other limiting aspects of DNA transfection methods are related to the dependence of internalization efficiency on the cell type, and the long working time required to reach optimal fluorescence expression. Direct internalization methods of fluorescent POI, such as microinjection and protein transduction, are challenging strategies to overcome the technical disadvantages of DNA transfection.

Lawrence and colleagues published a novel genetically modified GFP: the supernegative (−30)GFP [15]. Differently from the other genetically modified variants, this GFP form is optically similar to the original protein but has different surface electrostatic characteristics that alter its molecular interaction properties [15, 16]. From a molecular point of view, Lawrence and colleagues replaced 15 surface amino acids with negatively charged residues (glutamate and aspartate), changing the protein net charge from −7 to −30.

Zuris and colleagues demonstrated that highly anionic proteins can be delivered intracellularly by the same electrostatics-driven mechanism used for nucleic acid delivery. They exploited the anionic properties of the novel GFP to create anionic protein:cationic lipid complexes to import intracellularly (−30)GFP-based recombinant proteins *in vitro* and *in vivo*. To the best of our knowledge, the intracellular delivery of proteins able to penetrate the plasma membrane –also known as protein transduction—is a molecule internalization approach that has not yet been used for protein dynamics studies such as nucleocytoplasmic transport. In this work, our aim was to engineer a fluorescent variant of a transcription factor (TF) suitable for further protein localization and intracellular dynamics studies. We decided to exploit the advantages offered by the protein transduction technology by fusing the (−30)GFP to a myogenic TF, MyoD (Myoblast determination protein 1), generating a fluorescent variant of MyoD able to be delivered into the nucleus as a recombinant protein when complexed with cationic lipids. To demonstrate the advantages and the suitability of the supernegative TF for protein dynamics studies, we assessed the protein transduction efficiency and intranuclear localization by fluorescence confocal microscopy. We also compared the working time and the internalization efficiency for (−30)GFP-MyoD transduction and classical DNA transfection procedures. In addition, we evaluated the influence of the supernegative GFP on MyoD transcription promoting activity, measuring the gene expression of specific MyoD targets by real-time PCR. Finally, we investigated some aspects arousing concern about the influence of the (−30)GFP on MyoD functionality generating computational models and running molecular dynamics simulations. In both gene expression analyses and computational simulations, we used a photoactivatable variant of MyoD as a functional control complex.

2. Methods

2.1. Gene cloning

(−30)GFP sequence was amplified from pET-(−30) GFP-9xGGS-Cre-6xHis (Addgene, plasmid #32372) and subcloned into the prokaryotic vector pRSET A using the XbaI and BamHI sites of the vector multiple cloning site. The transcription factor MyoD sequence was subcloned into the plasmid pRSET A-(−30)GFP using NheI and HindIII sites.

The sequences coding for MyoD, (−30)GFP, (−30)GFP-MyoD were also inserted into the eukaryotic vector pcDNA 3 (resistant to kanamycin instead of ampicillin) using NheI and HindIII sites, NdeI and NheI sites, NdeI and HindIII sites, respectively.

To generate the plasmid coding for MyoD-PAGFP (pMyoD-PAGFP), we subcloned the MyoD sequence into the plasmid coding for PAGFP (pPAGFP-N1) (Addgene, plasmid #11909) using the BglII and EcorI restriction sites. We used TOP10 *E. coli* strain for gene cloning procedures.

2.2. Protein purification

Recombinant variants were expressed in the BL21(DE3) pLysS E. coli strain. Starter cultures were prepared by growing multiple colonies of E. coli cells carrying the recombinant plasmids in a 1 L flask containing Luria-Bertani medium added with ampicillin and chloramphenicol (respectively 100 μg ml^{-1} and 33 μg ml^{-1}); E. coli cells were grown with shaking until they reached an OD 600 nm ∼ 0.6, at which time 0.4 mM IPTG was added to induce protein expression. Cells expressing recombinant proteins were collected after 12 hours of growth with shaking at ∼17–23 °C. Protein purification required a first step of centrifugation at 6500 rpm, 4 °C

for 20 min The pellet was resuspended in lysis buffer (PBS-NaCl 2 M). Cells were disrupted by sonication on ice, using a Branson Sonifier 450 (Branson Ultrasonics, Danbury, CT, USA). The lysate was cleared by centrifuging at 12500 rpm, 4 °C for 30 min The recombinant proteins were recovered from supernatant by chromatography using a HIS-Select® Nickel Affinity Gel (Sigma-Aldrich). After separation on the chromatographic column, the proteins were eluted (NaCl 2 M, Imidazol 0.75 M, pH 8) and dialyzed (NaCl 0.5 M, PBS). Elution products were analyzed by spectrophotofluorometer to check the quality and quantity of the purified proteins.

2.3. Cell culture

Rat bone marrow mesenchymal stem cells (rBM-MSCs) were kindly provided by Dr Barbara Bonandrini (Dept. of Bionengineering, Mario Negri Institute for Pharmacological Research, Bergamo, Italy) [17]. Cells were cultured using α-MEM supplemented with 20% fetal bovine serum (v/v) (Gibco, Thermo Fisher Scientific), penicillin 100 IU/ml/streptomycin 100 μg ml^{-1} (Invitrogen, Thermo Fisher Scientific) and 0.5 mM L-glutamine (Invitrogen, Thermo Fisher Scientific). Cells were seeded in a six-well plate at a density of 35 $\times$ 10^3 cells/well for genetic assays. For fluorescence imaging analyses cells were seeded either in a 24-well plate at 8.5 $\times$ 10^3 cells/well or into 8-well Nunc® Lab-Tek® at 5 $\times$ 10^3 cells/well.

2.4. Protein Transduction

24 hours before transduction, cells were plated into a 24-well plate to reach 50% confluency. For transduction of (−30)GFP-MyoD we tested four transfection reagents, liposome-based and non-liposomal: Lipofectamine™ MessengerMAX, Lipofectamine™ 3000, JetPRIME®, Fugene® HD. We set specific conditions starting from the manufacturer's instructions (see supplementary table 1 is available online at stacks.iop.org/MAF/8/025007/mmedia). One hour before addition of the anionic protein:cationic molecule complex, we conditioned the cell culture with serum-free Optimem medium (Gibco, Thermo Fisher Scientific). Three hours after protein transduction we replaced the transduction medium with cell growth medium.

2.5. DNA transfection

DNA vectors were transfected using FuGENE® HD transfection reagent (Promega Corporation) according to the manufacturer's instructions. Cells were plated 24 hours before transfection into a six-well plate at a density of 35 $\times$ 10^3 cells/well. We selected 3:1 = FuGENE:DNA as the optimal transfection ratio.

2.6. Transduction\transfection efficiency

To assess the transduction or transfection efficiency of our constructs, we incubated cells in growth medium supplemented with Hoechst 33342 (1 μg ml^{-1}) (Thermo Fisher Scientific, USA). After 10 min cells were fixed with 4% paraformaldheyde (PAF) for 15 min, washed with PBS, then mounted on a glass slide. Transduction or transfection efficiency was measured 4 hours from protein addition or 48 hours from DNA transfection, and was calculated as:

$$\text{Efficiency (\%)} = (\text{transduced cells}) / (\text{total amount of nuclei})$$

Cells were considered transduced or transfected when at least one green fluorescent spot was detected.

2.7. Fluorescence microscopy and image analysis

Cell fluorescence was detected using a FluoView FV10i confocal laser scanning microscope (Olympus, Japan) equipped with 405 nm and 473 nm laser diodes for GFP and PAGFP acquisition and a water-immersion 60X phase contrast objective/NA 1.2. Live cell imaging was done at 37 °C, 5% CO_2. Fixed cell imaging was done at room temperature. The following collection ranges were adopted: 352–455 nm (Hoechst 33342), 489–550 nm ((−30)GFP and PAGFP). Images were analysed with ImageJ software.

2.8. Photoactivation

Protein photoactivation was done by fluorescence confocal microscopy (FluoView FV10i, Olympus, Japan). Photoactivation of MyoD-PAGFP was done by single cell irradiation with a 405 nm light laser enhancing the green light fluorescence emission. Photoactivated cells were visible by exciting them with 10% of the 473 nm laser power and acquiring the emission spectra in the same range as for (−30)GFP.

2.9. Gene expression

To measure the expression levels of mRNA target genes of MyoD, MSCs were treated into a six-well plate at a density of 35 $\times$ 10^3 cells/well. Following the literature [18], in the case of protein transduction, we extracted total mRNA nine hours after protein addition, thus six hours after optimal protein nuclear detection (activated state). In the case of DNA transfection, MSC culture samples were lysed for RNA extraction after 72 h. RNA was extracted from the MSC culture samples using a miRNeasy Mini Kit (Qiazol™) following the manufacturer's instructions. The quantity and purity of extracted RNA was measured by spectrophotometry (ND-1000; NanoDrop™). cDNA was synthesized from total RNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems™). We measured relative mRNA expression levels of myogenin (Myog), cadherin 15 (Cdh15), desmin (Des), Creatin Kinase M-Type (CKM) by quantitative real-time polymerase chain reaction (qRT-PCR) using TaqMan™ Reagents (Applied Biosystems, Thermo Fisher Scientific) on a 7900HT Fast Real-Time PCR System (Applied

Biosystems, Thermo Fisher Scientific). We used 18 s rRNA as reference gene. To measure the gene expression level, we employed the widely used Livak method, which calculates the normalized expression ratio respect to a negative control, the calibrator.

pMyoD was used as experimental positive control. As calibrator for pMyoD, we used a plasmid coding for a control protein with a similar size but with no regulatory function, such as transcriptional activity. We chose a commonly used enhanced form of GFP (pEGFP).

2.10. Molecular models

The 3D models of (−30)GFP and PAGFP were obtained by homology modeling starting from the photoactivatable GFP deposited in the RCSB Protein Data Bank (PDB code: 3GJ2). For alignment we used the NCBI-BLASTp tool, available on the NCBI website. 3D models for (−30)GFP and PAGFP were examined by surface electrostatic potential analyses using the APBS plugin of Visual Molecular Dynamics (VMD) software [19]. We took the atomic structure of MyoD BHLH domain (DNA-binding domain) from the model 1MDY in RCSB PDB, which is a homodimer bound to the DNA consensus sequence. We extracted the chain in yellow in supplementary figure 1A corresponding to the BHLH binding domain of a MyoD monomer and used this monomer to build the GFP-MyoD molecular complexes. To evaluate the influence of the lacking MyoD region, not available in PDB, we predicted the protein structure of the whole protein using PSI-blast based secondary structure PREDiction (PSIPRED v3.3) and DISOrder PREDiction (DISOPRED3) servers [20, 21].

2.11. MD configurations

After minimization and preliminary equilibration, the 3D structures of (−30)GFP and PAGFP were used in combination with the MyoD model to obtain GFP-MyoD complexes. Four configurations of each GFP-MyoD molecular complex were defined with different positions and orientations of the MyoD BHLH domain with respect to the GFP. To generate the 4 configurations, the center of mass of the GFP was assumed as the origin of the axes of the molecular system (X_{GFP}, Y_{GFP}, $Z_{GFP} = 0, 0, 0$). Then, in relation to the specific GFP, the MyoD domain was rigidly moved obtaining the following configurations:

Configuration n° 1: (X_M, Y_M, $Z_M = +3, +10, -30$)

Configuration n° 2: (X_M, Y_M, $Z_M = +28, +15, -5$)

Configuration n° 3: (X_M, Y_M, $Z_M = +0, +15, -28$)

Configuration n° 4: (X_M, Y_M, $Z_M = -27, +10, +5$)

where X_M, Y_M, Z_M are coordinates of the MyoD domain (see supplementary figure 1E).

2.12. MD simulation parameters

MD simulations were carried out with NAnoscale Molecular Dynamics 2.12 (NAMD) software together with the VMD software [19, 22]. The Force Field used was CHARMM 36 and all the MD simulations were done in explicit TIP3 water molecules, applying periodic boundary conditions.

(−30)GFP and PAGFP were first equilibrated for 1 ns under constant pressure and temperature (NPT ensemble) to relax the volume of the periodic boundary box. The equilibrated structures of GFPs were then used to build the GFP-MyoD complexes in the four configurations (see MD configurations set-up). To relax the volume of the periodic boundary box, the MD simulations were carried out under NPT conditions. The pressure was set at 1 atm, the temperature at 310 K, the time step was 2 fs and the non-bonded cutoff was 12 Å. Langevin dynamics were employed to control the temperature, maintaining the pressure at 1 atm, a period of 100 ns and a decay of 50 fs. Trajectories were recorded every 500 fs for 2,500,000 steps with 5 ns.

2.13. Non-bonded interaction energies

Non-bonded interaction energies were calculated with the NAMDEnergy plugin. First, we recorded the Van der Waals energy, electrostatic Coulomb energy and the total non-bonded interaction energy between the MyoD BHLH domain and the specific GFP. Then we examined the non-bonded interaction energies of each fluorescent protein with the BHLH domain separated in three elements: Helix 1 (residues from 105 to 136); Loop (residues from 137 to 146); Helix 2 (residues from 147 to 166). The energies were expressed in kcal/mol.

2.14. Statistical analyses

Experiments were performed in triplicate and the data were analysed by one-way analysis of variance (ANOVA), two-way ANOVA and post hoc tests, or with Student's t-test for direct comparison of two groups. Associations with $P < 0.05$ were considered significant. ns $P > 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; **** $P < 0.0001$.

3. Results

3.1. (−30)GFP-MyoD transduction and nuclear localization

To achieve the best (−30)GFP-MyoD transduction efficiency, we tested four cationic reagents generally used for DNA o RNA transfection: two lipid-based, Lipofectamine™ 3000 and Lipofectamine™ MessengerMAX, and two non-liposomal, Fugene® HD and JetPRIME® PolyPlus. We achieved optimal fluorescence detection 3–4 h after addition of the transduction complex. Then we assessed the intracellular (−30) GFP-MyoD delivery by fluorescence confocal

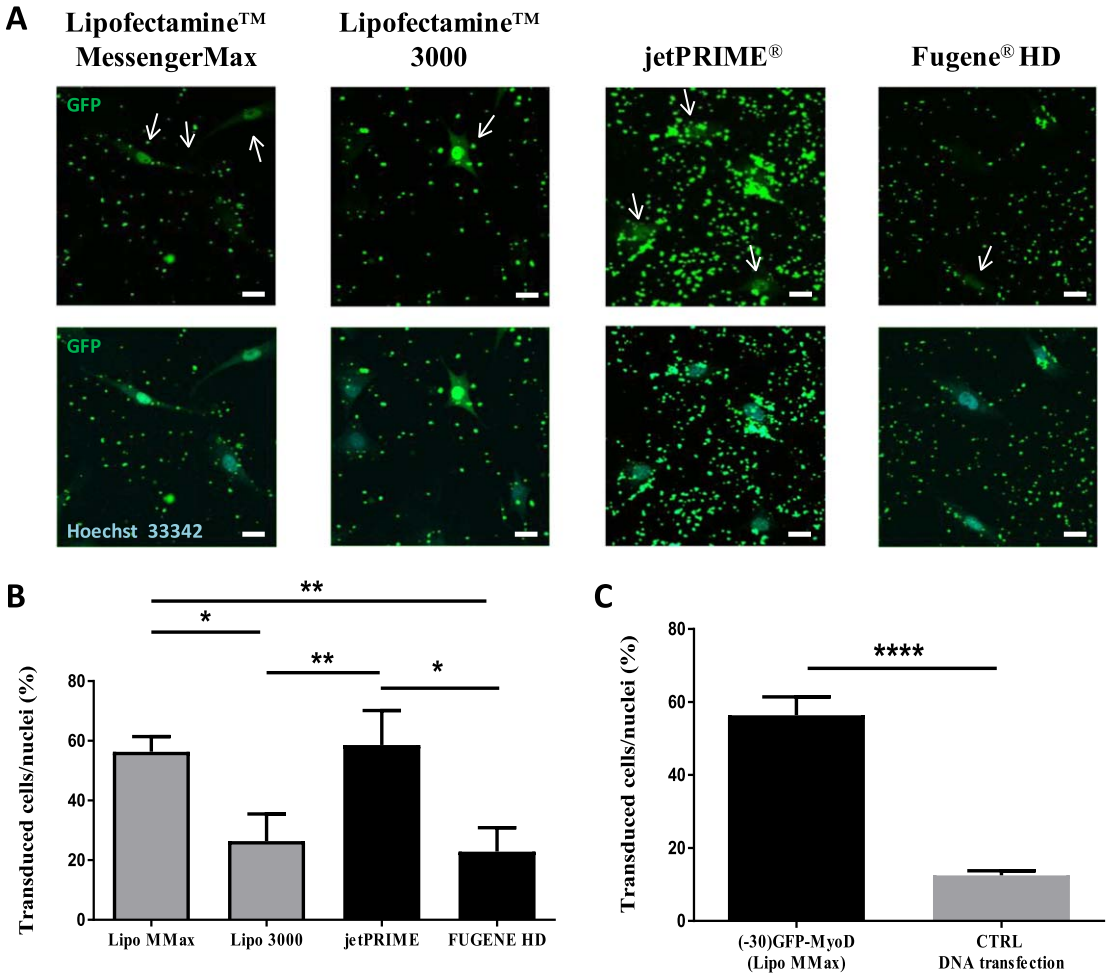

Figure 1. (−30)GFP-MyoD transduction efficiency. (A) Confocal images of MSCs transduced with (−30)GFP-MyoD when complexed with Lipofectamine™ MessengerMax (Lipo MMax), Lipofectamines™ 3000 (Lipo 3000), JetPRIME® Polyplus or Fugene® HD. Transduced cells are highlighted by the white arrows. The same images are shown superimposed on Hoechst33342 dye nuclear stain (lower images). Scale bar 20 μm; (B) Graphical representation of (−30)GFP-MyoD transduction efficiency using different cationic transfection reagents. Data were analyzed by one-way ANOVA test using Tukey's test for multiple comparisons; for each replicate we counted 85 cells. (C) Comparison of (−30)GFP-MyoD transduction and standard transfection of a plasmid coding for a control fluorescent protein (pmaxGFP). Data were analyzed by Student's t-test. For each replicate we counted 85 cells. Data are shown as mean ± SD.

microscopy. Most of the recombinant protein efficiently reached the nucleus, where the TF is active and physiologically localized (figure 1(A)). These results indicated that the (−30)GFP did not alter the localization of the native TF MyoD, facilitating the passage of both the cell membrane and the nuclear envelope. We calculated the protein transduction efficiency as the ratio of transduced cells to the total amount of MSC nuclei stained with Hoechst 33342 dye. The best (−30) GFP-MyoD transduction efficiency was reached by the complexation with both Lipofectamine™ MessengerMAX and JetPRIME® Polylus (figure 1(B)). The two reagents showed comparable transduction efficiency but, as shown in figure 1(A), complexation with JetPRIME® Polyplus generated a large amount of protein aggregates that could interfere with fluorescence confocal microscopy acquisitions and measurements. We therefore selected Lipofectamines™ MessengerMAX as the best cationic lipid-based transfection reagent for intracellular (−30)GFP-MyoD delivery.

Finally, we compared the efficiency of (−30)GFP-MyoD transduction with the transfection of a plasmid coding for a control fluorescent protein (pmaxGFP): efficiency was 5 times greater in the protein transduction-based internalization (figure 1(C)). The (−30) GFP-based transduction thus offers the possibility of intracellular detection of the POI after a few hours and with greater efficiency than the standard DNA transfection.

3.2. (−30)GFP inhibits MyoD transcription promoting activity

To further investigate the properties of the recombinant fluorescent tool and the influence of the super-negative GFP on MyoD, we examined its function measuring its transcription promoting activity by real-time PCR. We measured the mRNA expression levels

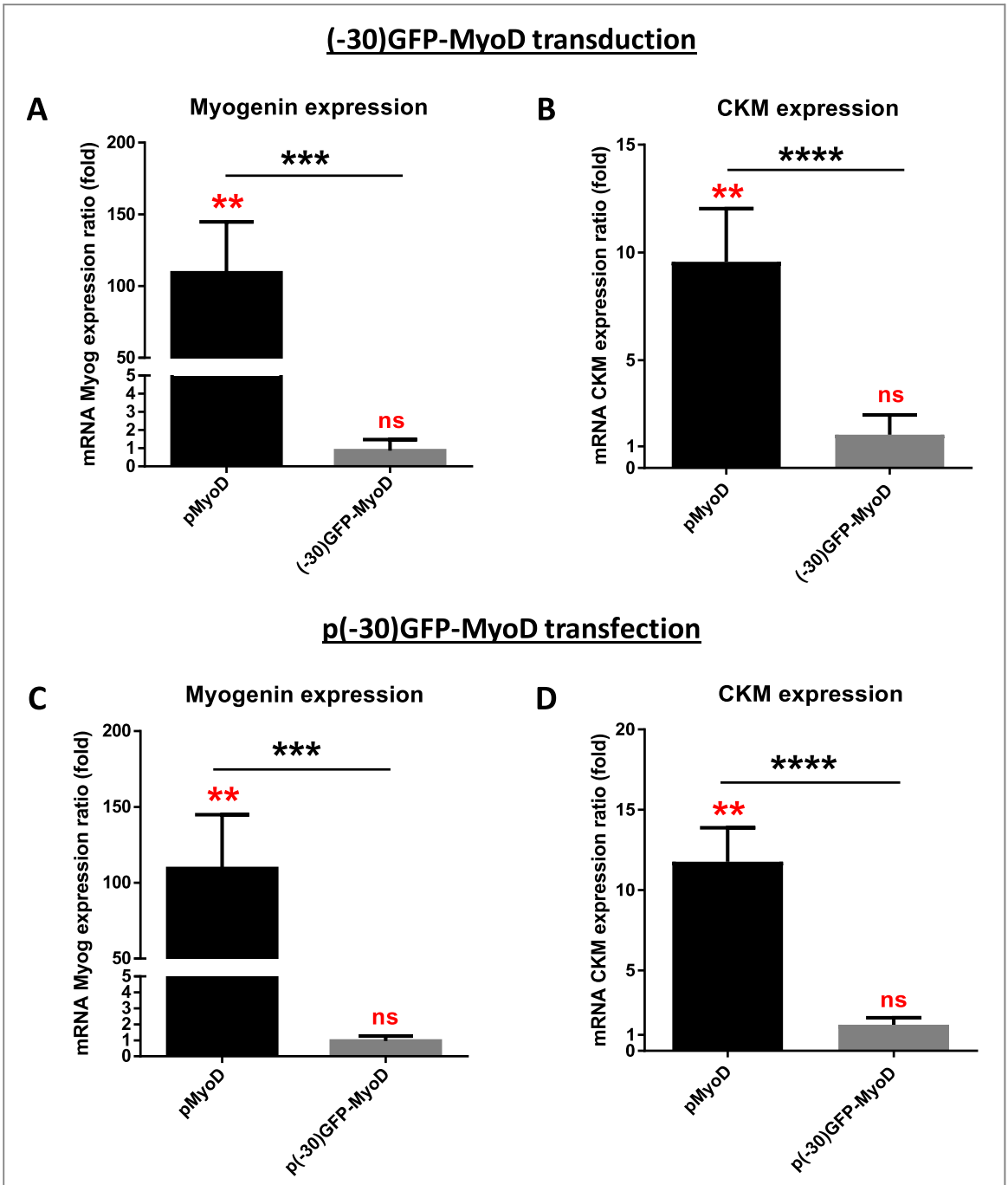

Figure 2. Graphical representation of (−30)GFP-MyoD transcription promoting activity. (A-B) (−30)GFP-MyoD represents MSCs transduced with the recombinant (−30)GFP-MyoD while pMyoD is the positive control, representing MSCs transfected with the plasmid coding for MyoD; (C-D) p(−30)GFP-MyoD shows the mRNA expression levels of MSCs transfected with a plasmid coding for (−30)GFP-MyoD. The experimental group including the calibrators was analysed by one-way ANOVA and Student's t-test was used to compare the experimental-paired groups. Data are mean ± SD. Significance symbols highlighted in red refer to the comparison of each group and its calibrator. Significance symbols in black refer to one-way ANOVA.

of four target genes of MyoD: myogenin (Myog), creatine kinase M-type (CKM), desmin and cadherin 15 (Cdh15). We extracted the total mRNA in MSCs overexpressing the native MyoD at four time points (24 h, 48 h, 72 h, 96 h) and measured the target genes expression. Among the analysed target genes, Myog and CKM were expressed at higher levels than the untransfected MSCs at 72 h (difference >90%). Thus, we selected these as the optimal genes to be analysed in order to evaluate the MyoD transcription promoting activity (see supplementary figure 2).

To analyze (−30)GFP-MyoD function, we transduced MSCs with the supernegative TF complexed with Lipofectamine™ MessengerMax. We compared the transcription promoting activity of (−30)GFP-MyoD with the function of MyoD without (−30)GFP. We used MSCs transiently transfected with the plasmid coding for the native MyoD (pMyoD) as experimental positive control. Figures 2(A) and (B) show the resulting mRNA expression levels of Myog and CKM. In both cases (−30)GFP-MyoD did not promote the transcription of any gene suggesting an

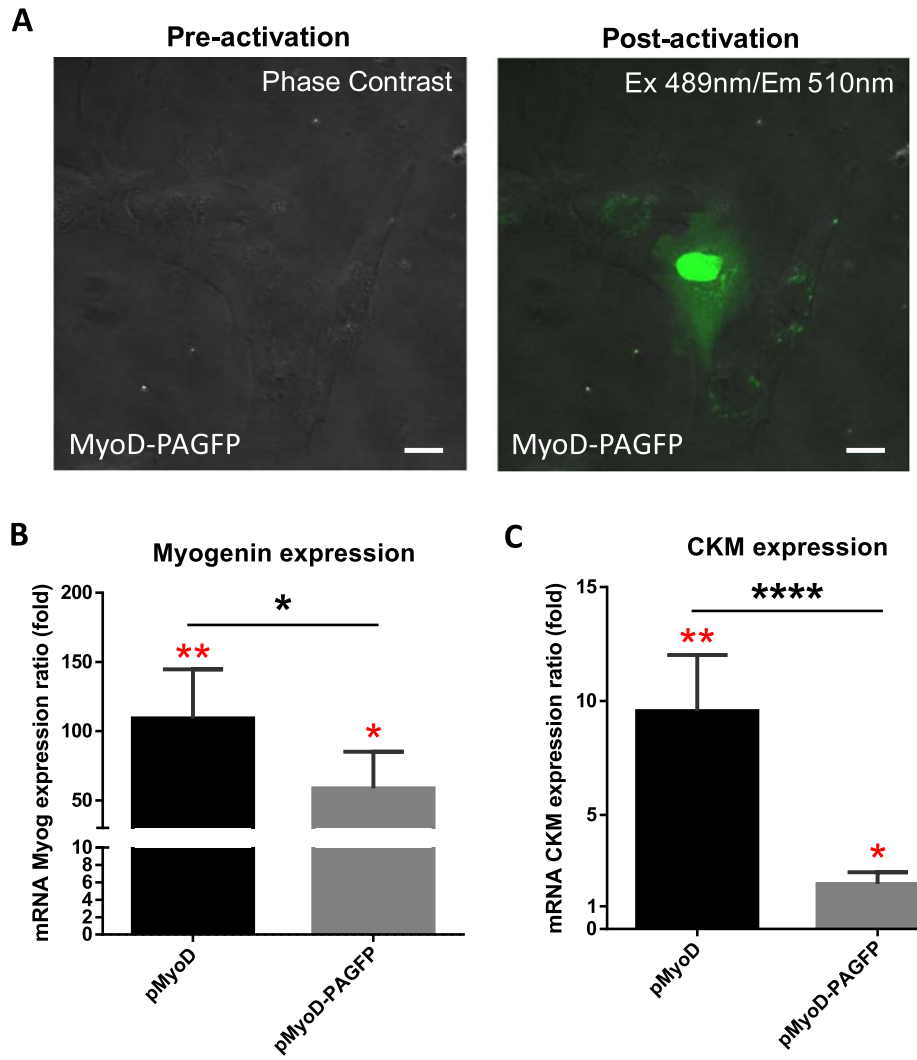

Figure 3. Transcription promoting activity of MyoD fused to PAGFP. (A) Confocal images of MSCs transfected with pMyoD-PAGFP. The left acquisition shows the pre-activated state of the photoactivatable variant of MyoD and the right image shows the fluorescent state after photo-irradiation at 405 nm; (B-C) graphical representation of Myog and CKM mRNA expression levels in MSCs transfected with the plasmids coding for MyoD or MyoD-PAGFP. The experimental group including the calibrators was analysed by one-way ANOVA and Student's t-test was used to compare the two experimental-paired groups. Data are mean ± SD. Significance symbols highlighted in red refer to the comparison of each group and its calibrator. Significance symbols in black refer to one-way ANOVA.

inhibitory effect of the (−30)GFP on MyoD activity. To exclude any problems related to protein manipulation methods, we transfected MSCs with a eukaryotic plasmid coding for the supernegative variant of MyoD and measured the mRNA expression levels 72 h after transfection. The results confirmed the inhibitory effect of the (−30)GFP on MyoD (figures 2(C) and (D)).

To check that the inhibitory effect was (−30)GFP-dependent, we fused MyoD to another GFP-based probe. We chose the photoactivatable GFP (PAGFP), a GFP variant that emits green fluorescence after photo-irradiation at 405 nm. Differently from the (−30)GFP, PAGFP cannot complex with cationic lipids and pass the plasma membrane. Thus, we obtained MSCs over-expressing the photoactivatable variant of MyoD by a standard DNA transfection procedure. We evaluated the quality of the fluorescent recombinant TF and assessed the maintenance of the physiological nuclear localization 48 h after DNA transfection (figure 3(A)). We extracted the total RNA 72 h after DNA transfection and measured its transcription promoting activity. The MyoD-PAGFP significantly promoted Myog and CKM transcription compared to PAGFP without MyoD (figures 3(B) and (C)). Even though at lower levels than the native TF (pMyoD), the activity of the photoactivatable MyoD confirmed that MyoD is suitable for engineering with fluorescent probes and that the inhibitory effect is specific to (−30)GFP.

3.3. Unraveling the inhibitory effects of (−30)GFP via molecular dynamics

We hypothesized that the (−30)GFP inhibitory effect on MyoD transcription promoting activity might be due to the interaction between the peculiar surface negative charges of the (−30)GFP, exposing a large

number of negative charges, and the positive charges characterizing the DNA binding domain of the TF, thus interfering with MyoD function. To verify this, we used molecular dynamics as a suitable computational approach for further investigation at the atomistic level. We generated two molecular models, one characterized by a MyoD BHLH domain (DNA binding domain) and (−30)GFP and the second consisting of a MyoD BHLH domain and PAGFP. Since MyoD-PAGFP was the fluorescent variant of MyoD that experimentally showed positive transcription promoting activity, the second computational model was used as control. We examined molecular dynamics trajectories by analyzing structural rearrangements and stability and non-bonded interaction energies between MyoD BHLH domain and the GFPs.

3.3.1. GFP and MyoD structural features

The 3D structures of the experimentally used (−30) GFP and PAGFP were obtained aligning the sequences to the amino acid sequence of the PAGFP deposited in the Protein Data Bank (PDB code 3GJ2). Since the sequence alignments revealed close identity in both cases (97% identity for PAGFP and 91% for (−30) GFP), we used the 3GJ2 structure as template for 3D homology modeling. Surface electrostatic potential analysis of the GFP molecular models showed a (−30) GFP surface charge of −28, close to the theoretical value of −30, and a weak negative charge of −6 for the PAGFP surface, biochemically similar to the standard GFP used by Lawrence and colleagues (−7) to generate the supernegative GFP [15].

The molecular model of the MyoD DNA binding domain was generated starting from the 1MDY.pdb available in the Protein Data Bank, which is a homodimer of the MyoD BHLH domain-DNA complex (see supplementary figure 1 A). We extracted the monomeric domain of MyoD and analyzed the residues in the dimerization domain (see supplementary figure 1B) and DNA binding domain (see supplementary figure 1 C). The residues involved in dimerization were mainly hydrophobic while amino acids that interact with DNA were mainly positively charged and belong for the most to the basic region (108–125aa) (see supplementary figure 1D).

The complete structure of MyoD has not yet been crystallized and solved by x-ray diffraction, so the 3D model of the entire structure is not available. Using the protein structure prediction servers PSI-blast-based secondary structure PREDiction (PSIPRED v3.3) and DISOrder PREdiction (DISOPRED3), we verified that the global structure of MyoD was predominantly disordered, except for the BHLH domain (data not shown).

We then generated two molecular models characterizing the (−30)GFP-MyoD and PAGFP-MyoD complexes. Since MyoD was characterized by the only structured functional core of the protein (BHLH domain), we created four complex configurations facing the MyoD domain in different positions in relation to the GFPs (see supplementary figure 1E). For each configuration we analyzed structural rearrangements pre- and post-simulation and structural stability by root mean square deviation (RMSD) calculation of trajectories and the features of the binding regions occurred between MyoD and related GFP. We also evaluated non-bonded interaction energies to quantify the interaction between the MyoD BHLH domain and the specific GFP.

3.3.2. Interaction sites and key residues

We analyzed each configuration, observing the complexes rearrangements in pre- and post-simulation conformation and the residues involved in contact regions between each GFP and the MyoD BHLH domain. Figure 4 shows the intermolecular contacts at the end of the MD simulation run of the first configuration of the MyoD-GFP complexes. The first noteworthy finding is that the monomer moved differently depending on the type of GFP in the system. In fact, as shown in the insets of figure 4 the MyoD BHLH monomer moved perpendicularly with respect to the PAGFP (figure 4(A)) but remained longitudinal and in proximity to the GFP molecule in the case of (−30)GFP (figures 4(B) and (C)). As highlighted in figures 4(A) and (B) both GFPs kept contact with the loop of MyoD, involving residues with different chemical characteristics (hydrophobic, polar and charged). Only (−30)GFP showed specific non-bonded interactions with the MyoD BHLH basic region (figure 4(C)). The interactions involved the MyoD positively charged Arg111, the (−30)GFP negative Glu170 and the (−30)GFP charge-neutral Gln169. As highlighted in supplementary figure 1 C, Arg111 is a key residue of the BHLH domain involved in the DNA binding. Except for the second configuration, in which the monomer moved around the GFPs without generating particular non-bonded interactions, the third and the fourth configurations confirmed what was seen in the first configuration. The residues of the MyoD DNA binding region involved in the interaction with the (−30)GFP of the third and the fourth configurations are Arg 110, Arg 111 and Arg 110, Arg 117, Arg 120, respectively.

3.3.3. GFP-MyoD non-bonded interaction energies

To quantify the charge-based attractive effect observed in structural analyses, we examined the non-bonded interaction energies of the molecular complexes. First we measured Van der Waals and electrostatic energies between the MyoD BHLH domain and their related GFP. The sum of the two values was considered the total non-bonded interaction energies. All the calculated energy values are reported in supplementary table 2. Figure 5(A) (and supplementary figure 3 A) described the mean energy values (Van der Waals and electrostatic energy) of the last (and first) 500 ps of the MD simulation run. As shown in figure 5(A), in each

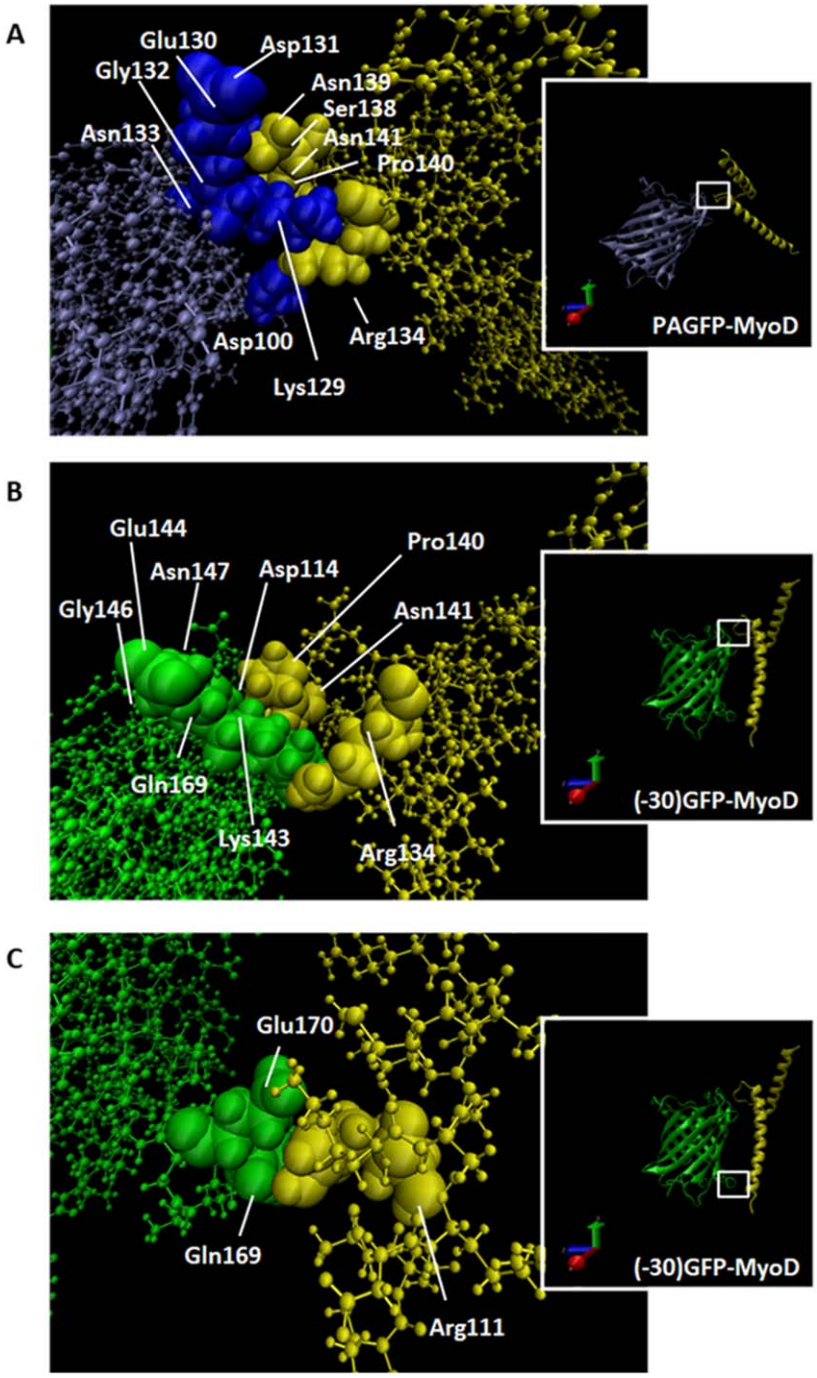

Figure 4. VMD representation of residues involved in intermolecular contacts between the fluorescent protein and MyoD after MD simulation of the first configuration. The square inside the insets on the right represents the region of the molecular model detailed on the left. PAGFP is represented in iceblue, (−30)GFP in green and the MyoD domain in yellow. PAGFP interacted with the MyoD domain at the loop region (A). (−30)GFP interacted with the MyoD domain at both the loop region (B) and the DNA-binding region (C).

of the four explored configurations, the interaction energies were negative and significantly lower in the MyoD domain and (−30)GFP than in MyoD and PAGFP. Considering that lower negative values for non-bonded interaction energies mean a greater attraction effect, these values confirmed that the MyoD domain interacted more strongly with the supernegative GFP than with the PAGFP. Analyzing the electrostatic energy contribution (blue segment) expressed by the percentage in the graph, in relation to Van der Waals contribution (orange segment), the charge-based interaction always contributed more in the case of (−30)GFP-MyoD complex. This supported the hypothesis that (−30)GFP has a strong interaction with MyoD mediated by charge-driven attractive contacts.

To understand better which part of the MyoD BHLH domain was involved in the charge-based

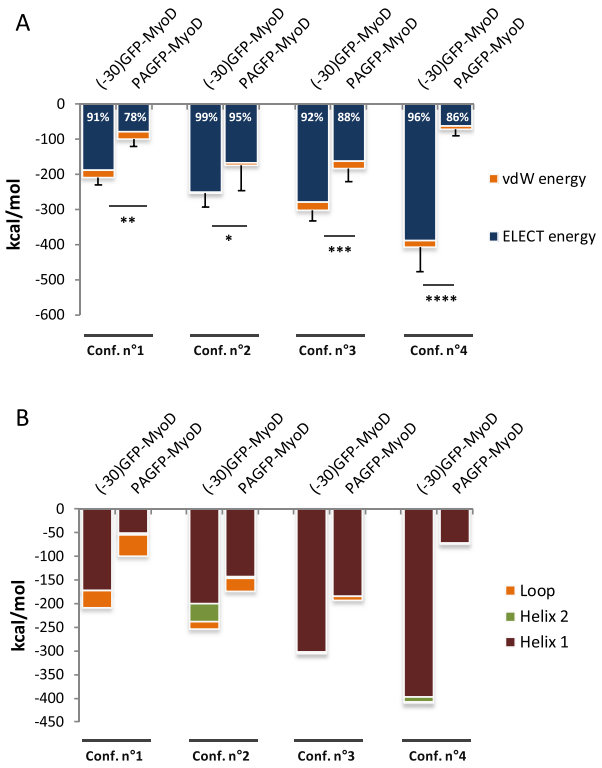

Figure 5. Comparison of interaction energies extracted from the last 500 picoseconds simulation. (A) The non-bonded interaction energy (electrostatic contribution in blue and Van der Waals contribution in orange) is calculated for the MyoD BHLH domain and the specific GFP. The percentage indicates the contribution of electrostatic energy to the total non-bonded energies. Groups were analyzed by two-way ANOVA followed by Sidak's multiple comparisons test. Data are mean ±SD. (B) Distribution of the non-bonded interaction energies calculated for the GFPs and the three characteristic elements of the MyoD BHLH domain. Helix 1 (DNA binding domain, residues 105–136); Helix 2 (dimerization domain, residues 147–166); Loop (residues 137–146). The energies were extracted from the last 500 ps simulation. Data are mean and the sum of the energies of the three elements gives the total non-bonded interaction energies between the BHLH domain and the specific GFP.

Table 1. Summary of structural analyses. X indicates when the specific MyoD region interacted with the (−30)GFP or PAGFP and if the contacts were charge-driven or not. RMSD values are mean ± SD and are calculated from last 500 ps simulation.

		INTERACTIONS				RMSD
CONFIG.	COMPLEX	DNA binding domain (Helix 1)	Loop	Dimerization domain (Helix 2)	Charge-driven	(last 500 ps)
1st	PAGFP-MyoD		X			8.08 ± 1.88
	(−30)GFP-MyoD	X	X		X	2.77 ± 0.42
2nd	PAGFP-MyoD		X			6.28 ± 0.49
	(−30)GFP-MyoD					9.94 ± 0.53
3rd	PAGFP-MyoD	X				3.68 ± 0.49
	(−30)GFP-MyoD	X		X	X	2.68 ± 0.24
4th	PAGFP-MyoD	X				7.10 ± 0.67
	(−30)GFP-MyoD	X			X	4.71 ± 0.29

contact with the (−30)GFP, we measured the non-bonded interaction energies separating the BHLH domain in the three main characteristic elements: Helix 1 (from residues 105–136), Helix 2 (from residues 147–166) and Loop (residues 137–146). As shown in figure 5(B) (and supplementary figure 3B for the first 500 ps), the MyoD characteristic element that contributed most to the interaction energy values was Helix 1, and thus the DNA binding region, confirming the structural analyses summarized in table 1.

4. Discussion

The standard procedure used to detect a POI fused to a GFP-based fluorescent probe requires the transfection of a coding DNA plasmid and its expression by eukaryotic cells. Protein transduction is an alternative to DNA transfection to internalize a purified protein and accelerate the fluorescence detection time up to 10-fold. Thanks to its surface electrostatic characteristics, (−30)GFP can be cell internalized in a few hours

and can be considered a good candidate for protein internalization. Despite the characteristics and advantages of protein transduction, featuring a reduced time for intracellular protein delivery and the control of cell loaded molecules, this internalization method has not yet been exploited to study intracellular protein tracking. With this purpose and considering (−30)GFP features, we developed a supernegative fluorescent variant of MyoD able to penetrate the plasma membrane when complexed with cationic molecules. We observed that the (−30)GFP-MyoD transduced in MSCs reached the nucleus in a few hours, with high transduction efficiency, significantly higher than standard DNA transfection methods. The working time required for (−30)GFP-MyoD transduction was comparable to other expertise-requiring protein internalization techniques, such as microinjection [23]. Moreover, our results are in line with the report by Zuris and colleagues in which they fused the (−30)GFP to different proteins and observed intracellular protein delivery both *in vitro* and *in vivo*. As Zuris demonstrated, the most efficient anionic protein: cationic lipid complex was based on the use of lipofectamines designed for RNA transfection [16]. Our data confirmed that the cationic reagent that was more efficient in complexing the (−30)GFP-MyoD was Lipofectamine™ MessengerMAX, which is specifically designed for mRNA intracellular delivery.

With a view to use this fluorescent labeling method to study rapid intracellular mechanisms, such as mechanotransduction events or protein-protein and protein-DNA interactions, we demonstrated that this (−30)GFP-based strategy considerably improves the measurement procedures by increasing the number of cells containing the fluorescent POI and reducing the experimental time. In fact, compared to standard GFP-based approaches, that require at least two working days for measurements, we were able to transduce MSCs and potentially measure protein dynamics the same day (3 h after protein transduction). As mentioned above, mechanotransduction is an example of rapid intracellular events since cells react to mechanical stimuli in a time-dependent way [24–28]. The (−30)GFP-based strategy could be useful to shorten the technical delay so as to measure the mechanotransduction event in a quick time after the mechanical input. For instance, the (−30)GFP-MyoD could be used to transduce mechanically stimulated cells and study intracellular dynamics, such as nucleocytoplasmic shuttling and nuclear permeability in response to mechanical stimuli. This approach could avoid the loss of fundamental information or unclear evaluation of the phenomenon.

To further clarify the functional properties of the supernegative recombinant factor, we analyzed the transcription promoting activity by real-time PCR. Unexpectedly, gene expression results showed an inhibitory effect of the (−30)GFP on the MyoD transcription promoting activity. Unlike the proteins fused to the (−30)GFP by Zuris and colleagues (Cas9, CRE and TALE-VP64), MyoD is a TF that can bind DNA with a specific BHLH domain characterized by a positively charged region. Our hypothesis was that the specific inhibitory effect of the (−30)GFP on the transcription promoting activity of MyoD was related to the charge-guided interference between the positive charges of the MyoD BHLH domain and the peculiar negatively charged surface of the supernegative GFP. To verify this, we used MD simulations to compare the (−30)GFP-MyoD BHLH domain complex with the PAGFP-MyoD complex; experimentally this produced a functional recombinant factor. Computational analyses confirmed that the MyoD domain is attracted and stabilized more by (−30)GFP than PAGFP through charge-driven interactions. This suggested an improper stabilization of MyoD in the presence of the (−30)GFP. Furthermore, the structural element that evidently gave the main contribution to the intermolecular interactions was the basic region of the MyoD domain, which is involved in the DNA binding. We therefore inferred that improper stabilization at the DNA binding region might alter its function by avoiding the DNA consensus sequence binding or the correct promotion of the transcriptional machinery's activity. Although the dimerization domain was not found to be negatively influenced by the supernegative GFP, the improper stabilization of the DNA binding region would not exclude an additional contribution due to a possible interference in the formation of the (−30)GFP-MyoD dimer necessary for the MyoD transcription promoting.

In light of the computational observations and experimental evidences, the DNA binding domain appears crucial for the transcriptional promoting activity of (−30)GFP-based factors. In the perspective of finding suitable differentiation TFs to be labeled with the (−30)GFP, it could be interesting to consider proteins with DNA-binding domains structurally different from the BHLH region. So far, the only (−30)GFP labeled proteins that resulted functionally active (Cas9, CRE recombinase and TALE-VP64), are characterized by peculiar and specific DNA-interacting regions, not commonly present in differentiation TFs [16]. For example, among the mesenchymal stem cell differentiation factors, two common DNA-binding domains, structurally different from the BHLH region, are the basic leucine zipper and the zinc finger domain [29]. In this respect, two factors that could represent good candidates for further development are the adipogenic CAAT/enhancer binding proteins (c/EBPs) and the osteogenic TF Osterix.

To reduce the inhibitory effect of the (−30)GFP on MyoD, different strategies are possible. Focusing on the surface electrostatic characteristics of the supernegative label, it could be interesting to engineer a GFP negative enough to complex with cationic lipids and sufficiently charge-inert not to interfere with transcription promoting activity. Actually, Zuris and

colleagues already proposed other polyanionic GFP variants with a lower net theoretical charge (−20 or −7). Both were able to promote cationic lipid-mediated delivery of a POI but in a charge-dependent manner and thus, at a lower efficiency respect to the (−30) GFP [16]. Therefore, we should optimize this strategy to achieve the best compromise between delivery efficiency and functionality influence. To this respect, to counteract the lower negative charge, an additional possibility may be to test or develop different cationic lipid nanocarrier systems with an increased net positive charge or enhanced cell membrane transducing efficiency [30].

Another alternative strategy to diminish the inhibitory effect could be to change the distribution of the 30 negative charges on the protein surface without reducing them. Following this idea, we should change the original modified residues of the starting GFP, so that the final protein would have a global net surface charge of −30 but with a different charge distribution and, thus potentially different MyoD-BHLH domain interaction dynamics. To this end, further MD studies may be useful for selecting the best negative charge distribution.

From the view point of finding a negative GFP that could be used as a label for both protein dynamics and activity-based investigations, we could screen different supernegative GFP using MD simulations to evaluate the influence of the GFP on MyoD activity and then doing experimental investigations to validate the computational results.

5. Conclusions

In conclusion, we have demonstrated that the supernegative GFP offers an efficient strategy for protein intracellular tracking studies. Although its inhibitory effect on MyoD activity could limit further functionality-based applications, the fluorescent tool does not interfere with the physiological localization of the fused POI and, in comparison to classical procedures, it markedly reduces the time needed for its detection, increasing the protein delivery efficiency. These properties make it suitable for clarifying several molecular mechanisms of unclear complex intracellular events.

Acknowledgments

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 646990 - NICHOID). The results reflect only the authors' views and the Agency is not responsible for any use that may be made of the information contained. We thank Judith Baggot for English editing.

Conflict of interest

The authors declare no conflict of interest pertaining this manuscript.

ORCID iDs

Lucia Boeri https://orcid.org/0000-0002-7334-837X

References

[1] Shimomura O, Johnson F H and Saiga Y 1962 Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea *J Cell Comp Physiol* **59** 223–39
[2] Chalfie M, Tu Y, Euskirchen G, Ward W W and Prasher D C 1994 Green fluorescent protein as a marker for gene expression *Science* **263** 802–5
[3] Day R N and Schaufele F 2008 Fluorescent protein tools for studying protein dynamics in living cells: a review *J. Biomed. Opt.* **13** 031202
[4] Toseland C P 2013 Fluorescent labeling and modification of proteins *J. Chem. Biol.* **6** 85–95
[5] Los G V *et al* 2008 HaloTag: a novel protein labeling technology for cell imaging and protein analysis *ACS Chem. Biol.* **3** 373–82
[6] Sun X *et al* 2011 Development of SNAP-tag fluorogenic probes for wash-free fluorescence imaging *Chem. Bio. Chem.* **12** 2217–26
[7] Michalet X *et al* 2005 Quantum dots for live cells, *in vivo* imaging, and diagnostics *Science* **307** 538–44
[8] Shaner N C, Patterson G H and Davidson M W 2007 Advances in fluorescent protein technology *J. Cell Sci.* **120** 4247–60
[9] Patterson G H and Lippincott-Schwartz J 2002 A photoactivatable GFP for selective photolabeling of proteins and cells *Science* **13** 1873–7
[10] Zhang M *et al* 2012 Rational design of true monomeric and bright photoactivatable fluorescent proteins *Nat. Methods* **9** 727–9
[11] Habuchi S *et al* 2005 Reversible single-molecule photoswitching in the GFP-like fluorescent protein Dronpa *Proc. Natl Acad. Sci. USA* **102** 9511–6
[12] Komatsu T *et al* 2011 Real-time measurements of protein dynamics using fluorescence activation-coupled protein labeling method *J. Am. Chem. Soc.* **133** 6745–51
[13] Loffreda A *et al* 2017 Live-cell p53 single-molecule binding is modulated by C-terminal acetylation and correlates with transcriptional activity *Nat. Commun.* **8** 313
[14] Dupont S *et al* 2011 Role of YAP/TAZ in mechanotransduction *Nature* **8** 179–83
[15] Lawrence M S, Phillips K J and Liu D 2007 Supercharging proteins can impart unusual resilience *J. Am. Chem. Soc.* **129** 10110–2
[16] Zuris J A *et al* 2015 Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo Nat. Biotechnol.* **33** 73–80
[17] Zoja C *et al* 2012 Mesenchymal stem cell therapy promotes renal repair by limiting glomerular podocyte and progenitor cell dysfunction in adriamycin-induced nephropathy *Am J Physiol Renal Physiol* **303** F1370–81
[18] Bergstrom D A, Penn B H, Strand A, Perry R L, Rudnicki M A and Tapscott S J 2002 Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression *Mol Cell* **9** 587–600
[19] Humphrey W, Dalke A and Schulten K 1996 VMD: visual molecular dynamics *J. Mol. Graph.* **14** 33–8
[20] Jones D T 1999 Protein secondary structure prediction based on position-specific scoring matrices *J. Mol. Biol.* **292** 195–202

[21] Jones D T and Cozzetto D 2015 DISOPRED3: precise disordered region predictions with annotated protein-binding activity *Bioinformatics* **31** 857–63
[22] Phillips J C *et al* 2005 Scalable molecular dynamics with NAMD *J. Comput. Chem.* **26** 1781–802
[23] Vandromme M, Cavadore J C, Bonnieu A, Froeschlé A, Lamb N and Fernandez A 1995 Two nuclear localization signals present in the basic-helix 1 domains of MyoD promote its active nuclear translocation and can function independently *Proc. Natl Acad. Sci. USA* **92** 4646–50
[24] Delaine-Smith R M and Reilly G C 2012 Mesenchymal stem cell responses to mechanical stimuli *Muscles Ligaments Tendons J* **2** 169–80 https:/ncbi.nlm.nih.gov/pmc/articles/PMC3666521/
[25] Zhang L *et al* 2008 Time-related changes in expression of collagen types I and III and of tenascin-C in rat bone mesenchymal stem cells under co-culture with ligament fibroblasts or uniaxial stretching *Cell Tissue Res* **332** 101–9
[26] Elosegui-Artola A *et al* 2017 Force triggers YAP nuclear entry by regulating transport across nuclear pores *Cell* **171** 1397–410
[27] Garcia A, Rodriguez Matas J F and Raimondi M T 2016 Modeling of the mechano-chemical behaviour of the nuclear pore complex: current research and perspectives *Integr Biol (Camb)* **8** 1011–21
[28] García-González A, Jacchetti E, Marotta R, Tunesi M, Rodríguez Matas J F and Raimondi M T 2018 The effect of cell morphology on the permeability of the nuclear envelope to diffusive factors *Front Physiol* **9** 925
[29] Boeri L, Albani D, Raimondi M T and Jacchetti E 2019 Mechanical regulation of nucleocytoplasmic translocation in mesenchymal stem cells: characterization and methods for investigation *Biophys. Rev.* **11** 817–31
[30] Dharmalingam P *et al* 2017 Green transfection: cationic lipid nanocarrier system derivatized from vegetable fat, palmstearin enhances nucleic acid transfections *ACS Omega.* **2** 7892–903