HAL
open science

Characterization of monomeric protein domains that bind specifically to a highly-conserved 100-Bp DNA target within rRNA genes

Elodie Carnus, Marie-Véronique Demattei, Sophie Casteret, Guillaume Carpentier, Fabien Palazzoli, Solenne Bire, Christophe Bressac, Yves Bigot

▶ **To cite this version:**

Elodie Carnus, Marie-Véronique Demattei, Sophie Casteret, Guillaume Carpentier, Fabien Palazzoli, et al.. Characterization of monomeric protein domains that bind specifically to a highly-conserved 100-Bp DNA target within rRNA genes. Gene Technology, 2012, 1 (1), pp.1-14. 10.4172/2329-6682.1000101 . hal-01129771

HAL Id: hal-01129771

https://hal.science/hal-01129771v1

Submitted on 29 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

rDNA Technology

Research Article et al. rDNA Technol 2012, 1:1
http://dx.doi.org/10.4172/rdt.1000101

Research Article Open Access

Characterization of Monomeric Protein Domains that Bind Specifically to a Highly-Conserved 100-Bp DNA Target within rRNA Genes

Elodie Carnus[2], Marie-Véronique Demattei[2], Sophie Casteret[1], Guillaume Carpentier[2], Fabien Palazzoli[2], Solenne Bire[1,2], Christophe Bressac[3] and Yves Bigot[1,2]*

[1]*UMR INRA-CNRS 7247, PRC, Centre INRA de Nouzilly, 37380 Nouzilly - France*
[2]*UMR CNRS 6239, Equipe Cellules usines, Ingénierie des protéines et Bio-informatique, UFR des Sciences et Techniques, Parc de Grandmont, 37200 Tours - France*
[3]*UMR CNRS 6035, IRBI, UFR des Sciences et Techniques, Parc de Grandmont, 37200 Tours - France*

Abstract

Proofs of concept have shown that chromosomal gene clusters encoding ribosomal RNA (rRNA) constitute gene delivery integration loci that are optimal for transgene expression. However, because homologous recombination is efficient to integrate DNA segments into these genes in animals, new molecular tools are required to construct systems able to target molecules in the immediate vicinity of the rRNA genes.

We investigated the properties of several DNA binding domains (DBDs) able to recognize specifically a motif within a 100-bp region of the rRNA genes that is 99-100% conserved among eukaryotes. Our findings demonstrate that two *Myb*-like DBDs originating from the endonucleases encoded by R2 non-LTR retrotransposons are promising candidates since they i) specifically recognize, with high affinity, a 20-bp binding site located within the expected genomic rDNA target, ii) act as monomers, iii) contain a nuclear localization signal, iv) remain functional when fused to another domain and, v) do not alter the functionality of the protein to which they are fused. However, results obtained *in vivo* with several R2DBD fusions reveal that two properties remain to be engineered before these DBDs can be integrated into a molecular targeting system directed into rRNA genes. The first concerns the ability of R2DBD to locate within the nucleolus, the organelle in which the rRNA genes reside. The second is the tendency of R2DBD to accumulate in certain parts of the nuclei, which limits its diffusion within nuclei. Solutions are discussed to circumvent these current limitations.

Our results supply important information concerning the R2DBD properties and the targeting of plasmid DNA within nuclei. They will need to be further analyzed from three aspects; the unexploited advantages of the R2DBDs, the possibilities and limitations of fusion peptides for targeting integrations of non-viral vector, and the alternatives to fusion peptides for targeting vectors.

Keywords: Polydactyl zinc finger domain; R2 non-LTR Retrotransposon; Recombination; Transposition; Targeting

Abbreviations: Bm- *Bombyx mori* ; bp- base pair; BB- Binding Buffer; BS- Binding Site; DBD- DNA Binding Domain; Dm-*Drosophila melanogaster* ; DMEM-Dulbecco's Modified Eagle Medium; DNA-Deoxyribonucleic Acid; DTT- Dithiothreitol; EDTA-Ethylenediaminetetraacetic acid; EMSA- Electrophoresis Mobility Shift Assay; FBS- Fetal Bovine Serum; His- Histidine; HR- Homologous Recombination; IPTG Isopropyl β-D-1-thiogalactopyranoside; ITR Inverted Terminal Repeat; GFP- Green Fluorescent Protein; Kd-"dissociation" constant; MBP- Maltose Ninding Protein; NeoR-Neomycin resistance gene; OD- Optical Density; ORF- Open Reading Frame; NLS- Nuclear Localization Signal; ^{32}P- phosphorus ^{32}; PAGE-Polyacrylamide Gel Electrophoresis; PB- *piggyback*; PBS- Phosphate Buffer Saline; PCR-Polymerase Chain Reaction; PEI- Polyethylenimine; pZFD- polydactyl Zinc Finger Domain; R2- R2 non-LTR retrotransposon ; RNA- Ribonucleic acid ; rRNA-ribosomal Ribonucleic Acid; SB- *Sleeping Beauty*; SB10, *Sleeping Beauty* transposase version 10;SDS- Sodium Dodecyl Sulfate; TACH-Temperature-Assisted Cyclic Hybridization; TBE- Tris Borate EDTA; Tpase-Transposase ; ZBS-Zinc Finger Domain Binding Site; ZF- Zinc Finger; ZFD- Zinc Finger Domain.

Introduction

The ability to use of DNA binding domains (DBDs) able to interact with a specific DNA target in order to develop synthetic proteins capable of modifying gene expression or the functioning of DNA recombination machinery, such as those involving endonucleases, integrases or transposases (Tpase), constitutes a major challenge in the post genomic era. From this perspective, a DBD added to an enzymatic system has to combine several qualities, including i) being bound to a DNA target large enough to ensure specific and high-affinity recognition, ii) having a controllable ability to assemble (or not) as homo-oligomer or with other cellular partners, and iii) having a known and modifiable propensity to locate to certain cell compartments when it has to be used *in vivo*.

Here, our focus was on the characterization of DBD candidates able to bind in monomer form to a 100-bp DNA segment that is 99-100% conserved among eukaryotes [Figure 1b], and located in the tandemly-repeated genes that encode the ribosomal 18S-5.8S-28S RNA (rRNA), [Figure 1a]. Whether intended for use in bioproduction or gene correction in unicellular eukaryotes [1,2] or in mammalian cells [3,4], these repeated genes have been shown to be good target candidates for the insertion of DNA fragments encoding recombinant proteins. The fact that only a few rRNA genes are disrupted accounts for the absence of deleterious effects in genetically-modified cells. Indeed, genetic investigations have demonstrated that the disruption of < 70% of the rRNA genes by integration of DNA segments is not lethal and does not impair the viability of invertebrate genomes [5,6]. The expression of the transgene also persists for long periods, because integrated DNA fragments are rarely subjected to expression silencing [7], and their expression can be carried out by the RNA Polymerase I promoter, but also by RNA Polymerase II and III promoters [8,5]. However, their interest has been considerably reduced in mammalian cells because the integrative vectors functioning with passive homologous recombination (HR) have integration efficiencies ranging

***Corresponding author:** Yves Bigot, UMR INRA-CNRS 7247, PRC, Centre INRA de Nouzilly, 37380 Nouzilly – France ; Tel: +33 2 47 36 70 63, Fax: +33 2 47 36 70 42, e-mail: yves.bigot@tours.inra.fr

Received March 06, 2012; **Accepted** March 31, 2012; **Published** April 02, 2012

Citation: Bigot Y, Carnus E, Demattei MV, Casteret S, Carpentier G, et al. (2012) Characterization of Monomeric Protein Domains that Bind Specifically to a Highly-Conserved 100-Bp DNA Target within rRNA Genes. rDNA Technol 1:101. doi:10.4172/rdt.1000101

Copyright: © 2012 Bigot Y, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Bigot Y, Carnus E, Demattei MV, Casteret S, Carpentier G, et al. (2012) Characterization of Monomeric Protein Domains that Bind Specifically to a Highly-Conserved 100-Bp DNA Target within rRNA Genes. rDNA Technol 1:101. doi:10.4172/rdt.1000101

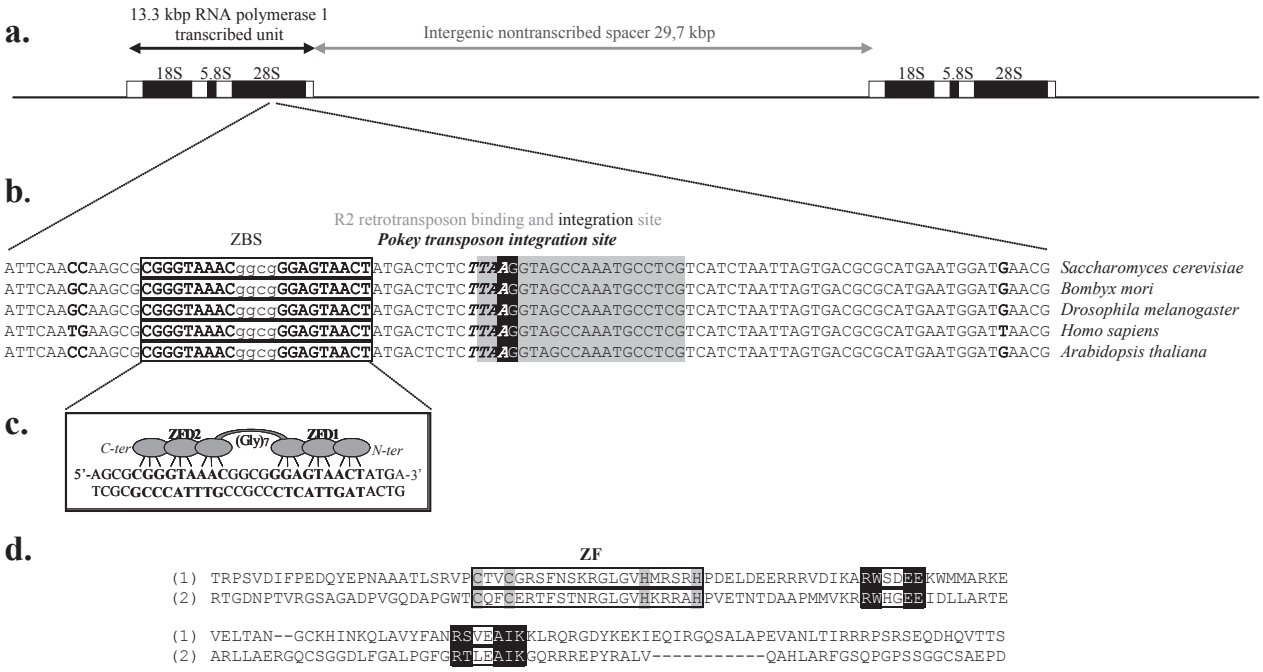

Figure 1: Features of rRNA genes, integration and cleavage sites, and the DNA binding domains targeted within the region encoding the 28S rRNA. (a) General map of the human ribosomal DNA repeat unit including the gene region transcribed by RNA polymerase I and the intergenic non-transcribed spacer [59]. In the transcribed region, white boxes correspond to transcribed spacers. Black boxes locate the region encoding the 18S, 5.8S and 28S rRNA. Dotted lines from a to b delineate the 100-bp conserved region. (b) Sequence comparison of the 100-bp region in the gene encoding the 28S rRNA that is 99 to 100% conserved from yeast to human beings. The TTAA tetranucleotide in which the *Pokey* DNA transposon integrates [11] is typed in italics and bold, and the binding site of the endonuclease encoded by the R2 non-LTR retrotransposon [28] is highlighted in gray, with the AG dinucleotide in which the endonuclease cleavage occurs highlighted in black and typed in white. The bipartite DNA binding site of the three pZFDs described in b and c are boxed and in bold, the linker between these sites being shown in lower-case letters. Since this 100-pb DNA segment contained the binding site of our pZFD (ZBS) and that of the R2 endonuclease, it was used as probe in our EMSA analysis, and designated the R2 site (c) Organization of the bipartite pZFD in which both ZFD trimers, ZFD1 and ZFD2, are separated by seven glycine residues. Their binding to the bipartite DNA 9-bp binding site is shown below the pZFD representation. (d) Amino acid sequences of the two DNA binding domains contained in the R2 non-LTR retrotransposons isolated from (1) *Drosophila melanogaster* and *(2) Bombyx mori* (Accession numbers: M16558 and X51967). The zinc fingers are boxed, and the residues involved in the C_2H_2 bonds highlighted in gray. The conserved motifs within the Myb sub-domain are also boxed, and the most highly conserved residues are highlighted in black and typed in white [29].

from 10^{-5} to 10^{-4} [8,5,7]. A way to enhance HR at this target could be to increase the concentration of DNA fragments for insertion within the nucleoli where rRNA genes are located in nuclei. Such a targeting might be achieved by using, for example, two fused DBDs [9] or an oligonucleotide [10] cross-linked to a DBD, these molecules being able on the one hand to bind or anneal specifically to the DNA fragment to be integrated and, on the other hand, to bind to the integration target. Vectors derived from certain transposons could also be suitable. The *Pokey* DNA transposon [11], which is an element distantly related to *piggyBac* (PB) [12], has been shown to integrate frequently into rDNA in its host of origin. However, *Pokey* vectors are unable to transpose in vertebrate cells (personal data). At least three innovative approaches have been developed to target transposon vectors, such as *piggyBac* [13] or *Sleeping Beauty* (SB) [14-16]. The first consists of fusing the Tpase to a DBD that specifically binds to the DNA integration target. The second and third approaches consist of using two fused protein domains in which the first domain specifically binds to the integration target, and the second to the transposon vector via a binding site contained in its sequence, or via an interaction with the Tpase bound to the end of the transposon (known as an inverted terminal repeat (ITR)).

Overall, whatever the species and strategy used to integrate a DNA fragment into rRNA genes, it is noticeable that it always requires a protein domain that can target the enzyme and/or the DNA fragment in the proximity of the rRNA genes. To date, we have identified only two kinds of DBD able to bind it as a monomer to a motif present within the conserved 100-bp DNA target, this motif being so large (at least 18-bp) as to be nearly unique in a mammalian genome. The first are synthetic DBDs corresponding to polydactyl zinc finger domains (pZFDs; reviewed in [17-19]. pZFDs consist of the fusion of n peptides from thirty amino acids, each of which specifically recognizes a three base-pair DNA motif. Research carried out by Professor Carlos Barbas 3rd and his colleagues has made it possible to develop a single ZF able to specifically recognize all GNNs, most CNNs and ANNs, and a few TNN DNA triplets [20]. Under these conditions, the fusion of n ZFDs theoretically permits the specific recognition of a 3n base-pair nucleotide motif [Figure 1c and S1]. In practice, the background knowledge that has been accumulating for more than 10 years now enables scientists to design pZFDs *in silico* using website facilities [20-22], or to manufacture them by *in-vitro* selection [23]. The second category consists of natural DBDs that originate from the endonuclease encoded by R2 mobile genetic elements, non-LTR retrotransposons [24]. The R2 endonuclease is related to type-IIS restriction enzymes [25]. It specifically allows the integration of R2 non-LTR retrotransposons into a 20-bp target motif located within the conserved 100 bp DNA segment contained in rRNA genes [Figure 1b] [26,27]. The minimum DBD of this enzyme (R2DBD), without its homo-dimerization activities, has recently been characterized [28]. It consists of one ZF associated with a *Myb*-like sub-domain [29] [Figure 1d].

In this paper, we investigated the potential of three pZFDs and two R2DBDs. We first evaluated their specific and non-specific binding site numbers in the genome of models currently used for gene therapy and bioproduction purposes. We then compared their binding specificities to about 20-bp binding sites contained within the 100-bp DNA target. Under these conditions, the best candidates were both R2DBDs. In a second step, we analyzed the activity of fusion peptides made from R2DBD *in vitro*, and *in vivo* in HeLa cells, using integration vector systems. Finally, we have found that R2DBDs combine several properties that ensure accurate nuclear traffic to rRNA genes within the nucleolus. Potential solutions to circumvent the current problems encountered are discussed.

Materials and Methods

pZFD design

An optimal motif was first calculated using the Zinc Finger Tools website (http://www.scripps.edu/mb/barbas/zfdesign/zfdesignhome.php) [20] to target a synthetic pZFD within the 100-bp region shown in Figure 1. This 22-bp motif is located 11-bp upstream of the R2 endonuclease binding site. DNA fragments encoding pZFD able to bind to this 22-bp motif were designed using facilities available at websites proposing three calculation methods. The first pZFD, designated ZFDBarbas, was designed using the Zinc Finger Tools website. Two other pZFDs, designated ZFDSangamo and ZFDToolgen respectively, were designed at the ZiFiT V3.0 website (http://bindr.gdcb.iastate.edu/ZiFiT/) [21], following the recommendations for optimization from the ZiFDB V1.0 website/

(http://bindr.gdcb.iastate.edu:8080/ZiFDB/controller/searchObjects) [22].

DNA fragment synthesis

DNA fragments containing the 100-bp segment including the binding sites of the pZFD and the R2DBD proteins [Figure 1b] were synthesized by ATGBiosynthetics (Merzhausen, Germany) and cloned in pUC57 (Fermentas, St. Leon-Rot, Germany). The fragments encoding pZFD and the R2DBD proteins [Figure 1d] were designed and synthesized with a codon usage avoiding codons that are rare in human cells. The R2DBD originated from sibling endonucleases encoded by the non-LTR R2 retrotransposon of *Bombyx mori* and *Drosophila melanogaster*. They were designated BmR2DBD and DmR2DBD respectively. The LexA binding site was constructed using two oligonucleotides

(LexAfix:5'- AATTCCATGGATCCG TCGACCTGTATATATATA-CACTCG AGATCTCCATGG -3' and LexAfix-Rev: 5'-AATTCCATG GAGATCTCGAGTGTATATATATACAGGTCGACGGATCCATGG -3') that were annealed, and then cloned into the *Eco*RI pBS-SK- plasmid. The DNA segment encoding the LexA-NLS was designed on the basis of data in the literature [14].

DNA probes

A 150-bp fragment containing the 100-bp segment with the protein binding sites was purified after plasmid digestion with *Kpn*I and *Sal*I, separation on an agarose gel and elution with the Wizard SV gel and PCR clean-up system kit (Promega, Madison WI, USA). This probe was designated the R2site. A variant fragment of 240-bp was obtained by PCR using M13 universal and reverse primers. A 51-bp fragment containing the LexA binding site (LexA-BS) was similarly purified using *Eco*RI plasmid digests. A 70-bp fragment containing the 3'-ITR of *Mos1* transposon was prepared as described [30]. DNA fragments were radiolabeled by filling using the DNA polymerase Klenow fragment (New England Biolabs France, Every, France), 30 µCi of [α-^{32}P] dATP (3000 Ci/mmole) and unlabeled dC-G-TNPs. The probes were purified by electrophoresis on a 6% native polyacrylamide gel and then eluted in Tris-HCl 10 mM pH8, 1 mM EDTA, and 50 mM NaCl buffer. After precipitation using tRNA as a carrier, the probes were resuspended in 50 µl of H_2O, and the DNA concentration was estimated by Hoechst assay [31] using a Mithras LB 940 (Berthold Technologies France, Thoiry, France).

Protein purification of bacterial recombinant protein

Each of the five fragments encoding a DBD was cloned into the pET14b vector (Novagen, Darmstadt, Germany), in a frame with a Histidine tag (His_6) at the amino terminal end. The pET14b constructs were transformed in BL21 *Escherichia coli* bacteria containing a pRARE plasmid (Novagen, Darmstadt, Germany), which encoded rare tRNA *E. coli* codons. They were also cloned into the pMalc2 system (New England Biolabs France, Every, France), and fused in frame with the Maltose Binding Protein (MBP) at the amino terminal end. The pMalc2 constructs were transformed in JM109 *E. coli* bacteria using the pRARE plasmid.

Bacteria were grown at 37°C and 250 rpm either until the OD reached 0.4 to 0.6 (proteins tagged with His_6) or for 6 hours (proteins tagged with MBP). Expression was then induced by supplying 1 mM IPTG for 3 hours at 37°C and 250 rpm. The cells were harvested by centrifuging, resuspended in buffer A (50 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM dithiothreitol (DTT)) and ruptured by enzymatic digestion with 1 mg/ml lysozyme, for 30 minutes at 4°C. Insoluble material was removed by centrifuging (15,000 g for 10 min, at 4°C) and the supernatant was used as a crude extract containing the expected fusion protein. The crude extracts were loaded in buffer A onto a HisLink Protein Purification Resin (Promega, Madison WI, USA) for His-tagged proteins and onto amylose resin (New England Biolabs France, Every, France) for MBP-fused protein. After washing the column with buffer A, the proteins were eluted with buffer A supplemented with 250 mM imidazole or 10 mM maltose, respectively. The protein concentration in each of the eluted fractions was determined by a Bradford assay using bovine serum albumin as the standard, and the purification quality was checked by electrophoresis on a sodium dodecyl sulfate - polyacrylamide gel (SDS-PAGE) with Coomassie blue staining [32]. Proteins fused with MBP have a cleavage site for factor Xa located between the tag and the recombinant protein. Proteolyses with factor Xa of the MBP fusions (New England Biolabs France, Every, France) were done according to the instructions of the enzyme supplier.

DNA plasmid transfection in HeLa cells

HeLa cells were cultured in DMEM supplemented with 10% fetal bovine serum (FBS). About 1 x 10^5 cells were seeded onto each well of a 24-well plate, one day prior to transfection. Cells were transfected with jetPEI™ (Polyplus Transfection, Illkirch, France) at an N/P rate of 5, according to Manufacturer's instructions. Briefly, plasmid DNA (0.2 to 0.5 µg) and jetPEI™ were each diluted in 50 µl of 150 mM NaCl, then mixed gently together and incubated for 15 min. Cells were then incubated for 2 to 4 h with the complexes dropped into 500 µL of Opti-MEM medium (Invitrogen/Life Technologies, Paisley, United Kingdom). The transfection solution was then discarded, replaced by fresh DMEM supplemented with 10% FBS, and incubated for 24 hours at 37°C.

Protein expression in HeLa cells

Each R2DBD fragment was cloned into the pCS2+ vector (Add-

gene). In each well of a 6-well-plate, 4x10^5 HeLa cells were plated one day prior to transfection, and were cultured in DMEM medium (PAA France, Les Mureaux, France) supplemented with 10% fetal bovine serum at 37°C and 5% CO_2. Two microgrammes of pCS2-R2DBD were transfected using JetPEI™ according to standard protocols (Polyplus Transfection, Illkirch, France). One day post-transfection, the cells were scraped off and washed with Phosphate Buffer Saline (PBS) 1X solution. Cells were resuspended in buffer A, and disrupted with four 5-minute freeze (80°C) and thaw (37°C) cycles, and sonicated (twice, for 10 sec at 50W). After centrifuging for 10 min at 15,000g and 4°C, the supernatant was used as a crude extract known to contain the protein of interest.

Integration assays in HeLa

The features of the plasmids used in integration assays, transfection, and microscopy analysis in HeLa cells are summarized in Table 1. About 1 x 10^5 HeLa cells were co-transfected with 500 ng of a 4 DNA plasmid mix. One fifth (100 ng) of the mix corresponded to the pGL3 plasmid (Promega, Madison WI, USA) that was used to check that transfection had occurred. The remaining 400 ng consisted of a mix composed of one of the 3 different Tpase sources (pCS2, pCS2-SB10 or pCS2-SB10-R2DBD), one of the 5 sources of a cassette encoding a resistance to G418 (pBS-NeoR, pBS-NeoR-LexA/BS, pBS-NeoR-ZBS, pT2-pSV40-NeoR, pT2-pSV40-NeoR-LexA/BS, pT2-pPol1h-NeoR), and one of the 8 target peptide variants (pCS2, pCS2-R2DBD, pCS2-R2DBD-LexA, pCS2-R2DBD-NLS-LexA, pCS2-R2DBD-N57SB10, pCS2-R2DBD-pZFD$^{bcr-abl}$, pCS2-NLS-LexA, pCS2-NLS-N57SB10). Two days after transfection, 1/3 of cells were used to check the effectiveness of transfection by evaluating the luciferase activity with the Luciferase Assay System Kit (Promega, Madison WI, USA). The remaining 2/3 of the cells were transferred to 100-mm plates followed by G418 sulfate selection (800 µg/mL, PAA France, Les Mureaux, France) for 14 days. Finally, the cells were fixed and stained with 70% EtOH - 0.5% methylene blue for 3h. Only colonies with a diameter > 0.5 mm were counted.

Microscopic analysis in HeLa cells

About 5 x 10^4 HeLa cells were co-transfected with 500 ng of a 2 DNA plasmid mix as described above. 50% of the mix corresponded to the plasmid expressing the R2DBD fused to the GFP [Table 1]; the remaining 50% consisted of a plasmid expressing a nucleolus marker fused to the DsRed (pDsRed-C1Hsnucleoline [33]). Twenty-four hours post-transfection, cells were washed three times with PBS 1X, and nuclear DNA was stained with 2 µg /mL DAPI for 2h. Three more PBS 1X washes were performed before microscopic observation. Cells were observed under an epifluorescence microscope (Olympus BX51).

Electrophoretic mobility gel assay (EMSA)

Binding reactions were carried out in binding buffer (BB) containing 50 mM Tris-HCl pH 8, 200 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 11% glycerol, 0.1 mg/ml of bovine serum albumin (BSA), and 0.01% Triton X-100. In each 20 µl of reaction mixture, 0.03 pmol of the 32P-labeled duplexes were mixed with appropriate amounts of purified proteins, and 1 µg of sheared DNA from herring sperm. Mixtures were incubated at 37°C for 1 hour. Reaction products were then separated using discontinuous 4 to 6 or 9% native polyacrylamide gels (depending on the protein size). Electrophoresis was carried out in 0.25X TBE at 200 V, dried and autoradiographed.

UV cross-linking of DNA-protein complexes

Complexes were assembled in 20 µl reaction mixtures containing 0.2 pmol of ^{32}P labeled R2 site probe (150-bp), and 5 pmol of protein His-BmR2DBD or 0.6 pmol of MBP-BmR2DBD in BB. After incubating for 1 hour at 37°C, samples were exposed to UV irradiation (312 nm) for 30 min at 4°C. Reactions were stopped by adding 4 µl

Plasmid name	Plasmid features
Transposase source for integration assays	
pCS2-SB10	Plasmid encoding the version 10 of the *sleeping beauty* transposase (SB10 [60])
pCS2-SB10-R2DBD	Plasmid encoding SB10 fused with R2DBD at its terminal end
Sources of NeoR cassette for integration assays	
pBS-NeoR	pBSKS- in which a gene encoding a resistance to G418 flanked at its 5' and 3' ends respectively by a SV40 promoter (pSV40) and a SV40 polyadenylation signal is cloned between the *EcoRI* and B*amHI* sites of the multicloning site (MCS)
pBS-NeoR-LexA/BS	pBS-NeoR in which a binding site to *LexA* is cloned at the *EcoRI* site in MCS
pBS-NeoR-ZBS	pBS-NeoR in which is cloned a binding site (12-mer) to the DNA binding domain ZFD$^{bcr-abl}$ [44] at the *EcoRI* site in the MCS
pT2-pSV40-NeoR	Plasmid coding an SB vector in which a NeoR cassette similar to that of the pBS-NeoR is cloned [60]
pT2-pSV40-NeoR-LexA/BS	pT2-pSV40-NeoR in which a binding site to LexA at the EcoRI site located upstream to the NeoR cassette is cloned
pT2-pPol1h-NeoR	pT2-pSV40-NeoR in which pSV40 was replaced by an RNA polymerase I promoter extracted from the pHENA plasmid [41]
Targeting peptide sources for integration assays	
pCS2-R2DBD-NLS	Plasmid encoding R2DBD fused to an SV40 NLS at its C-terminal end
pCS2-R2DBD-NLS-LexA/BS	Plasmid encoding R2DBD-NLS fused to the LexA DNA binding domain at its C-terminal end
pCS2-R2DBD-LexA	pCS2-R2DBD-NLS-LexA/BS in which the region encoding the NLS was mutagenized to remove it (from a PKKKRK motif to PKAIRK)
pCS2-R2DBD-NLS-N57SB10	Plasmid encoding R2DBD-NLS fused to the 57 first residues of SB10 (N57SB10) at its C-terminal end [28,45]
pCS2-R2DBD-N57SB10	pCS2-R2DBD-NLS-N57SB10 in which the region encoding the NLS was mutagenized to remove it (from a PKKKRK motif to PKAIRK)
pCS2-R2DBD-NLS- ZFD$^{bcr-abl}$	Plasmid encoding R2DBD-NLS fused to the DNA binding domain ZFD$^{bcr-abl}$
pCS2-NLS-LexA/BS	Plasmid encoding SV40 NLS fused to the LexA DNA binding domain at its C-terminal end
pCS2-NLS-N57SB10	Plasmid encoding SV40 NLS fused to N57SB10 domain at its C-terminal end
GFP or DsRed fusion proteins for epifluorescent microscopy analyses	
pCS2-R2DBD-NLS-GFP	Plasmid encoding R2DBD-NLS fused to the GFP protein at its C-terminal end
pCS2-R2DBD-GFP	pCS2-R2DBD-NLS-GFP in which the region encoding the NLS was mutagenized to remove it (from a PKKKRK motif to PKAIRK)
pCS2-MOS1-GFP	Plasmid encoding *Mos1* transposase fused to the GFP protein at its C-terminal end
pDsRed-C1Hsnucleoline	Plasmid encoding the Red fluorescent protein (DsRed) fused with the human nucleoline at its C-terminal end [34]

Table 1: Plasmids used for transfections into HeLa cells.

of 6X SDS-PAGE loading buffer. The samples were then boiled for 10 min.

Kd determination

Concentrated proteins or HeLa cell extracts were serially diluted and used in EMSA. Bound and unbound R2site probe was quantified by Instant Imager (Packard Instrument SA France, rungis, France). The apparent Kd for each binding reaction was estimated from the plot of bound/unbound probe versus binding protein concentration using PRISM 4 software (GraphPad Software, San Diego, CA USA). Data from 4 or 5 replicates were pooled to calculate each apparent Kd.

Analysis of the expressed proteins by immunoblotting

Cells recovered from the cultures were washed three times with 1X PBS. Total protein extracts were separated by electrophoresis, adding 20 µg of each sample to a sodium dodecyl sulfate 8% miniPAGE, and then electro-blotted onto nitrocellulose filters (Bio-Rad Laboratories). After blocking with 5% skim milk in 1X PBS for 1 h, the filters were incubated overnight with rabbit polyclonal anti-6-His (1: 20,000; Bethyl Laboratories Inc., Montgomery, TX USA) or anti-LexA (1:1,000; Sigma-Aldrish, St Louis, MO USA). The filters were then incubated with horseradish peroxidase-conjugated anti-rabbit IgG (Santa Cruz Biotechnology, Santa Cruz, CA) before being developed using enhanced chemiluminescence (Amersham Pharmacia Biotech, Sunnyvale, CA).

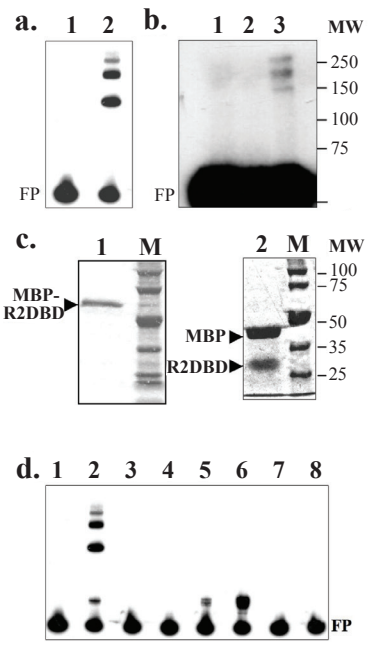

Figure 2: EMSA analysis of MBP-BmR2DBD. (a) Shifted complexes assembled with the R2site probe in the absence (lane 1) or presence of 10 nM MBP-BmR2DBD (lane 2). FP: Free R2site Probe. **(b)** SDS-PAGE (6%) analysis of UV cross-linked complexes resulting from the assembly of the MBP-BmR2DBD and the radiolabeled R2site probe. Lane 1: R2site probe after UV treatment; lane 2: complexes with no UV treatment; lane 3: UV-treated complexes. FP: Free R2site Probe. MW (in kDa) were scaled with the Precision Plus Protein Pre-stained Standards (BioRad France, Marnes la Coquette, France). **(c)** PAGE stained with colloidal Coomassie blue. Lane 1: MBP-BmR2DBD; Lane 2: MBP-BmR2DBD cleaved by Factor Xa. **(d)** Shifted complexes assembled with the R2site probe in the absence (lane 1) or presence (lane 2) of 5 nM MBP-BmR2DBD, 5, 50, 500 and 5000 nM of MBP-BmR2DBD cleaved with Factor Xa (for 2 hours at 25°C before complex assembly; lanes 3 to 6). Lanes 7 and 8 are negative controls obtained after incubating the R2site probe with MBP and Factor Xa proteins respectively. FP: Free Probe.

Nucleotide sequence analyzes

To make an inventory of the nucleotide motifs in the human genome, we used syntactic models to mine the database using WAPAM at the website http://genoweb.univ-rennes1.fr/Serveur-GPO/outils_acces.php3?id_syndic=185&lang=en. The syntactic models used were TAAGGTAGCCAAATGCCTCG to locate 100% conserved of R2 endonuclease binding sites, and CGGGTAAAC-x(0,12)-GGAGTAACT, for the different variants of the ZBS potentially bound by the pZFD used in this paper, and which has a variable spacer of 0 to 12 nucleotides between the 9-bp motifs bound by ZFD1 and ZFD2 respectively [Figure 1c]. The parameters were set to detect only exact matches (significance score of 100%). We confirmed results obtained with WAPAM by mining databases with BLASTn, using the NCBI facilities and non-default parameters that were the "Somewhat similar sequences (blastn)", a word size = 7, and an expected threshold of 100.

Results

Genomic distributions of pZFD (ZBS) and R2DBD (R2BS) binding sites

Before designing pZFD proteins able to bind specifically within the 100-bp region that is 99 to 100% conserved in the gene encoding 28S rRNA from yeasts to human beings [Figure 1], we used an *in silico* two-step procedure to locate an accurate binding site in this DNA segment. The first step consisted in finding out whether the 100-bp region contained a candidate nucleotide motif with sequence properties that matched with an accurate ZBS. The aim of the second step was to check whether the candidate ZBS occurred only within 100-bp region in the gene encoding 28S rRNA, or whether it was also present at other chromosomal locations in the genome of models frequently used for the purposes of gene therapy (mouse, rat, and human) or bioproduction (baker's yeast and chickens).

To target a unique sequence in the human genome, searches were performed to locate an appropriate 18-bp site within both strands of the 100-bp DNA segment. This was performed taking into account the fact that the pZFD had to contain two modules with three ZFs. In consequence, the aim of our search was to locate two 9-bp binding sites that were juxtaposed or separated by 1 to 12-bp, using the facilities of the Zinc Finger Tools website. Calculations revealed only one solution that included all the features required to design a bipartite pZFD within the 100-bp region. It consists of two 9-bp motifs separated by 4 bp [Figure 1b and c]. The distribution of this ZBS among the chromosomes of five models was determined on the one hand using WAPAM software with the CGGGTAAAC-x(4)-GGAGTAACT as a model to mine the database and, on the other hand, with BLASTn using the NCBI facilities, with a word size = 7, and an expected threshold of 100 as parameters. We obtained similar results with both types of software. The number of ZBSs located outside the chromosomal loci that contain the rRNA gene clusters in each model is summarized in Table 2a. The results revealed that there were a few ZBSs present outside the rDNA loci in the mammalian species, and none was detected in the chicken, drosophila or yeast genomes.

The number of R2BS located outside the rDNA loci was also evaluated using an *in silico* approach similar to that described above [Table 2b]. R2 non-LTR retrotransposons are able to integrate specifically and only in the "TAAGGTAGCCAAATGCCTCG" motif that is usually present in the genes encoding 28S rRNA. However, when identical

sites are present in a genome, as in *Bombyx mori*, in which R2BS occurs outside the rDNA loci, R2 non-LTR retrotransposons can also locate to and use these other sites for integration [34]. Here, although R2BSs were less abundant than ZBS, they were also found outside the rDNA loci.

Overall, these non-specific sites represented from 1% to 10% of the binding sites for both proteins. Such occurrence levels indicate that R2BS has, on average, only half as many binding targets as pZFD in the mammal genome. Since both domains have a population of specific and non-specific binding sites with equivalent quality in the different models, and given that previous studies [28,34] have shown that R2DBD was able to bind to R2BS, we limited our investments in designing and optimizing pZFD to *in silico* approaches to evaluate their potential so that functional pZFDS could be obtained quickly. With such an approach, the results obtained were less satisfactory as the binding efficiency of the three synthetic pZFDs was weak, whatever the *in silico* design strategy used [S1]. We therefore focused our work on the properties of the two R2DBDs.

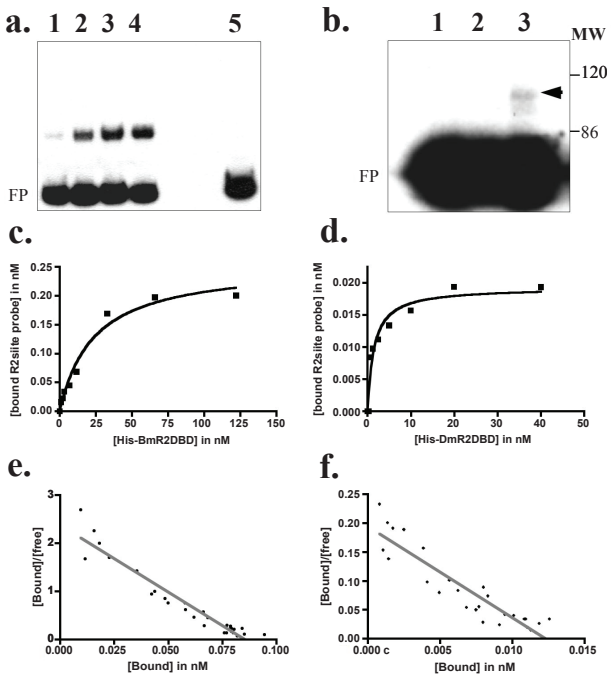

Figure 3: EMSA analysis of His-BmR2DBD. (a) Pattern of the complexes assembled with 50 nM (lane 1), 100 nM (lane 2), 200 nM (lane 3) and 250 nM (lane 4) of His-BmR2DBD and the R2site probe. Lane 5: R2site probe without protein. FP: R2site Free Probe. **(b)** SDS-PAGE (8%) analysis of UV cross-linked complexes resulting from the assembly of the His-BmR2DBD and the radiolabeled R2site probe. The fuzzy signal observed below the near 105 kDa complex originates from the dissociation of fragile complexes that occurs during electrophoresis, thus releasing free probe. Lane 1: R2site probe after UV treatment; lane 2: complexes with no UV treatment; lane 3: UV-treated complexes. FP: R2site Free Probe. MW in kDa were scaled with the Prestained Protein Molecular Weight Marker (Fermentas, St. Leon-Rot, Germany). **(c)** and **(d)** Saturation graphs at the binding equilibrium with a fixed probe concentration and variable concentrations of His-BmR2DBD and His-DmR2DBD, respectively. **(e)** and **(f)** Scatchard plots of the single complexes obtained with His-BmR2DBD and His-DmR2DBD, respectively. The bound/free (B/F) R2site probe was plotted *versus* concentration of bound R2site probe [nM], leading to the Scatchard plot. -1/Slope of the line gives the apparent Kd. The R^2 correlation coefficients calculated from the data used for the Scatchard linear regression were 0.92 for His-BmR2DBD and 0.86 for His-DmR2DBD.

Model and features of the rDNA loci‡	N° of ZBS	N° exact hits	Locations in chromosomes
Homo sapiens (GRCh37.p2 primary reference assembly) rDNA loci on chromosomes 13, 14, 15, 21 and 22. 350 rDNA copies/haploid genome	6	5	UGC-117088 UGC-161060 chr1-237766392 chr7- 68527469 chr11-77597559 chrX-108297433*
Mus musculus (strain C57BL/6) rDNA loci on chromosomes 12, 15, 16, 18 and 19. 100 rDNA copies/haploid genome	10	7	chr1-46132978 chr1-19294485 chr3-5860423 chr4-45359522 chr4-71689154 chr7-127897713* chr17-7146256 chr17-8145263 chr17-24802931* chr18-54858951*
Rattus norvegicus (strain BN/SsNHsdMCW) rDNA loci on chromosomes 3, 11, 12. 100-360 rDNA copies/haploid genome	3	3	chr5-24497292 chr12-34422429 chr16-68375450
Gallus gallus rDNA loci on chromosome 16. 261-331 rDNA copies/haploid genome	0	0	-
Drosophila melanogaster rDNA loci on chromosome X and Y. 240 rDNA copies/haploid genome	0	0	-
Saccharomyces cerevisae rDNA loci on chromosome XII. 150 rDNA copies/haploid genome	0	0	-

Table 2a: Number of ZBSs (CGGGTAAACggcgGGAGTAACT) outside the chromosomal loci that contain the rRNA genes.

Analysis of the R2DBD of the non-LTR retrotransposon endonucleases

The sequence encoding the DBD of the endonucleases encoded by the R2 non-LTR retrotransposons contained in the genome of the fruit fly *Drosophila melanogaster* [30] and the silkworm *Bombyx mori* [35] were located as described [25]. The codon usage was optimized *in silico* to avoid some restriction sites, and to allow their expression in mammalian cells. These DBDs were designated DmR2DBD and BmR2DBD, respectively. Genes were cloned to fuse them with an amino-terminal MBP or Histidine tag. The 4 fusion proteins, MBP-BmR2DBD, His-BmR2DBD, MBP-DmR2DBD and His-DmR2DBD, were produced in bacteria and purified by affinity chromatography. Preliminary EMSA showed that the four proteins were similarly able to bind specifically to the R2site probe [Figure 2a,d and Figure 3a]. Since the same results were obtained with BmR2DBD and DmR2DBD, only those obtained with BmR2DBD are illustrated in the figures. Data about DmR2DBD are supplied in [36].

Influence of tagging on the EMSA complexes assembled with MBP-R2DBD

EMSA analyses revealed that MBP-BmR2DBD and MBP-DmR2DBD assembled several different complexes with the R2site probe under our EMSA conditions [Figure 2a, lane 2]. Protein-DNA complexes were therefore assembled under similar conditions, cross-linked with UV, denatured by boiling, and analyzed by SDS-PAGE to determine how many proteins were involved in each of these complexes [Figure 2b]. The results revealed that there were three cross-linked complexes, with molecular weights indicating that they should correspond to one molecule of the R2site probe (MW = 99 kDa) linked to one protein molecule (MW = 99 + (1 x 59) = 158 kDa), two protein molecules (MW = 99 + (2 x 59) = 217 kDa), and three protein molecules (MW = 99 + (3 x 59) = 276 kDa), respectively. We also observed

that the number of complexes increased with the protein concentration, indicating that they resulted from protein oligomers, each being bound to a single R2site probe (data not shown). The fact that only one R2site probe was present in each complex was confirmed by performing short-long probe experiments, using two probes of 150-bp and 240-bp in length, one of which was radio-labeled, as previously described [36,37].

This made it impossible to determine the apparent Kd of the MBP-R2DBDs. Since fusion with the MBP allowed oligomerization to occur, the MBP tag of the MBP-BmR2DBD and MBP-DmR2DBD proteins was removed by specific site enzymatic cleavage with factor Xa at the hinge peptide between MBP and R2DBD. We first used PAGE to check that this enzymatic treatment did not degrade the peptides by proteolysis [Figure 2c]. EMSA analysis revealed that the binding activity of the BmR2DBD released was reduced 1000-fold as a result of the cleavage treatment [Figure 2d, lane 2 *versus* 3 to 6].

The apparent Kd of the cleaved protein was determined by EMSA. The saturation graph was plotted, and used to calculate a fixed concentration of 102 nM for the BmR2DBD protein, and of 116 nM for the DmR2DBD protein (data not shown). EMSA assays were then performed using the appropriate fixed concentration of the protein with variable concentrations of the R2 site probe. Under these assay conditions, the BmR2DBD and DmR2DBD were found to be low affinity proteins with an apparent Kd of more than 1 µM. Taking into account the data presented below, our results indicated that the release of the MBP tag did not seem to change the specific activity of either R2DBD to any great extent, but did modify their affinity, probably by destabilizing them or modifying their folding.

Properties of R2DBD

Under our laboratory conditions, and for all protein concentrations, His-BmR2DBD and His-DmR2DBD each assembled only in one complex with the R2site probe [Figure 3a]. To find out how many proteins were bound to each R2site molecule, UV cross-linking assays were carried out with both R2DBDs [Figure 3b]. The molecular weight of the only cross-linked formed complex indicated that only one protein was bound to each R2site probe. Both purified proteins were also used to determine the apparent Kd. Saturation graphs were plotted and used to calculate a fixed concentration of 6.6 nM for the His-BmR2DBD protein and of 5 nM for His-DmR2DBD [Figure 3c and d]. EMSA assays were then performed using a fixed concentration of protein and variable concentrations of the R2site. The apparent Kd of His-BmR2DBD for the R2site probe was estimated to be 30 nM, and that of His-DmR2DBD to be 18 nM [Figure 3e and f]. These two apparent Kd values were similar, and higher than those obtained using Factor Xa cleaved proteins. Overall, they showed that the tag used had a major impact on the binding of both R2DBDs to the R2site probe, and on their ability to oligomerize. Indeed, the 89 amino acid residues located upstream to R2DBD [28] were removed to suppress homo-dimerization. Our interpretation of the data herein and those of the literature is that the dimerization activity is contained within the amino acid residues located at the N-terminal end of R2DBD. These residues would be accurately folded and stabilized only when they are fused with certain peptides at their N-terminal end. Under this interpretation, the fusion of R2DBD to the MBP would allow the release of the dimerization activity.

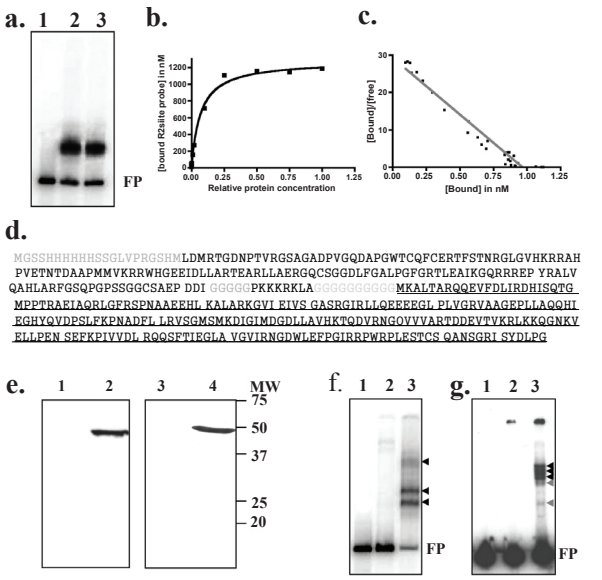

Figure 4: Features of His-BmR2DBD expressed in HeLa cells. (a) Shifted complexes assembled with the R2site probe in the absence of protein (lane 1), in the presence of purified protein produced in bacteria (lane 2) or HeLa cell extract expressing His-BmR2DBD (lane 3). FP: Free Probe. **(b)** Saturation graphs set at the binding equilibrium with a fixed concentration probe and variable amounts of HeLa cell extracts containing His-BmR2DBD. **(c)** Scatchard plot of the single complex obtained with His-BmR2DBD. The bound/Free (B/F) R2site probe was plotted *versus* concentration of bound R2site probe [nM], leading to the Scatchard plot. 1/Slope of the line gives the apparent Kd. The R^2 correlation coefficient calculated from data used for the Scatchard linear regression was 0.96. **(d)** Sequence of the His-BmR2DBD-LexA fusion. The peptide containing the Histidine tag is typed in dark gray and bold. The linker is typed in italics and bold, the flexible poly-Glycine tracts are typed in light gray, and the NLS in black. BmR2DBD and the LexA part are typed in black, the latter being underlined. **(e)** Western Blot analysis of the His-BmR2DBD-LexA protein expressed in HeLa cells. Lanes 1 and 3: HeLa cells extract transfected with pCS2. Lanes 2 and 4: HeLa cells extract transfected with pCS2-BmR2DBD-LexA vector. Revelation was done with an anti-Histidine tag antibody (lanes 1 and 2) and anti-LexA antibody (lanes 3 and 4). (**f and g**) His-BmR2DBD-LexA complexed with R2site **(f)** and LexA-BS **(g)** probes. Lane 1: probe alone; lane 2: protein extract from cells transfected with pCS2; lane 3: protein extract from cells transfected with pCS2-BmR2DBD-LexA. Black arrows indicate the shifted complexes. Those in gray indicate dissociated complexes that varied in intensity for the various experiments and binding conditions. FP: Free Probe.

Model***	N° of BS	Locations in chromosomes
Homo sapiens	3	chr2-133037437 chr8-70602505 chrX-108297465
Mus musculus	9	chr1-46132946 chr1-192944886 chr2-5300263 chr3-5860455 chr3-120952860 chr4-71689122 chr14-57312991 chr17-7146288 chr18-54858919
Rattus norvegicus	3	chr5-24497324 chr12-34422461 chr16-68375482
Gallus gallus	0	-
Drosophila melanogaster	0	-
Saccharomyces cerevisae	0	-

Table 2b: Number of R2DBD binding sites (TAAGGTAGCCAAATGCCTCG**) outside the chromosomal loci that contain the rRNA genes.

Target specificity of the His-BmR2DBD domain produced in eukaryotic cells

We found that under our operating conditions, the activity of the

Citation: Bigot Y, Carnus E, Demattei MV, Casteret S, Carpentier G, et al. (2012) Characterization of Monomeric Protein Domains that Bind Specifically to a Highly-Conserved 100-Bp DNA Target within rRNA Genes. rDNA Technol 1:101. doi:10.4172/rdt.1000101

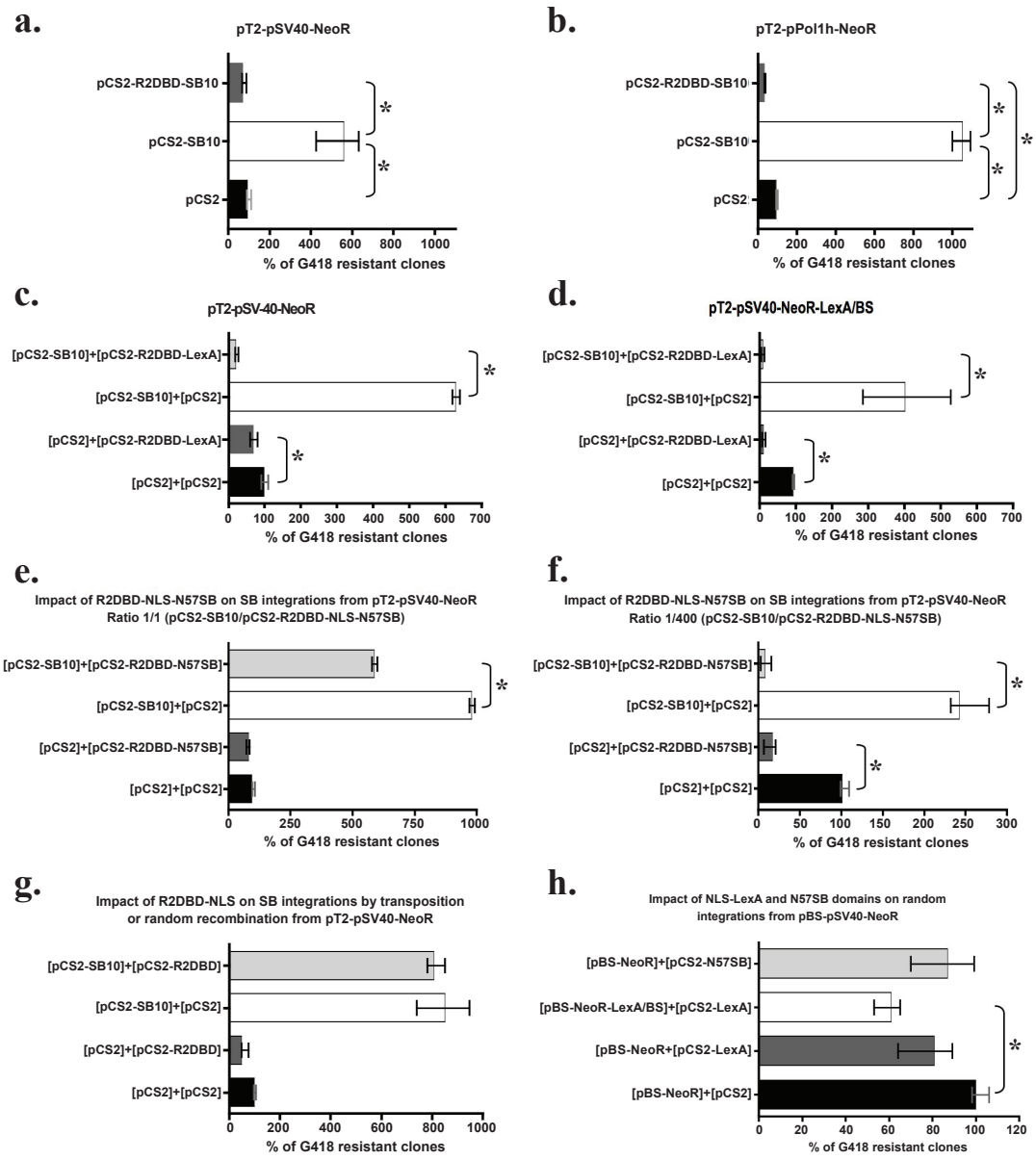

Figure 5: Impact of R2DBD peptides on *in vivo* NeoR cassette integration by random recombination of the plasmid carrier or transposition of the SB vector carrier. Transposition assays monitored in the absence (black bar) or presence of Tpase (SB10, white bar; R2DBD-SB10, dark gray bar) with a transposon source plasmid carrying a pSV40-NeoR **(a)** or a pPol1hNeoR cassette **(b)**. Impact of the R2DBD-NLS-LexA peptide in transposition assays monitored in the absence (black and dark gray bars) or presence of SB10 Tpase (white and light gray bars) with an SB transposon plasmid carrying a pSV40-NeoR cassette (**c and d**) and a LexA binding site (LexA/BS; **d**). (**e**) and (**f**), impact of the R2DBD-NLS-N57SB10 peptide in transposition assays monitored in the absence (black and dark gray bars) or presence of SB10 Tpase (white and light gray bars) with an SB transposon plasmid carrying a pSV40-NeoR cassette. 100 ng of pT2-pSV40-NeoR was transfected into 1x105 cells, but a 1/1 and 1/400 plasmid ratio between pCS2-SB10 and pCS2-R2DBD-N57SB10 were respectively used in (e) and (f) (100ng/100ng *versus* 1ng/400ng). **(g)**, impact of the R2DBD-NLS peptide in transposition assays monitored in the absence (black and dark gray bars) or presence of SB10 Tpase (white and light gray bars) with a transposon source plasmid carrying a pSV40-NeoR. **(h)**, impact of the NLS-LexA (black, dark gray and white bars) and N57SB10 peptides (light gray bar) on the integration by random recombination of pBS plasmids carrying a pSV40-NeoR cassette containing (white bars) or not a LexA/BS (white bars *versus* black, dark gray and light gray bars). In a, b, c, d, g and h, 100 ng of each plasmid was transfected into $1x10^5$ cells. Data were obtained from at least 6 replicates resulting from at least 2 different experiments. Percentages (%) of G418-resistant clones are calculated for each replicate using the sample represented by a black bar as a reference. They are represented with median and errors bars corresponding to quartiles 1 and 3. * and brackets indicate significant differences using a non-parametric, two-tailed Wilcoxon/Kruskal-Wallis test with a significance threshold of 0.05 (p>0.95).

His-BmDBD produced in bacteria was more stable than that of the His-DmR2DBD. It was therefore retained as the most promising candidate for investigation in HeLa cells. The size of His-BmR2DBD and its ability to bind to the R2site probe were first investigated using HeLa cell extracts in which the protein was transiently expressed. Western blot hybridized with an anti-His_6 monoclonal antibody revealed a single band with a molecular weight (18kDa), confirming His-BmR2DBD was expressed in HeLa cells (data not shown). EMSA experiments were therefore performed under conditions similar to those used with the proteins produced in bacteria. Similarly to previous findings with this protein,

only one main complex was assembled with the His-DmR2DBD and R2site probe, whatever the biological system in which it had been produced [Figure 4a, lane 2 *versus* 3]. HeLa cell extracts were then used to determine the apparent Kd of His-BmR2DBD. Saturation graphs were plotted to fix the concentration in HeLa cell extract [Figure 4b], and the apparent Kd was estimated to be 31 nM [Figure 4c]. This value was identical to that obtained with the protein produced in bacteria, indicating that expression in eukaryotic cells did not alter its activity.

Since we had previously observed that fusion of both R2DBDs with the MBP tag modified their ability to oligomerize, we investigated the properties of a fusion protein model that had previously been investigated as a potential way to target a nucleic acid vector close to its chromosomal integration target [14]. This consisted of fusing the BmR2DBD at its C-terminal end with a LexA DBD, and including a SV40 Nuclear Localization Signal (NLS) between the DBDs [Figure 4d]. The LexA DBD was an interesting model, because its functionality had previously been shown to depend on the properties of the domain to which it is fused [38], making it possible to find out whether His-BmR2DBD can modify the properties of a fused DBD. The profile of the R2DBD-LexA expressed in HeLa cells was verified. Western blot analysis revealed that the fusion was expressed as a single protein of about 46 kDa [Figure 4e; theoretical molecular weight = 45.6 kDa]. EMSA analyses done with the R2site probe revealed that the BmR2DBD in the fusion protein was active, but that three complexes were observed [Figure 4f, lane 3]. Similarly, EMSA indicated that the LexA part was also active, and once again, three complexes were also observed [Figure 4g, lane 3]. The presence of three complexes was very probably not due to BmR2DBD oligomerization, but to the fact that the LexA DBD can occur as a monomer, a dimer, or a tetramer in solution [38]. In contrast to MBP fusions, saturation graphs could be calculated for these proteins, although these fusions were able to assemble in several different complexes as a result of LexA DBD oligomerization. Similar UV cross-linking investigations and short-long probe EMSA indicated that they all contained a single DNA probe to which the fusion was oligomerized (data not shown). These controls made it possible to perform affinity analyses, which revealed that the values of the apparent Kd of both domains in the fusion proteins were similar to those of each isolated domain (about 20 nM; [38]).

Impact of R2DBD fusions on DNA integration in HeLa cells

The effects of several BmR2DBD fusions were assayed on the ability of *SB* vectors to integrate into chromosomes by transposition and that of a plasmid to integrate by random recombination. Integrations by transposition and random recombination occur at different rates during an assay monitored with a transposon vector system, depending on the cell lineage used and plasmid amounts transfected [39,40]. As a consequence, a NeoR cassette can integrate by transposing with the *SB* vector into chromosomes in presence of Tpase. A fraction of the plasmids carrying the *SB* vector will also integrate into chromosomes as a result of random recombination, but in general this happens less often than transposition. In the control assays carried out in absence of *SB* Tpase, only plasmid integrations by random recombination occur.

A first set of assays was monitored with an R2DBD fused to the N-terminal end of the *SB* Tpase (R2DBD-SB10; S2). One objective of these assays was to find out whether R2DBD-SB10 was able to mediate integration into rRNA genes. Two *SB* vectors were used (Table 1). The first vector, pT2-pSV40-NeoR, carried a NeoR cassette with expression that depended on an RNA polymerase II promoter, pSV40, and the second vector, pT2-pPol1h-NeoR, had NeoR cassette expression that depends on the human RNA polymerase I promoter, pPol1h [41]. The interest of this second construct is that pPol1h contains cryptic RNA polymerase II promoters that allow expression of NeoR cassettes integrated into nucleolar rRNA genes or elsewhere in the genome [42]. Our results indicated that R2DBD-SB10 was unable to achieve transposition, whatever the *SB* vector used (Figure 5a and b; white bars versus dark gray bars). These data agreed with those obtained with most of the *SB* Tpases so far fused to a DBD [14-16]. Interestingly, we found that the numbers of NeoR clones obtained in assays monitored with R2DBD-SB10 (Figure 5a and b, dark gray bars) were systematically lower than those in the controls done in the absence of Tpase (Figure. 5a and b, black bars). Statistical analyses indicated that this difference was only significant (p>95%) when pT2-pPol1hNeoR was used as the *SB* vector source.

In a second set of assays, we checked the impact of several peptides theoretically designed to target a DNA plasmid in the proximity of the rRNA genes [9,14]. First, we analyzed the effect of R2DBD-NLS-LexA peptide on the integration ability of *SB* vectors. Two plasmid sources of *SB* vectors were used: pT2-pSV40-NeoR and pT2-pSV40-NeoR-LexA/BS, in which a LexA binding site was cloned within the SB vector (Table 1). Our results indicated that the R2DBD-NLS-LexA expression in HeLa cells inhibited the transposition of both *SB* vectors (Figure. 5c and d, white bars versus light gray bars). Here too, we found that the numbers of NeoR clones obtained in the absence of SB10 Tpase, and in the presence of R2DBD-NLS-LexA (Figure. 5c and d, dark gray bars), were statistically lower than those of the control assays monitored in the absence of both proteins (Figure. 5c and d, black gray bars). We conclude that R2DBD-NLS-LexA was able to inhibit the *SB* vector transposition and the plasmid integration by random recombination. Because a similar effect of R2DBD-NLS-LexA was observed with both *SB* vector sources, we concluded that LexA interfered non-specifically with both plasmids independently of the presence of LexA/BS (Figure. 5c and d, top bars). Since some studies in the literature [43] have confirmed that LexA DBD has lower DNA binding specificity, we assayed a second peptide R2DBD-NLS-pZFD$^{bcr\text{-}abl}$ [44] in which pZFD$^{bcr\text{-}abl}$ is composed of a concatemer of four ZFs that recognizes a 12-bp motif (GACGCAGAAGCC; ZBS). ZBS is not found in the human genome, except in the genomes of most patients with chronic myelogenous leukemia (CML), and in some patients with acute lymphoblastic leukemia (ALL) or acute myelogenous leukemia (AML), in which the BRC and ABL genes are fused. Two pBS-NeoRs were used for the integration assays: pBS-NeoR and one pBS-NeoR derivative in which was cloned a ZBS to produce pBS-NeoR-ZBS. Whichever plasmid was used, our results indicated that R2DBD-NLS-pZFD$^{bcr\text{-}abl}$ also lacked binding specificity under our experimental conditions (S3). These results were confirmed for both peptides in experiments in which the pSV40-NeoR cassette was replaced by a pPol1h-NeoR cassette (data not shown).

The properties of a third peptide that theoretically works in a different way were therefore investigated. This peptide was composed of an R2DBD fused to a domain corresponding to the first 57 N-terminal residues of *SB*10 Tpase (N57SB; S4). N57SB has previously been proposed to be a Tpase-interacting domain that is able to interact with full-length SB Tpase complex bound to the *SB* vector ITRs [9,14]. The N57SB10-NLS-TetR peptide has also been demonstrated to be able to modify the integration preferences of SB vectors in the proximity of a TetR binding site [14]. For the integration assays, two ratios of the plasmids pCS2-SB10 and pCS2-R2DBD-NLS-N57SB were used. The results revealed that R2DBD-NLS-N57SB inhibited SB vector transposition, the intensity of inhibition depended on the amount of R2DBD-NLS-N57SB expressed in the cells (Figure. 5e and f, white bars versus light gray bars). Unexpectedly, controls done in the absence of SB10 Tpase

indicated that R2DBD-NLS-N57SB also inhibited plasmid integration by random recombination, at least under conditions in which R2DBD-NLS-N57SB was abundant (Figure 5f, black bars versus dark gray bars). Complementary experiments indicated that this effect was statistically significant from a plasmid ratio of 1:4 (data not shown). In agreement with structural data [45,46], our results therefore support the suggestion that N57SB is above all a DBD that binds to the SB vector ends (ITRs) contained in pT2-pSV40-NeoR.

In conclusion, R2DBD-NLS-LexA, R2DBD-NLS-pZFD$^{bcr\text{-}abl}$ and R2DBD-NLS-N57SB all have the ability to inhibit SB vector transposition and plasmid integration by random recombination. Preliminary results obtained with *piggyBac* vectors, R2DBD-NLS-LexA and R2DBD-NLS-pZFD$^{bcr\text{-}abl}$ led to similar conclusions, indicating that the effect was independent of the transposon used (data not shown; only one replicate of each of the transposition assays shown in Figure. 5 was done). In agreement with these conclusions, we also found that R2DBD-NLS-LexA, R2DBD-NLS-pZFD$^{bcr\text{-}abl}$ and R2DBD-NLS-N57SB have the ability to decrease the expression of a gene beard by a co-transfected plasmid (data not shown). These results were obtained using an internal control of our transposition assays (see Material and Methods). We checked 24 hours post transfection that plasmids were transfected by following the transient luciferase expression of the co-transfected pGL3 plasmid. We observed that luciferase activities were all 2 to 5-fold lower in HeLa cells transfected with plasmids expressing R2DBD-NLS-LexA or R2DBD-NLS-pZFD$^{bcr\text{-}abl}$, 5 to 20-fold lower in those expressing with R2DBD-NLS-N57SB, and similar to the controls in those expressing R2DBD-NLS, NLS-LexA or NLS-N57SB.

Overall, these inhibitory effects did not depend on the domain fused to R2DBD or on the promoter of the marker gene since we obtained similar results with all three peptides using a pBS-NeoR plasmid. However, their effects did depend on their cellular level of expression, which was to be expected since they do not have the same affinities for DNA binding sites. None of the domains fused here to R2DBD, i.e. LexA, pZFDbcr abl and N57SB, has previously been found to inhibit transposition and recombination [14-16] when associated with another DBD or enzyme. We therefore concluded that the inhibitory effect was allocated by R2DBD to the fusion peptides. Two hypotheses can be proposed to explain our results. The first supposes that the expression of R2DBD peptides has a toxic effect on HeLa cells, and so killed the transfected cell population. The second proposes that the nuclear trafficking of the plasmid to the target was affected, and that this impaired plasmid access to the transposition and recombination machineries.

Impact of R2DBD peptides on HeLa cells

To find out whether R2DBD peptides were toxic for HeLa cells, the effect of R2DBD-NLS was assayed on the transposition and plasmid integration of *SB* vectors. The results showed that R2DBD-NLS did not have any effect on transposition (Figure. 5g, white bars *versus* light gray bars) or recombination (Figure. 5g, black bars versus dark gray bars). The effects of the NLS-LexA and NLS-N57SB were also assayed. Our findings also showed that they had no effects (Figure. 5h), even though we did detect a weak effect with NLS-LexA when a plasmid containing a LexA/BS was used (Figure. 5h, black bars *versus* white bars). The toxicities of R2DBD and the peptides in which it was fused to LexA and N57SB were then verified using a viability assay evaluating cell survival during the 48 hours following the plasmid transfection (S5). We found that the viability of the HeLa cells was not altered, whatever R2DBD peptide was expressed. We therefore concluded that R2DBD peptides had no toxic effect on HeLa cells under our experimental conditions, and that their effects on transposition and recombination were attributable to some other cause.

The localization of R2DBD in nuclei was verified by transfecting HeLa cells with plasmids expressing an R2DBD-NLS-GFP fusion or a variant expressing an R2DBD-GFP fusion in which the SV40NLS had been removed by mutagenesis (PKKKRK to PKAIRK [47]). The expression profiles of both peptides were compared to those of HeLa cells transfected with control plasmids expressing GFP, a protein that passively diffuses into all cell compartments, or a MOS1-GFP, a protein fusion that is actively concentrated in the nucleus (personal data). Our findings indicated that fusions of both GFPs with R2DBD were actively accumulated in the nucleus (Figure. 6a). This indicated that the R2DBD sequence contained NLS information that was efficient enough to monitor a nuclear import as strong as that mediated by SV40 NLS. PSORT II facilities at http://psort.hgc.jp/form2.html allowed us to identify two NLS candidates (HKRR and PMMVKRR). The findings also revealed that the nuclear distribution of the green fluorescent signal was unexpected, since it was not homogenous with either of the GFP fusions, both of which assembled fluorescent "mottling" around and between the nucleoli. This feature was in striking contrast with the results obtained in the controls in which the GFP or the MOS1-GFP fluorescence were found to diffuse homogenously within the nuclei (Figure 6a). This

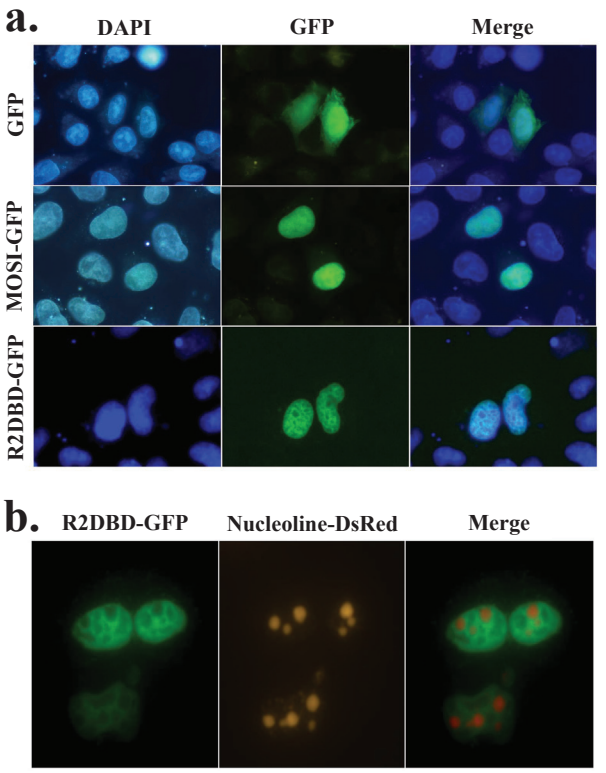

Figure 6: Location of R2DBD-NLS-GFP and R2DBD-GFP fusions expressed in HeLa cells. (a) Analysis of DAPI and GFP-fluorescence distribution in HeLa cells transfected with plasmids expressing GFP, MOS1-GFP or R2DBD-NLS-GFP proteins. GFP and MOS1-GFP were used as negative and positive controls of active nuclear import, respectively. The nuclear DNA was stained with DAPI. Merge pictures were used to confirm the co-localization between the nucleus and GFP proteins. Since the results were similar, only those obtained with R2DBD-NLS-GFP are shown. **(b)**. Analysis of GFP and DsRed-fluorescence distribution in HeLa cells co-transfected with plasmids each of which expressed the R2DBD-NLS-GFP and pDsRed-C1Hsnucleoline fusions. pDsRed-C1Hsnucleoline was used to label nucleoli. Merge pictures were used to verify the co-localization between nucleoli and GFP signal in nuclei. Since results were similar with both fusions, only those obtained with R2DBD-NLS-GFP are shown in a and b.

was therefore further investigated by co-expressing both R2DBD-GFP fusions with a nucleolar marker, human nucleoline [34]. The composite image created by merging the GFP and DsRED signals confirmed that R2DBD-GFP was indeed assembled as mottling organized around and between nucleoli, and that R2DBD-GFP did not co-localize with nucleolin within the nucleoli (Figure 6b). These observations led us to propose that R2DBD was located around and between the nucleoli, and probably around other small nucleolar accessory bodies [48]. Such a location could explain why the R2DBD-GFP signal formed mottled patterns, and suggests that R2DBD has an affinity for a nuclear protein that is abundant in nuclear organelles and molecular networks ensuring connections between them. Fibrillarin is a potential candidate [49].

Discussion

The positive and negative results reported here supply important information that needs to be analyzed from three aspects concerning i) the R2DBD properties, ii) the unexploited advantages of the R2DBDs, and iii) the limitations of fusion peptides for targeting integrations of non-viral vector and their alternatives.

Properties and current limitations of R2DBDs

Our objective was to characterize at least one DBD that has the properties required to specifically target rRNA genes, whatever the eukaryotic genome in which it could be used. The interests of the strategy that consist in targeting insertions of transgenic cassettes into rRNA genes in eukaryotic genomes are summarized in S6.

Here, our data show that R2BDBs have at least five of the properties that are required in such a peptide. Firstly, R2DBDs are able to bind specifically to a 20-bp motif located within the 100-bp region that is conserved among all eukaryotes. Moreover, the affinity of R2DBDs for the 20 bp target is rather high *in vitro*, their apparent Kds being around 20 nM. Secondly, R2DBDs bind to their DNA target in a monomeric form. This property is very important if R2DBDs are used to target other molecules, because it should prevent the assembly of homo-aggregates. Thirdly, the R2DBD sequences contain strong NLS information, which should avoid the addition of an NLS in the synthetic peptide. Fourthly, R2DBDs retain their DNA binding properties when there are fused to another protein or domain. Fifthly, the functionality of a protein fused to R2DBD is preserved, as illustrated here with LexA, pZFD$^{bcr\text{-}abl}$, N57SB and GFP. We found that the functionality of SB10 is in fact altered by its N-terminal fusion to R2DBD. However, this is very probably due to the fact that SB10 cannot tolerate N-terminal fusion, which is clearly illustrated in the literature with domains of various origins [14-16].

Our results also show that a few properties of the R2DBDs still need to be optimized before they become fully usable. Indeed, we observed that R2DBD-GFP fusions did not penetrate into all nuclear compartments, but were concentrated around nuclear organelles, nucleoli and other small nucleolar accessory bodies, that are connected to each other by proteins such as fibrillarin [48,49]. This property, that is probably due to the absence of a nucleolar localization signal (NoLS), means that in their present form R2DBDs are unable to reach the rRNA genes in the nucleoli *in vivo*. There are ways to circumvent both problems independently. Firstly, the localization of R2DBD in nucleoli can be determined by adding an NoLS at its N- or C-terminal ends, as described in [50]. Secondly, the potential affinity for the surface of nuclear organelles could be eliminated by screening a library of punctually mutated R2DBDs, using a single or a two yeast-hybrid system to locate mutants that have lost the ability to interfere, for example, with fibrillarin. This second point still needs to be engineered after adding a NoLS at one end of R2DBD. Even if it is relatively simple to think up ways of doing this, the implementation of such mutant screening will require considerable investment, since the properties of several mutants that do not assemble around nuclear organelles, will all need to be verified in order to identify one or more optimal DBD candidates.

Unexploited properties of R2DBD

Previous studies [28,34] have together demonstrated that R2DBD is able to bind specifically to a 20-bp rRNA target. The findings reported here indicate that R2DBD has no dual DNA binding activities, i.e. one DNA binding activity with high affinity for its 20-bp target, and a second activity characterized by the absence of specificity and low binding affinity. Indeed, we observed that R2DBD expressed alone *in vivo* does not inhibit the DNA plasmid integrations by transposition and recombination (Figure. 5g), and does not hamper the luciferase expression from a plasmid. The absence of dual DNA binding activities is not common in either natural or synthetic DBDs. Hence, the three DBDs used here to be fused to R2DBD, LexA, pZFD$^{bcr\text{-}abl}$ and N57SB, were also found to be able to interact non-specifically *in vivo* with plasmids that do not contain their specific binding target (Figure. 5h and S3). For LexA and N57SB DBDs, our observations confirm observations already reported in the literature, even though there are few such examples [43,51]. In the case of pZFD$^{bcr\text{-}abl}$, this was a priori not expected for a tetra-ZF domain, although dual DNA binding activities had already been identified for pZFDs bigger than tri-ZF. The objective of most molecular methods so far developed to obtain pZFDs is to produce tailored tri-ZFs. Fused to a nuclease or a recombinase catalytic domain, such tri-ZFs are sufficient to bind to a 18-bp binding target by dimerizing two ZF-nuclease or ZF-recombinase monomers [52]. Few molecular methods have been developed to manufacture hexa- and nona-ZF monomers to bind to 18- and 30-bp DNA motifs, respectively [53-56]. If tailored hexa- and nona-ZF monomers have nM affinity for their binding target, their organization into several blocks of di- or tri-ZF means that they are also able to bind with a lower affinity to smaller motifs similar to submotifs contained in their specific 18- and 30-bp DNA target. In an eukaryotic genome, the ability of a hexa-ZF monomer to recognize a unique, 18-bp chromosomal motif is therefore going to be a balance between its specificity for its target and its ability to bind with lower affinity to hexa- and nona-bp motifs that are much more abundant in eukaryotic genomes. These features of the pZFDs greater than tri-ZF explain why we found here that pZFD$^{bcr\text{-}abl}$ is able to interact non-specifically *in vivo* with plasmids that do not contain its binding site (S3).

Overall, these observations suggest that R2DBDs may have two very interesting properties for certain biotechnological purposes requiring highly specific DNA binding of a 20-bp target. First, R2DBDs seem to be unable to bind non-specifically to DNA. Second, R2DBDs seem unable to bind to DNA submotifs contained in the sequence of its 20-bp rDNA target. The 3D structure of R2DBDs is probably responsible for their binding as a single block to the 20-bp rDNA target.

To target one specific chromosomal locus requires that only one 18-bp DNA motif in eukaryote genomes is recognized. This means that pZFD monomers designed to target DNA motifs of at least 12-bp cannot ensure the uniqueness of their binding to their target as reliably as R2DBDs do to their 20-bp rDNA target. The core structure of R2DBDs consists of one ZF associated with a *Myb*-like sub-domain [28,29] (Figure. 1d). Starting from the R2DBD backbone, it might therefore be of interest to develop the molecular engineering of a new type of monomeric DBDs that would be tailored to the 20-bp target sequence to which they have to bind. The design of molecular methods to select

such tailored R2DBD derivatives might be facilitated by the fact that the ZF structure in R2DBD (modeled using I-TASSER [57]) is very similar to that used to design tailored ZF. Using data from the ZiFit database, it should therefore be possible to fit this ZF moiety to the nucleotide triplet to which it binds [21], and so "only" have to design a selection procedure on the *Myb*-like sub-domain. *Myb*-like sub-domains are composed of 3 main α helices in which only the third is involved in DNA binding [29]. Obtaining a crystal structure of an R2DBD bound to its 20-bp DNA target should help to confirm the viability of this approach.

Alternatives to fusion peptides to target vectors

Data recovered here about the functioning of the LexA, pZFD$^{bcr-abl}$ and N57SB confirm the need to have rigorous control procedures in place before attempting to use a protein domain in a synthetic peptide system. Nevertheless, our experience with such peptides suggests that these controls will very probably be far from sufficient to control the targeting. Indeed, whatever integration process is used (i.e. HR or transposition) and whatever locus is targeted (rRNA genes or other loci), the localization of the plasmid in the vicinity of the locus into which it must be targeted is going to depend to a large extent on the balance between two parameters: the peptide expression rate in the cell and the amount of plasmid able to reach the nuclei. A third parameter is that this balance is going to be complicated by the fact that the plasmid transfection efficiency is going to vary over about 4 orders of magnitude in a single cell population (this is strongly suggested by data obtained from the FACS analysis of a cell population transfected with a plasmid expressing the GFP). To date, although the third parameter seems to be difficult to control, the transfection of a pre-assembled complex (plasmid-targeting peptide) is a technical solution that should be kept in mind, since it could make it possible to circumvent the problems related to the two first parameters. One way to assemble such a complex could be to use a DNA oligonucleotide that is able, on the one hand, to be specifically annealed to the plasmid [10], and on the other hand, to be bound or cross-linked to a DBD. In such a procedure, the oligonucleotide might first be annealed to the plasmid. Depending on the presence of a biotin or SH group grafted at one end of the -oligonucleotide, a streptavidin-DBD peptide or a polyNQ-DBD peptide might be then bound or chemically cross-linked to the oligonucleotide/plasmid complex. The efficiency to anneal the oligonucleotide onto a plasmid, and of the binding of a monovalent streptavidin to a biotin [58], or of the cross-linking with the Sulfo-EMCS cross-linker between a SH group and a polyNQ tail mean that this might be a viable solution. As previously pointed out, even if it is relatively simple to think up such solutions, their implementation will require considerable investment.

Conclusion

R2DBDs are promising candidates to target DNA molecules in the immediate vicinity of the rRNA genes since they i) specifically recognize, with high affinity, a 20-bp binding site located within the expected genomic rDNA target, ii) act as monomers, iii) contain a nuclear localization signal, iv) remain functional when fused to another domain and, v) do not alter the functionality of the protein to which they are fused. However, R2DBDs have two properties remain to be engineered before these DBDs can be integrated into a molecular targeting system directed into rRNA genes. The first concerns the ability of R2DBD to locate within the nucleolus, the organelle in which the rRNA genes reside. The second is the tendency of R2DBD to accumulate in certain parts of the nuclei, which limits its diffusion within nuclei. Technological solutions are available to engineer these DBDs. The results reported here also supply important information concerning the R2DBD properties and the plasmid DNA targeting within nuclei that will need to be further analyzed from three aspects; the unexploited advantages of the R2DBDs, the possibilities and limitations of fusion peptides for targeting integrations of non-viral vector, and the alternatives to fusion peptides for targeting vectors.

Supplementary Information Files

S1. Supplementary Information 1. Design and properties of pZFD proteins

S2. Supplementary Information 2. Amino acid and nucleotide sequence features of R2DBD-SB10

S3. Supplementary Information 3. Impact of R2DBD-NLS-ZFD on the integration by random recombination of pBS-NeoR or pBS-NeoR-ZBS (Ratio 1/1 (pCS2-SB10/pCS2-R2DBD-NLS-ZFD)

S4. Supplementary Information 4. Amino acid sequence of the R2DBD-NLS-N57SB fusion

S5. Supplementary Information 5. Impact of R2DBD peptides on cell viability

S6. Supplementary Information 6. Interest of targeting insertions of exogenous DNA fragments into rRNA genes

Acknowledgements

We would like to thank Prof. Brian McStay for supplying human RNA polymerase I promoters and helpful scientific information and recommendations. The English text has been revised by Dr M. Ghosh. This work was supported by the C.N.R.S. and the François Rabelais University of Tours, and funded by grants from the European Commission (Project SyntheGeneDelivery, N°018716), the Ministère de l'Education Nationale, de la Recherche et de la Technologie, the Association Française contre la Myopathie and the Groupement de Recherche CNRS 2157.

References

1. Klabunde J, Kunze G, Gellisen G, Hollenberg CP (2004) Wide-range integrative expression vectors for fungi based on ribosomal DNA elements. In: Production of recombinant proteins Edited by Gellissen G. Weinheim. Wiley-VCH.
2. Dos Santos WG, Buck GA (2000) Simultaneous stable expression of neomycin phosphotransferase and green fluorescence protein genes in Trypanosoma cruzi. J Parasitol 86: 1281-1288.
3. Stewart S, MacDonald N, Perkins E, deJong G, Perez C, et al. (2002) Retrofitting of a satellite repeat DNA-based murine artificial chromosome (ACes) to contain loxP recombination sites. Gene Ther 9: 719-723.
4. Wen L, Liang S, Wu LQ, Cai F, Pan Q, et al. (2008) Efficient protein expression from the endogenous RNA polymerase I promoter using a human ribosomal DNA targeting vector. Biochem Biophys Res Commun 367: 846-851.
5. Jakubczak JL, Burke WD, Eickbush TH (1991) Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. Proc Natl Acad Sci USA 88: 3295-3299.
6. Gentile KL, Burke WD, Eickbush TH (2001) Multiple lineages of R1 retrotransposable elements can coexist in the rDNA loci of Drosophila. Mol Biol Evol 18: 235-245.
7. Robinett CC, O'Connor A, Dunaway M (1997) The repeat organizer, a specialized insulator element within the intergenic spacer of the Xenopus rRNA genes. Mol Cell Biol 17: 2866-2875.
8. Liu X, Liu M, Xue Z, Pan Q, Wu L, et al. (2007) Non-viral ex vivo transduction of human hepatocyte cells to express factor VIII using a human ribosomal DNA-targeting vector J Thromb Haemost 5: 347-351.
9. Voigt K, Izsvák Z, Ivics Z (2008) Targeted gene insertion for molecular medicine. J Mol Med 86: 1205-1219.
10. Chiou CC, Chen SW, Luo JD, Chien YT (2011) Monitoring triplex DNA formation with fluorescence resonance energy transfer between a fluorophore-labeled probe and intercalating dyes. Anal Biochem 416: 1-7.
11. Penton EH, Crease TJ (2004) Evolution of the transposable element Pokey in

the ribosomal DNA of species in the subgenus Daphnia (Crustacea: Cladocera). Mol Biol Evol 21: 1727-1739.

12. Wu SC, Meir YJ, Coates CJ, Handler AM, Pelczar P, et al. (2006) piggyBac is a flexible and highly active transposon as compared to Sleeping beauty, Tol2, and Mos1 in mammalian cells. Proc Natl Acad Sci USA 103:15008-15013.

13. Maragathavally KJ, Kaminski JM, Coates CJ (2006) Chimeric Mos1 and piggyBac transposases result in site-directed integration. FASEB J 20:1880-1882.

14. Ivics Z, Katzer A, Stüwe EE, Fiedler D, Knespel S, et al. (2007) Targeted Sleeping Beauty transposition in human cells. Mol Ther 15: 1137-1144.

15. Wilson MH, Kaminski JM, George AL Jr (2005) Functional zinc finger/sleeping beauty transposase chimeras exhibit attenuated overproduction inhibition. FEBS Lett 579:6205-6209

16. Yant SR, Huang Y, Akache B, Kay MA (2007) Site-directed transposon integration in human cells. Nucleic Acids Res 35: e50.

17. Nomura W, Sugiura Y (2007) Design and synthesis of artificial zinc finger proteins. Methods Mol Biol 352: 83-93.

18. Pabo CO, Peisach E, Grant RA (2001) Design and selection of novel Cys2His2 zinc finger proteins. Annu Rev Biochem 70: 313-340.

19. Palazzoli F, Carnus E, Wells DJ, Bigot Y (2008) Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration. Curr Gene Ther 8: 367-390.

20. Mandell JG, Barbas CF (2006) Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. Nucleic Acids Res 34: W516–W523.

21. Sander JD, Zaback P, Joung JK, Voytas DF, Dobbs D (2007) Zinc Finger Targeter (ZiFiT): an engineered zinc finger/target site design tool. Nucleic Acids Res 35: W599- W605.

22. Fu F, Sander JD, Maeder M, Thibodeau-Beganny S, Joung JK, et al. (2009) Zinc Finger Database (ZiFDB): a repository for information on C2H2 zinc fingers and engineered zinc-finger arrays. Nucleic Acids Res 37: D279-D283.

23. Wright DA, Thibodeau-Beganny S, Sander JD, Winfrey RJ, Hirsh AS, et al. (2006) Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. Nat Protoc 1: 1637-1652.

24. Han JS (2010) Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mob DNA 1: 15.

25. Yang J, Eickbush TH (1998) RNA-induced changes in the activity of the endonuclease encoded by the R2 retrotransposable element. Mol Cell Biol 18: 3455-3465.

26. Burke WD, Malik HS, Jones JP, Eickbush TH (1999) The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. Mol Biol Evol 16: 502-511.

27. Kojima KK, Fujiwara H (2004) Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. Mol Biol Evol 21: 207-217.

28. Christensen SM, Bibillo A, Eickbush TH (2005) Role of the Bombyx mori R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction. Nucleic Acids Res 33: 6461-6468.

29. Ogata K, Morikawa S, Nakamura H, Sekikawa A, Inoue T, et al. (1994) Solution structure of a specific DNA complex of the *Myb* DNA-binding domain with cooperative recognition helices. Cell 79: 639-648.

30. Roiha H, Glover DM (1981) Duplicated rDNA sequences of variable lengths flanking the short type I insertions in the rDNA of Drosophila melanogaster. Nucleic Acids Res 9: 5521-5532.

31. Rengarajan K, Cristol SM, Mehta M, Nickerson JM (2002) Quantifying DNA concentrations using fluorometry: a comparison of fluorophores. Mol Vis 8: 416-421.

32. Askar KA, Kudi AC, Moody AJ (2011) Purification of soluble acetylcholinesterase from sheep liver by affinity chromatography. Appl Biochem Biotechnol 165 : 336-346.

33. Becherel OJ, Gueven N, Birrell GW, Schreiber V, Suraweera A, et al. (2006) Nucleolar localization of aprataxin is dependent on interaction with nucleolin and on active ribosomal DNA transcription. Hum Mol Genet 15: 2239-2249.

34. Xiong Y, Burke WD, Jakubczak JL, Eickbush TH (1988) Ribosomal DNA insertion elements R1Bm and R2Bm can transpose in a sequence specific manner to locations outside the 28S genes. Nucleic Acids Res 16: 10561-10573.

35. Fujiwara H, Ogura T, Takada N, Miyajima N, Ishikawa H, et al. (1984) Introns and their flanking sequences of Bombyx mori rDNA. Nucleic Acids Res 12: 6861-6869.

36. Carnus E (2009) Conception d'un système protéique pour le ciblage de vecteurs non viraux au niveau des gènes codant les ARN ribosomiques Ressource électronique. University François Rabelais of Tours.

37. Augé-Gouillou C, Brillet B, Hamelin MH, Bigot Y (2005) Assembly of the mariner Mos1 synaptic complex. Mol Cell Biol 25: 2861-2870.

38. Schnarr M, Granger-Schnarr M, Hurstel S, Pouyet J (1988) The carboxy-terminal domain of the LexA repressor oligomerises essentially as the entire protein. FEBS Let 234: 56-60.

39. Izsvák Z, Ivics Z, Plasterk RH (2000) Sleeping Beauty a wide host-range transposon vector for genetic transformation in vertebrates. J Mol Biol 302: 93-102.

40. Grabundzija I, Irgang M, Mátés L, Belay E, Matrai J, et al. (2010) Comparative analysis of transposable element vector systems in human cells. Mol Ther 18: 1200-1209.

41. Palmer TD, Miller AD, Reeder RH, McStay (1993) Efficient expression of a protein coding gene under the control of an RNA polymerase I promoter. Nucleic Acids Res 21: 3451-3457.

42. Smale ST, Tjian R (1985) Transcription of Herpes Simplex Virus tk Sequences Under the Control of Wild-Type and Mutant Human RNA Polymerase I Promoters. Mol Cell Biol 5: 352-362.

43. Wade JT, Reppas NB, Church GM, Struhl K (2005) Genomic analysis of LexA binding reveals the permissive nature of the Escherichia coli genome and identifies unconventional target sites. Genes Dev 19: 2619-2630.

44. McNamara AR, Ford KG (2000) A novel four zinc-finger protein targeted against p190 (BcrAbl) fusion oncogene cDNA utilisation of zinc-finger recognition codes. Nucleic Acids Res 28: 4865-4872.

45. Yant SR, Park J, Huang Y, Mikkelsen JG, Kay MA (2004) Mutational analysis of the N-terminal DNA-binding domain of sleeping beauty transposase critical residues for DNA binding and hyperactivity in mammalian cells. Mol Cell Biol 24: 9239-9247.

46. Watkins S, van Pouderoyen G, Sixma TK (2004) Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. Nucleic Acids Res 32: 4306-4312.

47. Lange A, Mills RE, Lange CJ, Stewart M, Devine SE, et al. (2007) Classical nuclear localization signals: definition, function, and interaction with importin alpha. J Biol Chem 282: 5101-5105.

48. Nizami Z, Deryusheva S, Gall JG (2010) The Cajal body and histone locus body. Cold Spring Harb Perspect Biol 2: a000653.

49. Liu JL, Murphy C, Buszczak M, Clatterbuck S, Goodman R, et al. (2006) The Drosophila melanogaster Cajal body. J Cell Biol 172: 875-884.

50. Emmott E, Hiscox JA (2009) Nucleolar targeting: the hub of the matter. EMBO Rep 10: 231-238.

51. Daniel DC, Thompson M, Woodbury NW (2002) DNA-binding interactions and conformational fluctuations of Tc3 transposase DNA binding domain examined with single molecule fluorescence spectroscopy. J Biophys 82: 1654-1666.

52. Papworth M, Kolasinska P, Minczuk M (2006) Designer zinc-finger proteins and their applications. Gene 366: 27-38.

53. Liu Q, Segal DJ, Ghiara JB, Barbas CF 3rd (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. Proc Natl Acad Sci USA 94: 5525-5530.

54. Kamiuchi T, Abe E, Imanishi M, Kaji T, Nagaoka M, et al. (1998) Artificial nine zinc-finger peptide with 30 base pair binding sites. Biochemistry 37: 13827-13834.

55. Imanishi M, Yan W, Morisaki T, Sugiura Y (2005) An artificial six-zinc finger peptide with polyarginine linker selective binding to the discontinuous DNA sequences. Biochem Biophys Res Commun 333: 167-173.

56. Yan W, Imanishi M, Futaki S, Sugiura Y (2007) Alpha-helical linker of an artificial 6-zinc finger peptide contributes to selective DNA binding to a discontinuous recognition sequence. Biochemistry 46: 8517-8524.

57. Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9: 40.

58. Howarth M, Chinnapen DJ, Gerrow K, Dorrestein PC, Grandy MR, et al. (2006) A monovalent streptavidin with a single femtomolar biotin binding site. Nat Methods 3: 267-273.

59. Gonzalez IL, Sylvester JE (1995) Complete sequence of the 43-kb human ribosomal DNA repeat analysis of the intergenic spacer. Genomics 27: 320-328.

60. Ivics Z, Hackett PB, Plasterk RH, Izsvák Z (1997) Molecular reconstruction of Sleeping Beauty a Tc1-like transposon from fish and its transposition in human cells. Cell 91: 501-510.

61. Bloom SE, Bacon LD (1985) Linkage of the major histocompatibility (B) complex and the nucleolar organizer in the chicken. Assignment to a microchromosomes. J Heredity 76: 146-154.

62. Ide S, Miyazaki T, Maki H, Kobayashi T (2010) Abundance of ribosomal RNA gene copies maintains genome integrity. Science 327: 693-696.

63. Ito T, Osawa S, Shibata H, Kanda N (2007) Identification of rRNA Gene Loci in the Wild Mouse (Mus musculus molossinus) Captured at Hachioji, Tokyo. J Vet Med Sci 69: 1277-1279.

64. Su MH, Delany ME (1998) Ribosomal RNA Gene copy number and nucleolar-size polymorphisms within and among chicken lines selected for enhanced growth. Poult Sci 77:1748-1754.

65. Peterson CR, Cryar JR, Gaubatz JW (1984) Constancy of ribosomal RNA genes during aging of mouse heart cells and during serial passage of WI-38 cells. Arch Gerontol Geriatr 3: 115-125.

Submit your next manuscript and get advantages of OMICS Group submissions

Unique features:

- User friendly/feasible website-translation of your paper to 50 world's leading languages
- Audio Version of published paper
- Digital articles to share and explore

Special features:

- 200 Open Access Journals
- 15,000 editorial team
- 21 days rapid review process
- Quality and quick editorial, review and publication processing
- Indexing at PubMed (partial), Scopus, DOAJ, EBSCO, Index Copernicus and Google Scholar etc
- Sharing Option: Social Networking Enabled
- Authors, Reviewers and Editors rewarded with online Scientific Credits
- Better discount for your subsequent articles

Submit your manuscript at: http://www.omicsonline.org/submission/