Bioorganic & Medicinal Chemistry Letters 19 (2009) 3748–3751

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Systematic evaluation of split-fluorescent proteins for the direct detection of native and methylated DNA

Jennifer L. Furman [a], Ahmed H. Badran [a], Shengyi Shen [a], Cliff I. Stains [a], Jack Hannallah [a], David J. Segal [b], Indraneel Ghosh [a,*]

[a] Department of Chemistry, University of Arizona, 1306 E. University Blvd, Tucson, AZ 85721, USA
[b] UC Davis Genome Center and Department of Pharmacology, 4513 GBSF, 451 Health Sciences Dr, Davis, CA 95616, USA

ARTICLE INFO

Article history:
Received 16 February 2009
Revised 29 April 2009
Accepted 30 April 2009
Available online 5 May 2009

Keywords:
Split-protein
GFP
Zinc-finger
Sensor
DNA
Methylation
Detection

ABSTRACT

In order to directly detect nucleic acid polymers, we have designed biosensors comprising sequence-specific DNA binding proteins tethered to split-reporter proteins, which generate signal upon binding a predetermined nucleic acid target, in an approach termed SEquence-Enabled Reassembly (SEER). Herein we demonstrate that spectroscopically distinct split-fluorescent protein variants, GFPuv, EGFP, Venus, and mCherry, function effectively in the SEER system, providing sensitive DNA detection and the ability to simultaneously detect two target oligonucleotides. Additionally, a methylation-specific SEER-Venus system was generated, which was found to clearly distinguish between methylated versus non-methylated target DNA. These results will aid in refinement of the SEER system for the detection of user defined nucleic acid sequences and their chemical modifications as they relate to human disease.

Specific DNA sequences are typically detected by denaturation of double stranded (ds)DNA followed by hybridization with a labeled oligonucleotide probe, as is the case with Southern blotting, fluorescence in situ hybridization (FISH),[1] and DNA microarrays.[2] In contrast to these approaches, new methods have been developed to directly detect dsDNA, which include designed systems that rely on sequence-specific recognition in the grooves of the native DNA double helix, including hairpin polyamides[3] and triplex forming oligonucleotides (TFOs).[4] Currently the utility of these techniques is constrained by limitations on the length of dsDNA that can be targeted and restrictions in the constitution of detectable sequences, respectively.[5] In comparison to these design efforts, endogenous sequence-specific DNA binding proteins do not suffer from these drawbacks, while additionally providing a method for nucleic acid detection in its native environment. In our work we have focused on Cys_2-His_2 zinc fingers (ZFs), which comprise an important class of naturally occurring transcription factors and provide a programmable means of DNA detection.[6,7] In our method derived from split-protein assays also called protein-fragment complementation, fragments of a split-reporter protein are appended to DNA detection domains, and the binding event is monitored by signal generation arising from conditional reassembly of split-protein halves. Numerous proteins and enzymes have been genetically fragmented including ubiquitin,[8] dihydrofolate reductase,[9] green fluorescent protein (GFP),[10] β-lactamase,[11] and luciferase.[12] GFP is particularly appealing as it singularly lacks any requirement for cofactors or substrates.[13] By fusing split-GFP to sequence-specific ZF domains, recognition of target DNA induces GFP reassembly in a method called SEquence-Enabled Reassembly (SEER) (Fig. 1).[14]

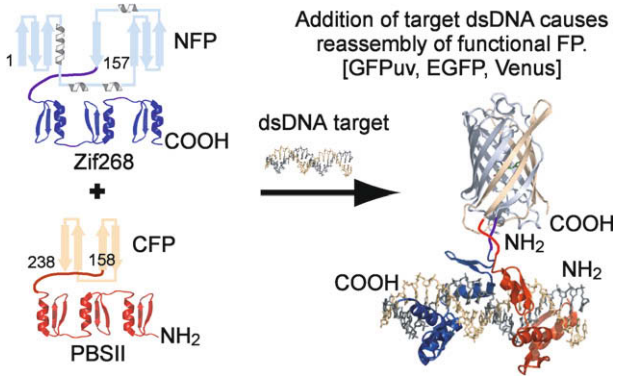

Figure 1. SEquence-Enabled Reassembly. The FP fusion constructs are initially non-fluorescent. The introduction of target DNA results in ZF binding, which induces reassembly of a productively fluorescent FP.

* Corresponding author. Tel.: +1 520 621 6331; fax: +1 520 621 8407.
E-mail address: ghosh@email.arizona.edu (I. Ghosh).

0960-894X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2009.04.141

Currently, our reported SEER systems can specifically detect the presence of a single oligonucleotide sequence at a time.[14–16] New SEER systems capable of simultaneously detecting multiple targets would be useful for numerous applications, including DNA profiling and ratiometric analysis. The GFP reporter affords the opportunity to accomplish this objective since GFP variants with distinct spectroscopic properties have been extensively studied.[17] At present the repertoire of proteins derived from wild type *Aequorea victoria* GFP includes blue, cyan, and yellow fluorescent proteins (BFP, CFP, and YFP, respectively). Although a red-only emitting FP has not been achieved using the *A. victoria* scaffold,[18] naturally occurring red fluorescent proteins (RFPs) have been identified. The so-called mFruits, monomeric DsRED derivatives, were recently selected and have since been employed as fusion tags in a variety of organisms, offering additional opportunities for protein complementation assays.[19,20] Thus, we constructed SEER systems incorporating five distinct fluorescent protein (FP) variants: a UV-excitable GFP, GFPuv;[21] a CFP, Cerulean;[22] an enhanced GFP, EGFP;[23] a YFP, Venus;[24] and a DsRED derived RFP, mCherry[19] (Fig. 2, left panels). Each N-terminal FP fragment (residues 1–157 for GFP-derived FPs and 1–168 for mCherry) was fused to ZF Zif268, while ZF PBSII was attached to each C-terminal FP fragment (residues 158–238 for GFP-derived FPs and 169–231 for mCherry). In all cases the split-FP domain was connected to the DNA binding domain via a flexible $[(Gly)_4\text{-}Ser]_3$ linker. Following isolation of these constructs, each SEER-FP system was assayed for preferential refolding in the presence of an optimized DNA target, *Zif268-0-PBSII*, by monitoring fluorescence emission of the reassembled complex. We were unable to generate a functional SEER-Cerulean system due to difficulties encountered during expression and purification of NCerulean-Zif268. However, all other tested split-FP constructs functioned effectively in the SEER context (Fig. 2). Of note, split-mCherry was capable of DNA templated reassembly, thus providing the first success of a DsRED derived FP in the SEER context. Each SEER-FP system was optimized with respect to protein concentration, and a DNA titration (Fig. 2, right panels) established sensitivities of the systems, which were qualitatively equivalent, and each reassembled FP produced signal over background at 10 nM (1.0 pmol) target DNA.

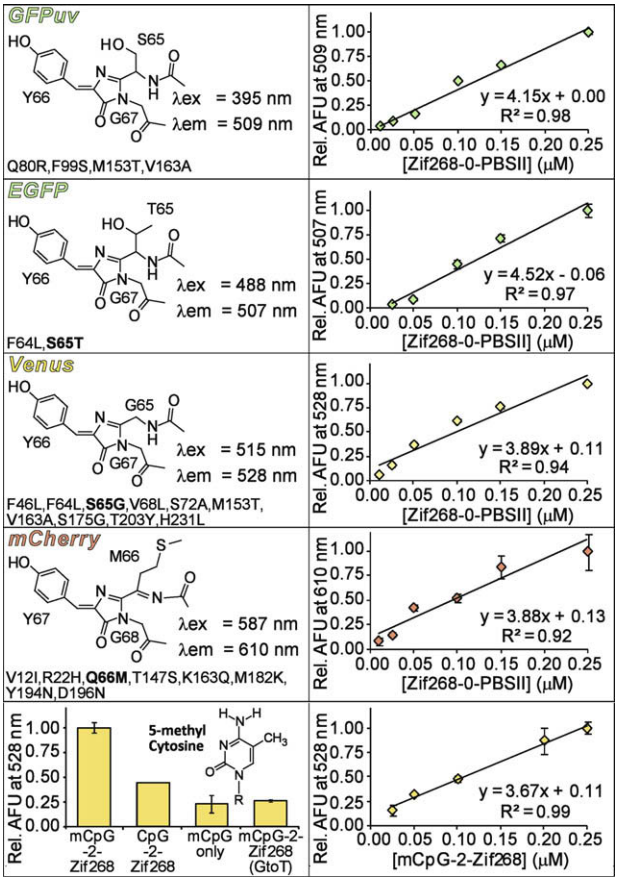

Figure 2. Properties of FPs utilized in SEER. The left panels display chromophore structures and wavelengths of maximum excitation (λ_{ex}) and emission (λ_{em}) for each of the full length FP variants. Mutations (*A. victoria* numbering) are listed below the corresponding structures with chromophore mutations indicated in bold. mCherry includes additional GFP-type residues on both its N- and C-terminus.[19] The right panels show a DNA titration for each SEER-FP system. The bottom panels show specificity data and a DNA titration for the mCpG-SEER system.

Due to the high fluorescence intensity of the SEER-Venus constructs, these proteins were selected to modify for detection of cytosine methylation, since this covalent DNA modification has emerged as a promising cancer biomarker.[25] By simply replacing ZF PBSII with a methyl cytosine-guanine dinucleotide (mCpG) binding domain, MBD2, we were able to generate a system called mCpG-SEER-Venus for site-specific evaluation of the methylation status at individual CpG islands. We previously characterized two mCpG-SEER systems that conferred specificity for methylation at CpG sites, one of which employed an *A. victoria* FP variant.[26,27] To assess the specificity of the new system, we refolded NVenus-Zif268 and CVenus-MBD2 in the presence of target and off-target oligonucleotides (Fig. 2, bottom panel). The mCpG-SEER-Venus system was shown to clearly distinguish between an mCpG site and its non-methylated equivalent with a 2.3-fold preference. This value seems slightly low since MBD2 has a 70-fold preference for binding methylated CpG islands (K_d = 2.7 nM) over the corresponding non-methylated site.[28] An improvement in specificity for methylated sites may be achieved through further optimization of the mCpG-SEER-Venus system. Additional controls revealed a 4.3-fold reduction in fluorescence signal upon removal of the *Zif268* binding site in the target (*mCpG* only) and a 3.8-fold reduction for a target with only a single guanine to thymine mutation in the *Zif268* binding site (*mCpG-2-Zif268 (G to T)*). These results demonstrate the ability of the mCpG-SEER-Venus system to discriminate between cognate target sites and non-methylated, non-specific, or mutated DNA sequences. Finally, a DNA titration curve was generated for the mCpG-SEER-Venus system, resulting in detection of at least 25 nM DNA target (Fig. 2, bottom panel).

The construction of several functional SEER-FP systems provides the potential for simultaneous detection of multiple DNA sequences, which we are currently pursuing with several new ZF DNA binding domains. An interesting alternative is the possibility of complementing different halves of the split-FPs to generate hybrid FPs with distinct spectral properties, which has previously been demonstrated for the detection of protein-protein interactions.[29] We first attempted to visualize two different DNA sequences by complementing NVenus with both CVenus and CGFPuv (Fig. 3A). The three fusion proteins were allowed to refold in the presence of one or two target oligonucleotides. Excitation and emission wavelengths were systematically scanned to identify major peaks. A fluorescence emission signal at 506 nm was attributed to the NVenus/CGFPuv hybrid and indicated the presence of the *Zif268-0-PBSII* target, while a 528 nm emission signal resulted from NVenus reassembling with CVenus in the presence of the *Zif268-0-PE8B* target (Fig. 3B). When both targets were simultaneously present in solution, two distinct fluorescence emission maxima were observed by alternating the wavelength of excitation from 395 to 515 nm, corresponding to the formation of two distinct reassembled FPs. Only a minimal degree of off-target fluorescence was observed as emission at 506 nm in the absence of the *Zif268-0-PBSII* target (<20%) and 528 nm emission in the absence of the *Zif268-0-PE8B* target (<15%). Therefore, by using Zif268 to