BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 8367–8375 8367
DOI: 10.1021/bi100448f

STPR, a 23-Amino Acid Tandem Repeat Domain, Found in the Human Function-Unknown Protein ZNF821[†]

Yasuhiro Nonaka,[‡] Hideki Muto,[§] Tomoyasu Aizawa,[∥] Etsuro Okabe,[‡] Shohei Myoba,[‡] Takuya Yokoyama,[‡] Shin Saito,[‡] Fumie Tatami,[∥] Yasuhiro Kumaki,[‡] Masakatsu Kamiya,[∥] Takashi Kikukawa,[‡] Mineyuki Mizuguchi,[⊥] Shigeharu Takiya,[#] Masataka Kinjo,[§] Makoto Demura,*[,∥] and Keiichi Kawano*[,‡]

[‡]*Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan,* [§]*Division of Cellular Life Science, Faculty of Advanced Life Science, Hokkaido University, Sapporo 001-0021, Japan,* [∥]*Division of Molecular Life Science, Faculty of Advanced Life Science, Hokkaido University, Sapporo 060-0810, Japan,* [⊥]*Faculty of Pharmaceutical Sciences, University of Toyama, Toyama 930-0194, Japan, and* [#]*Department of Biological Sciences and Center for Genome Dynamics, Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan*

Received March 25, 2010; Revised Manuscript Received August 9, 2010

ABSTRACT: The STPR motif is composed of 23-amino acid repeats aligned contiguously. STPR was originally reported as the DNA-binding domain of the silkworm protein FMBP-1. ZNF821, the human protein that contains the STPR domain, is a zinc finger protein of unknown function. In this study, we prepared peptides of silkworm FMBP-1 STPR (sSTPR) and human ZNF821 STPR (hSTPR) and compared their DNA binding behaviors. This revealed that hSTPR, like sSTPR, is a double-stranded DNA-binding domain. Sequence-independent DNA binding affinities and α-helix-rich DNA-bound structures were comparable between the two STPRs, although the specific DNA sequence of hSTPR is still unclear. In addition, a subcellular expression experiment showed that the hSTPR domain is responsible for the nuclear localization of ZNF821. ZNF821 showed a much slower diffusion rate in the nucleus, suggesting the possibility of interaction with chromosomal DNA. STPR sequences are found in many proteins from vertebrates, insects, and nematodes. Some of the consensus amino acid residues would be responsible for DNA binding and concomitant increases in α-helix structure content.

The one-score-and-three (i.e., 23)-amino acid peptide repeat (STPR)[1] domain was originally reported as the DNA-binding domain of fibroin modulator binding protein-1 (FMBP-1) (*1*), which is a transcription factor regulating heavy-chain fibroin gene expression in the silkworm (*Bombyx mori*). The silk protein production is tissue-specific and developmental stage-specific and is strictly controlled by FMBP-1 and several other proteins (*2*−*11*). FMBP-1 specifically recognizes the oligonucleotide sequence motif (5′-ATNTWTNTA-3′) that exists upstream from and in the intron of the heavy-chain fibroin gene (*1*).

The STPR domain derived from silkworm is composed of four 23-amino acid contiguous repeats (Figure S1 of the Supporting Information). The level of sequence identity among the four repeats is >80%. Previous CD and NMR experiments revealed the structural characteristics of the STPR domain (*12*, *13*). Without DNA, each repeat of STPR forms an α-helix at the N-terminus (residues 3−10), and thus, STPR consists of four short α-helices and four loop regions. It is thought that each α-helix does not interact with the others; in other words, there is no explicit tertiary structure. When STPR binds to a specific DNA element, the α-helical content of STPR more than doubles (from 31 to 76%). In addition, the STPR peptide bound to specific DNA forms a rigid structure, while DNA-free STPR is highly flexible, as shown by limited proteolysis. Interestingly, an amino acid substitution experiment indicated that the contributions of the four repeats to DNA binding differed from each other, in spite of the high level of sequence similarity among the repeats. A competition assay using DNA groove-binding agents suggested that STPR approaches DNA from the major groove (*14*). Mobility shifts of protein−DNA complexes showed that DNA bending is induced slightly by the STPR domain peptide and sharply by full-length FMBP-1. Experimental data describing STPR's complex structure have accumulated, although its structure at atomic resolution has not yet been determined.

In addition to FMBP-1, proteins with regions analogous to the STPR domain were found in other insects (*Drosophila* and mosquitoes), nematodes, and many vertebrates (*1*, *14*). Comparison of the full-length proteins that contain the STPR domain revealed that the amino acid sequences and molecular sizes are widely diversified, and these proteins are not necessarily homologous. It should be noted that the similarity of the 23 amino acids is greater among repeats of the same protein than between those of different STPR-containing proteins. However, Glu1 and Arg9, which were reported to stabilize the α-helix of each repeat in the

[†]This work was partly supported by the Program for the Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN), Japan.

*To whom correspondence should be addressed. K.K.: telephone, +81-11-70-2770; fax, +81-11-70-2770; e-mail, kawano@sci.hokudai.ac.jp. M.D.: telephone, +81-11-70-2771; fax, +81-11-70-2771; e-mail, demura@sci.hokudai.ac.jp.

[1]Abbreviations: STPR, one-score-and-three-amino acid peptide repeat; FMBP-1, fibroin modulator binding protein-1; CD, circular dichroism; ZNF821, zinc finger protein 821; hSTPR, human ZNF821 STPR; sSTPR, silkworm FMBP-1 STPR; dsDNA, double-stranded DNA; ssDNA, single-stranded DNA; NMR, nuclear magnetic resonance; HSQC, heteronuclear single-quantum coherence spectroscopy; EGFP, enhanced green fluorescent protein; LSM, laser scanning microscopy; FCS, fluorescence correlation spectroscopy; R6G, rhodamine 6G; SELEX, systematic evolution of ligands by exponential enrichment; NLS, nuclear localization signal.

Published on Web 08/26/2010

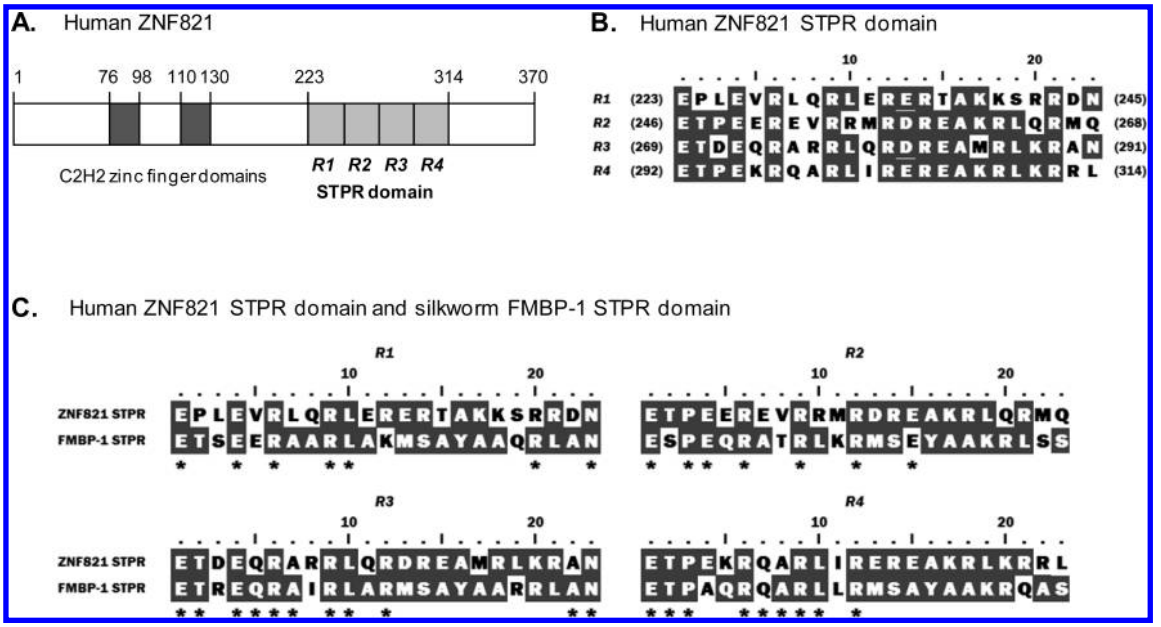

FIGURE 1: Human ZNF821 and STPR domain. (A) Schematic of ZNF821. Two zinc finger domains are colored dark gray, and the STPR domain is colored gray. The four repeats, designated R1−R4, are aligned contiguously. (B) Amino acid sequence of the STPR domain (Glu223−Lue314) in human ZNF821. Letters highlighted in dark gray are the conserved residues between at least two repeats. The identical sequences are also found in ZNF821 proteins of many other vertebrates (e.g., macaque, rat, cow, dog, platypus, and zebra finch). (C) Alignment of the STPR domains of human ZNF821 and silkworm FMBP-1 (see also Figure S1 of the Supporting Information). Letters highlighted in dark gray are the conserved residues between at least two repeats. Asterisks indicate the conserved residues between the two STPR domains.

STPR from silkworm (*12*), are completely conserved in all the repeats of these proteins. We expect that STPR represents a novel DNA-binding motif that generally exists in the animal kingdom. However, there is no evidence that an STPR domain other than that in FMBP-1 is a DNA-binding domain.

The human protein with an STPR sequence is designated as zinc finger protein 821 (ZNF821) (Figure 1) (*15*). Because of the existence of C2H2 zinc finger motifs, ZNF821 is expected to be a DNA-binding protein. The biological significance of this protein is, however, almost unknown. Orthologs of ZNF821 are found in a variety of vertebrates. The amino acid sequences of ZNF821s are well-conserved, especially in the STPR region (Figure S2 and Table S1 of the Supporting Information). STPR domain sequences from many species are completely or nearly identical to human ZNF821 STPR. The remarkable conservation implies that the STPR domain plays a significant role in this protein.

To understand the significance of the STPR domain, we have compared the properties of human ZNF821 STPR (hSTPR) and silkworm FMBP-1 STPR (sSTPR) peptides. In this study, we have confirmed that hSTPR, like sSTPR, has a sequence-independent DNA binding ability. Furthermore, we have examined the DNA-bound structure and the contribution of hSTPR to the subcellular localization. hSTPR and sSTPR domains share common properties, and STPR can be regarded as a novel DNA-binding motif.

EXPERIMENTAL PROCEDURES

Preparation of Recombinant STPR Peptides. Recombinant hSTPR and sSTPR peptides were prepared by an *Escherichia coli* expression system as described previously (*13*). The nucleotide sequence of the hSTPR domain was amplified using two oligonucleotide primers (5′-GGAATTCCATATGATGCGACAGAATGAGCCTTTGGAA-3′ and 5′-CCGCTCGAGCATTTTCTCCAGCCTCCTCTT-3′) and the pOTB7 plasmid including cDNA of human ZNF821 (LOC55565) (purchased from Invitrogen, Carlsbad, CA). The PCR products of the appropriate size were purified, digested with NdeI and XhoI, and subcloned into the pET-22b(+) vector (Novagen, Madison, WI), which was also digested with the same two restriction enzymes. The resulting construct included three residues on both sides of the STPR domain, and a histidine tag was attached to the C-terminus. The overexpressed peptide was purified with the Ni-NTA column and via reverse-phase high-performance liquid chromatography, and the purity was checked by reverse-phase high-performance liquid chromatography. For CD spectroscopy and the mobility shift assay, the peptide concentration was determined by monitoring the absorbance at 280 nm (sSTPR) (*16*) or 214 nm (hSTPR) (*17*).

Electrophoretic Mobility Shift Assay. The electrophoretic mobility shift assay was performed as described by Takiya et al. (*11*). Two kinds of 36-mer double-stranded DNA probes were prepared. The +290 probe (5′-AATTGATGAATCTATGTAAATACTGGGCAGACAATT-3′ and 5′-AATTGTCTGCCCAGTATTTACATAGATTCATCAATT-3′) contains a recognition element of sSTPR (underlined). This sequence originates from the +290 region of the heavy-chain fibroin gene. The random probe (5′-GGATCGAGATGGAGAGTGATCTCAGTCTCAGGTTCA-3′ and 5′-TGAACCTGAGACTGAGATCACTCTCCATCTCGATCC-3′) does not contain the recognition element of sSTPR. This was designed randomly but did not contain AT- and GC-rich regions.

Hybridization with step-by-step cooling was conducted after mixing two DNA strands. The strands were labeled with [γ-^{32}P]ATP (Perkin-Elmer, Wellesley, MA) and T4 kinase (Takara, Tokyo, Japan). Labeled DNA (5 nM) was incubated with protein in 10 μL of the reaction mixture [30 mM Tris-HCl, 7.5 mM $MgCl_2$, 60 mM NaCl, 0.06 mM EDTA, 0.6 mM DTT, 0.06% NP40, and 20% glycerol (pH 7.9)] for 20 min on ice. The protein−DNA complexes were separated on 10% polyacrylamide gels equilibrated with TBE buffer. The gels were dried, exposed to an imaging plate, and visualized using an FLA-7000 RGB imaging analyzer (Fuji Film, Tokyo, Japan).

Apparent dissociation constants were calculated using least-squares fitting to the following equation:

$$\%\ \text{of free DNA} = \frac{{K_d}^n}{{K_d}^n + [\text{STPR}_{\text{free}}]^n}$$

where K_d is the dissociation constant and n is the Hill coefficient. The values of % of free DNA and [$STPR_{free}$] were obtained from the EMSA band intensity. Densitometry of bands was performed using Multi Gauge version 3.0 (Fuji Film).

CD Spectroscopy. CD spectroscopy was performed using a JASCO J-725 spectropolarimeter (Jasco, Tokyo, Japan). Measurements were taken using a 1 mm quartz cell. The concentration of the STPR peptide was 10 μM, prepared in 50 mM sodium phosphate and 50 mM NaCl buffer (pH 7.0). To examine the structures when bound to DNA, double-stranded DNA used for the mobility shift assay (see above) was added. After incubation of the mixture of the peptide (10 μM) and DNA (12.5 μM) for 20 min on ice, the spectra were recorded at 25 °C. The secondary structure contents of peptides were calculated using CONTINLL (*18*) and CDSSTR (*19*) with the data set of SDP48 in the CDPro suite (*20*).

NMR Spectroscopy. NMR experiments were performed to acquire a $^1H-^{15}N$ HSQC (*21*) spectrum of uniformly ^{15}N-labeled hSTPR peptide, using a Bruker DRX-600 spectrometer equipped with a cryogenic probe. The peptide concentration was 0.13 mM dissolved in 50 mM sodium phosphate and 50 mM NaCl (pH 7.0), and the temperature was 293 K.

Live Cell Imaging and Fluorescence Correlation Spectroscopy Measurement. The nucleotide sequence for ZNF821 was amplified by PCR from cDNA of ZNF821 using two oligonucleotide primers: forward primer (5′-CCCAAGCTTCGGCCACCATGTCCCGTCGGAAACAGAC-3′) and reverse primer (5′-CGGGGTACCTCAGTGCAGAGAGCTGCTG-3′). The forward primer and reverse primer contain the *Hin*dIII site and the *Kpn*I site (underlined), respectively. PCR products were purified, digested with the restriction enzymes *Hin*dIII and *Kpn*I, and subcloned downstream of the EGFP coding region in the pEGFP-C1 vector (Clontech, Palo Alto, CA), which was previously digested with the same two restriction enzymes. To create the plasmid of the STPR-truncated mutant (EGFP-ZNF821-δhSTPR), site-directed mutagenesis was conducted using a QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA). Thus, the codon corresponding to the first residue of the hSTPR domain was converted to a stop codon in this mutant. To create the mutant that had only the hSTPR domain (EGFP-hSTPR), the nucleotide sequence for hSTPR was amplified by PCR and incorporated into the pEGFP-C1 vector, as well as full-length ZNF821.

For the transient expression of EGFP fusion proteins, HeLa cells were plated on Lab-Tek chambered coverslips with eight wells (Nalge Nunc International, Rochester, NY) and incubated in Dulbecco's modified Eagle's medium (DMEM, Sigma-Aldrich, St. Louis, MO) supplemented with 10% fetal bovine serum, 100 units/mL penicillin, and 10 mg/mL streptomycin in a 5% CO_2 humidified atmosphere at 37 °C for 24 h. Transfection was conducted with Optifect transfection reagent (Invitrogen) as indicated by the manufacturer. The cells were then grown in DMEM supplemented with 10% fetal bovine serum for 24 h in a 5% CO_2 humidified atmosphere. During LSM imaging and FCS measurement, HeLa cells were maintained in Opti-MEM reduced-serum medium (Invitrogen) at 37 °C in a 5% CO_2 humidified atmosphere.

LSM imaging and FCS measurement were performed using an LSM510 inverted confocal laser scanning microscope (Carl Zeiss, Jena, Germany), which consisted of a CW Ar+ laser, a water immersion objective [C-Apochromat, 40×, 1.2 NA (Carl Zeiss)], and a ConfoCor 3 (Carl Zeiss). EGFP was excited at a 488 nm laser line, and the emission signal through a 505−540 nm bandpass filter was detected. To visualize the nuclei, cells were stained with 25 μg/mL Hoechst 33342 (Invitrogen). Dye was excited by a 405 nm diode laser, and the emission signal through a 465−510 nm bandpass filter was detected.

FCS was measured, and data were processed in essentially the same manner as described previously (*22*, *23*) except for the measurement time (75 s in this study). For quantitative analysis, the fluorescence autocorrelation function $G(\tau)$ was calculated as follows.

$$G(\tau) = \frac{\langle I(t)I(t+\tau)\rangle}{\langle I(t)\rangle^2}$$

where τ represents the time delay, I is the fluorescence intensity, and the brackets denote the ensemble average. $G(\tau)$ was fitted with FCS Access fit (EVOTEC BioSystems, Hamburg, Germany) by a multicomponent model, as follows.

$$G(\tau) = 1 + \frac{1}{N}\sum_i y_i \left(1 + \frac{\tau}{\tau_i}\right)^{-1} \left(1 + \frac{\tau}{s^2\tau_i}\right)^{-1/2}$$

where y_i and τ_i are the fraction and diffusion time of component i, respectively, N is the number of fluorescent molecules in the detection volume element defined by beam waist w_0 and axial radius z_0, and s is the structure parameter representing the z_0/w_0 ratio. The structure parameter was calibrated by measurement of Rhodamine 6G (R6G). The diffusion constants of the samples were calculated from the published diffusion constant of R6G (280 $\mu m^2/s$) (*24*) and the measured diffusion times of R6G (τ_{R6G}) and of each sample (τ_{sample}), as follows.

$$D_{\text{sample}} = D_{\text{R6G}}\frac{\tau_{\text{R6G}}}{\tau_{\text{sample}}} = 280\frac{\tau_{\text{R6G}}}{\tau_{\text{sample}}}\ \mu\text{m}^2/\text{s}$$

The autocorrelation function was normalized by N for comparison of diffusional speed, as follows.

$$\text{normalized } G(\tau) = N[G(\tau) - 1]$$

RESULTS

DNA Binding Ability of hSTPR. The sequence similarity of hSTPR and sSTPR implies that the former can also be a DNA-binding domain (Figure 1C). We performed a mobility shift assay to examine whether hSTPR binds to DNA probes. Two probes were prepared: the +290 probe, which was derived from the intron of the heavy-chain fibroin gene containing a recognition element of sSTPR, and the random probe, which had no sSTPR recognition sequence (see Experimental Procedures). In the following experiments, we focused on both sequence-specific and sequence-independent DNA binding.

As expected, sSTPR bound to the +290 probe at a lower concentration than to the random probe (Figure 2A). A competition assay clearly showed the specific affinity of sSTPR for the +290 probe (Figure 2B). When a large amount of poly(dI-dC) was added, the free +290 probe was hardly detected, while the free random probe appeared. sSTPR−+290 probe binding is sequence-specific, while sSTPR−random probe binding is considered sequence-independent. Previous studies suggested

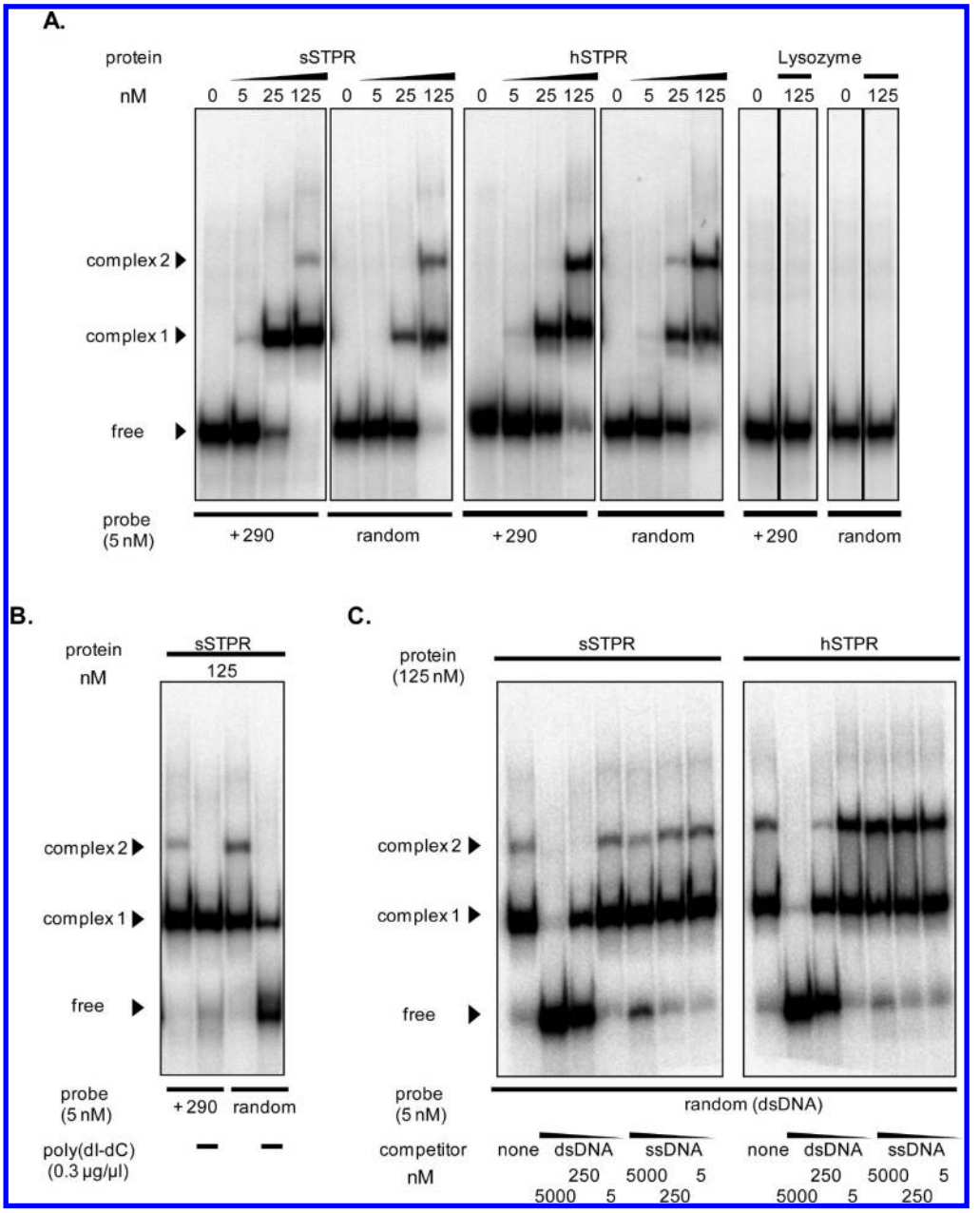

FIGURE 2: Electrophoretic mobility shift assay of sSTPR and hSTPR. (A) DNA binding of sSTPR, hSTPR, and lysozyme. A mixture of 5−125 nM protein and 5 nM DNA probe was incubated for 20 min on ice, followed by separation on a 10% polyacrylamide gel. The protein concentration (nanomolar) is denoted above the images. Images for the same protein are from the identical gel. (B) Competition assay with poly(dI-dC) against sSTPR−DNA probe complexes. The experimental condition was the same as that described above. (C) Competition assay with double-stranded (ds) or single-stranded (ss) random probe vs STPR−random probe complexes. The experimental conditions were the same as those described above. Both images are from the identical gel.

that sSTPR was associated with a recognition element as a monomer (*13*, *14*). Therefore, the second band from the bottom (denoted complex 1) would correspond to the complex of one peptide and one probe. The third band (complex 2) would correspond to the complex of two peptides and one probe, reflecting the fact that one sSTPR associates with a recognition site and another binds to the probe in a manner independent of sequence. The regions other than the recognition element were 9 and 18 bp (see Experimental Procedures), and therefore, the length required for the sequence-independent binding is not more than 18 bp. The length required for the sequence-specific binding is not more than 16 bp, because the 16-mer +290 probe is bound to sSTPR (*12*, *13*).

hSTPR had comparable affinity for the +290 and random probes (Figure 2A); that is, the recognition sequence of sSTPR was not specifically recognized by hSTPR. The DNA binding of hSTPR observed here would be sequence-independent. Complexes 1 and 2 of hSTPR migrated at rates similar to those of sSTPR, suggesting that these complexes consisted of one peptide with one probe and two peptides with one probe, respectively. Lysozyme, which possesses an excess of positively charged residues (pI 11), did not associate with the two probes under this condition. Therefore, the DNA binding of hSTPR observed here is not considered simply a nonspecific electrostatic interaction. Each dissociation constant was estimated with a serial titration experiment at 4 °C (Figure 3 and Table 1). The apparent dissociation constants of sSTPR-specific and sequence-independent bindings were ~7 and ~40 nM, respectively. That of hSTPR−DNA binding was ~100 nM.

To determine whether hSTPR recognizes the DNA duplex specifically, the binding abilities for double-stranded (ds) and single-stranded (ss) random probes were compared (Figure 2C). Consistent with the previous study in which ssDNA did not compete in the binding of FMBP-1 to the double-stranded +290 probe (*11*), a 1000-fold excess of ssDNA hardly competed in the binding of sSTPR to dsDNA. Likewise, ssDNA had little effect on the hSTPR−dsDNA complex. This result indicates that hSTPR, like sSTPR, prefers dsDNA to ssDNA.

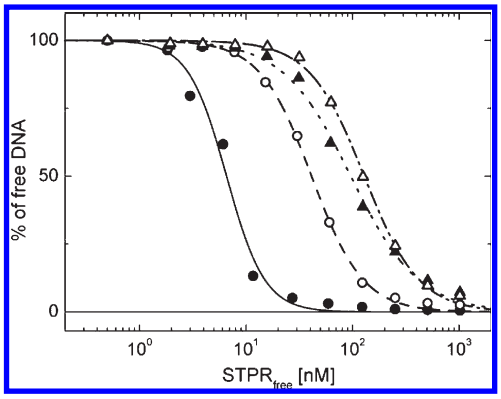

FIGURE 3: Titration of each STPR peptide to DNA. The percentages of the free DNA were estimated for the binding of sSTPR and the +290 probe (●, —), sSTPR and the random probe (○, ---), hSTPR and the +290 probe (▲, ···), and hSTPR and the random probe (△, −·−). The equation (see Experimental Procedures) was fitted to the plots by the least-squares method.

Table 1: Dissociation Constants and Hill Coefficients of STPR and DNA Probes

peptide	probe	K_d (nM)	n
sSTPR	+290	6.57	2.42
	random	41.3	1.79
hSTPR	+290	96.4	1.37
	random	129.0	1.69

Structures of DNA-Free and DNA-Bound hSTPR. We previously reported that, when sSTPR bound to a specific DNA sequence, significant structural changes in both protein and DNA occurred (*12*). In the study presented here, we observed a structural change upon sequence-independent DNA binding. The secondary structures of STPR peptides with and without DNA were monitored by far-UV CD spectroscopy (Figure 4A). In the absence of DNA, hSTPR adopted a typical α-helical structure. The helical content calculated from the spectra was 61.0%, which is greater than that of sSTPR (52.0%). Next, the structural changes in hSTPR and sSTPR upon binding to the random probe were monitored. The spectra showed that binding to the random probe enhanced the α-helical contents of both hSTPR and sSTPR to a comparable extent (77.9 and 82.8%, respectively). Interestingly, the peak intensity at 222 nm of the sSTPR−+290 probe complex (sequence-specific recognition) was larger than that of the sSTPR−random probe complex (sequence-independent recognition), while the estimated helical content was comparable (81.8%).

The spectra at the near-UV region predominantly reflect the DNA structure (Figure 4B). Previous studies reported that the DNA structure also changed upon specific DNA binding (*13*, *14*). The DNA spectrum was changed by the sequence-independent

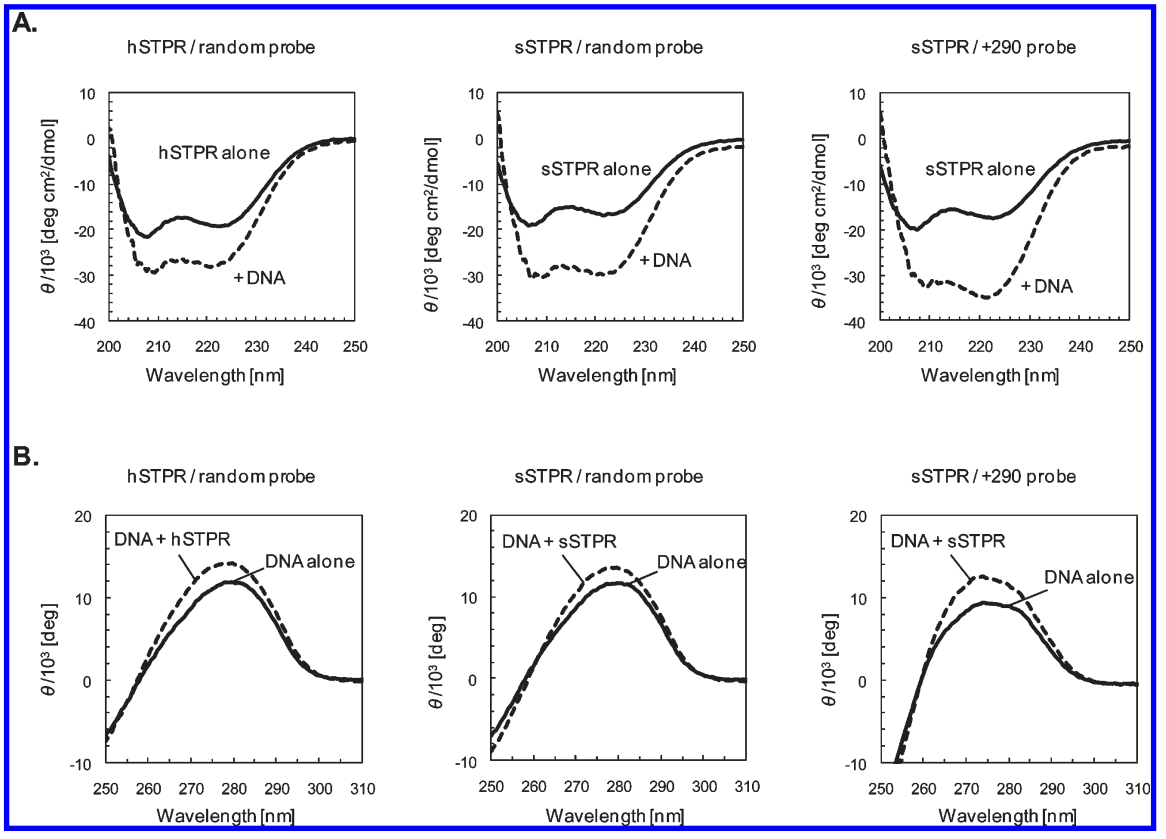

FIGURE 4: CD spectra of STPR−DNA complexes. (A) Far-UV CD spectra of the STPR peptide without (—) and with (---) the DNA probe. A mixture of 10 μM peptide and 12.5 μM DNA probe was incubated for 20 min on ice, followed by the measurement. The spectrum of the peptide with the DNA probe was obtained by subtracting the spectrum of the free DNA probe from that of the mixture. (B) Near-UV CD spectra of the DNA probe (—) and of the mixture of the probe and peptide (---). The experimental conditions were the same as those described above.

binding of hSTPR or sSTPR, in a manner similar to that of the sequence-specific binding of sSTPR, although the increment in the ellipticity of the specific binding was slightly larger than that of the sequence-independent binding. The far- and near-UV CD spectra suggest that the structure of the hSTPR−DNA complex would correspond to that of the sSTPR−DNA complex, at least in the case of sequence-independent binding.

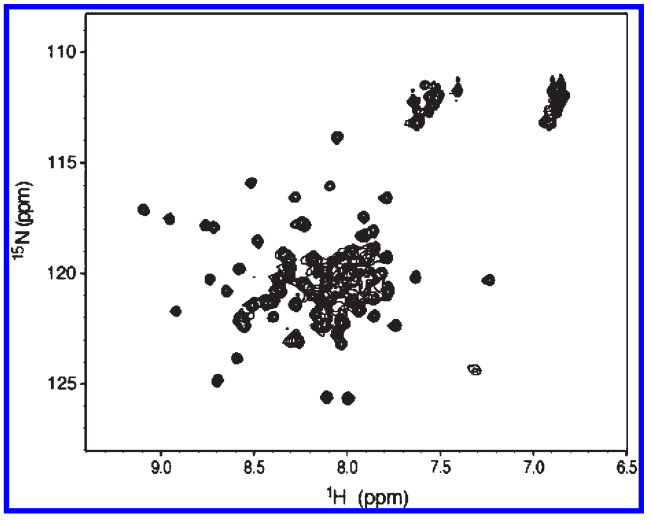

FIGURE 5: ^{1}H−^{15}N HSQC NMR spectrum of the hSTPR peptide. The spectrum was recorded at 293 K. The peptide concentration was 0.13 mM in 50 mM sodium phosphate and 50 mM NaCl buffer (pH 7.0).

To obtain more structural infomation, we acquired the ^{1}H−^{15}N HSQC NMR spectrum of hSTPR (Figure 5). Many resonances were not well-dispersed, and that suggested that hSTPR was largely unstructured. However, downfield-shifted proton resonances were observed around 9.0 ppm, suggesting that the DNA-free structure of hSTPR is also partially structured, like sSTPR (*12*). The solubility of the hSTPR−DNA complex was much lower; moreover, the spectrum showed few clear peaks (data not shown), so that the structural analysis of the complex was difficult.

Subcellular Localization of ZNF821 and Contribution of the hSTPR Domain. To further explore the function of STPR, in vivo expression was assessed. Full-length ZNF821, ZNF821 with the STPR domain truncated, and only the STPR domain were each fused to EGFP at N-termini (EGFP-ZNF821, EGFP-ZNF821−δhSTPR, and EGFP-hSTPR, respectively), and were expressed in HeLa cells (Figure 6). Green fluorescence of the expressed protein was observed by confocal fluorescence microscopy. EGFP-ZNF821 and EGFP-hSTPR were localized only in the nucleus, while EGFP-ZNF821-δhSTPR was not exclusively localized in the nucleus. This clearly indicates that the STPR domain of ZNF821 is necessary and sufficient for nuclear localization.

The diffusion constants of EGFP, EGFP-ZNF821, and EGFP-hSTPR in nuclei were estimated by fluorescence correlation spectroscopy (FCS) (Figure 7 and Table 2). The $G(\tau)$ for EGFP was fitted well by a one-component model. On the other hand, those for EGFP-ZNF821 and EGFP-hSTPR were fitted well not by a one-component model but by a two-component model (Figure S3 of the Supporting Information). The normalized

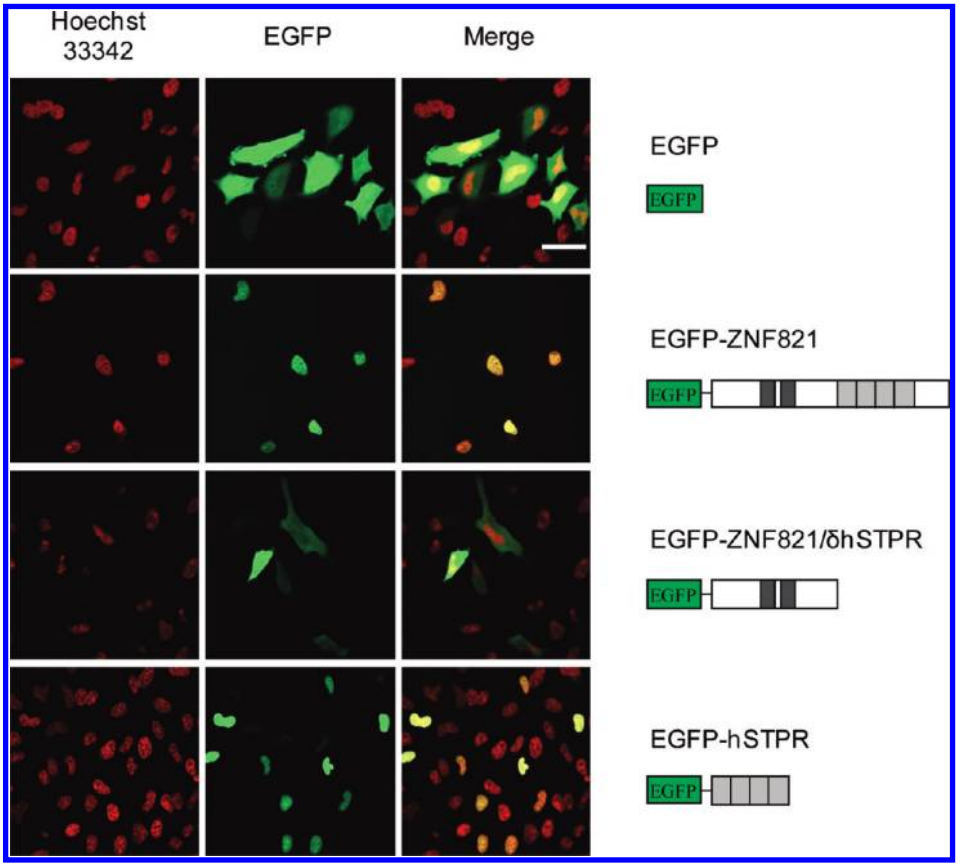

FIGURE 6: Subcellular localization of ZNF821, ZNF821−δhSTPR, and hSTPR, fused with EGFP at the N-termini. The proteins were transiently expressed in HeLa cells. Red color shows nuclei stained with Hoechst 33342 (left panels), and green color shows fluorescence from EGFP (middle panels). The scale bar is 50 μm. Schematic representations of the proteins are shown to the right of the microscopic images. Green, dark gray, and light gray boxes represent EGFP, predicted C2H2 zinc finger domains, and STPR repeats, respectively.

Table 2: Diffusion Constants of EGFP Fusion Proteins in Nuclei[a]

		first		second	calculated[d]	
sample	n[b]	fraction (%)[c]	D (μm^2/s)	D (μm^2/s)	M_w (kDa)	D_{sphere} (μm^2/s)
EGFP	13	100	20.61 ± 1.24	–	27	–
EGFP-ZNF821	36	72.89 ± 7.45	4.95 ± 0.81	0.18 ± 0.12	69	15.07
EGFP-hSTPR	18	51.57 ± 12.32	10.60 ± 3.65	1.18 ± 0.41	39	18.23

[a]Each autocorrelation function of EGFP was fitted by a one-component model, while those of EGFP-ZNF821 and EGFP-hSTPR were fitted by a two-component model (see Figure S3 of the Supporting Information). [b]Number of cells for FCS measurement in a nucleus. [c]Percentage of fast-moving molecules in the total molecules, N (see also Experimental Procedures). [d]M_w was calculated from amino acid composition. For the calculation of D_{sphere}, a measured value of D_{EGFP} (20.61 μm^2/s in this study) was used as in

$$D_{sphere} = D_{EGFP}\sqrt[3]{\frac{M_{wEGFP}}{M_w}} = 20.61\sqrt[3]{\frac{27}{M_w}}$$

In this equation, the shape of the protein was simply assumed to be a sphere and EGFP in a living cell was assumed to be free from the other molecules.

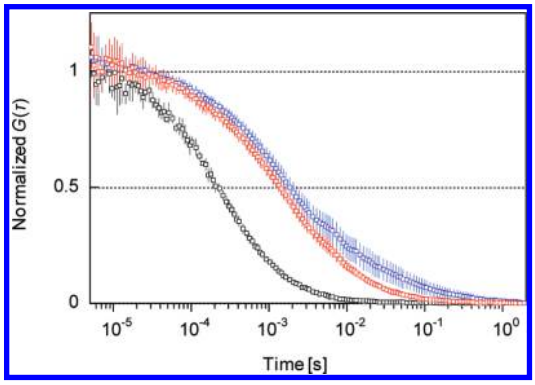

FIGURE 7: Normalized autocorrelation functions of EGFP (black), EGFP-ZNF821 (blue), and EGFP-hSTPR (red) in the nuclei. FCS measurements for EGFP, EGFP-ZNF821, and EGFP-hSTPR were obtained in the nuclei of HeLa cells. Each autocorrelation function of EGFP was fitted by a one-component model, while that of EGFP-ZNF821 and that of EGFP-hSTPR were fitted by a two-component model. They were then normalized by N to unity to compare the extent of diffusional speed (see also Experimental Procedures). The average of normalized autocorrelation functions is shown. The obtained diffusion constants are listed in Table 2.

autocorrelation functions [$G(\tau)$] of the fusion proteins (Figure 7, blue and red) were shifted to the right compared with that of EGFP (black). These shifts reflect the presence of slower diffusion components. Theoretical diffusion constants (Table 2) were calculated from the molecular masses and the diffusion constant of EGFP, assuming that these proteins diffused as globular proteins. EGFP-ZNF821 exhibited two slow diffusion constants (first and second) that were much smaller than the theoretical diffusion constant. Also, the two diffusion constants of EGFP-hSTPR were much smaller than the theoretical value. These experimental values can be explained by interaction with other large molecules rather than by conformational variation (*25*). In particular, the diffusion constants of the second components were quite small and exhibited considerably large apparent molecular masses (3.9×10^7 kDa for EGFP-ZNF821 and 1.4×10^5 kDa for EGFP-hSTPR). The second component can be attributed to the interaction with chromosomal DNA, suggesting that the STPR domain of ZNF821 acts as a DNA binding domain in vivo.

DISCUSSION

Several ZNF821 sequences were obtained from a variety of vertebrates using BLAST (*26*) and ALN (*27*) and were aligned using ClustalW (*28*) with the identity matrix (Figure S2 of the Supporting Information). The alignment scores between hSTPR and STPRs from the other vertebrates are considerably high, most being ~100, while those between full-length human ZNF821 and the other ZNF821s range from 60 to 100 (Table S1 of the Supporting Information). This remarkable conservation suggests that the hSTPR domain should be biologically significant.

The mobility shift assay (Figure 2) and CD spectroscopy (Figure 4) revealed that, like sSTPR, the hSTPR peptide recognized dsDNA. However, the recognition DNA sequences of sSTPR and hSTPR are clearly different (Figure 2A). Although SELEX has been conducted to obtain a recognition sequence of hSTPR, a sequence that specifically bound to hSTPR has not yet been obtained (data not shown). Whether hSTPR has a recognition sequence motif remains unclear.

With respect to the sequence-independent binding, sSTPR and hSTPR exhibit submicromolar dissociation constants (Table 1). The relatively high affinity of sSTPR for other sequences versus the recognition element may confer a disadvantage on the sequence discrimination of FMBP-1. Other regions of FMBP-1 could directly or indirectly contribute to the sequence discrimination. In the case of ZNF821, the existence of zinc finger domains implies the possibility that these domains are responsible for sequence recognition. Otherwise, ZNF821 may be a sequence-independent DNA-binding protein, like DNA repair proteins (*29*, *30*). The sequence-independent binding of the STPR domains could be biologically significant. It has been reported that many DNA-binding proteins, such as transcription factors, associate with nonspecific DNA sequences and subsequently search for their target sequences (*31*−*34*).

Comparison of hSTPR and sSTPR shows that the sequence in the N-terminal half of each repeat is more conserved than that in the C-terminal half (Figure 1C). An Ala scan of sSTPR repeat 3 showed that the important residues for the binding to the +290 probe were Glu1, Arg6, Arg9, Leu10, Met13, and Arg20 (*13*). Importantly, the former four residues are shared by hSTPR, and the latter two residues are not. Our previous studies showed that each repeat of DNA-free sSTPR consisted of an N-terminal α-helix and a C-terminal loop region (*12*). Glu1 and Arg9 form a salt bridge to stabilize the α-helix, which could be formed also in the DNA-bound form. The result here showed that the sequence-independent binding was accompanied by an increment in the α-helical content as well as by sequence-specific binding (Figure 4A). Therefore, Glu1, Arg6, Arg9, and Leu10 are expected to be

significant for both the sequence-specific and sequence-independent dsDNA binding. Unconserved residues in the C-terminal half could be responsible for the oligonucleotide sequence recognition.

Nuclear localization of ZNF821 (Figure 6) also strongly suggests that ZNF821 functions in the nucleus. Comparison of the subcellular localizations of full-length protein and truncated mutants indicated that the hSTPR domain contributed to the nuclear localization. Nuclear localization signal (NLS) predictions predicted that a portion of the hSTPR domain [Lys296−Leu314 by PredictNLS (*35*) and Arg280−Arg313 by NLStradamus with a cutoff of 0.8 (*36*)] would be the NLS for ZNF821. The nuclear import of an NLS-containing protein is mediated by transport factors such as importin α and β. To form a complex with the transport factors, an NLS generally possesses a basic amino acid cluster. Because the net charges of repeats 1−4 of hSTPR are +3, +2, +3, and +6, respectively, repeat 4 may be mainly recognized by the nuclear transport factors, consistent with the predictions. NLS of FMBP-1 has not been determined experimentally. FMBP-1 has the positively charged sSTPR domain and the hyperbasic region (*1*, *14*); these regions could be responsible for nuclear localization. The conservation of Arg residues in various STPR domains suggests that the STPR motif can generally be an NLS, like other DNA-binding domains that also play a role in nuclear localization (*37*).

EGFP-ZNF821 and EGFP-hSTPR each exhibited two diffusion constants, which were both much slower than the theoretical diffusion constants (Figure 7 and Table 2). These slower diffusion constants strongly suggest that ZNF821 and hSTPR interact with other molecules in the nuclei. Apparent molecular masses estimated from the first diffusion constants are $\sim 2.0 \times 10^3$ kDa for EGFP-ZNF821 and $\sim 2.0 \times 10^2$ kDa for EGFP-hSTPR. The first components can be attributed to the formation of a large complex (cf. transcription complex, RNA polymerase−mediator complex). It is noteworthy that hSTPR alone could have the ability to form the complex. The second components exhibited extremely small diffusion constants (with apparent molecular masses of $> 10^5$ kDa), suggesting that ZNF821 and hSTPR would interact with chromosomal DNA. Consistent with the in vitro experiments, hSTPR would associate with DNA also in the nucleus of a live cell. Interestingly, the second diffusion constant of EGFP-ZNF821 was significantly smaller than that of EGFP-hSTPR. Full-length ZNF821 would have a higher DNA binding affinity than hSTPR alone. As previously described, the two zinc finger domains may be responsible for sequence-specific DNA recognition. In conclusion, our results revealed the significant roles of the STPR domain in vivo: DNA binding, nuclear localization, and perhaps complex formation.

In this work, we characterized the STPR domain of ZNF821 in terms of structure and function. The characteristics of hSTPR and sSTPR presented here, such as the affinity of binding for dsDNA and the concomitant enhancement of the α-helical structure, are expected to be shared by the other STPR domains. The contribution to nuclear localization can also be a general property of STPR. From the alignment and experimental results, we can speculate about a DNA-binding motif $[EX(4)RX(2)RLX(13)]_4$. Further analyses will enable us to identify the amino acid sequence element that is essential for an "STPR motif".

SUPPORTING INFORMATION AVAILABLE

Amino acid sequence of silkworm FMBP-1 STPR (Figure S1), an alignment of various vertebrate ZNF821 proteins (Figure S2), comparison of the one- and two-component fittings of the autocorrelation functions (Figure S3), and alignment scores between human ZNF821 and other ZNF821 proteins (Table S1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Takiya, S., Ishikawa, T., Ohtsuka, K., Nishita, Y., and Suzuki, Y. (2005) Fibroin-modulator-binding protein-1 (FMBP-1) contains a novel DNA-binding domain, repeats of the score and three amino acid peptide (STP), conserved from *Caenorhabditis elegans* to humans. *Nucleic Acids Res. 33*, 786–795.
2. Durand, B., Drevet, J., and Couble, P. (1992) P25 gene regulation in *Bombyx mori* silk gland: Two promoter-binding factors have distinct tissue and developmental specificities. *Mol. Cell. Biol. 12*, 5768–5777.
3. Horard, B., Julien, E., Nony, P., Garel, A., and Couble, P. (1997) Differential binding of the *Bombyx* silk gland-specific factor SGFB to its target DNA sequence drives posterior-cell-restricted expression. *Mol. Cell. Biol. 17*, 1572–1579.
4. Hui, C. C., and Suzuki, Y. (1989) Enhancement of transcription from the Ad2 major late promoter by upstream elements of the fibroin- and sericin-1-encoding genes in silk gland extracts. *Gene 85*, 403–411.
5. Hui, C. C., Matsuno, K., and Suzuki, Y. (1990) Fibroin gene promoter contains a cluster of homeodomain binding sites that interact with three silk gland factors. *J. Mol. Biol. 213*, 651–670.
6. Matsuno, K., Hui, C. C., Takiya, S., Suzuki, T., Ueno, K., and Suzuki, Y. (1989) Transcription signals and protein binding sites for sericin gene transcription in vitro. *J. Biol. Chem. 264*, 18707–18713.
7. Suzuki, T., Matsuno, K., Takiya, S., Ohno, K., Ueno, K., and Suzuki, Y. (1991) Purification and characterization of an enhancer-binding protein of the fibroin gene. I. Complete purification of fibroin factor 1. *J. Biol. Chem. 266*, 16935–16941.
8. Suzuki, T., Takiya, S., Matsuno, K., Ohno, K., Ueno, K., and Suzuki, Y. (1991) Purification and characterization of an enhancer-binding protein of the fibroin gene. II. Functional analyses of fibroin factor 1. *J. Biol. Chem. 266*, 16942–16947.
9. Suzuki, Y., Tsuda, M., Hirose, S., and Takiya, S. (1986) Transcription signals and factors of the silk genes. *Adv. Biophys. 21*, 205–215.
10. Takiya, S., Hui, C. C., and Suzuki, Y. (1990) A contribution of the core-promoter and its surrounding regions to the preferential transcription of the fibroin gene in posterior silk gland extracts. *EMBO J. 9*, 489–496.
11. Takiya, S., Kokubo, H., and Suzuki, Y. (1997) Transcriptional regulatory elements in the upstream and intron of the fibroin gene bind three specific factors POU-M1, Bm Fkh and FMBP-1. *Biochem. J. 321* (Part 3), 645–653.
12. Saito, S., Aizawa, T., Kawaguchi, K., Yamaki, T., Matsumoto, D., Kamiya, M., Kumaki, Y., Mizuguchi, M., Takiya, S., Demura, M., and Kawano, K. (2007) Structural approach to a novel tandem repeat DNA-binding domain, STPR, by CD and NMR. *Biochemistry 46*, 1703–1713.
13. Saito, S., Yokoyama, T., Aizawa, T., Kawaguchi, K., Yamaki, T., Matsumoto, D., Kamijima, T., Kamiya, M., Kumaki, Y., Mizuguchi, M., Takiya, S., Demura, M., and Kawano, K. (2008) Structural properties of the DNA-bound form of a novel tandem repeat DNA-binding domain, STPR. *Proteins 72*, 414–426.
14. Takiya, S., Saito, S., Yokoyama, T., Matsumoto, D., Aizawa, T., Kamiya, M., Demura, M., and Kawano, K. (2009) DNA-binding property of the novel DNA-binding domain STPR in FMBP-1 of the silkworm *Bombyx mori*. *J. Biochem. 146*, 103–111.
15. Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F., Zeeberg, B., Buetow, K. H., Schaefer, C. F., Bhat, N. K., Hopkins, R. F., Jordan, H., Moore, T., Max, S. I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A. A., Rubin, G. M., Hong, L., Stapleton, M., Soares, M. B., Bonaldo, M. F., Casavant, T. L., Scheetz, T. E., Brownstein, M. J., Usdin, T. B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S. S., Loquellano, N. A., Peters, G. J., Abramson, R. D., Mullahy, S. J., Bosak, S. A., McEwan, P. J., McKernan, K. J., Malek, J. A., Gunaratne, P. H., Richards, S., Worley, K. C., Hale, S., Garcia, A. M., Gay, L. J., Hulyk, S. W., Villalon, D. K., Muzny, D. M., Sodergren, E. J., Lu, X., Gibbs, R. A., Fahey, J., Helton, E., Ketteman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Young, A. C., Shevchenko, Y., Bouffard, G. G., Blakesley, R. W., Touchman, J. W., Green, E. D., Dickson, M. C., Rodriguez, A. C., Grimwood, J., Schmutz, J., Myers, R. M., Butterfield, Y. S., Krzywinski, M. I., Skalska, U., Smailus,

D. E., Schnerch, A., Schein, J. E., Jones, S. J., Marra, M. A., and The Mammalian Gene Collection Program (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci. U.S.A. 99*, 16899–16903.
16. Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem. 182*, 319–326.
17. Moffatt, F., Senkans, P., and Ricketts, D. (2000) Approaches towards the quantitative analysis of peptides and proteins by reversed-phase high-performance liquid chromatography in the absence of a pure reference sample. *J. Chromatogr., A 891*, 235–242.
18. Provencher, S. W., and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry 20*, 33–37.
19. Johnson, W. C. (1999) Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins 35*, 307–312.
20. Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem. 287*, 252–260.
21. Palmer, A. G., III, Cavanagh, J., Wright, P. E., and Rance, M. (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy. *J. Magn. Reson. 93*, 151–170.
22. Pack, C., Saito, K., Tamura, M., and Kinjo, M. (2006) Microenvironment and effect of energy depletion in the nucleus analyzed by mobility of multiple oligomeric EGFPs. *Biophys. J. 91*, 3921–3936.
23. Saito, K., Ito, E., Takakuwa, Y., Tamura, M., and Kinjo, M. (2003) In situ observation of mobility and anchoring of PKCβI in plasma membrane. *FEBS Lett. 541*, 126–131.
24. Rigler, R., Mets, U., Widengren, J., and Kask, P. (1993) Fluorescence correlation spectroscopy with high count rate and low-background: analysis of translational diffusion. *Eur. Biophys. J. 22*, 169–175.
25. Mikuni, S., Pack, C., Tamura, M., and Kinjo, M. (2007) Diffusion analysis of glucocorticoid receptor and antagonist effect in living cell nucleus. *Exp. Mol. Pathol. 82*, 163–168.
26. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol. 215*, 403–410.
27. Gotoh, O. (2000) Homology-based gene structure prediction: Simplified matching algorithm using a translated codon (tron) and improved accuracy by allowing for long gaps. *Bioinformatics 16*, 190–202.
28. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res. 25*, 4876–4882.
29. Bernstein, D. A., Zittel, M. C., and Keck, J. L. (2003) High-resolution structure of the *E. coli* RecQ helicase catalytic core. *EMBO J. 22*, 4910–4921.
30. Daniels, D. S., Woo, T. T., Luu, K. X., Noll, D. M., Clarke, N. D., Pegg, A. E., and Tainer, J. A. (2004) DNA binding and nucleotide flipping by the human DNA repair protein AGT. *Nat. Struct. Mol. Biol. 11*, 714–720.
31. Dahirel, V., Paillusson, F., Jardat, M., Barbi, M., and Victor, J. M. (2009) Nonspecific DNA-protein interaction: Why proteins can diffuse along DNA. *Phys. Rev. Lett. 102*, 228101.
32. Halford, S. E., and Szczelkun, M. D. (2002) How to get from A to B: Strategies for analysing protein motion on DNA. *Eur. Biophys. J. 31*, 257–267.
33. Halford, S. E., and Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res. 32*, 3040–3052.
34. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes. *Science 305*, 386–389.
35. Cokol, M., Nair, R., and Rost, B. (2000) Finding nuclear localization signals. *EMBO Rep. 1*, 411–415.
36. Nguyen Ba, A. N., Pogoutse, A., Provart, N., and Moses, A. M. (2009) NLStradamus: A simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinf. 10*, 202.
37. LaCasse, E. C., and Lefebvre, Y. A. (1995) Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins. *Nucleic Acids Res. 23*, 1647–1656.