Published online 17 August 2007

Nucleic Acids Research, 2007, Vol. 35, No. 16 ***e107***
doi:10.1093/nar/gkm618

Live cell imaging of repetitive DNA sequences via GFP-tagged polydactyl zinc finger proteins

Beatrice I. Lindhout[1], Paul Fransz[2], Federico Tessadori[2], Tobias Meckel[3], Paul J.J. Hooykaas[1] and Bert J. van der Zaal[1,*]

[1]Institute of Biology Leiden, Department of Molecular and Developmental Genetics, Clusius Laboratory, Leiden University, Wassenaarseweg 64, 2333 AL, Leiden, [2]Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 SM, Amsterdam and [3]Physics of Life Processes, Leiden Institute of Physics, Leiden University, Niels Bohrweg 2, 2333 CA, Leiden, The Netherlands

Received June 13, 2007; Accepted July 28, 2007

ABSTRACT

Several techniques are available to study chromosomes or chromosomal domains in nuclei of chemically fixed or living cells. Current methods to detect DNA sequences *in vivo* are limited to *trans* interactions between a DNA sequence and a transcription factor from natural systems. Here, we expand live cell imaging tools using a novel approach based on zinc finger-DNA recognition codes. We constructed several polydactyl zinc finger (PZF) DNA-binding domains aimed to recognize specific DNA sequences in *Arabidopsis* and mouse and fused these with GFP. Plants and mouse cells expressing PZF:GFP proteins were subsequently analyzed by confocal microscopy. For *Arabidopsis*, we designed a PZF:GFP protein aimed to specifically recognize a 9-bp sequence within centromeric 180-bp repeat and monitored centromeres in living roots. Similarly, in mouse cells a PZF:GFP protein was targeted to a 9-bp sequence in the major satellite repeat. Both PZF:GFP proteins localized in chromocenters which represent heterochromatin domains containing centromere and other tandem repeats. The number of PZF:GFP molecules per centromere in *Arabidopsis*, quantified with near single-molecule precision, approximated the number of expected binding sites. Our data demonstrate that live cell imaging of specific DNA sequences can be achieved with artificial zinc finger proteins in different organisms.

INTRODUCTION

Microscopical detection of chromosomal DNA is an important tool in understanding the function of genomic sequences in the interphase nucleus. In chemically fixed cells or tissues we can visualize any chromosomal domain, including individual gene loci, using standard fluorescence hybridization methods (1). However, to observe the dynamics of chromosomes in living cells, such fixation-based techniques are not suitable. Consequently, alternative strategies have been developed allowing monitoring of chromosomal domains in living systems. The first reports on *in vivo* imaging of DNA targets involved systems in which fluorescent proteins, such as GFP, were fused to a DNA-binding domain of a transcription factor, in this case a *tet* or *lac* repressor (2–5). When introduced into cells, that contain a recombinant locus carrying a series of cognate binding sites, the GFP-tagged transcription factors bound to their targets and enabled *in vivo* localization of the recombinant locus. However, since introduction of an artificial target locus into the genome usually occurs at random positions, the technique is in fact inapplicable for *in vivo* imaging of specific endogenous genomic loci. As an alternative to this inherent problem, GFP fusions have been made with proteins that are known to be present in chromatin. For example, in various organisms GFP-tagged histones have been visualized in chromatin of living cells (6–11). In addition, more specific chromosomal loci were monitored *in vivo* with GFP-fused proteins that are known to localize at specific chromatin domains, such as the centromere (12–15).

The techniques mentioned above have clearly demonstrated the intriguing possibilities for *in vivo* visualization of DNA sequences and/or chromatin structures by means of fluorescent proteins. In order to expand the application of *in vivo* chromosome labeling, a larger variety of tools is required. In that respect, recent developments regarding the construction of artificial DNA-binding proteins based on Cys_2His_2 zinc finger (ZF) domains are of great interest. This rapidly developing technology is based upon the established 3-bp DNA recognition code of Cys_2His_2 ZFs (16) and the possibility to construct polydactyl zinc finger (PZF) proteins by fusion of

*To whom correspondence should be addressed. Tel: +31 71 527 4760; Fax: +31 71 527 4999; Email: b.j.v.d.zaal@biology.leidenuniv.nl

individual ZF moieties in such a manner that the number of fingers fused determines the length of the cognate DNA recognition site. For example, three-fingered (3ZF) PZF domains will recognize 9- bp target sequences (3×3 bp), while six-fingered (6ZF) PZF domains in principle interact specifically with 18-bp sites (6×3 bp). Although the complete recognition code for all possible 64 3-bp sequences has not yet been elucidated and not all ZF-DNA contacts are of high affinity and specificity, PZF technology already allows construction of artificial transcription factors as well as novel site-specific nucleases (17; http://www.zincfingers.org/). We hypothesized that construction and expression of sequence-specific DNA-binding PZF:GFP proteins should enable live cell imaging of repetitive DNA sequences within a given genome. Here, we present the use of PZF:GFP reporter proteins in the model plant *Arabidopsis thaliana* as well as in mouse cells. Centromeric repeats were readily observed in both model species. Furthermore, PZF:GFP content was quantified in living *Arabidopsis* cells via indirect comparison to the fluorescent signal of single eGFP molecules. The sensitivity and further applications of the technique are discussed.

MATERIALS AND METHODS

Constructs

PZFs for binding to the selected sequences were constructed in pSKN-SgrAI, as described previously (18) after which the PZF encoding sequences were cloned as SfiI fragments into the plant expression vector pRF-GFP or pcDNA-GFP (see subsequently). Vector pRF-GFP was obtained by modifications of pGPTV-KAN (19), including removal of NotI and SfiI sites from the vector backbone, replacing the promoterless GUS coding sequence by a 1.7-kb XmaI-SacI fragment containing the *RPS5A* promoter (20), and insertion of sequences providing the vector with an ATG translational start codon, a FLAG tag, a SV40 nuclear localization signal (NLS) and an in frame coding sequence of mGFP6 + (HIS)6, an enhanced GFP version (F64L, S65T) with a C-terminal HIS tag (21). For expression in mouse cells, an identical GFP-containing reading frame was introduced into the widely used expression vector pcDNA3 (Invitrogen), forming pcDNA-GFP. This involved the removal of the polylinker via HindIII/ApaI digestion, which was replaced by a PCR fragment containing an ATG translational start codon, a FLAG tag, a SV40 NLS, SfiI sites for directional cloning of PZF domains and GFP. Outlines of the constructs are given in Figure 1. Details of construction and the final vector sequence are available upon request. The construct encoding mRFP:HP1α (a gift of Martijn S. Luijsterburg) was based on the backbone encoding for pEGFP-C1 in which EGFP was replaced by mRFP (22). The HP1α sequence (23) was cloned in the EcoRI/BamHI sites of mRFP-C1, resulting in an mRFP:HP1α fusion protein.

Plant material and transformation

The plasmids encoding constructed PZF:GFP fusion proteins were mobilized into *Agrobacterium tumefaciens* strain AGL1 (24). PZF:GFP proteins for the centromeric repeat and the 5S rDNA repeat were expressed in *Arabidopsis* ecotype Columbia. Line A of ecotype Zurich (25), containing a silent hygromycin resistance locus, was used for experiments involving the *HPT* targeted PZF. Transformation of *Arabidopsis* was carried out by floral dip (26) and primary transformants were selected on solidified MA solid medium (27) without sucrose and containing kanamycin (50 mg/l) for selection and timentin and nystatin (both 100 mg/l) to inhibit bacterial and fungal growth, respectively. Primary transformants were further grown under greenhouse conditions and allowed to self pollinate.

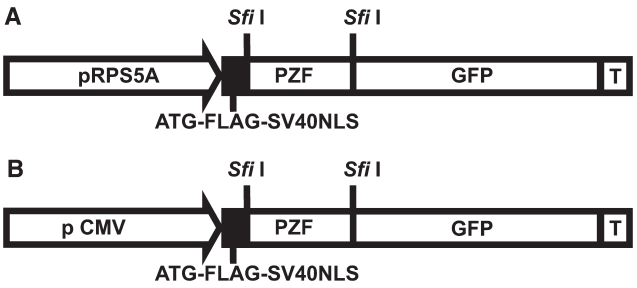

Figure 1. Schematic representation of PZF:GFP expression cassettes for *Arabidopsis* and mouse cells. (**A**) PZF:GFP fusion proteins are expressed from the *RPS5a* promoter in *Arabidopsis*. (**B**) For expression in mouse cell cultures PZF:GFP proteins are expressed from the *CMV* promoter. T = transcriptional terminator. Drawings are not to scale.

Mouse cell line and transfection

NIH 3T3 mouse cells were grown in Dulbecco's modified Eagle medium (DMEM; Invitrogen) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Invitrogen) and 100 units/ml of a penicillin/streptomycin antibiotic mixture (Invitrogen). Culture conditions were 37°C, 5% CO_2.Cells were co-transfected with plasmid DNA using Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions and were analyzed by confocal microscopy 24–48 h post transfection.

Live cell imaging of PZF:GFP fusion proteins

For analysis of PZF:GFP expression in *Arabidopsis*, T2 populations expressing *PZF:GFP* chimeric genes were sown on solidified 1/2 MS medium containing kanamycin (30 mg/l) and sucrose (10 g/l) and grown vertically for 5–8 days prior to analysis via confocal microscopy using either a BIO-RAD confocal microscope equipped with a combined Ar/Kr laser (excitation at 488 nm and emission filter 522DF32) or a Leica confocal setup (see subsequently). Mouse cells were examined 24–48 h post transfection. For *in vivo* DNA staining, DRAQ5 (Biostatus Ltd) was added to the culture medium to a final concentration of 1 μM. All nuclei were recorded with a Zeiss LSM 510 (Zeiss, Oberkochen, Germany) confocal laser scanning microscope. For eGFP, an argon 488 nm laser and a 505–550 nm band pass filter was used. For mRFP, a 543 nm helium-neon laser and a 585–615 nm band pass filter was used. DRAQ5 was

recorded using a 633 nm HeNe laser and an LP 650 emission filter.

Quantification of fluorescent signals

Fluorescent intensities in confocal images of plant root nuclei were quantified in relation to intensities of yellow–green fluorescent styrol spheres with a diameter of 20 nm (FluoSpheres, F8787, Molecular Probes, Leiden, The Netherlands). Nuclei and FluoSpheres were imaged with a confocal microscope (Leica TCS SP, Leica Microsystems GmbH, Heidelberg, Germany) directly one after another to ensure identical imaging conditions. Samples were excited at 488 nm and fluorescent emission was collected with a Leica PlanApo 100×1.4 NA objective between 505 and 650 nm using an RSP500 filter. In order to relate these intensities to discrete numbers of GFP molecules, intensities of individual FluoSpheres were compared to single recombinant eGFP (BioVision Inc., Mountain View, CA, USA). The setup for single-molecule imaging has been described in detail (28). Briefly, samples were mounted onto an inverted microscope (Axiovert 100TV, Zeiss, Oberkochen, Germany) equipped with a 100× oil-immersion objective (NA 1.4, Zeiss, Oberkochen, Germany) and illuminated for 5 ms by an ArKr-laser (Spectra Physics, Mountain View, CA, USA) at a wavelength of 488 nm. Illumination intensity was set to 2 kW/cm2. This permitted the detection of individual fluorophores by a nitrogen-cooled CCD-camera system (Spec-10:400B/LN, Princeton Instruments, Trenton, NY, USA).

Stock solutions of FluoSpheres and eGFP (2% solids and 1 μg/μl, respectively) were diluted 10^5 and 10^6 times in PBS, respectively, which contained 1% (w/v) poly vinyl alcohol (PVA, Roth Chemicals, Germany). Samples were then embedded in PVA films on cleaned glass slides by spin coating 50 μl of the dilutions on a spin coater (Model P6700, SCS, Indianapolis, IN, USA) with a two-step protocol: 10 s at 300 r.p.m. and 1 min at 2000 r.p.m. The thickness of these films was on average 1 μm, as revealed by confocal microscopy (29).

After image acquisition, a 2D Gaussian was fitted to the diffraction limited spots (Figure 5B) to obtain their integrated intensities. Histograms of the latter were then fitted with a sum of Gaussians of the form:

$$f(x) = \sum_{1}^{i} A_i e^{-1/2\left(x - ix_c/w\sqrt{i}\right)^2} \quad \mathbf{1}$$

to find the mean intensities for the single particles and their aggregates (Figure 5). Since the centromeric regions were larger than the confocal volume, their integrated intensities had to be determined from 3D scans. The axial step size was adjusted to

$$w_{FWHM_z}\sqrt{\frac{\pi}{(2\ln 4)}} \quad \mathbf{2}$$

in order not to overestimate their integrated signal. For each confocal slice of a centromeric region, the background signal outside the nucleus was subtracted from the integrated intensity of the region and summed up subsequently.

RESULTS AND DISCUSSION

PZF design for repetitive target sites

Although by now most of the amino acid sequences needed for the specific recognition of a particular 3-bp triplet by Cys_2His_2 ZFs are known, most practical applications still heavily lean on the use of the relatively robust ZF-GNN recognition code (16,17). Also in our laboratory, assembly of PZF domains based on the ZF-GNN code proved to be successful (18). To assess the potential of PZFs for live cell imaging, three repetitive loci from the *Arabidopsis* genome varying in length and copy number were chosen to be targeted by PZF modules. Within these loci, we selected the longest non-interrupted $(GNN)_n$ sequences as PZF binding sites, taking into account that a possible target site overlap should not occur (16).

The centromeric 180-bp repeat, also known as pAL1, Atcon or 178-bp repeat, was selected because of its high copy number presence exclusively in the 0.6–1.2 Mb centromeric regions of *Arabidopsis* chromosomes (30,31). Within the 180-bp repeat sequence, a single 9-bp target site 5′-GTTGCGGTT-3′ was chosen for 3ZF PZF design. Considering (i) that the centromeric region also contains multiple copies of the 106B repeat (32–33) and (ii) that the exact 9-bp target site for the PZF is not necessarily present in all 180-bp repeats, we estimate that the centromeric region accommodates about 2000 copies of the 9-bp target sequence of the PZF fusion protein, 180:GFP.

For the second repeat, we selected 5S ribosomal DNA (5S rDNA). In the accession Columbia, the 5S rDNA repeat predominantly consists of repeating units of 497 bp that are located in the peri-centromeric regions of chromosomes 3, 4 and 5, with an estimated total copy number of about 1000 per genome (32,34). Within the 497-bp sequence, we selected the sequence 5′-GCCGTGGGT-3′, located upstream of the transcribed region, as a target site for the PZF fusion protein 5S:GFP. We estimate that each of the three major 5S rDNA loci contain 300–350 copies of the 5S repeat. However, due to sequence variation between blocks of repeats at the different loci (35), we postulate a maximal number of 200 binding sites per 5S locus for the 5S:GFP fusion protein.

The third genomic PZF target that was chosen in *Arabidopsis* consisted of a transgenic locus which contains 5–10 intact copies of a silenced *HPT* gene, as well as multiple incomplete and rearranged copies (25). The repeat locus is visible as a heterochromatic knob on chromosome 1 (Fransz,P., unpublished data) and appears in interphase nuclei as a mini chromocenter (36). Within the *HPT* sequence, it was possible to select a $(GNN)_6$ target site: 5′-GTCGGAGACGCTGTCGAA-3′. Six-fingered 6ZF moieties generally have a much higher affinity for DNA targets than 3ZF domains (37). Anticipating that visualization of a relatively small number of repeating units might prove to be rather demanding, we chose to construct the corresponding 6ZF HPT:GFP in this case.

In order to test whether PZF:GFP-mediated chromosomal labeling was also possible in other higher eukaryotes, we constructed a 3ZF PZF domain aimed to bind

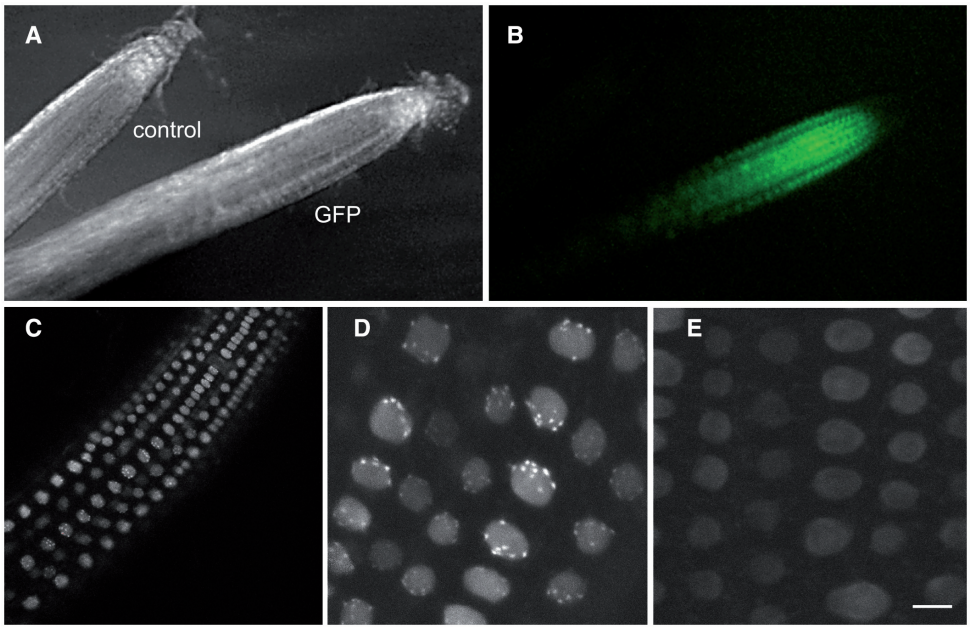

Figure 2. PZF:GFP expression in *Arabidopsis* roots. (**A**) Root tips from a GFP-transformant (GFP) and an untransformed (control) plant under white light and (**B**) under blue excitation. Only the transformant shows green fluorescence in the meristematic cells. (**C**) Confocal image of a root tip expressing 180:GFP, showing GFP-fluorescent nuclei. (**D**) Magnification of cells shown in (C) reveals up to 10 brightly fluorescing spots per nucleus. (**E**) In comparison, the control GFP construct without PZF domain shows a diffuse fluorescence throughout the nucleus. Scale bar = 5 microns. (C–E) are Z stacks of optical sections.

to the major satellite (MaSat) repeats within the pericentromeric region of mouse chromosomes (38). The 234-bp MaSat units, present in 1000–10 000 copies per chromosome (39) are rather AT-rich, but the majority of them should contain the $(GNN)_3$ target site 5′-GGCGAGGAA-3′.

Expression of PZF:GFP genes in *Arabidopsis* root tips

Transgenes encoding the PZF:GFP fusion proteins 180:GFP, 5S:GFP and HPT:GFP were expressed under the control of the *RPS5A* promoter, derived from the ribosomal *S5A* gene (20). This promoter is predominantly active in meristematic tissues, in particular in root tips. A major advantage of the root for live cell imaging is the absence of autofluorescent chloroplasts. An N-terminal SV40 nuclear localization signal ensured targeting of the PZF:GFP fusion proteins to the nucleus. For all constructs, several independent primary transformants were maintained to obtain T2 seeds. Seedlings of the T2 populations were used for fluorescence microscopy analysis. Lines containing a nuclear-targeted GFP construct without a PZF domain were used as a control.

All transfortmants, including the GFP control, showed a clear fluorescence in the root tip, especially in the meristematic region (Figure 2). Close examination revealed a specific nuclear localization of the signals (Figure 2C–E). Only in the 180:GFP transformants (Figure 2C and D) we observed bright spots of GFP fluorescence against a faint diffuse fluorescent background. Confocal microscopical examination revealed up to ten 180:GFP-fluorescent spots at the periphery of the nucleus (Figure 3A). The number and the spatial position of the spots matched very well with the number and position of chromocenters in *Arabidopsis*. These heterochromatic domains accommodate all major repeats including the centromeric 180-bp repeat and other pericentromeric repeats such as transposons (40). In order to establish co-localization of 180:GFP with chromocenters, we stained the root tip with propidium iodide. Confocal sectioning clearly demonstrated the co-localization of the 180:GFP spots with propidium iodide positive chromocenters (Figure 3B and C). This co-localization indicates that the three finger PZF of the 180:GFP construct recognizes the centromeric 180-bp repeats. In support of this, the 180:GFP spots are in a symmetrical position in several juxtaposed cells within a longitudinal cell file (Figure 2C and D). This is in agreement with the symmetrical position of centromeres in daughter cells immediately after mitosis. Since the 9-bp target sequence is only present as a repetitive array in the centromere region, we conclude that 180:GFP indeed binds to its target. A note of interest is disappearance of the fluorescent spots from the chromocenters, requiring a sequential procedure. This suggests that the binding of the PZFs to the centromere is not stable under these conditions. A brief fixation with formaldehyde could not prevent the dissociation of the 180:GFP molecules from the chromatin. We observed the same feature with other DNA dyes such as DAPI, Hoechst and Sytox® Orange (Invitrogen). Furthermore, the absence of a specific pattern in metaphase cells suggests that the 180:GFP construct does not bind to the centromere during mitosis.

The experiments with the 5S:GFP and the HPT:GFP fusion proteins did not generate a specific pattern of the

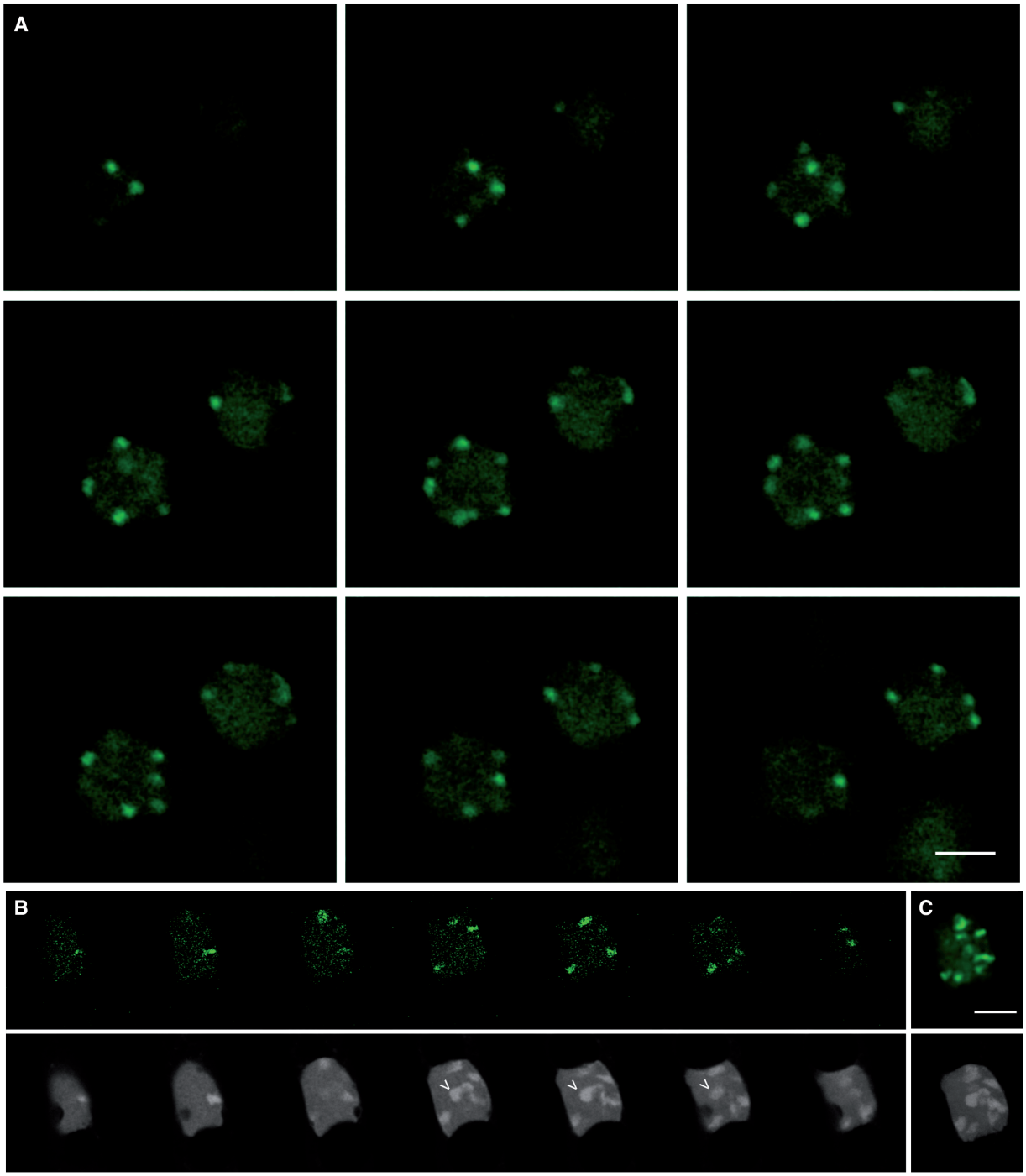

Figure 3. Live cell imaging of 180:GFP labeled nuclei. (**A**) Serial optical sections showing labeled centromeres at the periphery of nuclei. (**B**) 180:GFP fluorescent spots (upper file) correspond with chromocenters visualized with propidium iodide (lower file). The arrowhead (>) marks the central position of the propidium iodide-positive nucleolus. (**C**) Z-stack of the optical sections shown in (**B**), illustrating the co-localization of 180:GFP fluorescent signals and propidium iodide stained chromocenters. Scale bar = 5 microns.

nuclear GFP signal but only a diffuse fluorescent stain similar to the one in control transformants expressing a GFP protein lacking a PZF domain (Figure 2E). This suggests that either binding of the PZF construct to these targets does not occur or that the number of the target sites is too small to be discriminated from the diffuse background.

Expression of PZF:GFP fusion proteins in mouse cells

Arabidopsis proved to be an attractive system to investigate the potential of PZF:GFP in live cell imaging of repetitive sequences. Following the experiments in the plant model system, we addressed the question if the same approach is applicable to animals. We therefore used a

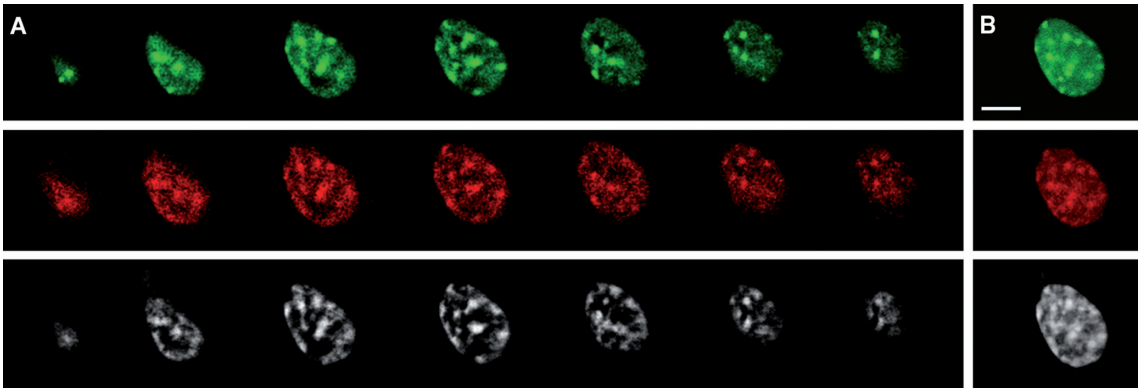

Figure 4. Live cell imaging of MaSat:GFP in NIH 3T3 mouse cells. (**A**) Serial optical sections of a 3T3 cell nucleus expressing MaSat:GFP (green) and mRFP-HP1α (red). The cell was stained with DRAQ5 (white) to visualize DNA. Note the co-localization of DRAQ5-intensely stained chromocenters with regions of high mRFP-HP1α and MaSat:GFP fluorescence. (**B**) Z-stack of the optical sections shown in (A). Scale bar = 5 micron.

three finger PZF:GFP to detect the major satellite (MaSat) repeat sequence in mouse. This pericentromeric sequence is present in 1000–10 000 copies of 234 nt units per chromosome (39) and localizes to mouse chromocenters. These heterochromatin regions are condensed DNA domains that are positive for the heterochromatin protein HP1α (38,41).

The *CMV*-driven MaSat:GFP fusion protein (Figure 1) was expressed in NIH 3T3 cells. Like in the *Arabidopsis* system, live cell imaging was readily established. Twenty to forty intense fluorescent spots per nucleus were observed for MaSat:GFP. To demonstrate the co-localization of MaSat:GFP with chromocenters, we co-transfected the cells with both MaSat:GFP and a plasmid expressing mRFP:HP1α. In addition, we stained the cells with the vital DNA stain DRAQ5. The GFP spots co-localized with both the mRFP:HP1α and the DRAQ5 positive domains (Figure 4), demonstrating that MaSat:GFP proteins bind to expected chromosomal targets. We conclude that the PZF-GFP method is applicable to animal cells. Moreover, this is the first report of *in vivo* visualization of the mouse major satellite repeat and hence opens new roads to live cell imaging of chromosomes.

Quantification of 180:GFP content per centromere

In order to further characterize the *in vivo* binding of PZF proteins, the number of 180:GFP molecules per centromere was quantified via the GFP signal. Since the CLSM setup used for observation of living cells did not provide single molecule sensitivity, we first used a single molecule microscopy (SMM) setup (28) to gauge the fluorescence intensities of 20 nm FluoSpheres against single eGFP molecules. Intensity histograms revealed that a FluoSphere and a single eGFP had an average intensity of 1971 ± 49 and 422 ± 4 counts, respectively (Figure 5C and D), thus setting the fluorescent signal of a FluoSphere equivalent to that of 4.67 ± 0.12 eGFP molecules. From these calibrations, the number of 180:GFP molecules in centromeric spots as observed with CLSM (such as in Figure 3) was calculated by comparing the fluorescent signals from centromeres and FluoSpheres present within the same sample. When measured with the Leica CLSM setup, the signal due to one FluoSphere resulted in 311 ± 2 counts (Figure 5A) and the average integrated intensity of centromeric spots was found to be $52\ 357 \pm 11\ 932$ counts ($n = 46$), thus equivalent to 168 ± 38 FluoSpheres and corresponding to a signal of 786 ± 179 eGFP molecules. In the same nuclei, the average integrated intensity of the non- structured background signal within the same volumes as occupied by the centromeric spots was found to be $17\ 522 \pm 6915$ counts, corresponding to 263 ± 104 GFP molecules. Assuming the same background also within the centromeres, about 500 (524 ± 207) extra 180:GFP molecules thus seemed to be bound to targets inside each of the centromeres. We thus conclude that a significant fraction of the estimated 2000 binding sites is accessible for 180:GFP binding. The use of FluoSpheres proved to be valuable to quantify GFP concentrations even on microscope setups lacking single molecule sensitivity.

Detection limits of PZF:GFP-mediated visualization of DNA sequences

In those meristematic cells where 180:GFP-labeled centromeric regions were clearly visible, the nuclear GFP background fluorescence corresponded to a mean GFP concentration of 0.56 ± 0.27 μM. Similar nuclear background values were found for seedlings expressing the 5S:GFP and HPT:GFP constructs. As could be inferred from CLSM pictures and 3D stacks, the average 180:GFP-labeled centromeric region resembled an elliptical cylinder with minor and major diameter and height of 304 ± 47, 792 ± 130, 1038 ± 172 nm, respectively, with an apparent volume of 0.78 ± 0.22 fl. With 786 ± 179 eGFP molecules present within this volume, the apparent eGFP concentration per centromeric region would be 1.66 ± 0.60 μM. The successful visualization of fluorescent signals from the centromeric 180-bp repeats was thus based upon an apparent 3-fold increase in concentration of 180:GFP molecules. Given this fact at a mean 0.56 μM concentration of nuclear PZF:GFP, lack of detectable

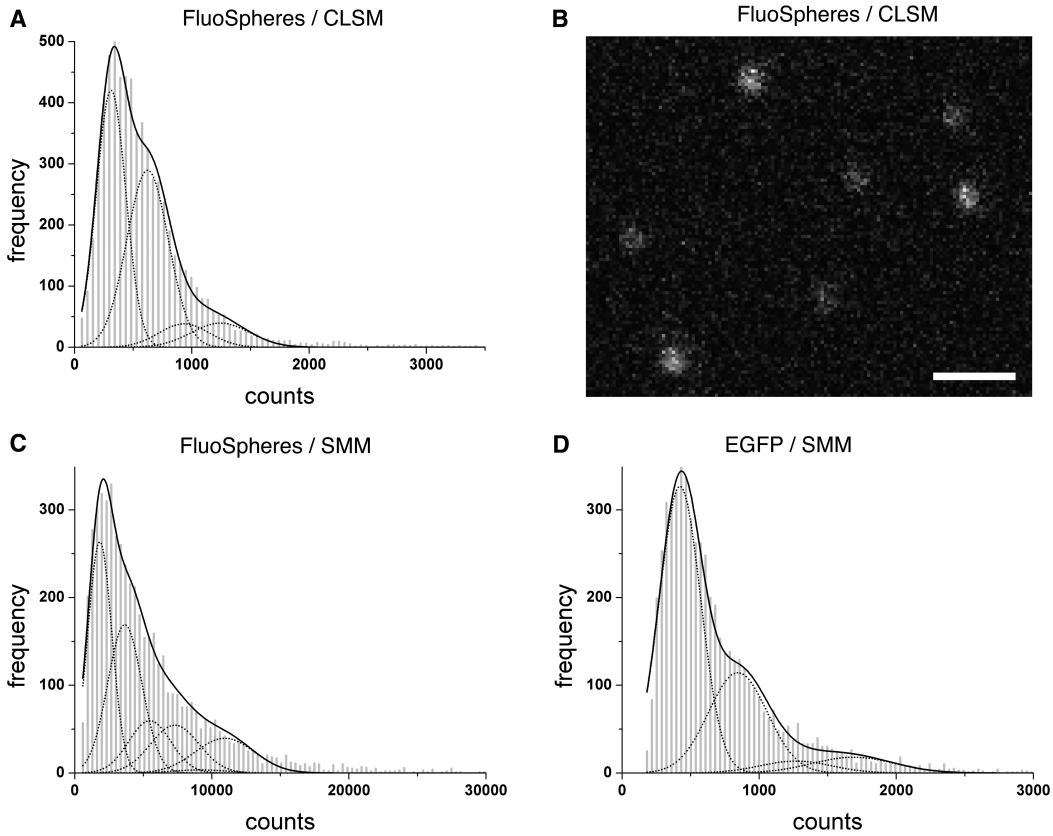

Figure 5. Quantification of fluorescent signals. (**A**) Intensity distributions of FluoSpheres with a diameter of 20 nm measures with CLSM and (**C**) with SMM. (**D**) Intensity distributions of single eGFP molecules measured with SMM. Intensity distributions were fit to a sum of Gaussians. Besides the signals of single particles, all distributions also clearly show signals of aggregates (i.e. dimers, tetramers and higher-order aggregates). The individual Gaussians (dotted lines) of the sum functions (solid lines) illustrate the separate contributions of monomers and aggregates to the distributions, The mean count rates for the respective single particles (i.e. the peak position of the first individual Gaussian), were found to be: 311 ± 2 (A), 1971 ± 49 (C) and 422 ± 4 (D). (**B**) A CLSM image of FluoSpheres illustrates the diffraction limited signals. Scale bar = 1 micron.

specific signals due to PZF:GFP binding to less abundantly repeated structures as the 5S rDNA and *HPT* sequences thus seemed to be a logical consequence of our experimental conditions.

When just considering the absolute physical concentration of DNA target sites within a condensed chromatin structure, thus with six nucleosomes and 1200 bp of DNA per 10 nm of the higher order solenoid 30 nm fiber (42), repetitive binding sites occurring once per nucleosome have a concentration of ~1.5 mM. Although such a high concentration suggests that very high signal to noise ratios should be obtained by means of PZF:GFP-mediated repeat labeling, the point spread function of CLSM setups unavoidably generates blurred pictures with volumes that are easily several hundred-fold larger than that of the real physical objects with a diameter of 30 nm. The observed 3-fold difference in fluorescent intensity of the 180:GFP-labeled centromeres compared to the background signal agrees very well with this optical phenomenon.

In theory, construction of higher affinity PZF domains should aid in signal detection when the nuclear PZF:GFP concentration is decreased accordingly. However, when applied for visualization of rather rare repeating DNA sequences it will also be more demanding to discriminate fewer bound molecules from those that are still freely diffusing.

Practical considerations regarding PZF:GFP-mediated DNA labeling

With dissociation constants for 3ZF PZF domains and their cognate 9-bp binding sites ranging between 10 and 75 nM (37), a nuclear PZF:GFP concentration around 0.5 μM seems to be a reasonable compromise to ensure that most of the accessible $(GNN)_3$ sequences can be bound by PZF:GFP molecules while still allowing visualization of highly repetitive DNA sequences. Higher GFP background levels, such as found for transient CaMV35S-driven GFP expression in protoplasts (43), should be avoided as they will obscure even the centromeric 180:GFP signals.

It is interesting to note that the REPRESSOR:GFP-mediated *in vivo* detection of transgenic *Arabidopsis* loci containing 112–256 copies of the cognate *tetO* or *lacO* binding sequences (5) apparently resulted in a similar signal to noise ratio as observed in our study of 180:GFP-mediated centromere labeling. The repeating units in these cases were 40–50 bp in length, about one quarter of the length of the 180-bp repeat, but in total 4- to 8-fold less

abundant per locus. Considering the available data and assuming nuclear PZF:GFP concentrations around 0.5 μM, we propose that it should be possible to detect PZF:GFP-mediated *in vivo* chromosomal labeling when the total number of binding sites at a specific locus divided by the length of the repeating sequence in base pairs is larger than 1. For the 180-bp repeat (2000 sites per locus divided by 180), this tentative formula gives a value of 11.1; for the repressors recognizing their operators, values are between 2.2 and 6.4. For the 5S repeat, even assuming the maximal 200 sites per locus, the value is 0.4, below the imaginary threshold.

Obviously, there should be a limit for the length of the repeating unit and this might be chosen at 200–300 bp, about once per one or two nucleosomes. A way to overcome problems with longer and more rare repeating units would be the simultaneous expression of several PZF:GFP fusion proteins, each recognizing a different sequence within the target site of interest, in this way providing the necessary number of binding sites for successful visualization. The more recently elucidated ZF recognition codes for ANN, CNN and some TNN triplets as well (17; http://www.zincfingers.org/), should offer ample opportunities for alternative PZF design, but since the usefulness of these codes is not yet supported by wider experimental evidence we refrained from using them in the present study.

CONCLUSION

A novel method is presented to establish live cell imaging of endogenous repetitive DNA sequences. The method employs sequence-specific polydactyl zinc finger (PZF) domains that can be designed to target any DNA sequence. In combination with a fluorescent tag, the PZF behavior can be monitored *in vivo*. In contrast to the *lac* or *tet* operator/repressor systems (2–5) the presented PZF system does not require the prior introduction of a transgenic target into the genome. PZF-mediated labeling can be applied to detect sequences in plants and animals. Using a 9-bp target, both the 180-bp repeat in *Arabidopsis* and the major satellite repeat in mouse were successfully detected *in vivo*.

Already in its present state the method is a valuable and flexible tool to study repetitive sequences and complement existing live cell imaging techniques in a wide variety of organisms. Although target regions containing moderate and low copy repeats were not visualized, refinement of the sensitivity using multiple PZF:GFP proteins and optimization of the protein concentration should enable the detection of theoretically any repetitive DNA sequence.

ACKNOWLEDGEMENTS

This work was supported by The Netherlands Organisation for Scientific Research (grant 050-10-123 from the NWO Programma Genomics), the Dutch CYTTRON consortium sponsored by the Ministry of Economic Affairs and by a grant to Tobias Meckel of the Deutsche Akademie der Naturforscher Leopoldina (BMBF-LPD 9901/8-124). We would like to thank Dr Jim Haseloff (MRC, Cambridge, UK) for the plasmid containing the mGFP6 sequence and Prof. Thomas Schmidt (LION, Leiden University) for access to the single-molecule setup. Furthermore we would like to thank Dr Aline Probst (Laboratory of Plant Genetics, University of Geneva), for providing line A ecotype Zurich, Martijn S. Luijsterburg and Daniël O. Warmerdam (Swammerdam Institute for Life Sciences, University of Amsterdam) for the mRFP:HP1α construct, Roger Y. Tsien (UC San Diego, CA, USA) for the mRFP sequence and Dr H. Schiessel (Instituut-Lorentz, Leiden University) for helpful discussions regarding chromatin structure. We are grateful to Gerda Lamers and Johan Pinas for technical assistance.

Conflict of interest statement. None declared.

REFERENCES

1. De Jong,H.J., Fransz,P. and Zabel,P. (1999) High resolution FISH in plants – techniques and applications. *Trends Plant Sci.*, **4**, 258–263.
2. Robinett,C.C., Straight,A., Li,G., Willhelm,C., Sudlow,G., Murray,A. and Belmont,A.S. (1996) *In vivo* localization of DNA sequences and visualization of large-scale chromatin organization using *lac* operator/repressor recognition. *J. Cell Biol.*, **135**, 1685–1700.
3. Kato,N. and Lam,E. (2001) Detection of chromosomes tagged with green fluorescent protein in live *Arabidopsis thaliana* plants. *Genome Biol.*, **2**, 0045.1–0045.10.
4. Gasser,S.M. (2002) Visualizing chromatin dynamics in interphase nuclei. *Science*, **296**, 1412–1416.
5. Matzke,A.J.M., Huettel,B., Van der Winden,J. and Matzke,M. (2005) Use of two-color fluorescence-tagged transgenes to study interphase chromosomes in living plants. *Plant Phys.*, **139**, 1586–1596.
6. Ahmad,K. and Henikoff,S. (2001) Centromeres are specialized replication domains in Heterochromatin. *J. Cell Biol.*, **153**, 101–110.
7. Boisnard-Lorig,C., Colon-Carmona,A., Bauch,M., Hodge,S., Doerner,P., Bancharel,E., Dumas,C., Haseloff,J. and Berger,F. (2001) Dynamic analyses of the expression of the HISTONE::YFP fusion protein in *Arabidopsis* show that syncytial endosperm is divided in mitotic domains. *Plant Cell*, **13**, 495–509.
8. Das,T., Payer,B., Cayouette,M. and Harris,W.A. (2003) *In vivo* time-lapse imaging of cell divisions during neurogenesis in the developing zebrafish retina. *Neuron*, **37**, 597–609.
9. Hadjantonakis,A.K. and Papaioannou,V.E. (2004) Dynamic in vivo imaging and cell tracking using a histone fluorescent protein fusion in mice. *BMC Biotechnol.*, **4**, 33–46.
10. Kanda,T., Sullivan,K.F. and Wahl,G.M. (1998) Histone–GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr. Biol.*, **8**, 377–385.
11. Savoian,M.S. and Rieder,C.L. (2002) Mitosis in primary cultures of *Drosophila melanogaster* larval neuroblasts. *J. Cell Sci.*, **115**, 3061–3072.
12. Fang,Y. and Spector,D.L. (2005) Centromere positioning and dynamics in living *Arabidopsis* plants. *Mol. Biol. Cell*, **16**, 5710–5718.
13. Henikoff,S., Ahmad,K., Platero,J.S. and Van Steensel,B. (2000) Heterochromatic deposition of centromeric histone H3-like proteins. *Proc. Natl Acad. Sci. USA*, **97**, 716–721.
14. Shelby,R.D., Vafa,O. and Sullivan,K.F. (1997) Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites. *J. Cell Biol.*, **136**, 501–513.
15. Talbert,P.B., Masuelli,R., Tyagi,A.P., Comai,L. and Henikoff,S. (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. *Plant Cell*, **14**, 1053–1066.

16. Segal,D.J., Dreier,B., Beerli,R.R. and Barbas,C.F.III (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl Acad. Sci. USA*, **96**, 2758–2763.
17. Blancafort,P., Segal,D.J. and Barbas,C.F.3rd. (2004) Designing transcription factor architectures for drug discovery. *Mol. Pharmacol.*, **66**, 1361–1371.
18. Neuteboom,L.W., Lindhout,B.I., Saman,I.L., Hooykaas,P.J. and Van der Zaal,B.J. (2006) Effects of different zinc finger transcription factors on genomic targets. *Biochem. Biophys. Res. Commun.*, **339**, 263–270.
19. Becker,D., Kemper,E., Schell,J. and Masterson,R. (1992) New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol. Biol.*, **20**, 1195–1197.
20. Weijers,D., Franke-van Dijk,M., Vencken,R.J., Quint,A., Hooykaas,P.J. and Offringa,R. (2001) An *Arabidopsis* minute-like phenotype caused by a semi-dominant mutation in a RIBOSOMAL PROTEIN S5 gene. *Development*, **128**, 4289–4299.
21. Haseloff,J., Dormand,E.L. and Brand,A.H. (1999) Live imaging with green fluorescent protein. *Methods Mol. Biol.*, **122**, 241–259.
22. Campbell,R.E., Tour,O., Palmer,A.E., Steinbach,P.A., Baird,G.S., Zacharias,D.A. and Tsien,R.Y. (2002) A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA*, **99**, 7877–7882.
23. Mateos-Langerak,J., Brink,M.C., Luijsterburg,M.S., Van der Kraan,I., Van Driel,R. and Verschure,P.J. (2007) Pericentromeric heterochromatin domains are maintained without accumulation of HP1. *Mol. Biol. Cell*, **18**, 1464–1471.
24. Lazo,G.R., Stein,P.A. and Ludwig,R.A. (1991) A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Biotechnology (N.Y.)*, **9**, 963–967.
25. Mittelsten Scheid,O., Afsar,K. and Paszkowski,J. (1998) Release of epigenetic gene silencing by trans-acting mutations in *Arabidopsis*. *Proc. Natl Acad. Sci. USA*, **95**, 632–637.
26. Clough,S.J. and Bent,A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.*, **16**, 735–743.
27. Masson,J. and Paszkowski,J. (1992) The culture response of *Arabidopsis thaliana* protoplasts is determined by the growth-conditions of donor plants. *Plant J.*, **2**, 829–833.
28. Schmidt,T., Schuetz,G.J., Baumgartner,W., Gruber,H.J. and Schindler,H. (1995) Characterization of photophysics and mobility of single molecules in a fluid lipid membrane. *J. Phys. Chem.*, **99**, 17662–17668.
29. Steinmeyer,R., Noskov,A., Krasel,C., Weber,I., Dees,C. and Harms,G.S. (2005) Improved fluorescent proteins for single-molecule research in molecular tracking and co-localization. *J. Fluoresc.*, **1**, 707–721.
30. Maluszynska,J. and Heslop-Harrison,J.S. (1991) Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. *Plant J.*, **1**, 159–166.
31. Martinez-Zapater,J.M., Estelle,M.A. and Somerville,C.R. (1986) A highly repeated DNA sequence in *Arabidopsis thaliana*. *Mol. Gen. Genet.*, **204**, 417–423.
32. Fransz,P., Armstrong,S.J., Alonso-Blanco,C., Fischer,T.C., Torres-Ruiz,R.A. and Jones,G. (1998) Cytogenetics for the model system *Arabidopsis thaliana*. *Plant J.*, **13**, 867–876.
33. Hall,S.E., Kettler,G. and Preuss,D. (2003) Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. *Genome Res.*, **13**, 195–205.
34. Campell,B.R., Song,Y., Posch,T.E., Cullis,C.A. and Town,C.D. (1992) Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*. *Gene*, **112**, 225–228.
35. Cloix,C., Tutois,S., Yukawa,Y., Mathieu,O., Cuvillier,C., Espagnol,M.C., Picard,G. and Tourmente,S. (2000) Analysis of 5S rDNA arrays in *Arabidopsis thaliana:* physical mapping and chromosome-specific polymorphisms. *Genome Res.*, **10**, 679–690.
36. Probst,A.V., Fransz,P.F., Paszkowski,J. and Mittelsten Scheid,O. (2003) Two means of transcriptional reactivation within heterochromatin. *Plant J.*, **33**, 743–749.
37. Beerli,R.R., Segal,D.J., Dreier,B. and Barbas,C.F.III (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl Acad. Sci. USA*, **95**, 14628–14633.
38. Lehnertz,B., Ueda,Y., Derijck,A.A.H.A., Braunschweig,U., Perez-Burgos,L., Kubicek,S., Chen,T., Li,E., Jenuwein,T. *et al.* (2003) *Suv39h*-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.*, **13**, 1192–1200.
39. Vissel,B. and Choo,K.H. (1989) Mouse major (γ) satellite DNA is highly conserved and organized into extremely long tandem arrays: implications for recombination between nonhomologous chromosomes. *Genomics*, **5**, 407–414.
40. Fransz,P., De Jong,J.H., Lysak,M., Castiglione,M.R. and Schubert,I. (2002) Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl Acad. Sci. USA*, **99**, 14584–14589.
41. Guenatri,M., Bailly,D., Maison,C. and Almouzni,G. (2004) Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J. Cell Biol.*, **166**, 493–505.
42. Finch,J.T. and Klug,A. (1976) Solenoidal model for superstructure in chromatin. *Proc. Natl Acad. Sci. USA*, **73**, 1897–1901.
43. Furtado,A. and Henry,R. (2002) Measurement of green fluorescent protein concentration in single cells by image analysis. *Anal. Biochem.*, **310**, 84–92.