 Nucleic Acids Research, 1999, Vol. 27, No. 24 **4703–4709**

Nuclear targeting determinants of the phage P1 Cre DNA recombinase

Yunzheng Le, Sara Gagneten, Donatella Tombaccini, Bruce Bethke and Brian Sauer*

Developmental Biology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA

Received September 9, 1999; Revised and Accepted November 3, 1999

ABSTRACT

The Cre DNA recombinase of bacteriophage P1 has become a useful tool for genomic manipulation in mice and other eukaryotes. Because Cre is of prokaryotic origin, the 38 kDa protein has been presumed to gain access to the eukaryotic nucleus simply because it is sufficiently small to pass through the nuclear pore by passive diffusion. Instead, we show here that Cre carries nuclear targeting determinants that efficiently direct Cre entry into the nucleus of mammalian cells. Fusions of Cre with green fluorescent protein (GFP) identified two regions that are necessary for nuclear localization. Region I contains a cluster of basic amino acids that is essential for nuclear localization and which resembles a bipartite-like nuclear localization signal. Region II exhibits a β-sheet structure with which the bipartite motif may interact. However, neither region is by itself sufficient for nuclear localization. Nuclear transport *in vitro* with a 98 kDa GFP–Cre fusion protein shows that Cre does not gain access to the nucleus by passive diffusion, but instead enters the nucleus by means of an energy-dependent process. Thus, Cre is one of the few prokaryotic proteins that have been shown to carry determinants that allow it to target the eukaryotic nucleus.

INTRODUCTION

Both gene targeting by homologous DNA recombination and transgenic methods have become powerful tools for the genetic manipulation of a variety of eukaryotic organisms. These approaches have been useful in generating simple null and partial loss- or gain-of-function mutations, and have been particularly useful in dissecting gene function in mice (1). In some cases, for example those in which a null mutation results in embryonic lethality, it would be desirable to be able to direct inactivation (or activation) of a gene in a tissue-specific or temporally defined manner.

An increasingly useful strategy for effecting conditional gene inactivation or activation in mammalian systems employs the Cre recombinase of phage P1 (2). In bacteria, the 38 kDa Cre protein catalyzes efficient site-specific recombination between 34 bp sites called *loxP*. Remarkably, Cre also catalyzes efficient DNA recombination *in vivo* at *loxP* sites placed into the genomes of yeast, plants and mammalian cells (3–5). Thus, a DNA segment flanked by directly repeated *loxP* sites can be excised from the genome in a Cre-dependent fashion, and this strategy has been used in mice to activate expression of a target gene (6), delete an endogenous gene (7) and to remove the selectable marker that remains after homologous targeting in embryonic stem cells (8,9). Spatial and temporal control of excision events can be achieved by imposing the desired control on either the expression (6,10,11) or activity (12) of Cre recombinase. Similarly, the related yeast site-specific DNA recombinases FLP (13) and R (14) have become useful in manipulating the genomes of higher eukaryotes.

For efficient site-specific DNA recombination in eukaryotes, not only must the prokaryotic Cre protein catalyze recombination at *loxP* sites located in chromatin, but it must also find its way to the nucleus. Initially the entry of Cre into the nucleus was hypothesized to occur by passive diffusion through the nuclear pore (3) since the 38 kDa Cre protein is smaller than the ~50 kDa upper size limit imposed by the nuclear pore on passive diffusion of macromolecules into the nucleus (15). Nevertheless, most nuclear proteins of eukaryotes, even those small enough to gain entry to the nucleus by passive diffusion, do carry specific nuclear localization signals (NLSs) that facilitate their entry into the nucleus (16). Since even small proteins can benefit from having an NLS (17), one strategy to increase access to the nucleus by Cre is to alter the protein to carry an exogenous NLS (8). In addition, Cre has been fused to the ligand-binding domain of steroid receptor proteins in order to confer to the fusion protein hormonal control of nuclear/cytoplasmic compartmentalization (12).

Recently we observed that fusion of Cre recombinase with green fluorescent protein (GFP) resulted in a 68 kDa protein that can enter the nucleus of transfected cells (18). This would be puzzling if the only means that Cre has for gaining entry to the nucleus is by passive diffusion through the nuclear pore. One possibility is that Cre relies upon dissolution of the nuclear envelope during mitosis for nuclear entry. If this was to be the case, it would impose limitations on the use of certain Cre vectors for effecting rapid conditional gene ablation in post-mitotic cells in gene-modified mice. Yet, wild-type

*To whom correspondence should be addressed. Tel: +1 405 271 8168; Fax: +1 405 271 7312; Email: sauerb@omrf.ouhsc.edu
Present addresses:
Sara Gagneten, Food and Drug Administration, Bethesda, MD 20892, USA
Donatella Tombaccini, Istituto di Patologia Generale, Universita di Firenze, Firenze I-50134, Italy
Bruce Bethke, Department of Biology, St Vincent College, Latrobe, PA 15650, USA

unmodified Cre protein does catalyze rapid DNA genomic recombination in post-mitotic mouse cells in both liver and brain (19–21), so the importance of this concern for conditional gene targeting strategies is unclear.

An alternative possibility is that Cre DNA recombinase itself might carry determinants that direct nuclear targeting. Only a few prokaryotic proteins have been shown to carry an NLS. The Ti plasmid of *Agrobacterium tumefaciens* carries two genes, *virD2*, a site-specific DNA endonuclease, and *virE2*, a single-stranded DNA binding protein, that carry NLS sequences (22). Both of these gene products are involved in conveying the T-DNA of the plasmid into the nucleus of host plant cells. Here we show that the Cre DNA recombinase of the horizontally transmissible P1 phage/plasmid also carries an NLS-like element and that wild-type Cre efficiently targets the nucleus of cultured mammalian cells. Use of a sensitive *in vivo* integration assay indicates that the addition of an exogenous NLS confers no increase in Cre recombination at a chromosomal target site. There appear to be two general regions of Cre important for nuclear localization. Although one of these regions carries a necessary nuclear targeting determinant that somewhat resembles the bipartite NLS (16) present in many eukaryotic nuclear proteins, this region by itself is not sufficient for nuclear targeting.

MATERIALS AND METHODS

DNA constructs

The RSV-*cre* plasmid pBS118 (23) was modified using synthetic oligonucleotides, (i) to fuse the SV40 NLS to the N-terminus of Cre and (ii) to provide an improved mammalian translational start, similar to that of the CMV-*cre* plasmid pBS185 (24). The sequence following the *Hin*dIII site at the RSV promoter in the resulting RSV-*NLScre* plasmid pBS317 is d(AGC TTA GGT **ATG ACA CCT AAG AAG AAG AGA AAG GTA TGC AGG TCA CAC**) where bold type indicates the coding region of the fused NLS to *cre*. The CMV-*NLScre* plasmid pBS391 was constructed by replacing the *Xho*I–*Mlu*I *cre* fragment of pBS185 with the *Hin*dIII–*Mlu*I *NLScre* fragment from pBS317, after filling in the *Hin*dIII and *Xho*I sites to allow blunt end ligation. The *loxP* integration plasmid pBS226 (25) has been described previously.

The C-terminal GFP–Cre truncation constructs pBS477 (*Ssp*I), pBS476 (*Cla*I), pBS550 (*Bsm*BI), pBS475 (*Eco*RV) and pBS474 (*Bam*HI) were made from the RSV-GFP–*cre* plasmid pBS448 (18) by inserting complementary oligonucleotides to give the sequence d(CGA TTG GCA GAT CTG TGA CTC TAG) (for *Ssp*I, *Cla*I and *Eco*RV), d(GGT GTG CAA GAT CTG TGA CTC TAG) (for *Bsm*BI) and d(GAT CCA AGA TCT GTG ACT CTA G) (for *Bam*HI) between the corresponding restriction sites in the *cre* structural gene and a unique 3′ *Xba*I site. All truncations thus carry a unique *Bgl*II site and an opal termination codon (underlined).

Amino acids 100–123 of Cre were fused to GFP by amplifying that region from the *cre* gene with the primers d(TCT CAG ACC ATG GCT CGT CGG TCC GGG CTG CC) and d(TCG TTG CTC TAG ATC ATT CTT TTC GGA TCC G) using *Pfu* DNA polymerase (Stratagene), and then replacing the *cre* gene between the *Nco*I and *Xba*I sites of pBS448 with the *Nco*I–*Xba*I region from the amplified fragment to generate pBS498. The N-terminal Cre truncation/GFP fusions pBS486 and pBS549 were constructed from pBS476 by substituting the region between the *Nco*I site (at amino acid position 1 of Cre) and the *Bam*HI site (at amino acid position 120 of Cre) with complementary synthetic oligonucleotides to give the sequence d(CAT GGC TCG ACG GAT C) for pBS486 and d(CAT GGC TCG ACG GAT C) for pBS549. The N-terminal Cre truncation/GFP fusion pBS530 was derived from pBS486 by substituting the GFP–Cre fusion portion of pBS498 on a *Hin*dIII–*Bam*HI fragment for the corresponding region of pBS486. Single amino acid substitutions in the *cre* coding region of GFP–*cre* were made from pBS448 using the QuikChange mutagenesis kit (Stratagene).

A tribrid glutathione-S-transferase (GST)–GFP–Cre fusion vector was derived from pBS505 (26), a eukaryotic expression vector coding for a fusion of *cre* with the translationally optimized GFP gene of pEGFP-C1 (Clontech). The GFP–*cre* fusion gene on a *Kpn*I–*Nhe*I fragment was inserted into a polylinker-modified (*Eco*RI–*Kpn*I–*Sma*I–*Xba*I–*Xho*I) version of pGEX-4T-1 (Pharmacia) to give the tribrid vector pBS608. All synthetic and amplified DNA and junctions were verified by DNA sequencing.

Cell culture

Chinese hamster ovary (CHO-K1) and HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum, 100 μM MEM non-essential amino acids, 50 U/ml penicillin and 50 μg/ml streptomycin. CHO 14-1-2 cells (25) were propagated in α medium as previously described. All cell culture reagents were from Life Technologies Inc. (Gaithersburg, MD).

DNA transfection, immunofluorescence and microscopy

CHO-K1 cells (6×10^4 cells/30 mm well) were plated on coverslips 1 day before transfection using calcium phosphate precipitation (27). After transfection with 3 μg of DNA, cells were incubated for 14–16 h, and then washed twice with DMEM and incubated with fresh media. At 44 h after transfection, cells were either examined for GFP fluorescence using both live and fixed (2% paraformaldehyde) cells (18) or processed for immunofluorescence. Between 100 and 300 productively transfected cells were examined in each transfection experiment and assigned to one of three categories: cytoplasmic, cytoplasmic + nuclear, or nuclear.

For indirect immunofluorescence labeling, cells were washed once with phosphate-buffered saline (PBS) and then fixed with 2% formaldehyde in PBS for 20 min. Cells were washed twice with PBS, permeabilized with 0.1% triton in PBS for 6 min, washed twice with PBS, and incubated with 2% normal goat serum (NGS) (Life Technologies Inc.) in PBS for 30 min. After incubation with primary antibodies for 30 min, cells were washed three times, incubated with secondary antibody for 30 min and then washed three times before mounting on glass slides. All washes and incubations were performed at room temperature with PBS containing 2% NGS. Cre was detected using a 1:500 dilution of rabbit anti-Cre antiserum (4) followed by a 1:100 dilution of rhodamine conjugated goat anti-rabbit immunoglobulin G (Jackson ImmunoResearch). Anti-Cre antiserum was preabsorbed to confluent fixed CHO-K1 cells for 30 min at room temperature to ensure reduced background staining.

Cells were monitored with a Nikon Optiphot-2 epifluorescence microscope, using Nikon filter blocks B2-A (DM510/BA 520–560 nm/excitation 450–490 nm) for GFP fluorescence and G-2A(DM580/BA 590 nm/excitation 510–560 nm) for rhodamine fluorescence.

Assay for nuclear transport

Digitonin-permeabilized HeLa cells were prepared and used for the nuclear import assay, as described (28,29). Briefly, permeabilized cells on a cover slip were incubated in a humidified chamber with 25 µl complete import buffer for 30 min at 37°C. Complete import buffer consisted of Buffer A (20 mM HEPES pH 7.3, 110 mM potassium acetate, 2 mM dithiothreitol, 1 µg/ml leupeptin, 1 µg/ml aprotinin, 1 µg/ml pepstatin) plus 20 U/ml creatine kinase, 1 mM magnesium ATP, 0.5 mM EGTA, 5 mM creatine phosphate, 10% untreated rabbit reticulocyte lysate (Promega) and 10 µg GST–GFP–Cre fusion protein. The 98 kDa GST–GFP–Cre fusion protein was purified according to the manufacturer's suggestions by elution from the Bulk GST Purification Module (Pharmacia) with 0.5 M NaCl in the elution buffer.

Cells were fixed with Buffer A plus 6.7% formaldehyde (Electron Microscopy Services) at 25°C and mounted on 5 µl 0.1% *p*-phenylenediamine/90% glycerol in PBS, and viewed immediately. Microscopy of transport reaction assays was with a 63× oil immersion objective on a Zeiss Axiovert 100TV microscope mounted with a Pentamax CCD video camera. Initial image analysis was performed on an OEI PC computer using the Metamorph Imaging system (Universal Imaging Corp., West Chester, PA).

Integration assay

Quantitative Cre-mediated integration was assayed in CHO 14-1-2 cells (25,30). Briefly, 5×10^5 to 1×10^6 cells were co-electroporated with 15 µg of targeting vector pBS226 and 0, 0.15 or 0.30 µg of Cre expression plasmid pBS185 or pBS391 using cytomix buffer (31). Electroporation was by a single pulse of 350 V at 1600 µfarads (field strength +875 V/cm) in a Cell Porator (Life Technologies Inc.) and colonies resistant to 400 µg/ml G-418 were scored 10 days later.

RESULTS

Nuclear localization of Cre

Recently we observed that fusion of the Cre protein with the GFP of *Aequorea victoria* (32) resulted in a 68 kDa hybrid protein that localized predominantly in the nucleus of transfected mammalian cells (18). This result suggested that Cre contains a signal(s) that targets Cre to the eukaryotic nucleus. Similarly, we have observed nuclear targeting of the GFP–Cre fusion protein in the yeast *Saccharomyces cerevisiae* (data not shown), suggesting that nuclear entry is unlikely to be the simple consequence of dissolution of the nuclear envelope during mitosis since the nucleus of yeast remains intact during mitosis (33).

To verify that the wild-type protein itself targets the mammalian nucleus, CHO-K1 cells were transfected with a CMV-*cre* vector and 2 days later the cellular distribution of Cre was determined by indirect immunofluorescence. Figure 1A and B shows that transfected cells exhibit exquisite

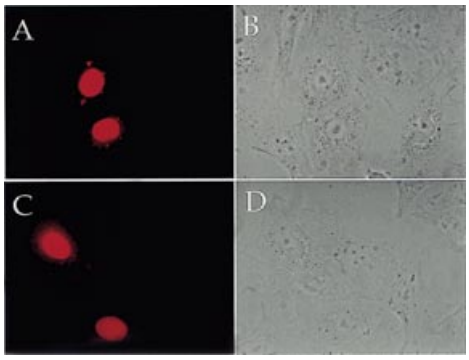

Figure 1. Detection of Cre in transfected cells using a specific anti-Cre antibody. CHO-K1 cells were transfected with either CMV-*cre* (**A**) and (**B**) or CMV-*NLScre* (**C**) and (**D**). Two days later cells were examined for immunofluorescence using specific anti-Cre antiserum (A) and (C) or by phase contrast (B) and (D) microscopy.

immunofluorescence in the cell nucleus with little or no immunofluorescence in the cytoplasm. For comparison, we fused the SV40 large T antigen NLS to the N-terminus of Cre and examined its cellular distribution in transfected cells. As expected, the NLS–Cre protein is localized almost exclusively in the cell nucleus (Fig. 1C and D). In particular, there is little if any difference between the cellular distribution of the wild-type Cre and the NLS–Cre proteins, indicating that Cre carries determinants that target Cre to the eukaryotic nucleus.

These results show that wild-type Cre resides chiefly in the nucleus of transfected cells. Yet, the addition of an exogenous NLS to Cre has been proposed to enhance site-specific DNA recombination in the eukaryotic environment (8). Perhaps this could occur because proteins with multiple NLSs may target the nucleus more efficiently (16). One way to detect Cre activity in eukaryotic cells is to monitor eviction of a *loxP*-flanked DNA fragment from the chromosome. However, such an assay is relatively insensitive to changes in Cre activity as only a single catalytic event is required per cell, and almost all cells actually taking up a Cre-expressing plasmid after transfection are committed to excisive recombination (11,18). In contrast to chromosomal excision, the Cre-mediated integration assay has a much greater dynamic range, showing a clear dose dependence on the amount of Cre available for recombination, and has been useful in clearly establishing that enhanced translatability of Cre mRNA results in markedly increased recombination (24).

We therefore used the *lox-neo* fusion integration strategy (25) shown in Figure 2A to compare recombination between wild-type Cre and NLS–Cre proteins. In this strategy a CHO cell line derivative carrying a single genomic *loxP* target is co-transfected with a large amount of a *loxP* targeting vector and increasing trace amounts of the Cre expression vector. Cre-mediated integration of the targeting vector at the genomic *loxP* site results in reconstruction of a functional *lox-neo* fusion gene so that the generation of $G418^R$ colonies is completely dependent on Cre activity. Figure 2B shows that both wild-type Cre and NLS–Cre under the control of the CMV promoter direct targeted integration in a dose-dependent manner. In particular, the efficiency of site-specific integration by NLS–Cre is approximately the same, or slightly less than, that for wild-type Cre. In additional experiments wild-type Cre

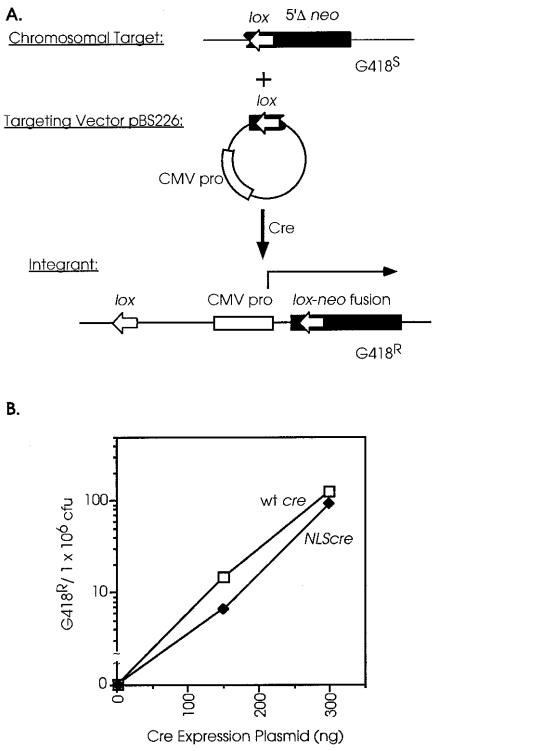

Figure 2. Assay of Cre-mediated site-specific DNA targeting in transfected cells. (**A**) Cre-mediated integration of pBS226 into the defective *lox-neo* chromosomal target in 14-1-2 cells results in restoration of a functional *neo* gene and confers resistance to G418. (**B**) The frequency of G418 resistant colonies per 1×10^6 colony forming units (cfu) was scored 10 days after co-transfection of 14-1-2 cells with 15 μg pBS226 and the indicated amount of either pBS185 (wt *cre*) or pBS391 (*NLScre*).

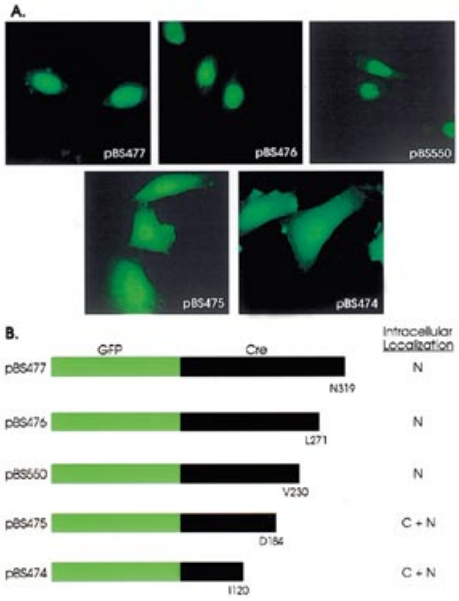

Figure 3. Nuclear localization of GFP–Cre C-terminal truncation proteins. (**A**) Representative GFP fluorescence observed in CHO-K1 cells 48 h after transfection with the indicated GFP–Cre truncation construct. (**B**) Map of GFP–Cre truncation constructs. The GFP portion is depicted in green, and the Cre portion in black, terminating at the Cre amino acid position shown. The indicated pattern of subcellular fluorescence was observed in >98% of the 100–300 fluorescent transiently transfected cells examined for each construct.

was consistently as efficient or superior to NLS–Cre in promoting targeted integration (data not shown). Thus, the addition of the SV40 NLS to the N-terminus of the Cre protein does not enhance Cre-mediated site-specific integration in CHO cells, a result that is consonant with the observation that Cre carries endogenous nuclear targeting determinants.

Mapping the endogenous Cre nuclear targeting determinant(s)

To facilitate identification of the region(s) of Cre important for nuclear targeting, we exploited the fluorescent and nuclear targeting properties of the 68 kDa GFP–*cre* fusion gene product. A series of C-terminal truncation derivatives of GFP–Cre were constructed, and the subcellular distribution of fluorescence for each truncated protein was determined directly in transfected cells (Fig. 3).

Removal of up to 113 amino acids from the C-terminus of the 343 amino acid Cre portion of the fusion protein resulted in no loss of nuclear targeting (Fig. 3B). However, removal of an additional 46 amino acids (pBS475) gave significantly increased cytoplasmic fluorescence, although nuclear accumulation was still apparent. Further removal of amino acids from the C-terminus to leave only the 123 N-terminal amino acids of Cre fused to GFP (pBS474) resulted in a protein that showed a similar pattern of increased cytoplasmic localization. The D184 truncation (pBS475) appeared to have a slightly more pronounced nuclear location than did the I120 truncation (pBS474) but only in ~40% of the transfected cells, suggesting that any difference in nuclear localization between these two proteins is quite modest (Fig. 3A). The C-terminal truncation results thus implicate the region between D184 and V230 of Cre as necessary for nuclear targeting.

The D184–V230 region is not sufficient for nuclear localization. Removal of the N-terminal 119 amino acids of Cre from the L271 C-terminal truncation (pBS486) resulted in enhanced cytoplasmic fluorescence reminiscent of the D184 C-terminal truncation (Fig. 4A). The point of fusion between GFP and Cre in pBS486 lies at I120 and is within a cluster of basic amino acids of which four of five are positively charged (Fig. 4B). Since similar clusters of basic amino acid residues have been shown to be important for nuclear targeting in a variety of eukaryotic proteins (16), the basic cluster was reconstructed to generate pBS549. Still, restoration of this arginine rich region was insufficient to restore nuclear targeting (Fig. 4A and B). Adding back an additional 18 amino acids to include a second basic cluster at R100 (pBS530), however, did restore nuclear targeting. The included region from R100 through R122 bears a resemblance to the bipartite NLS sequences of eukaryotic proteins that are comprised of two clusters of basic amino acids separated by 10–20 residues (16). To determine whether or not this bipartite basic region of Cre was in fact sufficient to act as an NLS, the region from R100 through E123 of Cre was fused to GFP (pBS498) and cells were transfected with this fusion (Fig. 4A). The presence of diffuse fluorescence throughout the cell indicated this bipartite basic region of Cre is insufficient for nuclear targeting. These results are consistent with results obtained with the C-terminal truncations. Thus, two distinct regions of Cre are necessary and are together, but not separately, sufficient for nuclear targeting. In particular, construct

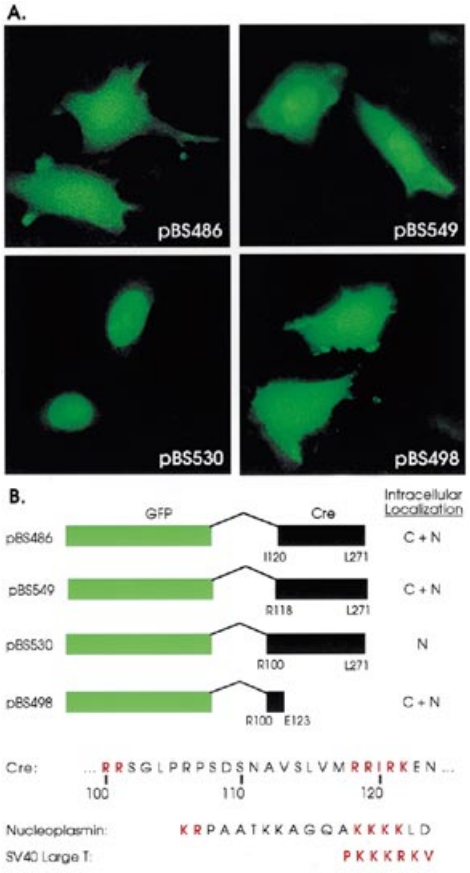

Figure 4. N-terminal mapping of Cre nuclear targeting determinants. (**A**) Representative GFP fluorescence observed in CHO-K1 cells 48 h after transfection with the indicated GFP–Cre N-terminal deletion construct. (**B**) Map of GFP–Cre N-terminal deletion constructs. GFP (green) is fused to Cre (black) at the indicated Cre amino acid position. The amino acid sequence of Cre between R100 and N124 is shown with the bipartite NLS-like sequence highlighted in red. Subcellular fluorescence was scored as described in Figure 3.

pBS530 from which both the N-terminal and the C-terminal DNA-binding regions of Cre (34) have been removed is unimpaired for nuclear targeting of the GFP fusion protein. Combined with the previous truncation data, this indicates that the 131 amino acid region from R100 to V230 of Cre carries determinants that allow Cre to efficiently enter the eukaryotic nucleus.

To further assess the importance of the nucleoplasmin-like bipartite basic region for nuclear targeting, each basic residue of the second basic region was systematically changed one at a time to alanine in the intact GFP–Cre fusion, and the subcellular location of the fusion protein was determined (Table 1). In each case, alteration of one of the positively charged residues resulted in a pronounced relaxation in nuclear targeting, not discernibly different from that observed for the truncation and deletion mutants. These results show that each basic residue in the basic cluster contributes to nuclear localization. Because it has been proposed that NLS sequences often overlap DNA- or RNA-binding domains of proteins (35), we also noted which of these residues appears to make a contact with DNA, based on the recent crystal structure of Cre (36). Three of the four basic residues of the second basic cluster of amino acids (RRIRK) in the bipartite NLS-like motif of Cre, in fact, contact DNA. Thus, there is an overlap of DNA contacts and necessary amino acid residues for nuclear targeting in this cluster.

Table 1. Single amino acid mutational analysis of the basic cluster of Cre

GFP–*cre* fusion	DNA contact removed[a]	Subcellular localization[b]
Wild-type (pBS448)	No	N
R118A	Yes	N + C
R119A	No	N + C
R121A	Yes	N + C
K122A	Yes	N + C

[a]From the X-ray crystal structure of Cre (36).
[b]Determined as described in Figure 3.

Nuclear import of GFP–Cre *in vitro*

To directly address whether or not Cre enters the mammalian nucleus by a mechanism other than passive diffusion, we assayed nuclear targeting using a permeabilized HeLa cell nuclear import assay. Under standard nuclear import conditions we observed that a 98 kDa GST–GFP–Cre fusion protein was rapidly (<30 min) and efficiently targeted to the nucleus (Fig. 5A). As expected for active transport, nuclear import was completely dependent on ATP (Fig. 5B), and was blocked by incubation at 4°C (Fig. 5C). Both of these results are inconsistent with the notion that Cre can only enter the nucleus by a passive diffusion mechanism. To confirm that nuclear entry by the GST–GFP–Cre fusion protein was proceeding by passage through the nuclear pore we also showed that nuclear entry (Fig. 5D) was blocked by incubation with the nuclear import inhibitor wheat germ agglutinin (37). Because the two regions of Cre which we have identified to be necessary for nuclear targeting may comprise a three-dimensional determinant important for targeting, we briefly heat-treated the GST–GFP–Cre fusion protein, a procedure that abolishes Cre activity (38), and investigated whether the fusion protein could still enter the nucleus. Although fluorescence was not markedly diminished by heat treatment, the fusion protein was no longer capable of nuclear import (Fig. 5E). This result, together with the observation that the bipartite basic region of Cre by itself is insufficient for nuclear localization, further supports the hypothesis that nuclear targeting of Cre involves an additional feature of region II in conjunction with the bipartite basic region of Cre.

DISCUSSION

We demonstrate here that Cre recombinase, although of prokaryotic origin, contains determinants that target the nucleus of eukaryotic cells. These determinants are sufficient to direct the nuclear localization of a 68 kDa GFP–Cre fusion protein. Addition of the SV40 T-Ag NLS to Cre does not enhance nuclear accumulation as measured both by indirect immunofluorescence and by a sensitive functional integration-based assay in CHO cells. Using an *in vitro* nuclear import assay we show that Cre rapidly enters the nucleus in an energy-dependent manner through the nuclear pore. This rapid nuclear entry by wild-type Cre protein helps explain why the addition

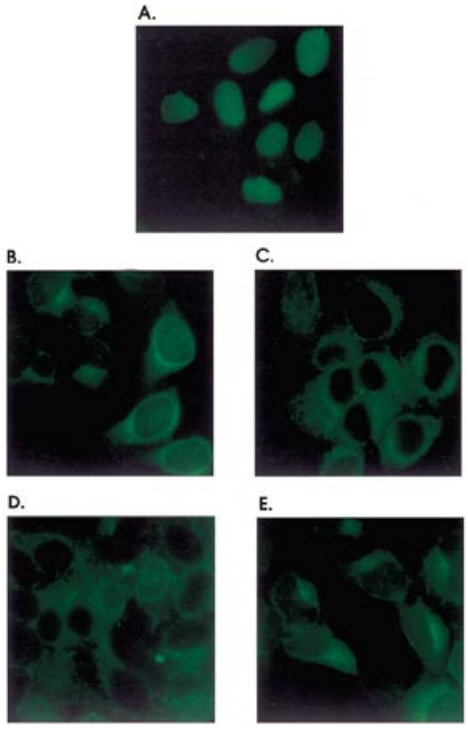

Figure 5. Nuclear transport *in vitro*. Nuclear import of the GST–GFP–Cre fusion protein in digitonin-permeabilized HeLa cells was assayed (**A**) using the complete transport system at 37°C, (**B**) in the absence of creatine kinase, creatine phosphate or ATP, but in the presence of 0.4 U/μl hexokinase and 6 mM glucose, (**C**) at 4°C, (**D**) in the presence of 0.4 mg/ml wheat germ agglutinin, (**E**) with heat-treated GST–GFP–Cre (5 min, 95°C).

of an exogenous NLS to Cre appears to be unnecessary for efficient targeting in eukaryotic cells. Cre is thus one of only a very few prokaryotic proteins that have been shown to possess the ability to enter the eukaryotic nucleus by a process other than simple passive diffusion.

Two distinct regions are necessary for nuclear targeting, one lying between R100 and D184 (Region I), and the other between D184 and V230 (Region II). Region I includes a motif (two clusters of basic amino acids separated by 16 residues) that is reminiscent of the bipartite NLS of nucleoplasmin. Unlike the bipartite NLS of nucleoplasmin, the bipartite motif of region I is necessary but not sufficient for nuclear targeting. It differs from the classical bipartite NLS motif of nucleoplasmin in two ways: (i) the basic clusters are predominantly arginine rather than lysine residues, and (ii) the spacer between the two clusters of basic amino acids is 16 amino acids instead of the 10 found in nucleoplasmin. It is unlikely that spacing by itself is important since increased spacing in the bipartite NLS of nucleoplasmin and other nuclear proteins does not abolish nuclear targeting (16). Moreover, the atypical NLSs of the integrases of the yeast Ty1 retrotransposon (39,40) and of the HIV-1 lentivirus (41) resemble the nucleoplasmin bipartite NLS, but have increased spacings of 22 and 29 amino acids, respectively, between the basic clusters. Whether arginine-rich bipartite sequences, such as the one in Cre, behave significantly differently from the more common lysine-biased ones is unknown.

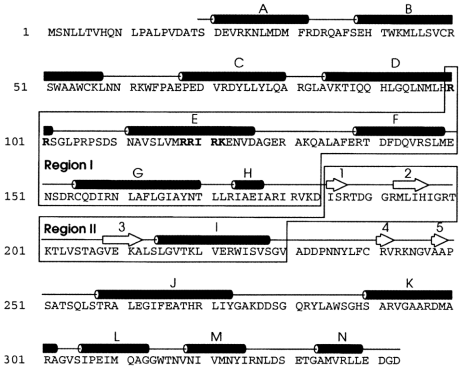

Figure 6. Alignment of nuclear targeting regions with Cre secondary structure. α-Helices are labeled A–N and β-strands are labeled 1–5 (36). Regions I and II are boxed and the bipartite NLS-like motif is indicated in bold type.

Although Cre is a DNA binding protein, it is unlikely that DNA binding is responsible for nuclear targeting. Each Cre monomer makes about 42 contacts with DNA distributed over the entire length of Cre (36), and most of these are with basic residues. The 131 amino acid region identified here that carries nuclear targeting activity includes only 14 of these contact residues. Moreover, it excludes both the N- and C-terminal regions of Cre that had previously been identified genetically as being important DNA binding regions (34). Since mutation of a single non-DNA-contacting arginine residue in the second basic region severely compromises nuclear targeting of the full-length protein, whereas removal of up to two-thirds of all DNA contacts has no effect, it is improbable that DNA binding plays a critical role in nuclear targeting by Cre.

The positions of Region I and Region II of Cre in the three-dimensional structure of Cre are intriguing. The co-crystal structure of Cre complexed with DNA (36) shows that Cre interacts with DNA much like a C-clamp. One arm of the clamp carries Region I with the N-terminal bipartite motif lying on two different α-helices (D + E) of an antiparallel three helix (C + D + E) bundle (Fig. 6). Region II, however, is located far away on the other arm of the C-clamp across the DNA helix and includes a small β-sheet (strands 1–3) along with helix I. Unlike Region I and classical NLS sequences, Region II does not exhibit basic amino acid-rich clusters. Interestingly, in the DNA synapsed tetrameric structure the C-terminus of helix E makes an intermolecular contact with the β-turn between strands 2 and 3. Since Cre is a monomer in solution (42), it will be enlightening to determine whether Cre makes a similar, but intramolecular, contact in the absence of DNA. Such a contact would help to explain the dual requirement for both Region I and Region II for nuclear targeting. Interaction may, for example, help form or stabilize an NLS-like surface that permits nuclear import (43–45), especially since in the absence of DNA the basic residues present in Region I should now be accessible by the nuclear import apparatus.

How is it that a prokaryotic protein like Cre would contain determinants that efficiently direct it to the eukaryotic nucleus? An initial impression is that the presence of nuclear targeting determinants within Cre may simply be fortuitous. Because Cre is a rather basic protein with a high incidence of arginine

and lysine residues (46), by chance there could occur a happy assemblage of basic residues that has nuclear targeting activity. The tendency for NLS sequences to overlap DNA-binding regions in proteins (and which is clearly the case here for the bipartite region in Cre) would provide a mechanism for maintaining the presence of the basic region (35). Although we have not determined the contribution of each of the amino acid residues in this region of Cre directly to DNA binding, we have determined that the integrity of this region is important for DNA recombinase activity (unpublished results). A speculative alternative possibility, though, is that Cre's nuclear targeting ability is the vestige of an ancient functionality, and arose from horizontal gene transfer. Cre is a member of the Int family of recombinases (47), a divergent group of proteins all of which are quite basic and most of which, like Cre, reside on mobile genetic elements. Some members of this family, like the FLP recombinase (48) of *S.cerevisiae*, are in fact eukaryotic nuclear proteins. Even so, these eukaryotic recombinases also fail to display an NLS of either the SV40 or bipartite nucleoplasmin type (16). Given that there is an abundance of horizontal gene transfer in the prokaryotic world, and that inter-kingdom genetic transfer is a way of life for *Agrobacterium* and clearly possible even for *Escherichia coli* (49), Cre or its ancestor may have been associated with eukaryotic cells long before the current era of molecular biologists.

ACKNOWLEDGEMENTS

We are grateful to A. Rüfer for his perceptive advice, to J. Hanover and M. Miller for useful discussions, to T. Sweitzer for guidance in setting up the *in vitro* system, to E. J. Cooper and J. Conaway for help with the Metamorph Imaging System and to M. Flynn for preparation of the manuscript.

REFERENCES

1. Gordon,J.W., Harold,G. and Leila,Y. (1993) *Hum. Cell*, **6**, 161–169.
2. Sternberg,N. (1978) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 1143–1146.
3. Sauer,B. (1987) *Mol. Cell. Biol.*, **7**, 2087–2096.
4. Sauer,B. and Henderson,N. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 5166–5170.
5. Odell,J., Caimi,P., Sauer,B. and Russell,S. (1990) *Mol. Gen. Genet.*, **223**, 369–378.
6. Lakso,M., Sauer,B., Mosinger,J.B., Lee,E.J., Manning,R.W., Yu,S.-H., Mulder,K.L. and Westphal,H. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 6232–6236.
7. Gu,H., Marth,J.D., Orban,P.C., Mossmann,H. and Rajewsky,K. (1994) *Science*, **265**, 103–106.
8. Gu,H., Zou,Y.R. and Rajewsky,K. (1993) *Cell*, **73**, 1155–1164.
9. Lakso,M., Pichel,J.G., Gorman,J.R., Sauer,B., Okamoto,Y., Lee,E., Alt,F.W. and Westphal,H. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 5860–5865.
10. Orban,P.C., Chui,D. and Marth,J.D. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 6861–6865.
11. Sauer,B. (1993) *Methods Enzymol.*, **225**, 890–900.
12. Metzger,D., Clifford,J., Chiba,H. and Chambon,P. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 6991–6995.
13. Golic,K.G. and Lindquist,S. (1989) *Cell*, **59**, 499–509.
14. Onouchi,H., Yokoi,K., Machida,C., Matsuzaki,H., Oshima,Y., Matsuoka,K., Nakamura,K. and Machida,Y. (1991) *Nucleic Acids Res.*, **19**, 6373–6378.
15. Peters,R. (1983) *J. Biol. Chem.*, **258**, 11427–11429.
16. Dingwall,C. and Laskey,R.A. (1991) *Trends Biochem. Sci.*, **16**, 478–481.
17. Breeuwer,M. and Goldfarb,D.S. (1990) *Cell*, **60**, 999–1008.
18. Gagneten,S., Le,Y., Miller,J. and Sauer,B. (1997) *Nucleic Acids Res.*, **25**, 3326–3331.
19. Wang,Y., Krushel,L.A. and Edelman,G.M. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 3932–3936.
20. Lee,Y.-H., Sauer,B., Johnson,P.F. and Gonzalez,F.J. (1997) *Mol. Cell. Biol.*, **17**, 6014–6022.
21. Yakar,S., Liu,J.-L., Stannard,B., Butler,A., Accili,D., Sauer,B. and LeRoith,D. (1999) *Proc. Natl Acad. Sci. USA*, **96**, 7324–7329.
22. Lartey,R. and Citovsky,V. (1997) *Genet. Engin.*, **19**, 201–214.
23. Sauer,B. and Henderson,N. (1989) *Nucleic Acids Res.*, **17**, 147–161.
24. Sauer,B. and Henderson,N. (1990) *New Biologist*, **2**, 441–449.
25. Fukushige,S. and Sauer,B. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 7905–7909.
26. Le,Y., Miller,J.L. and Sauer,B. (1999) *Anal. Biochem.*, **270**, 334–336.
27. Chen,C. and Okayama,H. (1987) *Mol. Cell. Biol.*, **7**, 2745–2752.
28. Adam,S.A., Marr,R.S. and Gerace,L. (1992) *Methods Enzymol.*, **219**, 97–110.
29. Sweitzer,T. and Hanover,J.A. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 14574–14579.
30. Bethke,B.D. and Sauer,B. (1999) In Kmiec,E. (ed.), *Methods in Molecular Biology: Strategies for Gene Targeting*. Humana Press, Totowa, NJ, in press.
31. van den Hoff,M.J.B., Moorman,A.F.M. and Lamars,W.H. (1992) *Nucleic Acids Res.*, **20**, 2902.
32. Prasher,D.C., Eckenrode,V.K., Ward,W.W., Prendergrast,F.G. and Cormier,M.J. (1992) *Gene*, **111**, 229–233.
33. Kubai,D.F. (1975) *Int. Rev. Cytol.*, **43**, 167–227.
34. Wierzbicki,A., Kendall,M., Abremski,K. and Hoess,R. (1987) *J. Mol. Biol.*, **195**, 785–794.
35. LaCasse,E.C. and Lefebvre,Y.A. (1995) *Nucleic Acids Res.*, **23**, 1647–1656.
36. Guo,F., Gopaul,D.N. and Van Duyne,G.D. (1997) *Nature*, **389**, 40–46.
37. Finlay,D.R., Newmeyer,D.D., Price,T.M. and Forbes,D.J. (1987) *J. Cell Biol.*, **104**, 189–200.
38. Abremski,K., Hoess,R. and Sternberg,N. (1983) *Cell*, **32**, 1301–1311.
39. Moore,S.P., Rinckel,L.A. and Garfinkel,D.J. (1998) *Mol. Cell. Biol.*, **18**, 1105–1114.
40. Kenna,M.A., Brachmann,C.B., Devine,S.E. and Boeke,J.D. (1998) *Mol. Cell. Biol.*, **18**, 1115–1124.
41. Gallay,P., Hope,T., Chin,D. and Trono,D. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 9825–9830.
42. Abremski,K. and Hoess,R. (1984) *J. Biol. Chem.*, **259**, 1509–1514.
43. Görlich,D. and Mattaj,I.W. (1996) *Science*, **271**, 1513–1518.
44. Nigg,E.A. (1997) *Nature*, **386**, 779–787.
45. Pemberton,L.F., Blobel,G. and Rosenblum,J.S. (1998) *Curr. Opin. Cell Biol.*, **10**, 392–399.
46. Sternberg,N., Sauer,B., Hoess,R. and Abremski,K. (1986) *J. Mol. Biol.*, **187**, 197–212.
47. Argos,P., Landy,A., Abremski,K., Egan,J.B., Ljungquist,E.H., Hoess,R.H., Kahn,M.L., Kalionis,B., Narayana,S.V.L., Pierson,L.S., Sternberg,N. and Leong,J.M. (1986) *EMBO J.*, **5**, 433–440.
48. Broach,J.R., Guarascio,V.R. and Jayaram,M. (1982) *Cell*, **29**, 227–234.
49. Sprague,G.F.,Jr (1991) *Curr. Opin. Genet. Dev. Dev.*, **1**, 530–533.