Project Title:

CGSENS: Visualization of CG methylation using a fluorescence protein biosensor

Report Title:

6month report of CGSENS: Visualization of CG methylation using a fluorescence protein biosensor

Summary

Cytosine DNA methylation is an epigenetic mark critical in diverse biological functions, such as gene regulation or genome stability. So far, the cellular epigenetic landscape has been pictured using *in vitro* approaches that provide information about DNA methylation at specific loci or genome-wide, but lacking an *in vivo* spatiotemporal resolution.

We propose the design of a **new generation of biosensor (CGSENS) based on bimolecular fluorescence complementation (BiFC) to study the global CG DNA methylation dynamics**. CGSENS will consist of a methyl-CpG binding domain (MBD) linked to the amino (CGSENS-N) or carboxyl (CGSENS-C) end of a split VENUS. We address these tasks:

1.- Design and chemically synthesize CGSENS
2.- Express and purify CGSENS recombinant protein from *Escherichia coli*
3.- Characterise the interaction of CGSENS and DNA by *in vitro* binding experiments

So far, we have achieved 2/3 of the objectives. We have successfully demonstrated the viability of CGSENS as biosensor. Using enriched fractions of CGSENS-C and CGSENS-N, we are able to detect a specific fluorescence signal emitted by the reassembled CGSENS bound to methylated CG. We want to improve the purification protocol to repeat the *in vitro* assays quantifying the interaction of equimolecular amounts of both proteins and methylated genomic DNA.

Report and outcomes

AIM: To test the *in vitro* interaction based on BiFC of CGSENS and methylated DNA to study global CG DNA methylation.

OBJECTIVE 1.- **Design and chemically synthesize CGSENS.**

Following the principle of VENUS-based BiFC[1] we designed two constructs named **CGSENS-N** and **CGSENS-C** (Figure 1.A). These constructs encode the methylated DNA (methyl-CpG) binding domain (MBD) of hsMBD10 fused in-frame with the N-terminal (CGSENS-N, 1–172 aa) and C-terminal (CGSENS-C, 155–238 aa) fragments of VENUS (Figure 1). We chose the human MBD domain instead of the Arabidopsis one because it has been published as functional when connected to a linker and a split luciferase[2]. The MBD domains are conserved across kingdoms, and synthetic proteins from plants have been proved to be fully functional in mammals[3]. The MBD domain was joined to its corresponding VENUS fragment by Glycine-Serine linker $(GGGGS)_6$. We included a His tag to purify the proteins by affinity chromatography. We also included a Factor Xa recognition sequence to facilitate tag removal after purification to avoid unspecific interactions (Figure 1.B).

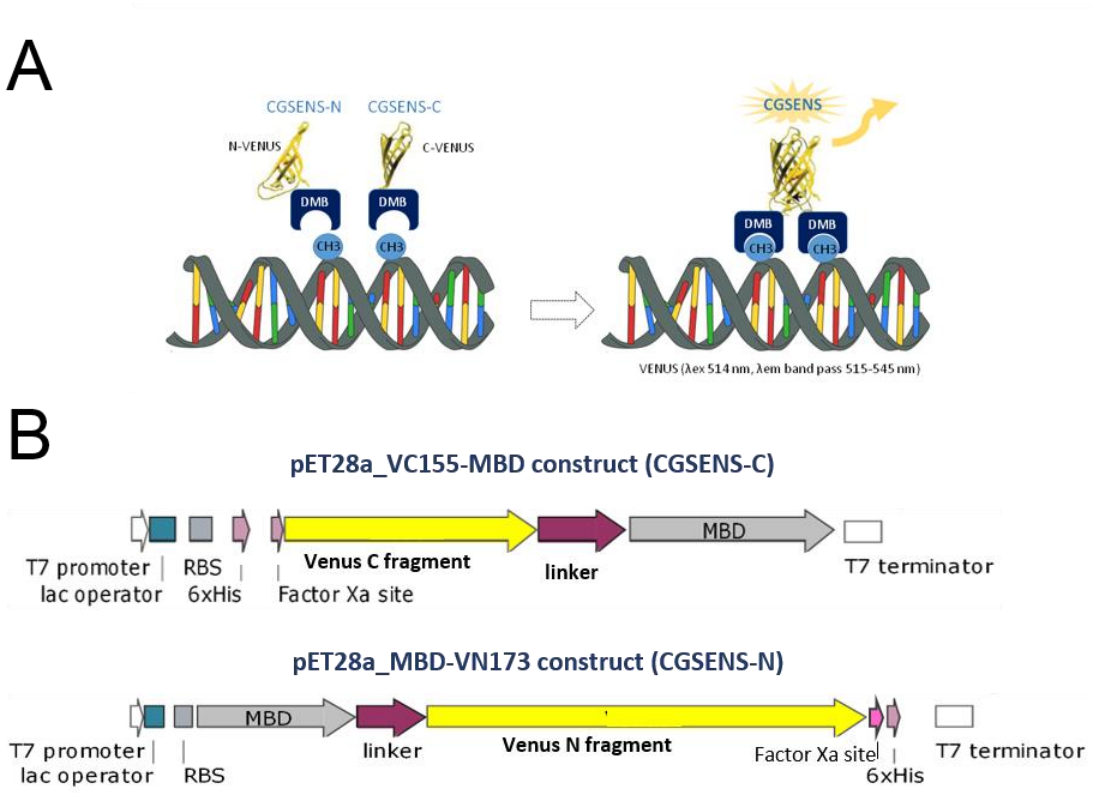

Figure 1- Schematic representation of the biosensor CGSENS.

1.A- Recognition of methylated cytosines in CG context by both CGSENS-C and CGSENS-N. When methylated CG are in close proximity, CGSENS-C and CGSENS-N will be close enough for the VENUS domains to interact and reassemble a functional protein CGSENS, that will emit fluorescence when excited at the adequate wavelength. 1.B- Schematic representation of the transcriptional units cloned in expression plasmid pET28a to obtain (upper panel) plasmid pET28a_VC155-MBD coding for CGSENS-C and (bottom panel) plasmid pET28a_VN173-MBD coding for CGSENS-N.

Both constructs were synthesized and codon optimized for expression in *E. coli,* and subcloned into vector pET28a (Novagen) by BaseClear (The Netherlands). We realised that the price drop in gene synthesis made moneywise to order the plasmids from BaseClear already subcloned in the expression vector. A summary of the constructs, and some physical and chemical properties of the fusion proteins are summarised in Table1.

Name	Construct	aa	MW (KDa)	Theoretical pI	Extinction coefficients*
CGSENS-C	pET28a_VC155-MBD	206	22	7.80	21555
CGSENS-N	pET28a_MBD-VN173	289	31	7.19	33140

Table 1- Physical and chemical properties of CGSENS-C and CGSENS-N obtained using the ProtParam tool from Expasy. (https://web.expasy.org/protparam). *Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm wavelength water.

OBJECTIVE 2.- **Expression and purification of recombinant CGSENS proteins from *E. coli*.**

We have expressed the recombinant proteins in *E. coli*, inducing the expression at 22 C° by adding IPTG at early exponential cultures. After induction, overexpression of proteins was checked on SDS-PAGE (Figure 2.A). We were able to detect CGSENS-N but we could not identify CGSENS-C, indicating that the protein expression was low. As both CGSENS-C and CGSEN-N contain a His tag (Figure 1.B), we used an affinity chromatography (Ni^{2+}NTA column) to purify the recombinant proteins. This type of chromatography is very specific so we expected to detect CGSENS-C after elution from the column, despite its low expression. We successfully purified CGSENS-N, but for CGSENS-C we could not see an enriched band. We did another test to check for the presence of CGSENS-C in the cell extract, using the same purification protocol but loading equivalent volumes of cell extracts containing CGSENS-C and CGSENS-N in the affinity column. We were able not only to demonstrate the presence of CGSENS-C in the cell extract but also, and most importantly, that both proteins were correctly folded and able to reassemble VENUS when they are in close proximity (Figure 2.B).

In order to validate this result, eluates from columns loaded with crude extracts of CGSENS-C, CGSENS-N and the combination of CGSENS-C + CGSENS-N were tested using a fluorescence plate reader. Fluorescence signal was obtained only for those fractions were both crude extracts were loaded in the column, confirming VENUS reassembly (Figure 2.C). We could further confirm the presence of both CGSENS-C and CGSENS-N by SDS-PAGE (Figure 2.D).

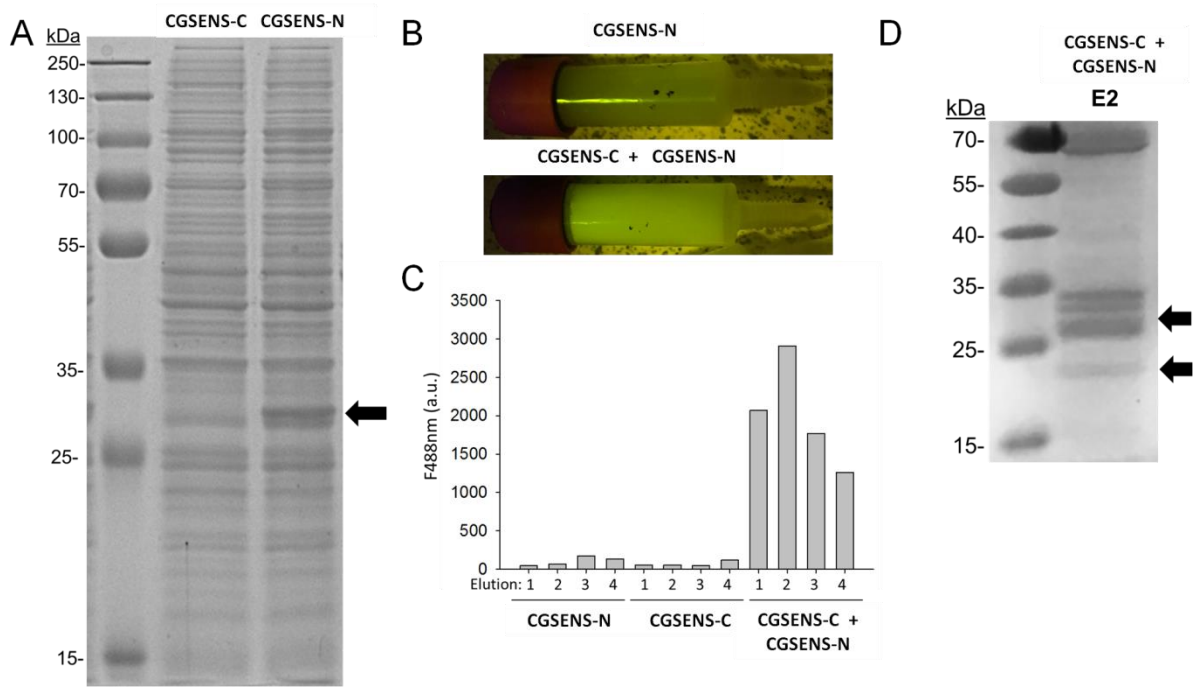

Figure 2- Purification and reassembly test of CGSENS-C and CGSENS-N.

2.A- 1ml of induced cultures of *E. coli* containing plasmid pET28a_MBD-VN173 (CGSENS-N) or pET28a_VC155-MBD (CGSENS-C) were resuspended in 100 µl 6x PAGE sample buffer and 10 µL were loaded on a 4-12% SDS-PAGE. We detected a band of the expected size of CGSENS-N (MW=31kDa) but not of CGSENS-C (MW= 22 kDa). 2.B- Cells were lysated using the EmulsiFlex-C5 Centrifuge at 10,000 x g for 20–30 min at 4°C to pellet the cellular debris. The supernatant was loaded into HisTrap Columns repacked with precharged Ni Sepharose. After excitation at 514 nm we did not detect any fluorescence from the

protein loaded in the column for CGSENS-N (upper panel) or CGSENS-C (not shown) on their own, but when both crude extracts were loaded in the same column fluorescence signal was detected (bottom panel). 2.C- Columns loaded with crude extracts of CGSENSE-N, CGSENSE-C and CGSENSE-N + CGSENS-C were eluted using an imidazole gradient (20 mM-500 mM). At 350-400 nM imidazole, a peak at 280 nm indicated the presence of protein (elution fractions 1-4). 100 µL of each of these fractions were assayed in a fluorescence plate reader (VENUS-λex 514 nm, λem band pass 515-545 nm). 2.D –SDS-PAGE of the eluate 2 (20 µL) from the column loaded with the combination of crude extract from CGSENSE-N + CGSENS-C, showing two bands with expected sizes of CGSENS-C (MW= 22 KDa) and CGSENS-N (MW=31KDa).

We hypothesize that the purification of CGSENS-C is improved in the presence of CGSENS-N because the protein is retained in the column not only by affinity of the His tag but also by affinity with the complementary VENUS domain. We will monitor the presence of the protein by Western blot assays of the crude extracts, flow through and eluates. Once we have troubleshooted the purification protocol, if the purification level is not satisfactory we may need to redesign the plasmid for CGSENS-C. We will deposit the final versions of the plasmids in Addgene at that point.

OBJECTIVE 3.- **Characterisation of the interaction of CGSENS and DNA by *in vitro* binding experiments**.

We were not successful in purifying both proteins to homogeneity, but we were able to reassemble CGSENS (Figure 2.B, 2.C) and therefore to validate the presence of both proteins, correctly folded, in their corresponding crude extracts. Thus, we decided to perform the DNA *in vitro* binding experiments with crude extracts of *E. coli* overexpressing CGSENS-N or CGSENS-C.

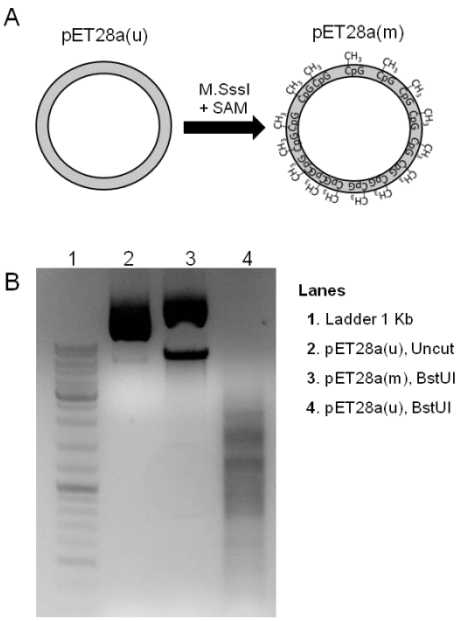

Figure 3- DNA target preparation: *In vitro* methylation of plasmid DNA.
3.A- Natively-methylated plasmid pET28a(u) was treated with M.SssI a CpG methyltransferase and S-adenosylmethionine (SAM) to yield a fully methylated plasmid, pET28a(m). 3.B- 1µg pET28a(u) and 1µg pET28a(m) were digested with *BstU*I endonuclease, which cleaves the symmetric 5'-CGCG-3' site only in the absence of methylation. *BstU*I fully degrades unmodified pET(u), while the fully methylated pET(m) is not digested.

As target DNA we used plasmid DNA. The naturally-methylated plasmid was methylated using a CpG methyltransferase (Figure 3.A). The process was monitored using the methylation sensitive endonuclease *BstU*I. When the plasmid has been fully methylated, *BstU*I is not able to cleave its target

sequence (Figure 3.B). Using this protocol we were able to generate a plasmid fully methylated (m) or carrying only native methylation (u). From the restriction pattern obtained from *BstU*I treatment, we can conclude that the amount of native methylation in our assay is practically null.

We initially tested the biosensor for its ability to bind methylated CG in the presence of unmethylated or fully methylated plasmid DNA. We incubated equivalent amounts of extracts containing CGSENS-C or CGSENS-N in the presence of unmethylated or fully methylated DNA, as described previously, and recorded fluorescence for 7 hours on a fluorescence plate reader (Figure 4.A). Fluorescence emission from CGSENS in the presence of unmethylated DNA was equivalent to the signal obtained from the buffer, meaning that not fluorescence emission is produced. On the other hand, when CGSENS was incubated in the presence of methylated DNA, the fluorescence signal rapidly increased, indicating that CGSENS is able to recognise specifically methylated DNA. The maximum difference between the signals obtained from CGSENS in the presence of unmethylated or methylated DNA was achieved at 2h and remained practically constant until the end of experiment, indicating that the time frame of our experiment was correct and probably could be reduced to 2-3h (Figure 4.B).

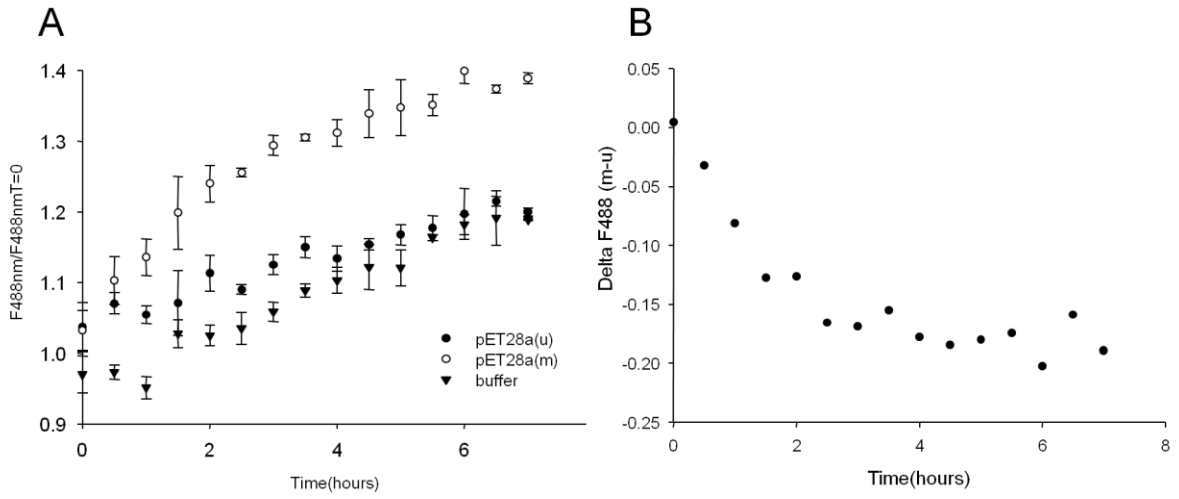

Figure 4- DNA target preparation: *In vitro* methylation of plasmid DNA.
4.A- 1 µg of unmethylated (u) or methylated (m)pET28a plasmid were incubated with equivalent amounts of crude extracts of bacteria overexpressing CGSENS-C and CGSENS–N in a final volume of 200 µL for 7 hours and monitored on a plate reader. 4.B- Fluorescence variation between methylated and unmethylated DNA in the presence of CGSENS was plotted against time. The maximum difference was reached at about 2 h and remained practically constant for the rest of the experiment.

Having purified recombinant GCSENS-N and CGSENS-C, we will test the biosensor for its ability to bind methylated CG in the presence of unmethylated or fully methylated genomic DNA. Titration experiments will be done to evaluate the functional dynamic range of CGSENS using a mix of different proportions of fully methylated and unmethylated DNA to obtain a range from 0-100% methylated DNA.

Expenditure

PRODUCT	COMPANY	CODE	PRICE
Gene synthesis for order #093516 (two plasmids and subcloning)	Baseclear	93516	GBP 412.00
NiSO4	SIGMA	656895-106	GBP 44.70
Thrombin	SIGMA	T4648-1KU	GBP 44.10
M. Sss1	NEB	M0226S	GBP 62.00
***BstU*I**	NEB	R0518S	GBP 53.00
PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa	ThermoFisher	26619	GBP 110.00
InstantBlue Protein Stain	EXPEDION	ISB1L	GBP 105.00
CpGenome™ Rat Methylated and Unmethylated Genomic DNA Standard Set	Merck	S7865	GBP 299.00
Ni-NTA Spin Kit (50 Spin columns)	Qiagen	31314	GBP 367.00
Ni-NTA Superflow (25 ml nickel-charged resin)	Qiagen	30450	GBP 332.00
NuPAGE™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 10-well (2 boxes)	Invitrogen	NP0321BOX	GBP 244.00
Anti-6X His tag® antibody [HIS.H8]	Abcam	Ab18184	GBP 285.00
Factor- Xa Protease 250 μg (1mg/ml)	NEB	P8010L	GBP 228.00
Xa Removal Resin (2 x 2.5 ml)	Qiagen	33213	GBP 350.00

Table 2- Summary of reagents used so far in the project and their cost.

Follow on Plans

As it has been stated in the report, the purification of CGSENS-C has been challenging. We expect to troubleshoot the protocol and be able to obtain enough amount of protein to perform the titration experiments with genomic DNA methylated at different levels. We may need to redesign the original plasmid. Our initial results show that CGSENS is functional and detects methylated DNA *in vitro*. After confirming these results, we would like to try *in vivo* experiments. To do so, we would like to perform *in vivo* transient expression in *Nicotiana benthamiana* and monitor the emission of fluorescence by confocal microscopy. We still have some remaining funds but **we would like to apply for the extra GBP 1000** to cover some additional reagents that we may need to buy (see table below), to synthesise two plasmids for *in vivo* assays and to pay for confocal microscopy service. We expect to have data about the functional dynamic range of CGSENS with genomic DNA, and the results of *in vivo* assays in plants in **six months**. Below (Table 3) is the intended used of the remaining and additional funds.

PRODUCT	COMPANY	CODE	PRICE
Gilson PIPETMAN Neo 8 channel P8x20N 2-20uL	Gilson	Ab18184	GBP 285.00
Gene synthesis for order #093516 (two plasmids and subcloning)	Baseclear		
CpGenome™ Rat Methylated and Unmethylated Genomic DNA Standard Set	Merck	S7865	GBP 299.00
NuPAGE™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 10-well (2 boxes)	Invitrogen	NP0321BOX	GBP 244.00
PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa	ThermoFisher	26619	GBP 110.00

Table 3- Intended used of the remaining and additional funds.

Changes to team

No changes were made to the original team.

References

1. Ohashi, K., Kiuchi, T., Shoji, K., Sampei, K. & Mizuno, K. Visualization of cofilin-actin and Ras-Raf interactions by bimolecular fluorescence complementation assays using a new pair of split Venus fragments. *Biotechniques* **52,** 45–50 (2012).

2. Badran, A. H. *et al.* Evaluating the Global CpG Methylation Status of Native DNA Utilizing a Bipartite Split-Luciferase Sensor. / *Anal. Chem* **83,** 7151–7157 (2011).

3. Ingouff, M. *et al.* Live-cell analysis of DNA methylation during sexual reproduction in Arabidopsis reveals context and sex-specific dynamics controlled by noncanonical RdDM. *Genes Dev.* **31,** 72–83 (2017).