(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 October 2009 (29.10.2009)

PCT

(10) International Publication Number
WO 2009/130697 A1

(51) **International Patent Classification**:
C12N 15/82 (2006.01) *C07K 14/415* (2006.01)
C12N 15/62 (2006.01)

(21) **International Application Number**:
PCT/IL2009/000434

(22) **International Filing Date**:
21 April 2009 (21.04.2009)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
61/071,298 21 April 2008 (21.04.2008) US

(71) **Applicants** *(for all designated States except US)*: **YEDA RESEARCH AND DEVELOPMENT CO. LTD.** [IL/IL]; at the Weizmann Institute of Science, P.O. Box 95, 76100 Rechovot (IL). **BEN GURION UNIVERSITY OF THE NEGEV RESEARCH AND DEVELOPMENT AUTHORITY** [IL/IL]; P.O. Box 653, 84105 Beer-Sheva (IL).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **GRAFI, Gideon** [IL/IL]; 155 HaGefen Street, 85360 Maboim (IL). **ZEMACH, Assaf** [IL/IL]; 9 Burla Street, 69364 Tel-Aviv (IL).

(74) **Agents**: **G.E EHRLICH (1995) LTD.** et al.; 11 Menachem Begin Street, 52521 Ramat Gan (IL).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2009/130697 A1

(54) **Title**: POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY AND USE THEREOF IN INCREASING BINDING AFFINITY OF DNA BINDING PROTEINS

(57) **Abstract**: Polynucleotides and polypeptides for increasing affinity of a DNA-binding protein to DNA are provided. Such polypeptides may be attached to the DNA binding, thereby increasing the affinity of the DNA-binding protein to DNA.

POLYNUCLEOTIDES AND POLYPEPTIDES ENCODED THEREBY AND USE THEREOF IN INCREASING BINDING AFFINITY OF DNA BINDING PROTEINS

FIELD AND BACKGROUND OF THE INVENTION

The present invention, in some embodiments thereof, relates to polynucleotides and polypeptides encoded thereby and use of same for enhancing binding affinity of nuclear proteins to their genomic target sites.

During cell division and differentiation many nuclear proteins are involved in metabolizing genomic DNA in various vital processes, such as replication, transcription, repair and recombination. In order to achieve accurate genomic activity, nuclear proteins need to specifically bind their DNA binding sites. Another important feature contributing to the accurate activity of nuclear proteins is the stringency level of their chromatin-binding affinity. In general, DNA-binding activity possessing a low Kd (i.e. high binding affinity) is associated with enhanced transcriptional activity [Otsuka et al., J. Biol. Chem. (2000) 275: 34122–34130; Zhang et al., J. Biol. Chem. (2000) 275: 33850–33860]. However, several cases have shown the opposite, hence, a weak protein-DNA binding was associated with a more efficient gene modulation [Lund et al., J. Mol. Biol. (2004) 340: 599–613]. Nonetheless, improper binding of nuclear proteins, such as transcription factors, can cause abnormal gene expression and consequently may lead to the formation of various human diseases such as cancer, dwarfism and congenital severe combined immunodeficiency [Latchman, N. Engl. J. Med. (1996) 334:28–33].

Several methods have been contemplated to enhance DNA-binding affinity of nuclear proteins. According to one method researchers have modified the coding regions of each nuclear protein separately [Visser et al., Adv. Genet. (2006) 56:131-161]. By another method, researchers have fused multiple DNA-binding domains, such as zinc-finger (ZF) binding modules, to construct artificial transcription factors and nucleases to enhance chromatin-binding affinity [Mandell and Barbas, Nuc. Acids Res. (2006) 34: W516-23]. Alternatively, researchers utilized artificial molecules (e.g. synthetic tripyrrole-peptide conjugates, dendrons) that are integrated chemically or physically inside cells [Blanco et al., Chemistry (2005) 11: 4171-4178; Kostiainen et al., Angew Chem Int Ed Engl. (2005) 44: 2556-2559].

U.S. Pat. No. 5,578,444 describes a method for altering the binding characteristics of a DNA-binding protein to a double-stranded DNA sequence by using a

small molecule (e.g. distamycin). This molecule may be used to enhance or inhibit the binding of a DNA-binding protein (e.g. DNA replication factor or transcription factor) to its binding site.

U.S. Pat. No. US6007988 disclosed libraries of DNA sequences encoding zinc finger binding motifs which may be incorporated in chimeric transcription factors, recombinases or nucleases. These chimeric proteins can specifically recognize and bind genomic DNA sequences.

SUMMARY OF THE INVENTION

According to an aspect of some embodiments of the present invention there is provided an isolated polynucleotide comprising a nucleic acid sequence at least 90 % identical to the nucleic acid sequence set forth in SEQ ID NO: 1, the nucleic acid sequence being no more than 250 bases in length.

According to an aspect of some embodiments of the present invention there is provided an isolated polynucleotide comprising a nucleic acid sequence encoding an amino acid sequence at least 90 % homologous to the amino acid sequence set forth in SEQ ID NO: 2, the nucleic acid sequence being no more than 250 bases in length.

According to an aspect of some embodiments of the present invention there is provided an isolated polynucleotide comprising the nucleic acid sequence of the present invention and at least one heterologous nucleic acid sequence encoding at least one polypeptide of interest, the at least one heterologous nucleic acid sequence being fused in-frame to the nucleic acid sequence.

According to an aspect of some embodiments of the present invention there is provided a nucleic acid expression construct comprising the isolated polynucleotide of the present invention operably attached to a cis regulatory element.

According to an aspect of some embodiments of the present invention there is provided an isolated polynucleotide comprising a nucleic acid sequence which specifically hybridize to the isolated polynucleotide of the present invention.

According to an aspect of some embodiments of the present invention there is provided an isolated polypeptide comprising an amino acid sequence at least 90 % homologous to the amino acid sequence set forth in SEQ ID NO: 2 the amino acid sequence being no longer than 80 amino acids in length.

According to an aspect of some embodiments of the present invention there is provided an isolated polypeptide comprising the amino acid sequence of the present invention and at least one heterologous nucleic acid sequence encoding at least one polypeptide of interest.

According to an aspect of some embodiments of the present invention there is provided an isolated antibody comprising an antigen recognition domain capable of specifically binding to the polypeptide of the present invention.

According to an aspect of some embodiments of the present invention there is provided an isolated cell exogenously expressing the polynucleotide of the present invention.

According to an aspect of some embodiments of the present invention there is provided a method of increasing affinity of a DNA-binding protein to DNA, the method comprising attaching the DNA binding protein to the polypeptide of the present invention, thereby increasing the affinity of the DNA-binding protein to DNA.

According to an aspect of some embodiments of the present invention there is provided a method of changing a target sequence in a genomic DNA of a eukaryotic cell, the method comprising introducing at least one chimeric nuclease into the cell, wherein the chimeric nuclease comprises (i) a DNA binding domain; (ii) a cleavage domain; (iii) a nuclear localization domain; and (iv) the amino acid sequence of the present invention, thereby changing the target sequence in the genomic DNA of the eukaryotic cell.

According to an aspect of some embodiments of the present invention there is provided a method of identifying a target sequence of a DNA binding protein, the method comprising: (a) contacting DNA binding protein bound to the isolated polypeptide of the present invention with a nucleic acid sequence which comprises at least one target sequence of the DNA binding protein under conditions which allow complex formation between the DNA binding protein and the at least one target sequence; and (b) identifying the at least one target sequence in the complex; thereby identifying the target sequence of the DNA binding protein.

According to some embodiments of the invention, the nucleic acid sequence is as set forth in 1, 34, 35, 39, 40, 42, 44 or 45.

According to some embodiments of the invention, the at least one polypeptide of interest is selected from the group consisting or a reporter polypeptide, a transcription factor, a nuclease, a DNA binding domain and a ligase.

According to some embodiments of the invention, the nucleic acid sequence encodes an amino acid sequence which increases a binding affinity of a DNA binding protein to a target sequence but does not alter binding specificity.

According to some embodiments of the invention, the amino acid sequence increases a binding affinity of a DNA binding protein to a target sequence but does not alter binding specificity.

According to some embodiments of the invention, said amino acid sequence is at least 41 amino acids length.

According to some embodiments of the invention, said amino acid sequence comprises a negatively charged amino acid.

According to some embodiments of the invention, said negatively charged amino acid is glutamic acid.

According to some embodiments of the invention, the amino acid sequence is as set forth in SEQ ID NO: 2, 46, 47, 51, 52, 54, 56 or 57.

According to some embodiments of the invention, the amino acid sequence increases polypeptide binding affinity to DNA and does not alter binding specificity.

According to some embodiments of the invention, the identifying is effected using a Chip on Chip assay.

Unless otherwise defined, all technical and/or scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of embodiments of the invention, exemplary methods and/or materials are described below. In case of conflict, the patent specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be necessarily limiting.

BRIEF DESCRIPTION OF THE DRAWINGS

Some embodiments of the invention are herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the

drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of embodiments of the invention. In this regard, the description taken with the drawings makes apparent to those skilled in the art how embodiments of the invention may be practiced.

In the drawings:

FIG. 1 is the Sticky-C (C-terminal domain; CTD) amino acid sequence (SEQ ID NO: 2). The nuclear localization signal (NLS) is indicated.

FIGs. 2A-G are images depicting the C-terminal domain (CTD) of AtMBD7 which comprises a strong chromatin binding affinity. Figure 2A is a schematic representation of the full AtMBD7 and its truncated forms fused to GFP. Closed boxes represent MBD and circles represent CTD; Figures 2B-E depict subnuclear localization of transiently expressed AtMBD7 and its truncated constructs in Arabidopsis protoplasts. The full length AtMBD7 (Figure 2B) as well as the MBD motifs 103-306 (Figure 2C) and 173-306 (Figure 2D), are localized to chromocenters, whereas the CTD (Figure 2E) is localized at speckles throughout the nucleus. Bar = 2 μm; Figure 2F depicts FRAP analysis. The graph indicates the average fluorescence recovery kinetics of five independent experiments of each of the indicated GFP fusion proteins; Figure 2G depicts the $t_{1/2}$ time (the time for 50 % recovery in the fluorescent signals derived from the FRAP experiments).

FIGs. 3A-O are images depicting subnuclear distribution of AtMBD7 which is controlled by CTD in CpG-hypomethylated *ddm1* nuclei. The indicated GFP/mRFP fusion proteins were coexpressed in *ddm1-2* protoplasts and were inspected for their colocalization. Figures 3A-C depict speckle-like distribution of AtMBD7 (green), whereas AtMBD5 was evenly dispersed within the nucleus; Figures 3D-F depict a CTD-deficient AtMBD7 truncated protein lacking its CTD which was evenly spread like AtMBD5; Figures 3G-I depict the speckle-like pattern of localization of AtMBD7 C-terminal domain; Figures 3J-L depict AtMBD2 chromocentric localization which has been used to verify low level of methylated-CpG in tested nuclei; Figures 3M-O depict AtMBD7 speckle-like distribution which was mediated by the C-terminal domain (as verified by the co-localization).

FIGs. 4A-F are images depicting CTD immobilized AtMBD5 and SlLHP1 at their chromosomal targets. Figures 4A-B show subnuclear localization of SlLHP1

without and with CTD (respectively) in Arabidopsis protoplasts; Figure 4C depicts FRAP analysis of five independent experiments of the SlLHP1 GFP fusion proteins; Figures 4D-E show subnuclear localization of AtMBD5 without and with CTD (respectively) in Arabidopsis protoplasts; Figure 4F depicts FRAP analysis of five independent experiments of the AtMBD5 GFP fusion proteins. Of note, fusion of CTD does not replaced AtMBD5 or SlLHP1 from their subnuclear targets, e.g. chromocenters (arrows) and nucleolus (asterisk). Bar = 2 μm. Also note the significant reduction in the recovery rate of the CTD fusion proteins.

FIGs. 4G-J are images depicting CTD immobilized p53 in human HeLa cells. Figures 4G-I show subnuclear localization of transiently expressed p53 (Figure 4G), p53-CTD (Figure 4H) and CTD (Figure 4I) as GFP fusion proteins in human HeLa cells. Of note, fusion of CTD does not significantly influence p53 subnuclear localization. Bar = 5 μm; Figure 4J shows FRAP analysis of five independent experiments of the indicated GFP fusion proteins. Note the significant reduction in the recovery rate of the CTD-p53 fusion protein.

FIGs. 5A-F are images depicting the effect of Sticky-C (CTD) on localization of the LacI repressor. Figures 5A-C show U2OS cells transfected with pEYFP-lacI; and Figures 5D-F show U2OS cells transfected with p7CTD-EYFP-lacI. Of note, both showed localization to a single spot. Also note that the Sticky-C reduced the background fluorescence of EYFP-lacI repressor protein. MS2-RFP was used as a counterstain labeling the transcription site and often nucleoli. Bar = 5 μm.

FIGs. 6A-B show the functional activity of truncated StkC proteins revealed by FRAP analysis in Arabidopsis protoplasts. Figure 6A, Schematic representation of the StkC truncated forms. The first and last four amino acids of each truncated sequence are indicated. Figure 6B, Selected images were taken at the indicated times after a chromocenter (indicated by arrow) was photobleached, showing the fluorescence recovery of AtMBD5 with the full length StkC (AtMBD5-StkC-FL) or with StkC truncated forms. The mobility scored is indicated on the right. M indicates the protein is mobile and IM indicates the protein is immobile.

FIGs. 7A-B depict conversion of glutamic acid at position 303 to alanine abolished StkC activity. Figure 7A - Representation of the five C-terminal amino acids and lysine residues (KKVK) of StkC that were converted to alanine. Figure 7B - FRAP

analysis. Selected images were taken at the indicated times after a chromocenter (indicated by arrow) was photobleached, showing the fluorescence recovery of AtMBD5 attached to the authentic StkC48 or its mutated forms (FRAP of 10 nuclei were performed for each construct). The mobility scored after 10 sec is indicated on the right. M indicates the protein is mobile, IM, immobile and SM indicates slightly mobile.

FIGs. 8A-C are schematic illustrations depicting homologous recombination stimulated by zinc finger nucleases (ZFN). Figure 8A shows that the ZFN is composed of FokI nuclease domain (blue circle) and three sequence specific zinc-finger DNA binding domains (green boxes). Furthermore, ZFNs function as dimers. The FokI nuclease is targeted to the specific genomic locus by the zinc-finger domains where it cleaves the DNA. Next, the repair vector integrates into the double-strand break (DSB) by homologous recombination; Figures 8B-C show a model of ZFN intra-nuclear targeting. Figure 8B shows a nucleus comprising regular engineered ZFN which is localized to its genomic target site (red circle) as well as to off target sites. Figure 8C shows a nucleus comprising ZFN fused to Sticky-C (purple circles) which is specifically localized only to its genomic target locus.

DESCRIPTION OF EMBODIMENTS OF THE INVENTION

The present invention, in some embodiments thereof, relates to polynucleotides and polypeptides encoded thereby, which can be used to increase binding affinity of DNA binding proteins to their cognate target sequences.

Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not necessarily limited in its application to the details set forth in the following description or exemplified by the Examples. The invention is capable of other embodiments or of being practiced or carried out in various ways.

While reducing the present invention to practice, the present inventors have uncovered a chromatin binding domain located at the C-terminus of AtMBD7 (GenBank Accession Number NM_125372; Gene ID At5g59800), which increases the affinity of the AtMBD7 to the target chromosomal sites. The present inventors have shown that this sequence is capable of increasing the binding affinity of heterologous DNA-binding proteins without affecting their target specificity suggesting its use in research, industrial (e.g., agronomical, microbiology) and clinical applications.

As is illustrated hereinbelow and in the Examples section which follows, the present inventors have shown that the C terminal domain (hereinafter, CTD or sticky C) of AtMBD7 comprises a strong chromain binding affinity. The CTD directs the AtMBD7 to its native binding site on the chromatin (See Example 1). Due to its affinity enhancement abilities, the CTD may even direct protein binding into hypomethylated sites, again supporting its affinity enhancement capabilities (see Example 2). Sticky C can also increase binding affinity of heterologous nuclear proteins which are fused thereto, altogether supporting it's role as a biological glue.

Thus, according to one aspect of the present invention there is provided an isolated polynucleotide comprising a nucleic acid sequence at least about 70 %, at least about 80 %, at least about 90 %, at least about 95 % or say 100 % identical to the nucleic acid sequence set forth in SEQ ID NO: 1, said nucleic acid sequence being no more than about 500, 400, 300, 250, 230 bases in length.

Identity (e.g., percent homology) can be determined using any homology comparison software, including for example, the BlastN software of the National Center of Biotechnology Information (NCBI) such as by using default parameters.

As used herein the phrase "an isolated polynucleotide" refers to a single or double stranded nucleic acid sequences which is isolated from it's natural environment and provided in the form of an RNA sequence, a complementary polynucleotide sequence (cDNA), a genomic polynucleotide sequence and/or a composite polynucleotide sequences (e.g., a combination of the above).

As used herein the phrase "complementary polynucleotide sequence" refers to a sequence, which results from reverse transcription of messenger RNA using a reverse transcriptase or any other RNA dependent DNA polymerase. Such a sequence can be subsequently amplified *in vivo* or *in vitro* using a DNA dependent DNA polymerase.

As used herein the phrase "genomic polynucleotide sequence" refers to a sequence derived (isolated) from a chromosome and thus it represents a contiguous portion of a chromosome.

As used herein the phrase "composite polynucleotide sequence" refers to a sequence, which is at least partially complementary and at least partially genomic. A composite sequence can include some exonal sequences required to encode the polypeptide of the present invention, as well as some intronic sequences interposing

therebetween. The intronic sequences can be of any source, including of other genes, and typically will include conserved splicing signal sequences. Such intronic sequences may further include cis acting expression regulatory elements.

According to some embodiments of this aspect of the present invention, the isolated polynucleotide comprises a nucleic acid sequence encoding an amino acid sequence at least about 70, at least about 80 %, at least about 90 %, at least about 95 % or say 100 % homologous to the amino acid sequence set forth in SEQ ID NO: 2.

Homology (e.g., percent homology) can be determined using any homology comparison software, including for example, the BlastP or TBLASTN softwares of the National Center of Biotechnology Information (NCBI) such as by using default parameters, when starting from a polypeptide sequence; or the tBLASTX algorithm (available via the NCBI) such as by using default parameters, which compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database.

According to an exemplary embodiment the isolated polynucleotide comprises the nucleic acid sequence set forth in SEQ ID NO: 1.

According to some embodiments of this aspect of the present invention, the isolated polynucleotide comprises a nucleic acid sequence which increases the binding affinity of a DNA binding protein to a target sequence without affecting or altering the binding specificity, hence it's term "Sticky C" or CTD, which is interchangeably used herein with polypeptides of the present invention.

As used herein the phrase "DNA binding protein" refers to a protein such as a nuclear protein, which has a specific or general affinity to DNA. The binding may be direct affinity to the nucleic acid sequence or indirect affinity such as that mediated by epigenetic modifications or histones (chromatin).

As used herein the phrase "target sequence" refers to the DNA binding site recognizable by the DNA binding protein. The target site may be contiguous or non-contiguous. Typicaly target sequence are comprised in promoters, enhancers, repressor and silencing elements, insulators, boundary elements and sequences that control DNA replication

Specificity of binding may be identified using general methods known in the art such as DNA foot printing and ChIP-on-Chip, as will be described below.

As used herein an "increase" in binding affinity refers to at least about 2 fold, about 5 fold, about 10 fold, about 50 fold, about 100 fold, about 200 fold, about 500 fold, about 1000 fold increase in binding affinity.

Methods of assaying binding affinity are well known in the art and exemplified below (see e.g., mobilization assay and FRAP assay). See for example, Zhang et al. (2007) PNAS 104:9:3061-3066 "Quantifying DNA-protein binding specificities by using oligonucleotide mass tags and mass spectroscopy".

According to specific embodiments of the invention, the polynucleotides described herein encode for polypeptides at least 40 amino acids length e.g., 40-100, 40-90, 40-80, 40-70, 40-60 40-50 e.g., 41, 42, 43, 44, 45, 46, 48 and the like

According to specific embodiments of the invention, the amino acid sequence comprises a negatively charged amino acid such as a glutamic acid, aspartic acid or mimetics thereof.

According to specific embodiments of the invention, the nucleic acid sequence is as set forth in SEQ ID NO: 1, 34, 35, 39, 40, 42, 44 or 45.

Thus, the invention encompasses isolated polynucleotides described hereinabove; fragments thereof, sequences hybridizable therewith (under stringent conditions), sequences homologous thereto (e.g., comprising conservative amino acid substitutions and sequences isolated from other species and organisms, based on sequence and functional homology), sequences encoding similar polypeptides with different codon usage, altered sequences characterized by mutations, such as deletion, insertion or substitution of one or more nucleotides, either naturally occurring or man induced, either randomly or in a targeted fashion.

The presently identified Sticky C polynucleotides of the present invention encode previously uncharacterized polypeptides.

Thus, according to another aspect of the present invention there is provided an isolated polypeptide comprising an amino acid sequence at least about 70, at least about 80 %, at least about 90 %, at least about 95 % or say 100 % homologous to the amino acid sequence set forth in SEQ ID NO: 2, the amino acid sequence being no longer than about 100 or 80 amino acids in length (see above provided exemplary length ranges).

According to an exemplary embodiment, the isolated polypeptide comprises an amino acid sequence as set forth in SEQ ID NO: 2, 46, 47, 51, 52, 54, 56 or 57.

As mentioned, the Sticky C polypeptides of the present invention increase binding affinity and immobilize nuclear proteins at their target chromosomal sites. See Example 3 of the Examples section which follows.

Thus, according to yet another aspect of the present invention there is provided a method of increasing affinity of a DNA-binding protein to a DNA target sequence, the method comprising attaching the DNA binding protein to the polypeptide of claim 4-5.1, thereby increasing the affinity of the DNA-binding protein to DNA.

Examples of DNA binding proteins which can be used in accordance with the present invention include, but are not limited to transcription factors, nucleases (E.C. 3.1.4) DNA binding domains, histones and ligases (E.C. 6).

As used herein the phrase "DNA binding domain" refers to a native or synthetic amino acid sequence such as of a protein motif that binds to double- or single-stranded DNA with affinity to a specific sequence or set thereof.

Examples of DNA binding domains include, but are not limited to, helix-turn-helix, leucine zipper (ZIP) domain, winged helix (WH) domain, winged helix turn helix domain (wHTH), helix-loop-helix and zinc finger domain. It should be noted that transcription factors are often classified based on the similarity of their DNA binding domains.

Attachment of the DNA binding protein to the Sticky C polypeptides of the present invention (e.g., SEQ ID NO: 2), is generally effected using recombinant DNA technology.

Thus, according to some embodiments of the present invention there is provided a chimeric polynucleotide comprising at least one of the sticky C nucleic acid sequences described above (e.g., at least two, at least 3, at least 4 nucleic acid sequences as set forth in SEQ ID NO: 1) and at least one heterologous nucleic acid sequence encoding at least one polypeptide of interest e.g., DNA binding protein and/or reporter molecule. In exemplary embodiments the at least one heterologous nucleic acid sequence is fused in-frame to the nucleic acid sequence.

As used herein the phrase “reporter polypeptide” refers to a polypeptide which can be detected in a cell. Preferably, the reporter polypeptide of this aspect of the present invention can be directly detected in the by exerting a detectable signal which can be viewed preferably in living cells (e.g., using a fluorescent microscope). Non-

limiting examples of a nucleic acid sequence encoding a reporter polypeptide according to this aspect of the present invention include fluorescent proteins such as the red fluorescent protein or the green fluorescent protein.

Alternatively, the reporter polypeptide can be indirectly detected such as when the reporter polypeptide is an epitope tag. Indirect detection can be effected by introducing a detectable moiety (labeled antibody) having an affinity to the reporter or when the reporter is an enzyme by introducing a labeled substrate. For example, the reporter polypeptide can be an antigen which is recognized by and binds to a specific antibody. Preferably, when such a reporter polypeptide is utilized, the antibody or the polypeptide capable of binding the reporter protein is labeled (e.g., by covalently attaching to a label such as a fluorescent dye).

To express polynucleotides in cells, the polynucleotide sequence is preferably ligated into a nucleic acid construct suitable for expression. Such a nucleic acid construct includes a cis acting regulatory sequence such as a promoter sequence operably attached to the sticky C polynucleotide of the present invention for directing transcription of the polynucleotide sequence in the cell in a constitutive or inducible manner.

The cell can be a prokaryotic cell (e.g., bacteria) or a eukaryotic cell (e.g., yeast, insect, plant, mammalian)

Constitutive promoters suitable for use with the present invention are promoter sequences which are active under most environmental conditions and most types of cells such as the cytomegalovirus (CMV) and Rous sarcoma virus (RSV). Inducible promoters suitable for use with the present invention include for example the tetracycline-inducible promoter (Zabala M, et al., Cancer Res. 2004, 64(8): 2799-804).

The nucleic acid construct (also referred to herein as an "expression vector") of the present invention may include additional sequences which render this vector suitable for replication and integration in prokaryotes, eukaryotes, or preferably both (e.g., shuttle vectors). In addition, a typical cloning vector may also contain a transcription and translation initiation sequence, transcription and translation terminator and a polyadenylation signal.

Eukaryotic promoters typically contain two types of recognition sequences, the TATA box and upstream promoter elements. The TATA box, located 25-30 base pairs

upstream of the transcription initiation site, is thought to be involved in directing RNA polymerase to begin RNA synthesis. The other upstream promoter elements determine the rate at which transcription is initiated.

Enhancer elements can stimulate transcription up to 1,000 fold from linked homologous or heterologous promoters. Enhancers are active when placed downstream or upstream from the transcription initiation site. Many enhancer elements derived from viruses have a broad host range and are active in a variety of tissues. For example, the SV40 early gene enhancer is suitable for many cell types. Other enhancer/promoter combinations that are suitable for the present invention include those derived from polyoma virus, human or murine cytomegalovirus (CMV), the long term repeat from various retroviruses such as murine leukemia virus, murine or Rous sarcoma virus and HIV. See, Enhancers and Eukaryotic Expression, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. 1983, which is incorporated herein by reference.

In the construction of the expression vector, the promoter is preferably positioned approximately the same distance from the heterologous transcription start site as it is from the transcription start site in its natural setting. As is known in the art, however, some variation in this distance can be accommodated without loss of promoter function.

Polyadenylation sequences can also be added to the expression vector in order to increase the efficiency of mRNA translation. Two distinct sequence elements are required for accurate and efficient polyadenylation: GU or U rich sequences located downstream from the polyadenylation site and a highly conserved sequence of six nucleotides, AAUAAA, located 11-30 nucleotides upstream. Termination and polyadenylation signals that are suitable for the present invention include those derived from SV40.

In addition to the elements already described, the expression vector of the present invention may typically contain other specialized elements intended to increase the level of expression of cloned nucleic acids or to facilitate the identification of cells that carry the recombinant DNA. For example, a number of animal viruses contain DNA sequences that promote the extra chromosomal replication of the viral genome in permissive cell types. Plasmids bearing these viral replicons are replicated episomally

as long as the appropriate factors are provided by genes either carried on the plasmid or with the genome of the host cell.

The vector may or may not include a eukaryotic replicon. If a eukaryotic replicon is present, then the vector is amplifiable in eukaryotic cells using the appropriate selectable marker. If the vector does not comprise a eukaryotic replicon, no episomal amplification is possible. Instead, the recombinant DNA integrates into the genome of the engineered cell, where the promoter directs expression of the desired nucleic acid.

Examples for mammalian expression vectors include, but are not limited to, pcDNA3, pcDNA3.1(+/-), pGL3, pZeoSV2(+/-), pSecTag2, pDisplay, pEF/myc/cyto, pCMV/myc/cyto, pCR3.1, pSinRep5, DH26S, DHBB, pNMT1, pNMT41, pNMT81, which are available from Invitrogen, pCI which is available from Promega, pMbac, pPbac, pBK-RSV and pBK-CMV which are available from Strategene, pTRES which is available from Clontech, and their derivatives.

Expression vectors containing regulatory elements from eukaryotic viruses such as retroviruses can be also used. SV40 vectors include pSVT7 and pMT2. Vectors derived from bovine papilloma virus include pBV-1MTHA, and vectors derived from Epstein Bar virus include pHEBO, and p2O5. Other exemplary vectors include pMSG, pAV009/A$^+$, pMTO10/A$^+$, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the SV-40 early promoter, SV-40 later promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

As described above, viruses are very specialized infectious agents that have evolved, in many cases, to elude host defense mechanisms. Typically, viruses infect and propagate in specific cell types. The targeting specificity of viral vectors utilizes its natural specificity to specifically target predetermined cell types and thereby introduce a recombinant gene into the infected cell. Thus, the type of vector used by the present invention will depend on the cell type transformed. The ability to select suitable vectors according to the cell type transformed is well within the capabilities of the ordinary skilled artisan and as such no general description of selection consideration is provided herein. For example, bone marrow cells can be targeted using the human T cell

leukemia virus type I (HTLV-I) and kidney cells may be targeted using the heterologous promoter present in the baculovirus Autographa californica nucleopolyhedrovirus (AcMNPV) as described in Liang CY et al., 2004 (Arch Virol. 149: 51-60).

Various methods can be used to introduce the expression vector of the present invention into prokaryotic and mammalian cells. Such methods are generally described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Springs Harbor Laboratory, New York (1989, 1992), in Ausubel et al., Current Protocols in Molecular Biology, John Wiley and Sons, Baltimore, Md. (1989), Chang et al., Somatic Gene Therapy, CRC Press, Ann Arbor, Mich. (1995), Vega et al., Gene Targeting, CRC Press, Ann Arbor Mich. (1995), Vectors: A Survey of Molecular Cloning Vectors and Their Uses, Butterworths, Boston Mass. (1988) and Gilboa et at. [Biotechniques 4 (6): 504-512, 1986] and include, for example, stable or transient transfection, lipofection, electroporation and infection with recombinant viral vectors. In addition, see U.S. Pat. Nos. 5,464,764 and 5,487,992 for positive-negative selection methods.

Introduction of nucleic acids by viral infection offers several advantages over other methods such as lipofection and electroporation, since higher transfection efficiency can be obtained due to the infectious nature of viruses.

As mentioned above the nucleic acid construct of the present invention can be utilized to transform plant cells.

The term "'plant" as used herein encompasses whole plants, ancestors and progeny of the plants and plant parts, including seeds, shoots, stems, roots (including tubers), and plant cells, tissues and organs. The plant may be in any form including suspension cultures, embryos, meristematic regions, callus tissue, leaves, gametophytes, sporophytes, pollen, and microspores.

Plant cells may be transformed stably or transiently with the nucleic acid constructs of the present invention. In stable transformation, the nucleic acid molecule of the present invention is integrated into the plant genome and as such it represents a stable and inherited trait. In transient transformation, the nucleic acid molecule is expressed by the cell transformed but it is not integrated into the genome and as such it represents a transient trait.

There are various methods of introducing foreign genes into both monocotyledonous and dicotyledonous plants (Potrykus, I., Annu. Rev. Plant.

Physiol., Plant. Mol. Biol. (1991) 42:205-225; Shimamoto et al., Nature (1989) 338:274-276).

The principle methods of causing stable integration of exogenous DNA into plant genomic DNA include two main approaches:

(i) Agrobacterium-mediated gene transfer: Klee et al. (1987) Annu. Rev. Plant Physiol. 38:467-486; Klee and Rogers in Cell Culture and Somatic Cell Genetics of Plants, Vol. 6, Molecular Biology of Plant Nuclear Genes, eds. Schell, J., and Vasil, L. K., Academic Publishers, San Diego, Calif. (1989) p. 2-25; Gatenby, in Plant Biotechnology, eds. Kung, S. and Arntzen, C. J., Butterworth Publishers, Boston, Mass. (1989) p. 93-112.

(ii) direct DNA uptake: Paszkowski et al., in Cell Culture and Somatic Cell Genetics of Plants, Vol. 6, Molecular Biology of Plant Nuclear Genes eds. Schell, J., and Vasil, L. K., Academic Publishers, San Diego, Calif. (1989) p. 52-68; including methods for direct uptake of DNA into protoplasts, Toriyama, K. et al. (1988) Bio/Technology 6:1072-1074. DNA uptake induced by brief electric shock of plant cells: Zhang et al. Plant Cell Rep. (1988) 7:379-384. Fromm et al. Nature (1986) 319:791-793. DNA injection into plant cells or tissues by particle bombardment, Klein et al. Bio/Technology (1988) 6:559-563; McCabe et al. Bio/Technology (1988) 6:923-926; Sanford, Physiol. Plant. (1990) 79:206-209; by the use of micropipette systems: Neuhaus et al., Theor. Appl. Genet. (1987) 75:30-36; Neuhaus and Spangenberg, Physiol. Plant. (1990) 79:213-217; glass fibers or silicon carbide whisker transformation of cell cultures, embryos or callus tissue, U.S. Pat. No. 5,464,765 or by the direct incubation of DNA with germinating pollen, DeWet et al. in Experimental Manipulation of Ovule Tissue, eds. Chapman, G. P. and Mantell, S. H. and Daniels, W. Longman, London, (1985) p. 197-209; and Ohta, Proc. Natl. Acad. Sci. USA (1986) 83:715-719.

The Agrobacterium system includes the use of plasmid vectors that contain defined DNA segments that integrate into the plant genomic DNA. Methods of inoculation of the plant tissue vary depending upon the plant species and the Agrobacterium delivery system. A widely used approach is the leaf disc procedure which can be performed with any tissue explant that provides a good source for initiation of whole plant differentiation. Horsch et al. in Plant Molecular Biology

Manual A5, Kluwer Academic Publishers, Dordrecht (1988) p. 1-9. A supplementary approach employs the Agrobacterium delivery system in combination with vacuum infiltration. The Agrobacterium system is especially viable in the creation of transgenic dicotyledenous plants.

There are various methods of direct DNA transfer into plant cells. In electroporation, the protoplasts are briefly exposed to a strong electric field. In microinjection, the DNA is mechanically injected directly into the cells using very small micropipettes. In microparticle bombardment, the DNA is adsorbed on microprojectiles such as magnesium sulfate crystals or tungsten particles, and the microprojectiles are physically accelerated into cells or plant tissues.

Following stable transformation plant propagation is exercised. The most common method of plant propagation is by seed. Regeneration by seed propagation, however, has the deficiency that due to heterozygosity there is a lack of uniformity in the crop, since seeds are produced by plants according to the genetic variances governed by Mendelian rules. Basically, each seed is genetically different and each will grow with its own specific traits. Therefore, it is preferred that the transformed plant be produced such that the regenerated plant has the identical traits and characteristics of the parent transgenic plant. Therefore, it is preferred that the transformed plant be regenerated by micropropagation which provides a rapid, consistent reproduction of the transformed plants.

Micropropagation is a process of growing new generation plants from a single piece of tissue that has been excised from a selected parent plant or cultivar. This process permits the mass reproduction of plants having the preferred tissue expressing the fusion protein. The new generation plants which are produced are genetically identical to, and have all of the characteristics of, the original plant. Micropropagation allows mass production of quality plant material in a short period of time and offers a rapid multiplication of selected cultivars in the preservation of the characteristics of the original transgenic or transformed plant. The advantages of cloning plants are the speed of plant multiplication and the quality and uniformity of plants produced.

Micropropagation is a multi-stage procedure that requires alteration of culture medium or growth conditions between stages. Thus, the micropropagation process involves four basic stages: Stage one, initial tissue culturing; stage two, tissue culture

multiplication; stage three, differentiation and plant formation; and stage four, greenhouse culturing and hardening. During stage one, initial tissue culturing, the tissue culture is established and certified contaminant-free. During stage two, the initial tissue culture is multiplied until a sufficient number of tissue samples are produced to meet production goals. During stage three, the tissue samples grown in stage two are divided and grown into individual plantlets. At stage four, the transformed plantlets are transferred to a greenhouse for hardening where the plants' tolerance to light is gradually increased so that it can be grown in the natural environment.

Although stable transformation is presently preferred, transient transformation of leaf cells, meristematic cells or the whole plant is also envisaged by the present invention.

Transient transformation can be effected by any of the direct DNA transfer methods described above or by viral infection using modified plant viruses.

Viruses that have been shown to be useful for the transformation of plant hosts include CaMV, TMV and BV. Transformation of plants using plant viruses is described in U.S. Pat. No. 4,855,237 (BGV), EP-A 67,553 (TMV), Japanese Published Application No. 63-14693 (TMV), EPA 194,809 (BV), EPA 278,667 (BV); and Gluzman, Y. et al., Communications in Molecular Biology: Viral Vectors, Cold Spring Harbor Laboratory, New York, pp. 172-189 (1988). Pseudovirus particles for use in expressing foreign DNA in many hosts, including plants, is described in WO 87/06261.

Construction of plant RNA viruses for the introduction and expression of non-viral exogenous nucleic acid sequences in plants is demonstrated by the above references as well as by Dawson, W. O. et al., Virology (1989) 172:285-292; Takamatsu et al. EMBO J. (1987) 6:307-311; French et al. Science (1986) 231:1294-1297; and Takamatsu et al. FEBS Letters (1990) 269:73-76.

When the virus is a DNA virus, suitable modifications can be made to the virus itself. Alternatively, the virus can first be cloned into a bacterial plasmid for ease of constructing the desired viral vector with the foreign DNA. The virus can then be excised from the plasmid. If the virus is a DNA virus, a bacterial origin of replication can be attached to the viral DNA, which is then replicated by the bacteria. Transcription and translation of this DNA will produce the coat protein which will encapsidate the viral DNA. If the virus is an RNA virus, the virus is generally cloned as

a cDNA and inserted into a plasmid. The plasmid is then used to make all of the constructions. The RNA virus is then produced by transcribing the viral sequence of the plasmid and translation of the viral genes to produce the coat protein(s) which encapsidate the viral RNA.

Construction of plant RNA viruses for the introduction and expression in plants of non-viral exogenous nucleic acid sequences such as those included in the construct of the present invention is demonstrated by the above references as well as in U.S. Pat. No. 5,316,931.

In one embodiment, a plant viral nucleic acid is provided in which the native coat protein coding sequence has been deleted from a viral nucleic acid, a non-native plant viral coat protein coding sequence and a non-native promoter, preferably the subgenomic promoter of the non-native coat protein coding sequence, capable of expression in the plant host, packaging of the recombinant plant viral nucleic acid, and ensuring a systemic infection of the host by the recombinant plant viral nucleic acid, has been inserted. Alternatively, the coat protein gene may be inactivated by insertion of the non-native nucleic acid sequence within it, such that a protein is produced. The recombinant plant viral nucleic acid may contain one or more additional non-native subgenomic promoters. Each non-native subgenomic promoter is capable of transcribing or expressing adjacent genes or nucleic acid sequences in the plant host and incapable of recombination with each other and with native subgenomic promoters. Non-native (foreign) nucleic acid sequences may be inserted adjacent the native plant viral subgenomic promoter or the native and a non-native plant viral subgenomic promoters if more than one nucleic acid sequence is included. The non-native nucleic acid sequences are transcribed or expressed in the host plant under control of the subgenomic promoter to produce the desired products.

In a second embodiment, a recombinant plant viral nucleic acid is provided as in the first embodiment except that the native coat protein coding sequence is placed adjacent one of the non-native coat protein subgenomic promoters instead of a non-native coat protein coding sequence.

In a third embodiment, a recombinant plant viral nucleic acid is provided in which the native coat protein gene is adjacent its subgenomic promoter and one or more non-native subgenomic promoters have been inserted into the viral nucleic acid. The

inserted non-native subgenomic promoters are capable of transcribing or expressing adjacent genes in a plant host and are incapable of recombination with each other and with native subgenomic promoters. Non-native nucleic acid sequences may be inserted adjacent the non-native subgenomic plant viral promoters such that said sequences are transcribed or expressed in the host plant under control of the subgenomic promoters to produce the desired product.

In a fourth embodiment, a recombinant plant viral nucleic acid is provided as in the third embodiment except that the native coat protein coding sequence is replaced by a non-native coat protein coding sequence.

The viral vectors are encapsidated by the coat proteins encoded by the recombinant plant viral nucleic acid to produce a recombinant plant virus. The recombinant plant viral nucleic acid or recombinant plant virus is used to infect appropriate host plants. The recombinant plant viral nucleic acid is capable of replication in the host, systemic spread in the host, and transcription or expression of foreign gene(s) (isolated nucleic acid) in the host to produce the desired protein.

In addition to the above, the nucleic acid molecule of the present invention can also be introduced into a chloroplast genome thereby enabling chloroplast expression.

A technique for introducing exogenous nucleic acid sequences to the genome of the chloroplasts is known. This technique involves the following procedures. First, plant cells are chemically treated so as to reduce the number of chloroplasts per cell to about one. Then, the exogenous nucleic acid is introduced via particle bombardment into the cells with the aim of introducing at least one exogenous nucleic acid molecule into the chloroplasts. The exogenous nucleic acid is selected such that it is integratable into the chloroplast's genome via homologous recombination which is readily effected by enzymes inherent to the chloroplast. To this end, the exogenous nucleic acid includes, in addition to a gene of interest, at least one nucleic acid stretch which is derived from the chloroplast's genome. In addition, the exogenous nucleic acid includes a selectable marker, which serves by sequential selection procedures to ascertain that all or substantially all of the copies of the chloroplast genomes following such selection will include the exogenous nucleic acid. Further details relating to this technique are found in U.S. Pat. Nos. 4,945,050; and 5,693,507 which are incorporated herein by

reference. A polypeptide can thus be produced by the protein expression system of the chloroplast and become integrated into the chloroplast's inner membrane.

Based on the present results showing the ability of the polypeptides of the present invention (e.g., SEQ ID NO: 2) to increase the DNA-binding affinity of many nuclear proteins to their native genomic loci (see Table 1 below), it is suggested that such an enhancement can be harnessed towards the identification and characterization of DNA binding sites of DNA binding proteins.

Thus, according to still another aspect of the present invention there is provided a method of identifying a target sequence of a DNA binding protein. The method comprising: contacting DNA binding protein bound to the polypeptide of the present invention (i.e., sticky C, SEQ ID NO: 2) with a nucleic acid sequence (e.g., genomic sequence) which comprises at least one target sequence of the DNA binding protein under conditions which allow complex formation between the DNA binding protein and the at least one target sequence; and identifying the at least one target sequence in said complex; thereby identifying the target sequence of the DNA binding protein.

Once complexes are formed, the identification of the target sequence can be effected using the traditional DNA finger printing assay.

In certain embodiments the assay may be effected using a Chip-on-Chip configuration. A specific example is provided in Example 8 of the Examples section which follows.

ChIP-on-chip is a technique that combines chromatin immunoprecipitation (ChIP) with microarray technology (chip). Thus, the DNA binding protein of interest fused in a chimeric configuration to sticky C is cross-linked with the DNA site it binds to such as in an in vivo environment. Typically, this done by gentle formaldehyde fixation that is reversible by heat.

Then the cells are lysed and the DNA is sheared by sonnication or using micrococcal nuclease. This results in double-stranded DNA fragments, normally 1 Kb or less in length. Those which are cross-linked to the DNA binding protein form the above mentioned complex.

In the next step, complexes are isolated using an antibody directed at the DNA-binding protein or an anti-sticky c antibody.

Thus, the present invention further provides an antibody which comprises an antigen recognition domain capable of specifically binding sticky C polypeptides such as provided in SEQ ID NO: 2. Detailed description of the term antibody and how to generate the same is provided hereinbelow.

The antibody may be attached to a solid support using methods which are well known in the art. Once the complexes are isolated they are reverse cross-linked and the DNA is purified. After an amplification and denaturation step, the single stranded DNA fragments are labeled with a fluorescent tag suchas Cy5 or Alexa647. Finally, fragments are hybridized over a surface of a DNA microarray which is spotted with short, single stranded sequences that cover the genomic portion of interest. Hybridization will form a double-stranded DNA fragment.

Antibodies used for ChIP-on-chip are selected having ChIP-grade (essentially capable of recognizing the antibody in solution and under fixed conditions). Antibodies to Sticky C are preferably selected having these features using methods which are well known in the art. Methods of generating anti Sticky C antibodies are further described hereinbelow and in Example 8 of the Examples section which follows.

The term "antibody" as used in this invention includes intact molecules as well as functional fragments thereof, such as Fab, F(ab')2, and Fv that are capable of binding to macrophages. These functional antibody fragments are defined as follows: (1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule, can be produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) Fab', the fragment of an antibody molecule that can be obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule; (3) (Fab')2, the fragment of the antibody that can be obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; F(ab')2 is a dimer of two Fab' fragments held together by two disulfide bonds; (4) Fv, defined as a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (5) Single chain antibody ("SCA"), a genetically engineered molecule containing the variable region of the light chain and the variable region of the heavy

chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule.

Methods of producing polyclonal and monoclonal antibodies as well as fragments thereof are well known in the art (See for example, Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York, 1988, incorporated herein by reference).

Antibody fragments according to the present invention can be prepared by proteolytic hydrolysis of the antibody or by expression in E. coli or mammalian cells (e.g. Chinese hamster ovary cell culture or other protein expression systems) of DNA encoding the fragment. Antibody fragments can be obtained by pepsin or papain digestion of whole antibodies by conventional methods. For example, antibody fragments can be produced by enzymatic cleavage of antibodies with pepsin to provide a 5S fragment denoted F(ab')2. This fragment can be further cleaved using a thiol reducing agent, and optionally a blocking group for the sulfhydryl groups resulting from cleavage of disulfide linkages, to produce 3.5S Fab' monovalent fragments. Alternatively, an enzymatic cleavage using pepsin produces two monovalent Fab' fragments and an Fc fragment directly. These methods are described, for example, by Goldenberg, U.S. Pat. Nos. 4,036,945 and 4,331,647, and references contained therein, which patents are hereby incorporated by reference in their entirety. See also Porter, R. R. [Biochem. J. 73: 119-126 (1959)]. Other methods of cleaving antibodies, such as separation of heavy chains to form monovalent light-heavy chain fragments, further cleavage of fragments, or other enzymatic, chemical, or genetic techniques may also be used, so long as the fragments bind to the antigen that is recognized by the intact antibody.

Fv fragments comprise an association of VH and VL chains. This association may be noncovalent, as described in Inbar et al. [Proc. Nat'l Acad. Sci. USA 69:2659-62 (19720]. Alternatively, the variable chains can be linked by an intermolecular disulfide bond or cross-linked by chemicals such as glutaraldehyde. Preferably, the Fv fragments comprise VH and VL chains connected by a peptide linker. These single-chain antigen binding proteins (sFv) are prepared by constructing a structural gene comprising DNA sequences encoding the VH and VL domains connected by an oligonucleotide. The structural gene is inserted into an expression vector, which is subsequently introduced

into a host cell such as E. coli. The recombinant host cells synthesize a single polypeptide chain with a linker peptide bridging the two V domains. Methods for producing sFvs are described, for example, by [Whitlow and Filpula, Methods 2: 97-105 (1991); Bird et al., Science 242:423-426 (1988); Pack et al., Bio/Technology 11:1271-77 (1993); and U.S. Pat. No. 4,946,778, which is hereby incorporated by reference in its entirety.

Another form of an antibody fragment is a peptide coding for a single complementarity-determining region (CDR). CDR peptides ("minimal recognition units") can be obtained by constructing genes encoding the CDR of an antibody of interest. Such genes are prepared, for example, by using the polymerase chain reaction to synthesize the variable region from RNA of antibody-producing cells. See, for example, Larrick and Fry [Methods, 2: 106-10 (1991)].

The present teachings may also be used for targeted mutations of genomic sequences. Gene targeting is a method to repair or inactivate any desired gene of interest. Gene targeting strategies use the introduction of a double-stranded break (DSB) into a genomic locus to enhance the efficiency of recombination with an exogenously introduced homologous DNA "repair template" (Figures 6A-B). DSBs can stimulate recombination efficiency several thousand-fold, approaching gene targeting frequencies as high as 20 %. Early experiments utilized highly specific homing endonucleases, enzymes that bind and cleave extended DNA sequences, to introduce DSBs into specific genomic loci. Unfortunately, since robust techniques to alter the DNA-binding specificities of these enzymes do not currently exist, the use of homing endonucleases to enhance gene targeting is limited to loci into which the target cleavage sequence can be introduced.

Zinc finger nucleases (ZFNs) provide an alternative to homing endonucleases for introducing site-specific DSBs. ZFNs consist of a DNA-binding zinc finger domain (composed of either three or four fingers) covalently linked to the non-specific DNA cleavage domain of the bacterial FokI restriction endonuclease (Figure 8A). ZFNs can bind as dimers to their target DNA sites, with each monomer using its zinc finger domain to recognize a "half-site". Dimerization of ZFNs is mediated by the FokI cleavage domain which cleaves within a five or six base pair "spacer" sequence that separates the two inverted "half sites". Importantly, because the DNA-binding

specificities of zinc finger domains can be re-engineered using various methods, customized ZFNs can theoretically be constructed to target nearly any gene sequence.

Recent work has shown that ZFNs can be used to direct gene targeting events to specific endogenous loci or genes in insect, plant, and human cells. However, this technology should be improved since off-site DSBs may occur, thus creating undesirable non-targeted mutations (Wright et al. 2005 Plant Journal 44:493-705).

Thus, according to an additional aspect of the present invention, there is provided a method of changing a target sequence in a genomic DNA of a eukaryotic cell. The method comprising introducing at least one chimeric nuclease into the cell, wherein said chimeric nuclease comprises:

(i) a DNA binding domain;

(ii) a cleavage domain;

(iii) a nuclear localization domain; and

(iv) an amino acid sequence of sticky C such as set forth in SEQ ID NO: 2,

thereby changing the target sequence in the genomic DNA of the eukaryotic cell.

As used herein the phrase "changing a target sequence in a genomic DNA" refers to a mutagenesis, which may include nucleotide(s) insertion, deletion, replacement, missense mutations, nonsense mutations and the like.

According to this aspect of the present invention, the DNA binding domains confer the DNA binding specificity, while the cleavage domains confer the double-stranded break activity. A variety of DNA binding domains are known in the art (e.g., see above description), and any DNA binding domain that recognizes the desired site with sufficient specificity may be employed. In exemplary embodiments, the DNA binding domains include zinc finger binding domains.

Cleavage domains may be derived from any nuclease that has DNA cleavage activity. Examples of protein types having cleavage domains include restriction enzymes, topoisomerases, recombinases, integrases and DNAses. For example, the cleavage domain may be derived from a type IIs restriction endonuclease, such as the cleavage domain of the FokI restriction enzyme ("Fn"). Enzymes of this group generally have separate cleavage and sequence recognition domains. Thus, in a particular embodiment, the chimeric nucleases are fusion proteins comprising specific zinc finger

binding domains and the cleavage domain of the FokI restriction enzyme (also referred to herein as the FokI cleavage domain).

The Cys.sub.2His.sub.2 zinc fingers are of particular interest in this regard. Each individual finger contacts primarily three consecutive base pairs of DNA in a modular fashion (Pavletich et al., 1991, Science, 252:809-817; Berg et al., 1996, Science, 271:1081-1085). By manipulating the number of fingers and the nature of critical amino acid residues that contact DNA directly, binding domains with novel specificities can be evolved and selected (see, e.g., Desjarlais et al., 1992, Proc. Natl. Acad. Sci. USA, 89:7345-7349; Rebar et al., 1994, Science, 263:671-673; Greisman et al., 1997, Science, 275:657-661; Segal et al., 1999, Proc. Natl. Acad. Sci. USA, 96:2758-2763). In principle, a very broad range of DNA sequences can serve as specific recognition targets for zinc finger proteins. Chimeric nucleases with several different specificities based on zinc finger recognition have already been constructed and characterized (see, e.g., Huang et al., 1996, J. Protein Chem., 15:481-489; Kim et al., 1998, Biol. Chem., 379:489-495).

In an exemplary embodiment, chimeric nucleases of the present invention comprise a nuclear localization signal (NLS) which facilitates the nuclear transport of the chimeric nucleases.

Design of chimeric nucleases which can be used in accordance with the present teachings is described in details in U.S. Patent Application Number 20050026157.

Dependent on the intended use a repair substrate may also be introduced into the target eukaryotic cell. Methods of introducing nucleic acid sequences into target cells are described above.

Polynucleotides, polypeptides, nucleic acid constructs, antibodies and compositions comprising same of some embodiments of the present invention may be if desired, be presented in a pack or dispenser device or kit. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for use.

It is expected that during the life of a patent maturing from this application many relevant sticky C polypeptide will be developed and the scope of the term is intended to include all such new technologies *a priori*.

As used herein the term "about" refers to ± 10 %.

The terms "comprises", "comprising", "includes", "including", "having" and their conjugates mean "including but not limited to". This term encompasses the terms "consisting of" and "consisting essentially of".

The phrase "consisting essentially of" means that the composition or method may include additional ingredients and/or steps, but only if the additional ingredients and/or steps do not materially alter the basic and novel characteristics of the claimed composition or method.

As used herein, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a compound" or "at least one compound" may include a plurality of compounds, including mixtures thereof.

Throughout this application, various embodiments of this invention may be presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 3, 4, 5, and 6. This applies regardless of the breadth of the range.

Whenever a numerical range is indicated herein, it is meant to include any cited numeral (fractional or integral) within the indicated range. The phrases "ranging/ranges between" a first indicate number and a second indicate number and "ranging/ranges from" a first indicate number "to" a second indicate number are used herein interchangeably and are meant to include the first and second indicated numbers and all the fractional and integral numerals therebetween.

As used herein the term "method" refers to manners, means, techniques and procedures for accomplishing a given task including, but not limited to, those manners, means, techniques and procedures either known to, or readily developed from known manners, means, techniques and procedures by practitioners of the chemical, pharmacological, biological, biochemical and medical arts.

It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable subcombination or as suitable in any other described embodiment of the invention. Certain features described in the context of various embodiments are not to be considered essential features of those embodiments, unless the embodiment is inoperative without those elements.

Various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below find experimental support in the following examples.

EXAMPLES

Reference is now made to the following examples, which together with the above descriptions, illustrate some embodiments of the invention in a non limiting fashion.

Generally, the nomenclature used herein and the laboratory procedures utilized in the present invention include molecular, biochemical, microbiological and recombinant DNA techniques. Such techniques are thoroughly explained in the literature. See, for example, "Molecular Cloning: A laboratory Manual" Sambrook et al., (1989); "Current Protocols in Molecular Biology" Volumes I-III Ausubel, R. M., ed. (1994); Ausubel et al., "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson et al., "Recombinant DNA", Scientific American Books, New York; Birren et al. (eds) "Genome Analysis: A Laboratory Manual Series", Vols. 1-4, Cold Spring Harbor Laboratory Press, New York (1998); methodologies as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057; "Cell Biology: A Laboratory Handbook", Volumes I-III Cellis, J. E., ed. (1994); "Current Protocols in Immunology" Volumes I-III Coligan J. E., ed. (1994); Stites et al. (eds), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, CT (1994); Mishell and Shiigi (eds), "Selected Methods in Cellular Immunology", W. H. Freeman and Co., New York (1980); available immunoassays are extensively described in the patent and scientific literature, see, for example, U.S. Pat.

Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 and 5,281,521; "Oligonucleotide Synthesis" Gait, M. J., ed. (1984); "Nucleic Acid Hybridization" Hames, B. D., and Higgins S. J., eds. (1985); "Transcription and Translation" Hames, B. D., and Higgins S. J., Eds. (1984); "Animal Cell Culture" Freshney, R. I., ed. (1986); "Immobilized Cells and Enzymes" IRL Press, (1986); "A Practical Guide to Molecular Cloning" Perbal, B., (1984) and "Methods in Enzymology" Vol. 1-317, Academic Press; "PCR Protocols: A Guide To Methods And Applications", Academic Press, San Diego, CA (1990); Marshak et al., "Strategies for Protein Purification and Characterization - A Laboratory Course Manual" CSHL Press (1996); all of which are incorporated by reference as if fully set forth herein. Other general references are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

EXAMPLE 1

The C-terminal domain of AtMBD7 (Sticky-C) comprises a strong chromatin binding activity

MATERIALS AND EXPERIMENTAL PROCEDURES

Construction of AtMBD-GFP plasmids

Fusion of AtMBD5, AtMBD6 and AtMBD7 to GFP was completed as previously described [Zemach et al., Plant Cell (2005) 17, 1549-1558]. For the generation of AtMBD5 and AtMBD7 derivatives fused to GFP, the various fragments were first amplified by PCR using the following primers: for AtMBD5/1-92, MBD5-sense 5' TGATATCAGATCTATGTCGAACGGCACGGATCAG (SEQ ID NO: 3) and MBD5(1-92aa)-antisense 5' TCTCCCCGGGTTCCGCTGTCTTGACGGAC (SEQ ID NO: 4); for AtMBD7/1-231, MBD7-sense 5'-GAGAAGATCTAGAATGCAGACGAGATCCTCTTCCTCTCC (SEQ ID NO: 5) and MBD7(231aa)-antisense 5' TCTCCCCGGGGTCAAGTGTAATGTTCCCG (SEQ ID NO: 6); for AtMBD7/103-231, MBD7(103aa)-sense 5' GAGGGATCCATGGATCACTGCGGAGTTGAATATG (SEQ ID NO: 7) and

MBD7(231aa)-antisense (SEQ ID NO: 6, see above); for AtMBD7/173-231, MBD7(173aa)-sense 5' GAGAAGATCTATGCGCCGTGGCCATTCCAAA (SEQ ID NO: 8), and MBD7(231aa)-antisense (SEQ ID NO: 6, see above). The various AtMBD PCR products were digested either with BamHI and SmaI or with BglII and SmaI and cloned into BglII-SmaI sites of pUC19-35S-GFP.

AtMBD5x2-GFP was constructed by PCR of AtMBD5-GFP using p35S-MBD5-GFP plasmid as template and the following primers: MBD5-sense (SEQ ID NO: 3, see above) and GFP-antisense 5' TGCCTGCAGTCAGGTCGACTTGTATAGTTC (SEQ ID NO: 9). The AtMBD5-GFP PCR product was digested with EcoRV and PstI and cloned into SmaI and PstI sites of p35S-MBD5.

The integrity of each construct was verified by sequencing. These constructs were used for protoplast transformation experiments.

Protoplast transformation

Arabidopsis thaliana ('Columbia' ecotype) plants were grown under short day conditions at 20 °C. At 4-6 weeks after germination, rosette leaves were collected for the isolation and transformation of protoplasts essentially as previously described [Sheen J. (2002) http://genetics(dot)mgh(dot)harvard(dot)edu/sheenweb]. The GFP signal was detected 24 hours after transfection using a laser confocal microscope (Olympus Fluoview FV500). Images for GFP were obtained using an excitation wavelength of 488 nm and 543 nm and collection through 505–525 mm and 610-700 filters, respectively.

Fluorescence recovery after photobleaching (FRAP)

FRAP experiments were carried out using a laser confocal microscope (Olympus IX70, Fluoview FV500; Hamburg, Germany). GFP fluorescence images were obtained using an excitation wavelength of 488 nm and signals were collected through 505-525 nm filter. In the FRAP experiments, images of 196 x 96 pixels were collected at maximum speed of approximately 180 ms using 1 % power laser for 2 seconds before the bleach and up to 10 seconds afterwards. The bleach pulse was produced by 100 % power laser for 0.1 seconds in a rectangular area (approximately 1 μm^2) of the nucleus. The relative fluorescence intensity was normalized to the nonbleached signal after subtraction of the background signal. Values were averages of at least 5 cells from three independent experiments. Fluorescence recovery curves were performed using Excel

software (Microsoft, Redmond, WA). Statistical significance was determined using T-TEST.

RESULTS

FRAP experiments revealed that besides its three MBD motifs, AtMBD7 comprises a C-terminal domain (CTD) having chromatin binding activity (data not shown). As illustrated in Figure 1 (and in SEQ ID NO: 2), CTD comprises 74 amino acids containing a putative bipartite nuclear localization signal (NLS). In order to examine the chromatin-binding activity of the CTD, subnuclear localization and mobility experiments were performed with truncated forms of AtMBD7 containing the CTD either alone (AtMBD7/232-306), or with one MBD motif (AtMBD7/173-306), or with two MBD motifs (AtMBD7/103-306) (Figure 2A). While in the context of the either the full length AtMBD7 molecule or the truncated forms, the CTD had no effect on subnuclear localization (Figures 2B-D), when fused alone with GFP it showed two types of intranuclear patterns, namely cloudy and speckle (Figure 2E). Thus, CTD directed GFP into many small sites within the nucleus. Fluorescence recovery after photobleaching (FRAP) demonstrated a very strong chromatin binding affinity of CTD, which appeared to be higher than the cumulative effects of the three MBD motifs (Figures 2F-G). Thus, the C-terminal domain (CTD) of AtMBD7 was designated 'sticky-C'. This feature raised the idea that sticky-C may be used as a biological 'glue' that fastens nuclear proteins to their chromosomal sites.

EXAMPLE 2

The C-terminal domain targets AtMBD7 in CpG hypomethylated nuclei

MATERIALS AND EXPERIMENTAL PROCEDURES

Construction of mRFP plasmids

AtMBD5-mRFP was constructed by PCR of mRFP1 using pRSETB-mRFP1 plasmid [Campbell et al., Proc Natl Acad Sci U.S.A. (2002) 99, 7877-7882] as a template and the following primers: mRFP-sense 5' GAGCCCGGGGCCTCCTCCGAGGACGTCATC (SEQ ID NO: 10) and mRFP-antisense 5' TCTCATGCATTTAGGCG CCGGTGGAGTGGCGG (SEQ ID NO: 11).

The mRFP1 PCR product was digested with *Sma*I and *Nsi*I and cloned into *Sma*I and *Pst*I sites of p35S-MBD5.

Construction of HisAtMBD6 plasmid, nuclear extraction and GST pull-down assays

HisAtMBD6 was constructed by PCR using pGEX-MBD6 [Zemach and Grafi, Plant J (2003) 34(5), 565-572] as a template and the following primers: MBD6-sense 5'- GAGGGATCCTCAGATTCTGTGGCCGGCG (SEQ ID NO: 12) and MBD6-antisense 5'- CTCCCCGGGTCAAGCCGACACTTTACTA (SEQ ID NO: 13). The PCR product was digested with BamHI and SmaI and cloned into the same sites of pQE30 (QIAGEN). Expression in E. coli bacteria (BL21) and purification using nickel agarose columns of HisAtMBD6 were performed according to the manufacturers' protocol (QIAGEN).

In order to isolate Arabidopsis ('Columbia') nuclei, 1 g of leaves (from 4-week-old plants) was ground in 10 ml NIB (10 mM Tris-HCl, pH 8.0, 1.14 M sucrose, 5 mM MgCl2, 2 mM DTT). After 30 minutes of shaking, the mixture was filtered using 50 µm nylon filter and centrifuged for 10 minutes at 500 x g. The pellet was washed once with NIB and a small fraction was taken for DAPI staining for testing nuclei concentration and quality. For nuclear extraction, isolated nuclei were resuspended in 0.4 ml NETN buffer (100 mM NaCl, 1 mM EDTA, 20 mM Tris, pH 8, and 0.5 % NP-40) supplemented with protease inhibitor cocktail (Sigma) followed by sonication for 20 seconds x 3 times at 50 Watts (XL-2000, Misonix). After centrifugation at 11,000 x g for 5 minutes, the supernatant was tested for protein concentration by the Bradford reagent. All steps (except DAPI staining and Bradford assay) were completed at 4 °C.

Purified HisAtMBD6 (5 µg) or nuclear extract (100 µg) derived from transgenic Arabidopsis plants expressing AtMBD5-GFP were subjected to GST pull-down assay as previously described [Grafi et al., Proc Natl Acad Sci U.S.A. (1996) 93(17): 8962-8967] using GST alone or GST-AtMBDs previously described [Zemach and Grafi, Plant J (2003) 34(5): 565-572]. Precipitated proteins were resolved by SDS-PAGE and immunoblotted using either anti-6-His monoclonal antibody (Covance) or anti-GFP (B-2) monoclonal antibody (Santa Cruz Biotechnology).

RESULTS

Under normal DNA methylation conditions, AtMBD7 is being strongly targeted by its three MBD motifs to the chromocenters. Previous reports have shown that in DNA hypomethylation mutants, *met1* and *ddm1*, AtMBD7 is no longer localized to chromocenters, but instead is dispersed in a speckle like manner within the nucleus [Zemach et al., (2005), supra], a localization pattern reminiscent that of CTD (see Figure 2E). To elaborate on this observation, colocalization experiments were performed using hypomethylation *ddm1* mutant cells. While AtMBD5-mRFP showed dispersion within the nucleus, AtMBD7-GFP demonstrated a speckle-like pattern (Figures 3A-C), similarly to the pattern displayed by the CTD alone (Figure 2E). To corroborate that this distribution pattern was contributed by the CTD and not by the MBD motifs, the distribution pattern of AtMBD7/103-231-GFP, containing two MBD motifs, but lacking the CTD, was examined in *ddm1* nuclei. In these nuclei AtMBD7/103-231 was mainly dispersed throughout the nucleus similarly to the distribution pattern of AtMBD5-mRFP [Figures 3D-F], suggesting that the MBD motifs did not contribute to the speckle-like distribution pattern of AtMBD7 in *ddm1* nuclei. In these nuclei, the CTD alone fused to GFP showed a speckle-like pattern as the full length AtMBD7 protein [Figures 3G-I]. Indeed, co-transfection of AtMBD7-mRFP with AtMBD/232-306-GFP in *ddm1* mutant cells showed that both proteins co-localized showing speckle-like distribution within the nucleus (Figure 3M-O). In conclusion, these results verified that the MBDs motifs were indeed guiding AtMBD7 within the methylated nucleus, while in hypo-CpG-methylated environment, the CTD was the major element directing the protein into its sites within the nucleus.

EXAMPLE 3

Sticky-C enhances chromatin binding affinity of nuclear proteins without effecting their nuclear target sites

MATERIALS AND EXPERIMENTAL PROCEDURES

Construction of AtMBD5-StickyC, SlLHP1-StickyC and p53-StickyC

AtMBD5/1-92-StickyC and SlLHP1/1-306-StickyC were constructed by fusion of the AtMBD7/232-306 PCR product (described hereinabove) digested with *Eco*RV-

*Pst*I and cloned into p35S-MBD5/1-92-GFP and p35S-SlLHP1/1-306-GFP plasmids, respectively [as described in Zemach et al., Plant Cell (2006) 18:133-45].

StickyC-GFP and StickyC-p53-GFP were constructed by PCR of AtMBD/232-306aa using p35S-MBD7-GFP as a template and the following primers StickyC-sense 5' GAAGATCTCGAGGCCACCATGGAGTCGGTTTCTATGGTGCATTC (SEQ ID NO: 14) and StickyC-antisense 5' TCGTCGACTCGGAT CCAGAGCGGTCTTCGATCAGTG (SEQ ID NO: 15). The StickyC PCR product was digested with either BglII and SalI or with XhoI and BamHI and cloned into BglII-SalI sites of pp53-GFP or XhoI-BamHI sites of pEGFP-N1 (ClonTech).

Cell culture, transfection and live imaging

Human U2OS osteosarcoma cells containing a single integration site of 256 copies of *lac* operator (lacO) sequence were maintained in low glucose Dulbecco's modified Eagle's medium (DMEM, Biological Industrial Israel) containing 10 % fetal bovine serum (FBS, HyClone). For transient transfection, U2OS cells were transfected with 1-5 μg of plasmid DNA using $CaCl_2$ methodology. The YFP signal was detected 24 hours after transfection using a laser confocal microscope (Olympus IX70, Fluoview FV500; Hamburg, Germany). Images for YFP were obtained using an excitation wavelength of 488 nm and collection through 505–525 mm filters.

Human HeLa cells were maintained in low glucose Dulbecco's modified Eagle's medium (DMEM, Biological Industrial Israel) containing 10 % fetal bovine serum (FBS, HyClone). For transient transfection, HeLa cells were transfected with 1-5 μg plasmid DNA using $CaCl_2$ methodology.

Plasmids

EYFP-lac repressor fusion construct (pYFP-lacI) was obtained.

Cloning of StickyC-YFP-lacI was done by two PCR reactions using the following primers sets:

Sticky-C-sense 5' GAGAGCTAGCGCCACCATGGAGTCGGTTT (SEQ ID NO: 16), Sticky-C-YFP-antisense 5' GCCCTTGCTCACAGAGCGGTCTTC (SEQ ID NO: 17), Sticky-C-YFP-sense 5' ATCGAAGACCGCTCTGTGAGCAAGGGC (SEQ ID NO: 18), YFP-antisense 5' GACTTGTACAGCTCGTCCATG (SEQ ID NO: 19). The following plasmids were used as templates for PCR pStickyC-EGFP and pYFP-lacI. Both PCR products were then used as templates for PCR using Sticky-C-sense and

YFP-antisense primers. This latter PCR product was digested with NheI and BsrGI and ligated into pYFP-lacI digested at the same restriction sites.

RESULTS

To confirm sticky-C's activity as a glue protein, CTD (SEQ ID NO: 1) was fused to nuclear proteins derived from plants and mammals (see table 1, hereinbelow), including the plant nuclear proteins AtMBD5 (GeneBank Accession No. NP_190242.1) and tomato like-heterochromatin protein 1 (SlLHP1, GeneBank Accession No. AF428244), both highly mobile proteins. The effect of CTD on the nuclear proteins' subnuclear localization and mobility was tested. As illustrated in Figures 4A-F, fusion of CTD to either AtMBD5 or SlLHP1 did not alter their native subnuclear localization, but significantly reduced their mobility. Hence, CTD significantly enhances the chromatin binding affinity of nuclear proteins and at different subnuclear compartments, e.g. chromocenters and nucleoli.

The capacity of CTD of AtMBD7 (Sticky-C) to confer immobility in mammalian cells when fused to mammalian proteins was also examined (see table 1, hereinbelow). CTD (SEQ ID NO: 1) was fused to the tumor suppressor protein p53 (GeneBank Accession No. NP_000537) and expressed in human HeLa cells. As illustrated in Figures 4G-J, p53-CTD displayed a nuclear distribution pattern similar to that of the wild-type p53, yet its mobility was significantly reduced compared to wild-type p53, thus confirming the general activity of sticky-C as a 'glue' protein in various eukaryotic systems.

To make sure that Sticky-C did not effect the subnuclear localization of nuclear proteins to which it was attached, a lac operator/lac repressor system was used as was previously described [Janicki et al., Cell (2004) 116, 683-698]. This system comprised 256 copies of the lac operator sequence and was stably integrated into the genome of U2OS cells. The lac repressor binding protein was fused with the yellow fluorescent protein (YFP-lac repressor) and when it was expressed in the abovementioned U2OS cells the integration site fluoresces was visualized as a single spot. Consequently, this system was used to examine the effect of CTD on subnuclear localization of the EYFP-lacI repressor protein. U2OS cells were transfected with either pEYFP-lacI or with pCTD-EYFP-lacI and the cells were inspected after 24 hours using a laser confocal

microscope. As shown in Figures 5A-F, CTD did not effect the subnuclear localization of the lacI repressor protein displaying localization to a single spot. Furthermore, the background fluorescence of lacI within the nucleus was reduced when fused with CTD, thus, further demonstrating the strong binding affinity conferred by this peptide.

Table 1: nuclear proteins fused to sticky-C

Gene	Organism	Subnuclear localization	Tested tissue
AtMBD7	Arabidopsis	10 centromeres	Arabidopsis leaf cells
AtMBD5	Arabidopsis	2 perinucleolar centromeres	Arabidopsis leaf cells
SlLHP1	Tomato	Centromeres, speckles, nucleolus	Arabidopsis leaf cells
p53	Human	Spread	Human HeLa cells
LacI	E. Coli	A single spot	Human U2OS cells

Taken together, these results illustrate the major value of Sticky-C as an enhancer of chromatin binding affinity of a wide variety of nuclear proteins, at any nuclear environment, without affecting their native chromatin targets. Thus, Sticky-C may be used broadly in the regulation of specific endogenous genes.

EXAMPLE 4

Structural-functional analysis of Sticky-C

MATERIALS AND EXPERIMENTAL PROCEDURES

Construction of the various truncated StkC was performed using PCR. Initially, three StkC N-terminal deletion derivatives, namely, StkC58, StkC41 and StkC20 were fused downstream from AtMBD5(1-92) in frame with GFP. To generate these constructs, pUC19-35S-AtMBD7(232-306)-GFP plasmid was used as a template to amplify the three StkC derivatives fused with GFP using the following forward primers: StkC58-F 5'-TGA ACG ATA TCG GTA TCA GA TTT CAG AGC GAA G (SEQ ID NO: 20); StkC41-F 5'-CCG AAG GAT ATC AAG TGG GTT CTT ACC GGT TC (SEQ ID NO: 21); StkC20-F 5'-GAT CGG ATA TCT CTA GCT TGG TTA AAC

ATA CAT GG (SEQ ID NO: 22), and GFP-R 5'- TGC CTG CAG TCA GGT CGA CTT GTA TAG TTC (SEQ ID NO: 23) was used as reverse primer. The PCR products were digested with *Eco*RV and *Pst*I and cloned into *Sma*I and *Pst*I sites of pUC19-35S-MBD5(1-92). Additional constructs, namely, StkC38, StkC43, and StkC48 as well as StkC-I301A, StkC-E302A, StkC-D303A, StkC-R304A, StkC-S305A and StkC-E302AD303A were generated by PCR using pUC19-35S-AtMBD7(232-306)-GFP plasmid as a template, StkC48-F 5'-ATA AGA TCT GAT ATC CCA AAC CCA CCG AAG AAA G (SEQ ID NO: 24) flanked with *Bgl*II and *Eco*RV was used as a forward primer with each of the following reverse primers flanked with *Sma*I site: StkC-38-R 5' ATA CCC GGG CTC GGA CCA TGT ATG TTT (SEQ ID NO: 25); StkC-43-R 5'-ATC CCG GGC AGT GAA ACA AAT GC (SEQ ID NO: 26); StkC-(I301A)-R 5'-ATC CCG GGA GAG CGG TCT TCG GCC AGT GAA AC (SEQ ID NO: 27); StkC-(E302A)-R 5'-ATC CCG GGA GAG CGG TCC GCG ATC AGT GAA AC (SEQ ID NO: 28); StkC-(D303A)-R 5'-ATC CCG GGA GAG CGC GCT TCG ATC AGT GAA AC (SEQ ID NO: 29); StkC-(R304A)-R 5'-ATC CCG GGA GAC GCG TCT TCG ATC AGT GAA AC (SEQ ID NO: 30); StkC-(S305A)-R 5'-ATC CCG GGG GCG CGG TCT TCG ATC AG (SEQ ID NO: 31) and the PCR products were digested with *Bgl*II and *Sma*I and cloned into the same sites of pUC19-35S-GFP. The corresponding pUC19-35S-StkC-GFP plasmids were then digested with *Eco*RV and *Pst*I and the StkC-GFP fragments were subcloned into *Sma*I and *Pst*I sites of pUC19-35S-MBD5 (1-92) to generate the desired constructs. The construct pUC19-35S-MBD5(1-92)-StkC(AAVA)-GFP in which lysine residues within the StkC motif were converted to alanine (KKVK263-266AAVA) was generated by PCR using pUC19-35S-AtMBD7 (232-306aa)-GFP plasmid as a template and StkC-AAVA-F 5' AGA TAT CGC AGC AGT AGC ATG GGT TCT TAC CGG T' (SEQ ID NO: 32) and GFP-R primer. The PCR product was digested with *Eco*RV and *Pst*I and cloned into the *Sma*I and *Pst*I sites of pUC-35S-MBD(1-92). All constructs were sequenced to ensure in frame fusion of all sequences.

FRAP analysis - effected as described above.

RESULTS

To pinpoint the amino acid residues essential for StkC (also referred to herein as Sticky-C) activity, truncated forms of StkC fused at the C-terminus of AtMBD5

upstream from the GFP sequence and controlled by the 35S promoter (pUC-35SP-AtMBD5-StkC-GFP). The various constructs (Figure 6*A*) were transformed into Arabidopsis protoplasts and analyzed by FRAP for their capability to confer immobility at their chromosomal sites. Results showed (Figure 6*B*) that a peptide containing the 41 amino acids of the C-terminus (StkC-41) retained the StkC activity and was sufficient for reducing the mobility of AtMBD5, while a peptide containing the 20 amino acids of the C-terminus (StkC-20) did not. Interestingly, deletion of the proximal 10 amino acids of the C-terminus (StkC-38) abolished the StkC activity, suggesting that these amino acids are required though not sufficient for nuclear protein immobilization. Thereafter, the StkC was modified, whereby the 5 amino acids of the C-terminus (IEDRS) were deleted (StkC43) and analyzed its capacity to confer intranuclear immobility was analyzed. Results showed that deletion of IEDRS (SEQ ID NO: 33) abolished StkC activity rendering MBD5 protein mobile within the nucleus.

To pinpoint the amino acid residue(s) required for the StkC motif immobility function, each of the 5 amino acids at the C terminus was converted to alanine (Figure 7A), and the ability of the mutated StkC to confer nuclear immobility was analyzed. As shown in Figure 7B, conversion of serine and arginine into alanine (R305A and S306A) did not affect the StkC immobility function; slight increase in mobility observed for I-302A and for D304A constructs. However, the highest effect was observed for E303A pointing to the importance of the glutamic acid in StkC immobility function.

Because glutamic acid has a hydrophilic acidic group with strong negative charge, it has the potential of binding positively charged amino acid residues such as lysine residues (262-KKVK-266) located at the StkC motif (NLS) to promote formation of a functional StkC three-dimensional structure. Therefore KKVK was converted to AAVA (see Figure 7A) and the effect of this StkC mutant on AtMBD5 nuclear mobility was analyzed. The results showed (Figure 7B) that conversion of lysine residues to alanine (AAVA) had no effect on StkC immobility function. Table 2 below summarizes the results described above.

Table 2

StkC Derivative	SEQ ID NO (nucleic acid/amino acid):	Activity (immobility)
StkC (FL) 75 aa	1 / 2	+
StkC58	34 / 46	+
StkC41	35 / 47	+
StkC20	36 / 48	-
StkC38	37 / 49	-
StkC43	38 / 50	-
StkC48	39 / 51	+
StkC48 I302A	40 / 52	+-
StkC48 E303A	41 / 53	-
StkC48 D304A	42 / 54	+-
StkC48 R305A	43 / 55	-
StkC48 S306A	44 / 56	+
StkC48 AAVA	45 / 57	+

EXAMPLE 5

Unraveling genomic binding sites of p53 using Sticky-C

The ability to derive a whole genome map of genomic binding sites of DNA binding proteins [e.g. transcription-factor binding sites (TFBS)] is vital for identifying functions of various transcriptional regulators and identifying their target genes during cellular processes, organ development and disease progression. A common and powerful approach for isolating and identifying DNA sequences occupied by specific DNA binding proteins in cells is chromatin immunoprecipitation coupled with microarray analysis (ChIP-on-chip), also known as genome-wide location analysis [Radonjic et al., Mol. Cell. (2005) 18: 171-83]. The identified binding sites may also be used as a basis for annotating functional elements in genomes. The types of functional elements that may be identified using ChIP-on-chip include promoters, enhancers, repressor and silencing elements, insulators, boundary elements and sequences that control DNA replication.

As an important transcription factor, p53 regulates the expression of genes involved in a variety of cellular functions, including cell-cycle arrest, DNA repair, and apoptosis [Vogelstein et al., Nature (2000) 408, 307–310]. Several efforts have been made to identify p53-targeted genes through various techniques that were found to be inconsistent in the number of binding sites as well as their genomic location [Cawley et

al., Cell (2004) 116: 499–509; Mirza et al., Oncogene (2003) 22: 3645–3654; Wei et al., Cell (2006) 124: 207–219]. By fusing Sticky-C to p53 in ChIP-on-ChIP experiments we are able to enhance p53's DNA binding affinity and to identify p53 target genes. Furthermore, the use of Sticky-C enables to distinguish between real binding sites and noise in ChIP-on-chip analysis without the need for additional molecular analyses.

MATERIALS AND EXPERIMENTAL PROCEDURES

ChIP-on-chip analysis

ChIP-on-chip analysis is performed as previously described [www(dot)chiponchip(dot)org], specifically:

1) Sticky-C (SEQ ID NO: 1) is fused to p53 (GeneBank Accession No. NP_000537) as explained in detail in example 7 hereinabove. Human cell lines (e.g. HeLa, H1299 and PC3 cells) are transformed with the p53 or p53-Sticky-C constructs.

2) The p53 or p53-Sticky-C is cross-linked with DNA sites within the genomic DNA by a gentle formaldehyde fixation that is reversible with heat. Specifically, 10^9 cells suspended in growth medium are placed in tubes on ice for 10 minutes, cross-linking solution (11 % formaldehyde, 0.1 M NaCl, 1 mM Na-EDTA, 0.5 mM Na-EGTA, 50 mM Hepes, pH 8.0) is added directly to each tube for 10 minutes (on ice) and a 2.5 M glycine solution is added to each tube to stop the cross-linking reaction. The cells are harvested by centrifugation at 2000 x g for 10 minutes at 4 °C. The cell pellets are re-suspended in cold PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 2 mM KH2PO4) and washed twice.

3) The transformed cells are lysed by chilled Lysis Buffer [0.05 M Hepes-KOH, pH 7.5, 0.14 M NaCl, 1mM EDTA, 10 % glycerol, 0.5 % NP-40, 0.2 5% Triton X-100, with protease inhibitor cocktail (Roche Applied Science)] by pipetting and mixing for 10 minutes at 4 °C on a rocking platform. After centrifugation at 2000 x g for 10 minutes at 4 °C, the cell pellet is re-suspended in 24 ml Lysis Buffer [0.2 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris pH 8, protease inhibitor cocktail] by pipetting and mixed gently at room temperature for 10 min on a rocking platform. After centrifugation at 2000 x g for 10 minutes at 4 °C, the pellet is resuspended in 5 ml of Lysis Buffer [1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH 8, protease inhibitor cocktail].

The mixture is divided into aliquots in conical tubes and placed on ice. A sonicator (Branson Sonifier 450) is used to disrupt the cell and nuclear membranes and fragment the chromatin. The tube is immediately placed on ice for at least 1 minute to avoid over-heating the sample. Sonication is repeated until the chromatin fragments are of the desired length (normally 1 kb or less in length). Finally, the chromatin solution is adjusted to 0.5 % Sarkosyl (sodium lauryl sarcosine) and gently mixed for 10 minutes at room temperature on a rocking platform. The chromatin solution is then transferred to a centrifuge tube and spun for 10 min at 10,000 x g to remove cell debris. The supernatant (comprising p53-DNA or p53-Sticky-C-DNA complexes) is collected for chromatin immunoprecipitation.

4) The p53-DNA or p53-Sticky-C-DNA complexes are filtered out of the set of DNA fragments by immunoprecipitation (IP), using antibodies specific for Sticky-C (see detailed description hereinbelow). The antibodies are attached to magnetic beads (Dynal).

To prepare the magnetic beads, 100 µl of sheep anti-rabbit IgG-conjugated Dynabeads (Dynal) are first washed three times with cold PBS containing 5 mg/ml Bovine Serum Albumin (BSA) and then resuspended in 5 ml of cold PBS. The Sticky-C antibodies are added to the mixture and incubated overnight on a rotating platform at 4 °C. After collecting the magnetic beads by centrifugation and washing three times with cold PBS containing 5 mg/ml BSA, the beads are re-suspended in 100 µl of cold PBS with 5 mg/ml BSA and are ready for immunoprecipitation.

For IP, the soluble chromatin are first added to an IP mixture (1 % Triton X-100, 0.1 % sodium deoxycholate, protease inhibitor cocktail and 1 X TE) and are then mixed with 100 µl of magnetic beads pre-bound to the specific antibody. The mixture is incubated at 4 °C overnight on a rotating platform. The magnetic beads are then collected using a magnet (Dynal) and the supernatant is removed by aspiration. To remove material non-specifically bound to the beads, RIPA buffer is added to the tube, and the beads are gently re-suspended by removing magnet and inverting by hand. The magnetic beads are again collected with the magnet and washed with RIPA buffer a total of 8 times. After washing once with TE, the beads are precipitated with magnet and the suspension is removed. The beads are then collected by centrifugation at 2000 x g for 3 minutes and re-suspended in elution buffer (10 mM Tris pH 8, 1 mM EDTA, 1

% SDS). To elute precipitated chromatin from the beads, the tubes are incubated at 65 °C for 10 minutes with constant agitation then centrifuged for 30 seconds at maximum speed in microcentrifuge (14000 rpm). Supernatants are removed and mixed with TE with 1 % SDS.

5) The p53-Sticky-C-DNA complexes are reverse cross-linked by incubation at 65 °C overnight and the DNA fragments are purified. Proteins in the DNA sample are removed by incubation with 150 µl of Proteinase K solution (2 % glycogen, 5 % Proteinase K stock solution, and TE) for 2 hours at 37 °C. The sample is then extracted twice with phenol and once with 24:1 chloroform/isoamyl alcohol. The sample is adjusted to 200 mM NaCl. After ethanol precipitation, the DNA is dissolved in 30 µl of TE containing 10 µg of DNase-free RNase A and incubated for 2 hours at 37 °C. The DNA at this step can be further purified with a Qiaquick PCR clean-up kit (Qiagen).

DNA fragments may be further amplified (by a ligation-mediated PCR procedure) and is then labeled with a fluorescent tag such as Cy3, Cy5 or Alexa 647.

6) The fragments are poured over the surface of the genomic micro-array chip which is spotted with short, single-stranded sequences of DNA comprising promoter sequence which p53 is capable of binding to (e.g. GeneChip® Human Promoter 1.0R Array, Affymetrix). Hybridization is carried out in a hybridization chamber (Corning) at 60 °C overnight in a water bath. The hybridization step is carried out in duplicated (i.e. two chips for each protein). After hybridization, the micro-array chip is washed once with wash buffer (2 x SSC, 0.1 % SDS) followed by a wash with a second buffer (0.2 X SSC, 0.1 %SDS) for 10 minutes at room temperature and three times with a third buffer (0.2 X SSC), 1 minute each, at room temperature. The chip is then dried by a brief spin at 1000 x g in a table-top centrifuge. Whenever a labeled fragment "finds" a complementary fragment on the array, they will hybridize and form again a double-stranded DNA fragment.

7) Following hybridization, the array is illuminated with fluorescence light. Those probes on the array that are hybridized to one of the labeled fragments will emit a light signal which can be captured by a camera (e.g. GenePix 4000B scanner from Axon Instruments). The data is normalized and analyzed.

8) For positive control, known targets of wild type p53 (e.g. p21 and MDM2 promoters) are used. In order to further differentiate between a true p53 binding site

and non-specific binding, the ChIP-on-chip experiments are performed using a fusion protein of Sticky-C and mutant-p53 (e.g. 273H, 248P or 245S), which is mutated at its DNA-binding domain.

9) Each novel DNA binding site found using the ChIP-on-chip analysis is then validated by PCR on precipitated DNA from the ChIP-on-chip analyses of all three proteins p53, p53-Sticky-C and Sticky-C. Additionally, the direct binding between p53 and the novel DNA recognition sequences is tested by electro-mobility shift assay (EMSA).

Generation of Sticky-C monoclonal antibodies

Sticky-C is fused to GST (Glutathione S-transferase) and expressed in E.coli bacteria. Following purification it is injected as an antigen into mice. Spleen cells of positive mice (i.e. recognized Sticky-C in the bleeding tests) are then fused to myeloma cells. Single hybridoma cells are then screened for their ability to recognize Sticky-C epitopes.

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims.

All publications, patents and patent applications mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the present invention. To the extent that section headings are used, they should not be construed as necessarily limiting.

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleic acid sequence at least 90 % identical to the nucleic acid sequence set forth in SEQ ID NO: 1, said nucleic acid sequence being no more than 250 bases in length.

2. An isolated polynucleotide comprising a nucleic acid sequence encoding an amino acid sequence at least 90 % homologous to the amino acid sequence set forth in SEQ ID NO: 2, said nucleic acid sequence being no more than 250 bases in length.

3. The isolated polynucleotide of claim 2, wherein said amino acid sequence is at least 41 amino acids length.

4. The isolated polynucleotide of claim 2, wherein said amino acid sequence comprises a negatively charged amino acid.

5. The isolated polynucleotide of claim 4, wherein said negatively charged amino acid is glutamic acid.

6. The isolated polynucleotide of claim 1 or 2, wherein said nucleic acid sequence is as set forth in SEQ ID NO: 1, 34, 35, 39, 40, 42, 44 or 45.

7. An isolated polynucleotide comprising the nucleic acid sequence of claim 1-6 and at least one heterologous nucleic acid sequence encoding at least one polypeptide of interest, said at least one heterologous nucleic acid sequence being fused in-frame to said nucleic acid sequence.

8. The isolated polynucleotide of claim 7, wherein said at least one polypeptide of interest is selected from the group consisting or a reporter polypeptide, a transcription factor, a nuclease, a DNA binding domain and a ligase.

9. The isolated polynucleotide of claim 1 or 7, wherein said nucleic acid sequence encodes an amino acid sequence which increases a binding affinity of a DNA binding protein to a target sequence but does not alter binding specificity.

10. The isolated polynucleotide of claim 2, wherein said amino acid sequence increases a binding affinity of a DNA binding protein to a target sequence but does not alter binding specificity.

11. A nucleic acid expression construct comprising the isolated polynucleotide of claim 1-10 operably attached to a cis regulatory element.

12. An isolated polynucleotide comprising a nucleic acid sequence which specifically hybridize to the isolated polynucleotide of claim 1-10.

13. An isolated polypeptide comprising an amino acid sequence at least 90 % homologous to the amino acid sequence set forth in SEQ ID NO: 2 said amino acid sequence being no longer than 80 amino acids in length.

14. The isolated polypeptide of claim 13, wherein said amino acid sequence is at least 41 amino acids length.

15. The isolated polypeptide of claim 13, wherein said amino acid sequence comprises a negatively charged amino acid.

16. The isolated polypeptide of claim 13, wherein said negatively charged amino acid is glutamic acid.

17. The isolated polypeptide of claim 13, wherein the amino acid sequence is as set forth in SEQ ID NO: 2, 46, 47, 51, 52, 54, 56 or 57.

18. The isolated polypeptide of claim 13, wherein said amino acid sequence increases polypeptide binding affinity to DNA and does not alter binding specificity.

19. An isolated polypeptide comprising the amino acid sequence of claim 13-18 and at least one heterologous nucleic acid sequence encoding at least one polypeptide of interest.

20. An isolated antibody comprising an antigen recognition domain capable of specifically binding to the polypeptide of claim 13, 17 or 18.

21. An isolated cell exogenously expressing the polynucleotide of claim 1, 2 or 10.

22. A method of increasing affinity of a DNA-binding protein to DNA, the method comprising attaching the DNA binding protein to the polypeptide of claim 13-18, thereby increasing the affinity of the DNA-binding protein to DNA.

23. A method of changing a target sequence in a genomic DNA of a eukaryotic cell, the method comprising introducing at least one chimeric nuclease into the cell, wherein said chimeric nuclease comprises:

(i) a DNA binding domain;

(ii) a cleavage domain;

(iii) a nuclear localization domain; and

(iv) the amino acid sequence of claim 13-18,

thereby changing the target sequence in the genomic DNA of the eukaryotic cell.

24. A method of identifying a target sequence of a DNA binding protein, the method comprising:

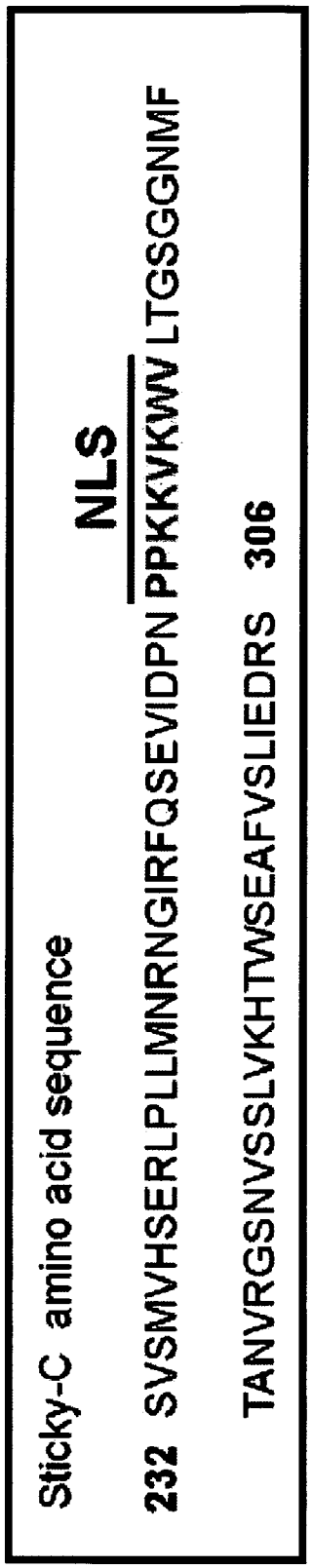
(a) contacting DNA binding protein bound to the isolated polypeptide of claim 13-18 with a nucleic acid sequence which comprises at least one target sequence of said DNA binding protein under conditions which allow complex formation between said DNA binding protein and said at least one target sequence; and

(b) identifying said at least one target sequence in said complex;

thereby identifying the target sequence of the DNA binding protein.

25. The method of claim 24, wherein said identifying is effected using a Chip on Chip assay.

Figure 1



SEQ ID NO: 2

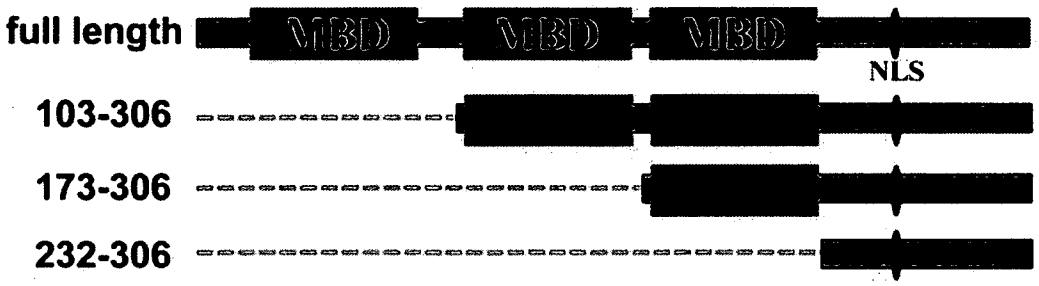


FIG. 2A

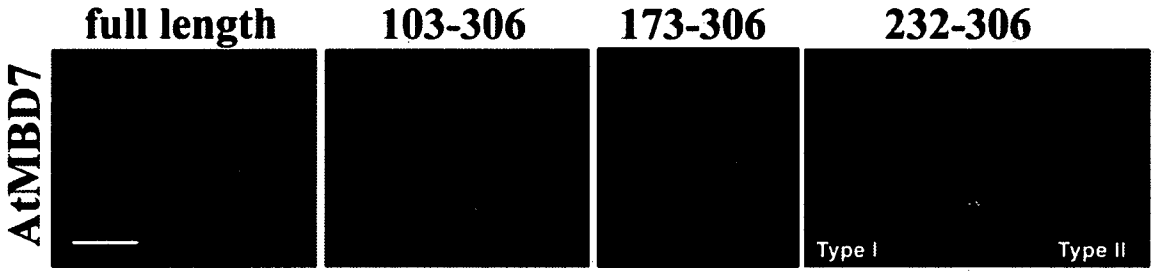


FIG. 2B FIG. 2C FIG. 2D FIG. 2E

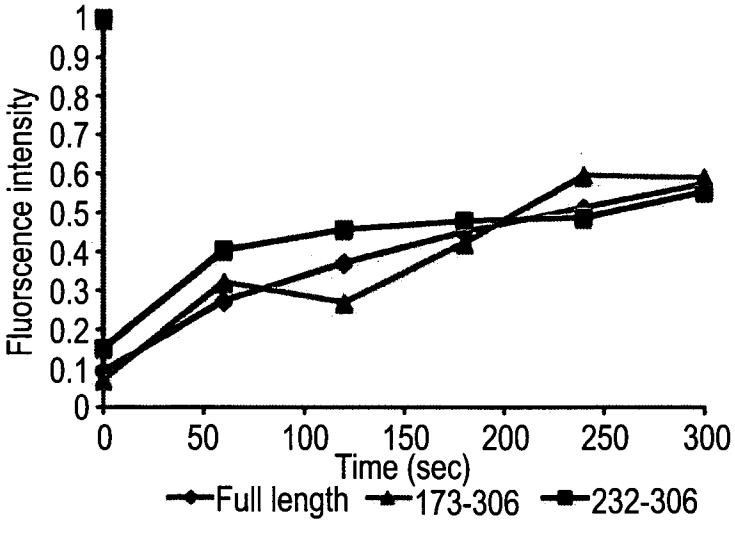


FIG. 2F

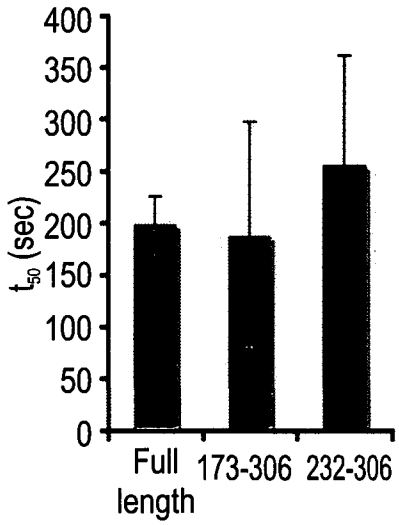


FIG. 2G

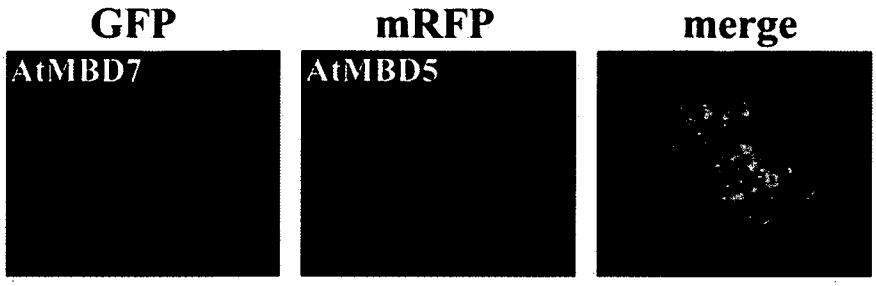


FIG. 3A FIG. 3B FIG. 3C

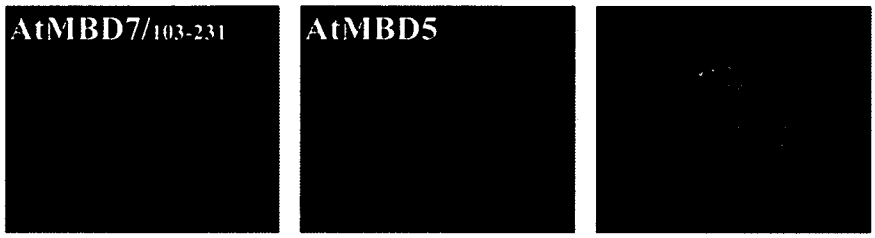


FIG. 3D FIG. 3E FIG. 3F

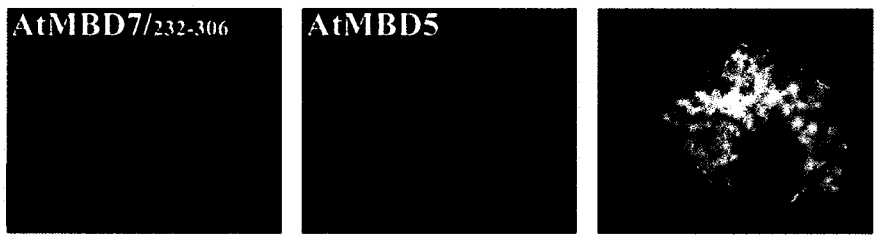


FIG. 3G FIG. 3H FIG. 3I

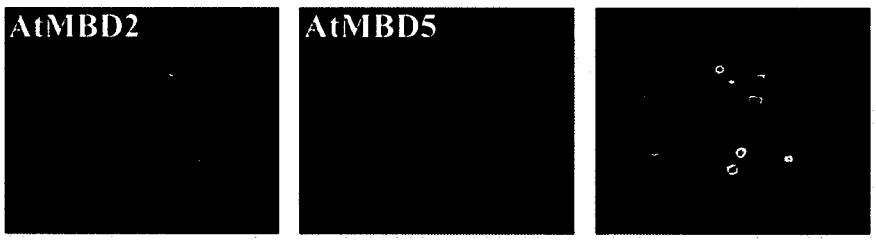


FIG. 3J FIG. 3K FIG. 3L

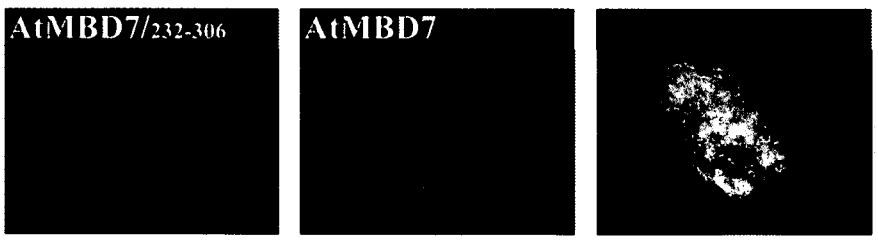


FIG. 3M FIG. 3N FIG. 3O

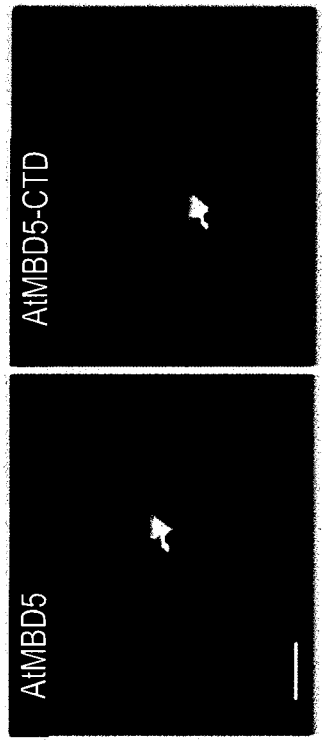


FIG. 4A

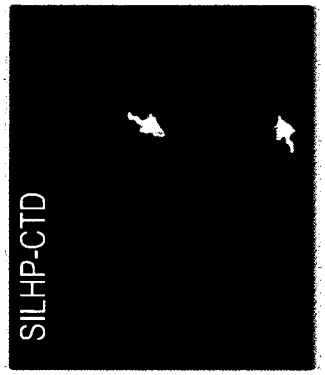


FIG. 4B

FIG. 4D

FIG. 4E

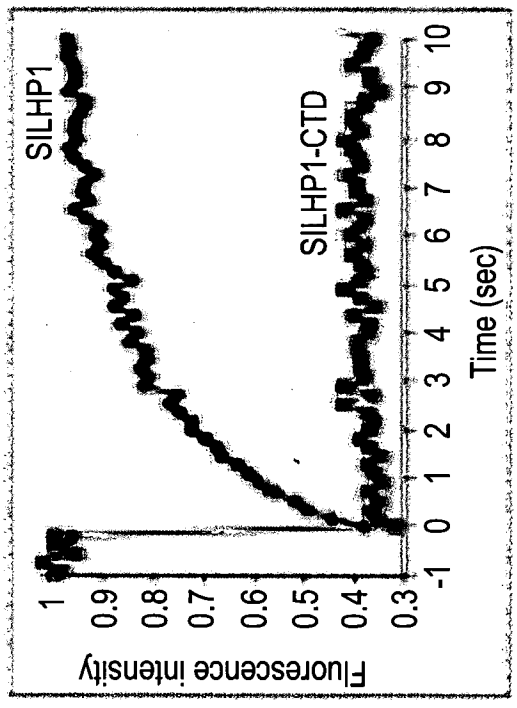


FIG. 4C

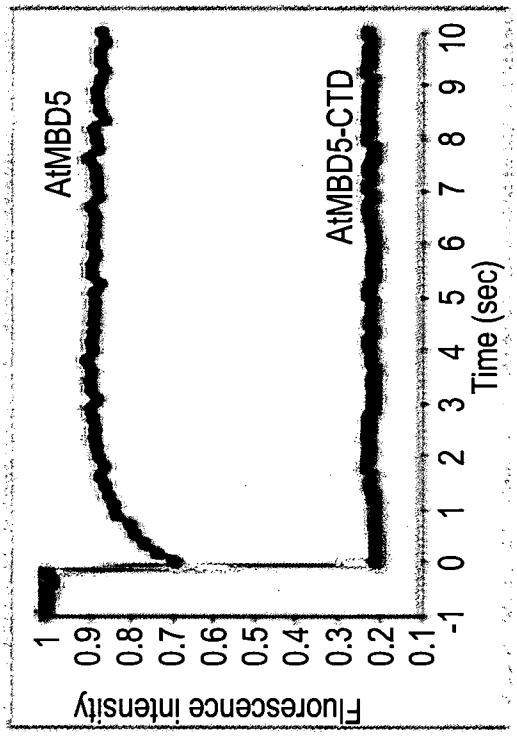


FIG. 4F

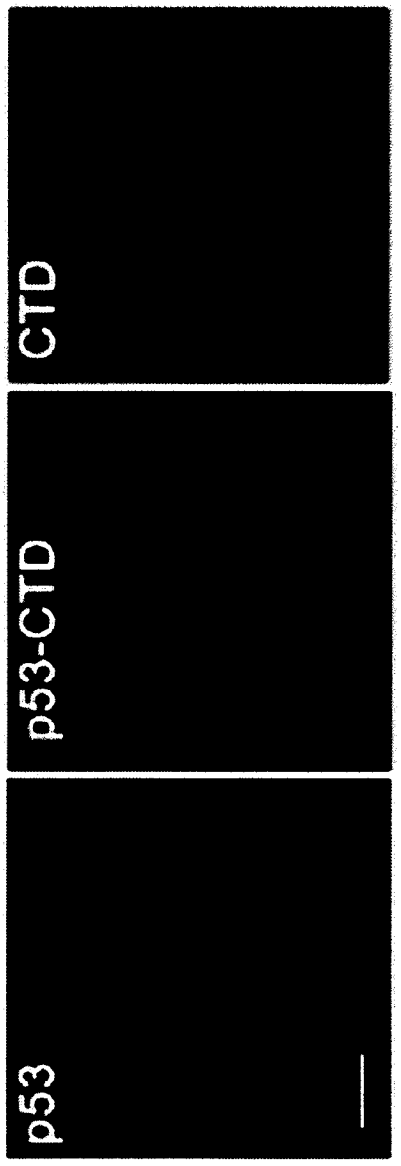


FIG. 4G

FIG. 4H

FIG. 4I

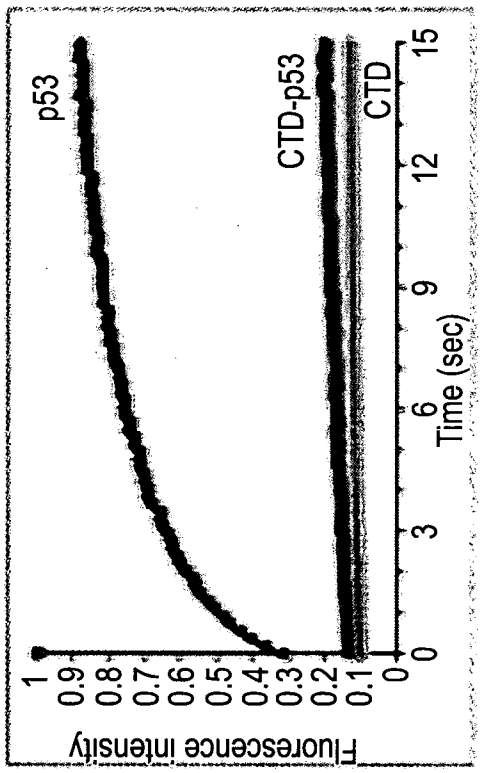


FIG. 4J

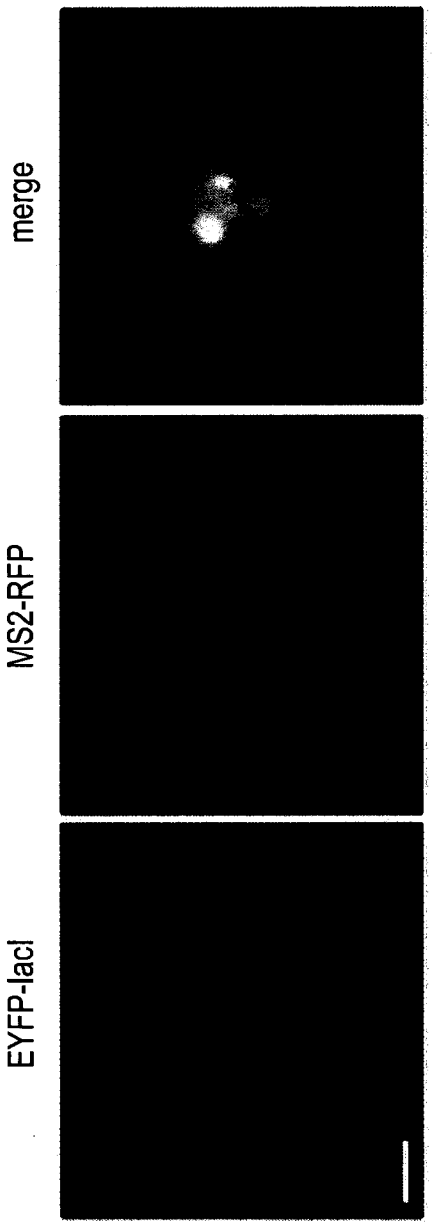


FIG. 5A

FIG. 5B

FIG. 5C

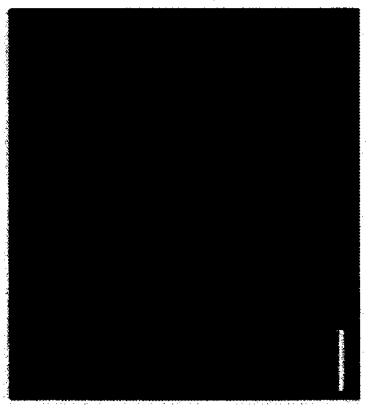


FIG. 5D

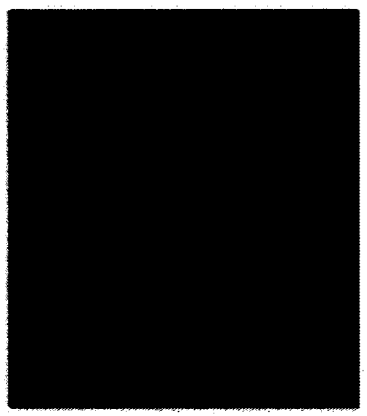


FIG. 5E

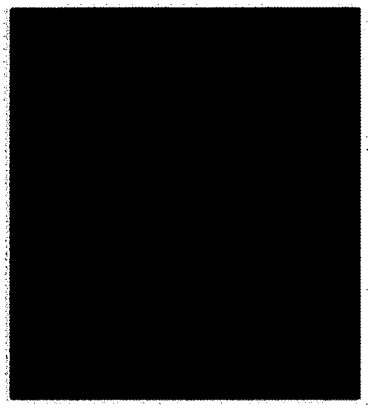


FIG. 5F

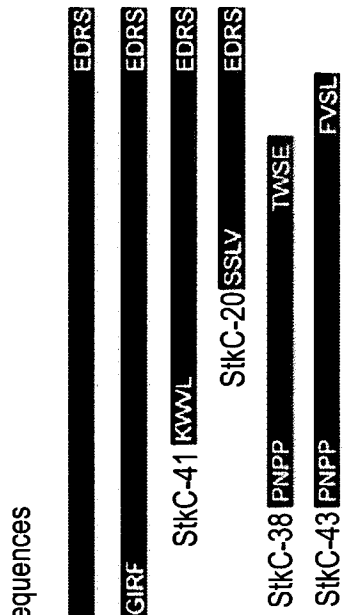

StkC truncated sequences
StkC (FL) 75aa
SYSM
EDRS
StkC-58
GIRF
EDRS
StkC-41
KWVL
EDRS
StkC-20
SSLV
EDRS
StkC-38
PNPP
TWSE
StkC-43
PNPP
FVSL

FIG. 6A

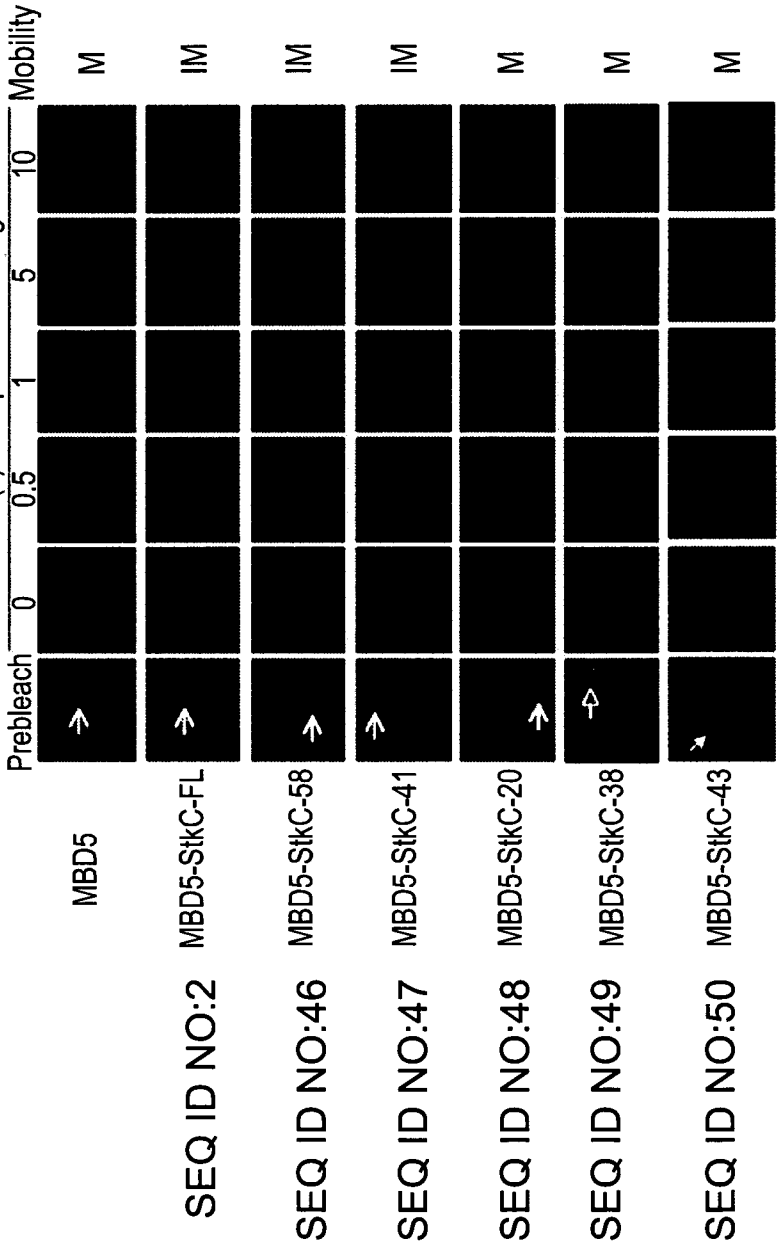

Time (s) after photobleaching
Prebleach
0
0.5
1
5
10
Mobility
MBD5
M
MBD5-StkC-FL
IM
SEQ ID NO:2
MBD5-StkC-58
IM
SEQ ID NO:46
MBD5-StkC-41
IM
SEQ ID NO:47
MBD5-StkC-20
M
SEQ ID NO:48
MBD5-StkC-38
M
SEQ ID NO:49
MBD5-StkC-43
M
SEQ ID NO:50

FIG. 6B

FIG. 7A

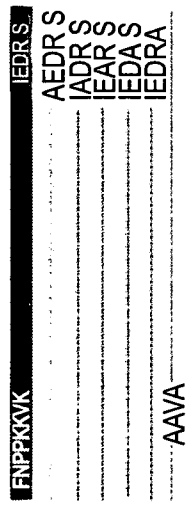
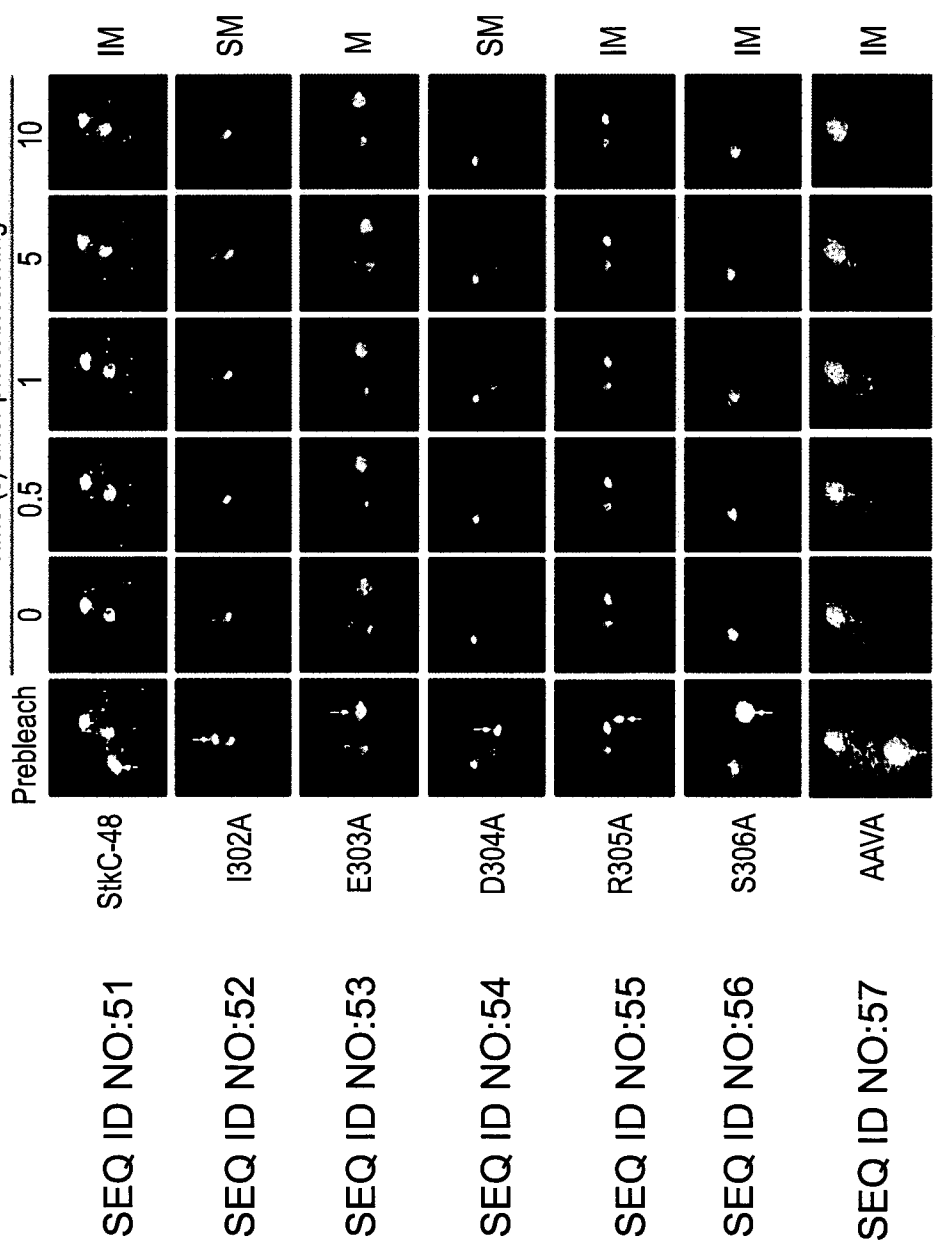

StkC-48
I302A
E303A
D304A
R305A
S306A
AAVA
FNPPKKVK
IEDRS
AEDRS
IADRS
IEARS
IEDAS
IEDRA
AAVA

FIG. 7B


Prebleach
0
0.5
1
5
10
Time (s) after photobleaching
IM
SM
M
SM
IM
IM
IM
StkC-48
I302A
E303A
D304A
R305A
S306A
AAVA
SEQ ID NO:51
SEQ ID NO:52
SEQ ID NO:53
SEQ ID NO:54
SEQ ID NO:55
SEQ ID NO:56
SEQ ID NO:57

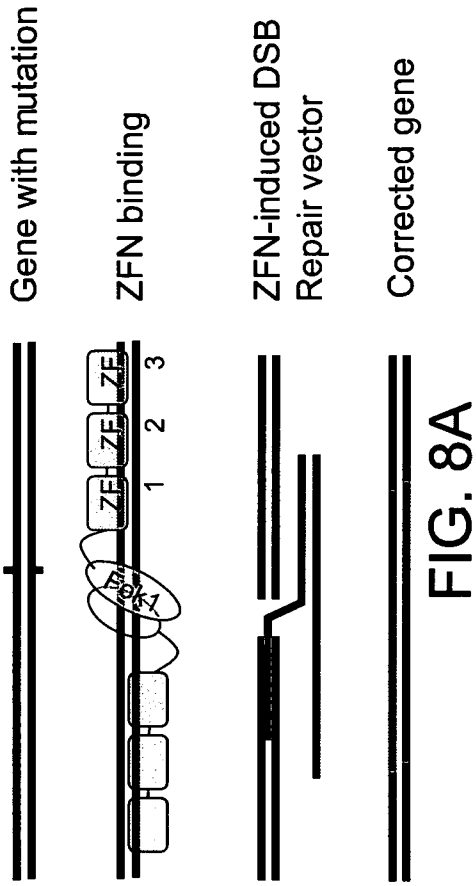


FIG. 8A

FIG. 8B

FIG. 8C

INTERNATIONAL SEARCH REPORT

International application No
PCT/IL2009/000434

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/82 C12N15/62 C07K14/415

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, EMBL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ZEMACH ASSAF ET AL: "The three methyl-CpG-binding domains of AtMBD7 control its subnuclear localization and mobility" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 283, no. 13, March 2008 (2008-03), pages 8406-8411, XP002541613 ISSN: 0021-9258 the whole document	1-10, 13-21

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 August 2009	28/08/2009

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Fax: (+31–70) 340–3016	Espen, Josée

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCEBBA FRANCESCA ET AL: "PRMT11: a new Arabidopsis MBD7 protein partner with arginine methyltransferase activity" PLANT JOURNAL, vol. 52, no. 2, October 2007 (2007-10), pages 210-222, XP002541614 ISSN: 0960-7412	11,12
Y	page 212, right-hand column, last paragraph - page 213, left-hand column, paragraph 1; figures 2-4	1-10, 13-21
Y	VODA M ET AL: "Cloning, expression, purification and protein binding studies of AtMBD5 from Arabidopsis thaliana" COMPARATIVE BIOCHEMISTRY AND PHYSIOLOGY PART A MOLECULAR & INTEGRATIVE PHYSIOLOGY, vol. 141, no. 3, Suppl. S, July 2005 (2005-07), page S260, XP009121418 & ANNUAL MEETING OF THE SOCIETY-FOR-EXPERIMENTAL-BIOLOGY; BARCELONA, SPAIN; JULY 11 -15, 2005 ISSN: 1095-6433 abstract	1-10, 13-21
Y	DATABASE EMBL [Online] 24 February 2006 (2006-02-24), ALEXANDROV N A ET AL: "Arabidopsis thaliana cDNA clone 4349" XP002541616 Database accession no. DR370706 EST	1-10, 13-21
Y	DATABASE GENESEQ [Online] 28 December 2007 (2007-12-28), CERES INC.: "Plant protein fragment" XP002541617 Database accession no. ALK03686 SEQ ID NO: 70938 & EP 1 586 645 A (CERES INC [US]) 19 October 2005 (2005-10-19)	1-10, 13-21

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IL2009/000434

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1586645 A	19-10-2005	US 2006084796 A1	20-04-2006