The Nucleoplasmin Nuclear Location Sequence Is Larger and More Complex than That of SV-40 Large T Antigen

Colin Dingwall, Jacqueline Robbins, Stephen M. Dilworth, Bruce Roberts,* and William D. Richardson‡

Cancer Research Campaign, Molecular Embryology Research Group, Zoology Department, Cambridge CB2 3EJ, England; *Integrated Genetics, Framingham, Massachusetts 01701; and ‡Department of Biology, Medawar Building, University College London, London WC1E 6BT, England

Abstract. The carboxy-terminal tail of nucleoplasmin, which specifies entry into the cell nucleus, contains four short sequences that are similar to previously identified nuclear location sequences. We show that none of these is able to locate chicken muscle pyruvate kinase to the cell nucleus. Deletion analysis was used to determine the limits of a nuclear location sequence and indicated that a 14-amino acid segment (RPAATKKAGQAKKK) should function as a minimal nuclear location sequence. When tested directly, however, this sequence was unable to locate pyruvate kinase to the cell nucleus. Restoration of three amino acids of nucleoplasmin sequence at either end of this sequence generated sequences that were able to locate pyruvate kinase to the cell nucleus.

The 14-amino acid proposed minimal nuclear location sequence is present in the functional sequences, AVKRPAATKKAGQAKKK, RPAATKKAGQAKKKKLD, and the sequence AVKRPAATKKAGQAKKKKLD, which has additional amino acids at both ends. The minimal sequence element is therefore necessary but not sufficient for transport into the cell nucleus. This unusual feature of the nucleoplasmin nuclear location sequence suggests ways in which it could interact with the nuclear transport mechanism.

In addition to its established role in chromatin assembly and histone storage in the *Xenopus* oocyte and egg (4, 16, 18, 21), nucleoplasmin has emerged as the protein of choice for studies of protein transport into the cell nucleus (reviewed in reference 5). Investigations using this acidic, thermostable, pentameric protein provided the decisive evidence that a mechanism of uptake involving selective entry into the nucleus operates in oocytes (8).

Nucleoplasmin has also been used to demonstrate that entry into the *Xenopus* oocyte nucleus occurs through the nuclear pore complex (9), to evaluate the transport properties of nuclei reconstituted in vitro, and to show that ATP is required for the transport process (28).

Primary amino acid sequences specifying entry into the nucleus have been identified in a number of proteins (5, 23, 25, 30). The sequence that has been studied in greatest molecular detail is the SV-40 large T antigen nuclear location sequence (3, 11, 14, 15, 19, 20, 33, 37). The earlier studies in this series demonstrated that the amino acid sequence PKK^{128}KRKV is able to translocate chicken muscle pyruvate kinase to the cell nucleus (15).

The chicken muscle pyruvate kinase expression vector system has also been used to map the polyoma virus large T nuclear location sequences (32) and more recently to investigate the effect of protein context upon the function of the SV-40 large T antigen nuclear location sequence (33). Therefore we chose this system to identify the nuclear location sequence or sequences in the nucleoplasmin "tail" region since a large body of data exists against which we can interpret our results.

We have sequenced nucleoplasmin and shown that the tail region which is both necessary and sufficient for selective entry of nucleoplasmin into the nucleus corresponds to the carboxy-terminal 50 amino acids (6, 7). In addition we have pointed out the presence of four sequences in the tail region which show significant degrees of similarity to other nuclear location sequences (6, 7).

Recently Burglin and De Robertis (1) have shown that one-half of the tail region of nucleoplasmin can translocate *Escherichia coli* β-galactosidase to the *Xenopus* oocyte nucleus. This region includes a sequence which bears a striking similarity to the SV-40 large T antigen nuclear location sequence, differing only by two amino acid substitutions. Each of these substitutions alone still allows the SV-40 large T antigen sequence to locate the protein to the cell nucleus (37).

Here we describe the construction of recombinant plasmids in which portions of the nucleoplasmin tail region are linked to the chicken muscle pyruvate kinase cDNA. We confirm that the entire nucleoplasmin tail region can translocate a cytoplasmic protein to the cell nucleus (in agreement

C. Dingwall's present address is Medical Research Council, Laboratory of Molecular Biology, Cambridge CB2 2QH, England.

with reference 1) and that the sequence that is responsible for this activity lies in the amino-terminal half of the tail. We find, however that the sequence in this half of the tail region which most strongly resembles the SV-40 nuclear location sequence is unable to locate chicken muscle pyruvate kinase to the cell nucleus when fused to the carboxy or amino terminus of the protein. In fact none of the four sequences similar to other nuclear location sequences that we have observed and discussed previously (1, 6, 7) is capable of translocating pyruvate kinase to the cell nucleus.

Deletion analysis reveals that a single sequence in the nucleoplasmin tail region is capable of translocating pyruvate kinase to the nucleus. However this sequence is longer and more complex than previously identified nuclear location sequences. The unusual properties of this sequence suggest a model for the interaction of the nuclear location sequence with the putative receptor in the transport mechanism.

Materials and Methods

Microinjection

Vero (African Green Monkey) cells were grown on glass coverslips and microinjected using glass capillaries as described previously (12, 31). Plasmid DNAs were injected at a concentration of 100 μg/ml into the nucleus delivering ~1,000 molecules/cell. Proteins at a concentration of 1 mg/ml were injected into the nucleus or cytoplasm of the cell delivering on the order of 10^6 molecules/cell.

Immunofluorescence Microscopy

Nucleoplasmin/pyruvate kinase fusion proteins were detected with rabbit polyclonal antiserum raised against purified chicken muscle type M1 pyruvate kinase (22) followed by FITC-conjugated goat anti-rabbit IgG (Nordic Immunological Labs. Ltd., Maidenhead, UK) or with a mixture of monoclonal antibodies to the tail region of nucleoplasmin (7) followed by FITC-conjugated rabbit anti-mouse IgG.

Proteins

Nucleoplasmin, the trypsin resistant core of nucleoplasmin and the 12-kD tail fragment of nucleoplasmin were all prepared using previously published procedures (8).

Plasmid Construction

Vector Fragments. The materials and methods for plasmid construction are identical to those described previously (15, 32, 33).

Nucleoplasmin Fragments. An M13 clone containing a full-length nucleoplasmin cDNA (7) with Eco R1 linkers was used as the start material for the production of nucleoplasmin DNA fragments. All procedures used were conventional techniques described in detail elsewhere (24).

pPK-NP1. A fragment encompassing the tail region (amino acids 149–200) was prepared by digestion of 1 μg of the M13 clone (RF) with Hinf I, repaired with DNA polymerase (Klenow) and addition of an Xho I linker. The required Xho–Pk fragment was purified from an acrylamide gel and ligated to the appropriate vector fragment (see below) to produce plasmid pPK-NP1. The fragment was also ligated into a pUC 13 vector in which an Xho linker had been inserted into the Hinc II site in the polylinker (a gift from T. J. Mohun, Zoology, Cambridge, UK) to produce pUC-NP1.

pPK-NP2. 1 μg of the pUC13 subclone described above (pUC13-NP1) was digested with Sac I, repaired with the T4 DNA polymerase and an Xho I linker added. The required Xho–Eco RI fragment encoding amino acids 178–200 was purified from an acrylamide gel and ligated to the appropriate vector fragments.

pPK-NP3. 1 μg of pUC13-NP1 was digested with Xho I and Sac I and the DNA fragment encoding amino acids 149–176 of nucleoplasmin was purified from a polyacrylamide gel and ligated into suitably cleaved pUC 13. The subclone produced was digested with Sac I, and repaired with T4 DNA polymerase. Universal translation terminator linkers (GCTTAATTAATTAAGC; Pharmacia, Inc., Piscataway, NJ) were added to the repaired DNA. The linear plasmid DNA was purified from an agarose gel and then ligated to form circular molecules. A recombinant plasmid (pUC 13-NP3) was identified which had the required nucleoplasmin fragment and a universal terminator linker. After digestion with Xho I and Eco RI this fragment was purified from an acrylamide gel and ligated to the appropriate vector fragments to produce pPK-NP3.

pPK-NP4. 1 μg of the plasmid pUC13-NP3 was digested with Bst NI, repaired with DNA polymerase large fragment, and Xho linkers were added. The DNA was digested with Xho I and Eco RI and the DNA fragment encoding amino acids 165–176 of nucleoplasmin was purified from an acrylamide gel and ligated to the appropriate vector fragments to produce pPK-NP4.

pNP4-PK. An Xho I–Eco RI fragment encoding amino acids 165–176 was ligated in a three fragment ligation with a Bam HI–Xho I fragment from XSRL 40 and an Eco RI–Bam HI fragment from RL18 PK12 to link the nucleoplasmin sequence at the amino terminus of pyruvate kinase.

pPK-NP5. 1 μg of the plasmid pUC13-NP3 was digested with Bst NI, repaired with the DNA polymerase large fragment, and a universal translation terminator linker and an Eco RI linker were added to the repaired ends. The resulting Xho I–Eco RI fragment encoding amino acids 149–165 was purified from an acrylamide gel and ligated to the appropriate vector fragment to produce pPK-NP5.

pPK-NP6. 1 μg of the plasmid pUC13-NP1 was digested with Bst NI, repaired with DNA polymerase large fragment, and an Xho I linker was added to the repaired ends. The DNA preparation was then digested with Eco RI and the appropriate Xho I–Eco RI fragment encoding amino acids 165–200 was purified from an acrylamide gel and ligated to the appropriate vector fragment to produce pPK-NP6.

Bal 31 Deletion. 10 μg of the plasmid pUC13-NP3 was linearized with Eco RI or Xho I. The linear DNA molecule was digested with Bal 31 nuclease (New England Biolabs, Beverly, MA) at 30°C for 8 min. Appropriate linkers were added to the digested DNA and excess linkers were removed by restriction enzyme digestion. The linear plasmid was purified from an agarose gel, ligated to form circular molecules, and used to transform competent *E. coli* cmk 603. DNA was prepared from colonies of transformed cells and screened by restriction enzyme digestion. Fragments showing a reduction in size were sequenced and suitable deletions were ligated to the pyruvate kinase vector fragment to give the pPK-Δ series of plasmids.

Minimal Sequence Constructs

The Xho I–Eco RI fragments encoding the chosen deletions were subcloned into pUC13 and grown in NM545 (thr$^-$ leuB rpsL dam$^-$ dcm$^-$ [GM 48 Sm^R]). The Xho I–Eco RII and Eco RII–Eco RI fragments from the chosen original deletions were ligated to the appropriate pyruvate kinase vector fragment in a three fragment ligation. An Xho I–Eco RI fragment was excised from the recombinants and sequenced to confirm their identity.

Results

Selective Entry Model for Nuclear Accumulation of Proteins Is Applicable to Mammalian Cells

We wished to determine if the selective entry model for nuclear accumulation established in *Xenopus* oocytes (8) applies to cultured mammalian cells since these cells were used in the analysis of the nuclear location sequence (see below). Therefore, as a prelude to these studies, the transport mechanism for nucleoplasmin was investigated by microinjection of unlabeled protein followed by detection with antibodies.

Nucleoplasmin accumulates rapidly in the cell nucleus after microinjection into the cytoplasm (Fig. 1 *a*). To determine the initial rate of nucleoplasmin uptake, cells were cooled to 20°C, injected with protein, and rapidly fixed (Fig. 1 *b*). This would correspond to ~1 min after the injection of the first cell in a group of cells. A number of cells show protein accumulation within the nucleus while in others the protein is predominantly cytoplasmic. This indicates that under normal conditions the bulk of the injected nucleoplasmin

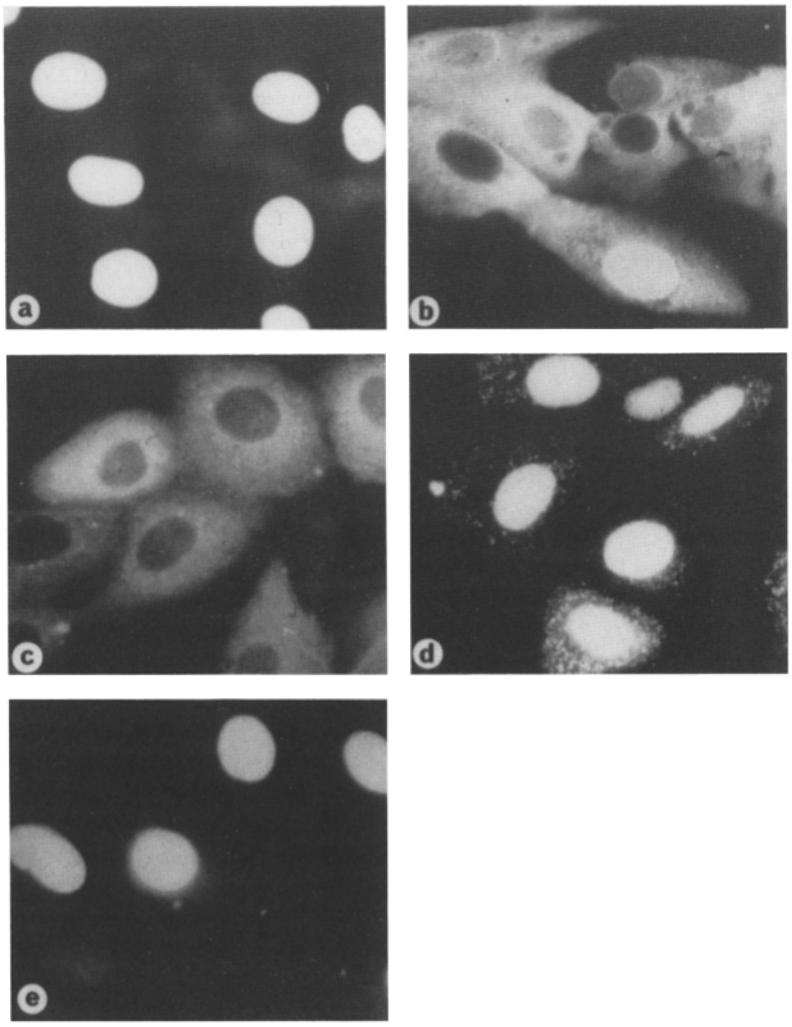


Figure 1. Subcellular distribution of nucleoplasmin, the trypsin-resistant core, and tail regions. Vero cells were microinjected with (*a*) nucleoplasmin into the cytoplasm, fixed, and stained after 1 h at 37°C; (*b*) nucleoplasmin into the cytoplasm, fixed, and stained within 1 min after injection at 20°C; (*c*) trypsin-resistant core of nucleoplasmin injected into the cytoplasm, fixed, and stained 1 h after injection at 37°C; (*d*) the 12,000-M_r tail fragment of nucleoplasmin injected into the cytoplasm, fixed, and stained after 1 h at 37°C; (*e*) the trypsin-resistant core of nucleoplasmin injected directly into the nucleus, fixed, and stained after 1 h at 37°C.

would accumulate in the nucleus in <1 min. In some instances a distinct, sharply defined zone of increased staining is observed at the level of the nuclear envelope (Fig. 1 *b*). We and others have shown elsewhere, that inhibition of ATP generation reversibly blocks intranuclear accumulation but not perinuclear association, and on this basis have proposed a two step mechanism for the accumulation of proteins in the cell nucleus (27, 35). The trypsin-resistant core of nucleoplasmin does not enter the nucleus (Fig. 1 *c*) while the purified tail fragment does (Fig. 1 *d*). In the latter case distinct punctate cytoplasmic fluorescence is observed and contrasts with the absence of cytoplasmic staining in cells injected with intact nucleoplasmin (compare Fig. 1, *a* and *d*). The tail fragment may be unstable in these cells as it is in *Xenopus* oocytes and the fluorescent spots may represent regions of protein turnover.

In the *Xenopus* oocyte the decisive evidence for a selective entry mechanism for nuclear accumulation was the retention within the nucleus of the trypsin-resistant core of nucleoplasmin after direct microinjection into the nucleus (8). When the same experiment was performed in Vero cells the nucleoplasmin core molecule remains stably in the nucleus (Fig. 1 *e*).

These experiments demonstrate that as in the *Xenopus* oocyte nucleoplasmin accumulates extremely rapidly in the cell nucleus by a selective entry mechanism and not by free diffusion and binding within the nucleus.

Identification of a Region Encoding a Nuclear Location Signal (NLS)

As mentioned above the sequence of a full-length nucleoplasmin cDNA has been determined and we have shown that the nucleoplasmin polypeptide is 200 amino acids long and the carboxy-terminal 50 amino acids constitute the tail region that specifies entry into the nucleus (7). In this region there are four sequences which are similar to nuclear location signal sequences that have been identified in other nuclear proteins (6, 7). Two regions show close similarity to the SV-40 large T antigen NLS (37) and two are similar to the consensus sequence identified in the MAT α2 protein of yeast (13).

Fig. 2 *a* shows the location of these putative nuclear location sequences in the tail region and the position of selected restriction enzyme cleavage sites. These sites were used to construct recombinant plasmids (Fig. 2 *b*) in which these se-

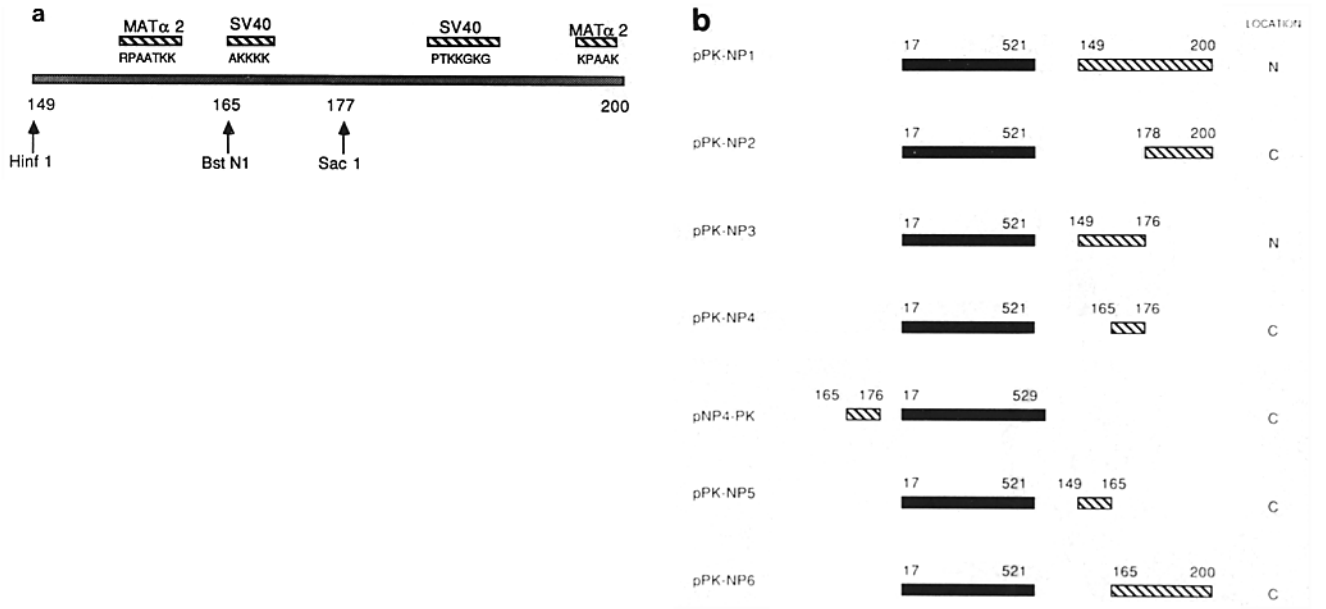


Figure 2. (*a*) Diagramatic representation of the tail region of nucleoplasmin. The bar represents amino acids 149–200 of the nucleoplasmin tail (7). Residue 200 is the carboxy terminus of nucleoplasmin. Protein sequences similar to the MAT α2 consensus or the SV-40 nuclear location sequence are shown above. Restriction enzyme cleavage sites in the DNA are indicated by arrows below. (*b*) Structure and subcellular location of nucleoplasmin/pyruvate kinase fusion proteins. The plasmid vectors listed on the left were made as described in Materials and Methods. The fusion proteins consist of a major part of pyruvate kinase (*solid bars*) to which are linked sequences from the nucleoplasmin tail region (*hatched bars*). The subcellular location of the fusion protein listed on the right was determined by immunofluorescence assay after microinjection of plasmid DNA into Vero cells. *N*, nuclear location; *C*, cytoplasmic location.

quences are linked to the COOH terminus or NH_2 terminus of chicken muscle pyruvate kinase using the plasmid vector system described previously (15, 33). The intracellular distribution of the fusion protein expressed from each plasmid was examined by indirect immunofluorescence with antiserum against pyruvate kinase and where possible with monoclonal antibodies against the tail region of nucleoplasmin.

When the carboxy-terminal 51 amino acids of nucleoplasmin (residues 149–200) that constitute the entire tail region are linked to the carboxyterminus of pyruvate kinase (plasmid pPK-NP1), the fusion protein becomes located in the cell nucleus (Fig. 3). Both an antiserum against pyruvate kinase and a mixture of three monoclonal antibodies specific for the tail region of nucleoplasmin showed strong nuclear fluorescence (Fig. 3, *a* and *a'*). This result indicates that the entire nucleoplasmin tail sequence can locate PK to the cell nucleus.

In pPK-NP2 the carboxy-terminal 23 amino acids of nucleoplasmin (residues 178–200) are linked to the carboxy terminus of PK. The fusion protein expressed from this plasmid was unable to locate to the cell nucleus as shown by the presence of cytoplasmic fluorescence with antiserum to pyruvate kinase (Fig. 3 *b*). The three tail-specific monoclonal antibodies did not give a fluorescent signal with the fusion protein. This lack of reaction and the inability of the fusion protein to locate to the cell nucleus is not due to linkage of the two DNA fragments in the incorrect reading frame as the same result was obtained when the nucleoplasmin fragment was linked to pyruvate kinase in all three reading frames (data not shown) and agrees with the findings of Burglin and De Robertis (1) using β-galactosidase.

Therefore the sequence lying between amino acids 178 and 200 of the nucleoplasmin tail region is unable to locate PK to the cell nucleus even though it contains sequences that resemble previously identified nuclear location sequences (see above). Also the three monoclonal antibodies to nucleoplasmin do not recognise an epitope in the 23 amino acids at the extreme COOH terminus of nucleoplasmin.

In the plasmid pPK-NP3, amino acids 149–176 of nucleoplasmin are linked to the COOH terminus of pyruvate kinase. The fusion protein expressed from this plasmid is located in the cell nucleus as indicated by staining with both the antiserum to pyruvate kinase and the tail-specific anti-nucleoplasmin monoclonal antibodies (Fig. 3, *c* and *c'*). This demonstrates that this region of nucleoplasmin contains a sequence which can locate pyruvate kinase to the cell nucleus. This region has also been shown to locate β-galactosidase to the *Xenopus* oocyte nucleus (1). In this 27-amino acid segment there is a sequence (residues 156–162) that shows a reasonable similarity to the yeast MAT α2 consensus and a sequence (residues 166–172) that is strikingly similar to the SV-40 large T antigen nuclear location sequence. These two sequences encoded within the nucleoplasmin fragment in pPK-NP3 were separately linked to the COOH terminus of pyruvate kinase to give two plasmids, pPK-NP4 and pPK-NP5 (Fig. 2 *b*). Surprisingly the fusion protein expressed from either of these plasmids is located in the cell cytoplasm (Fig. 3, *d* and *d'*). The nucleoplasmin fragment in pPK-NP4 which encodes the sequence with closest similarity to the SV-40 nuclear location sequence is also unable to locate the fusion protein to the cell nucleus when linked to the amino terminus (pNP4-PK) rather than the carboxy terminus (Fig. 2; Fig. 3, *e* and *e'*).

These fusions are directly comparable with those used to

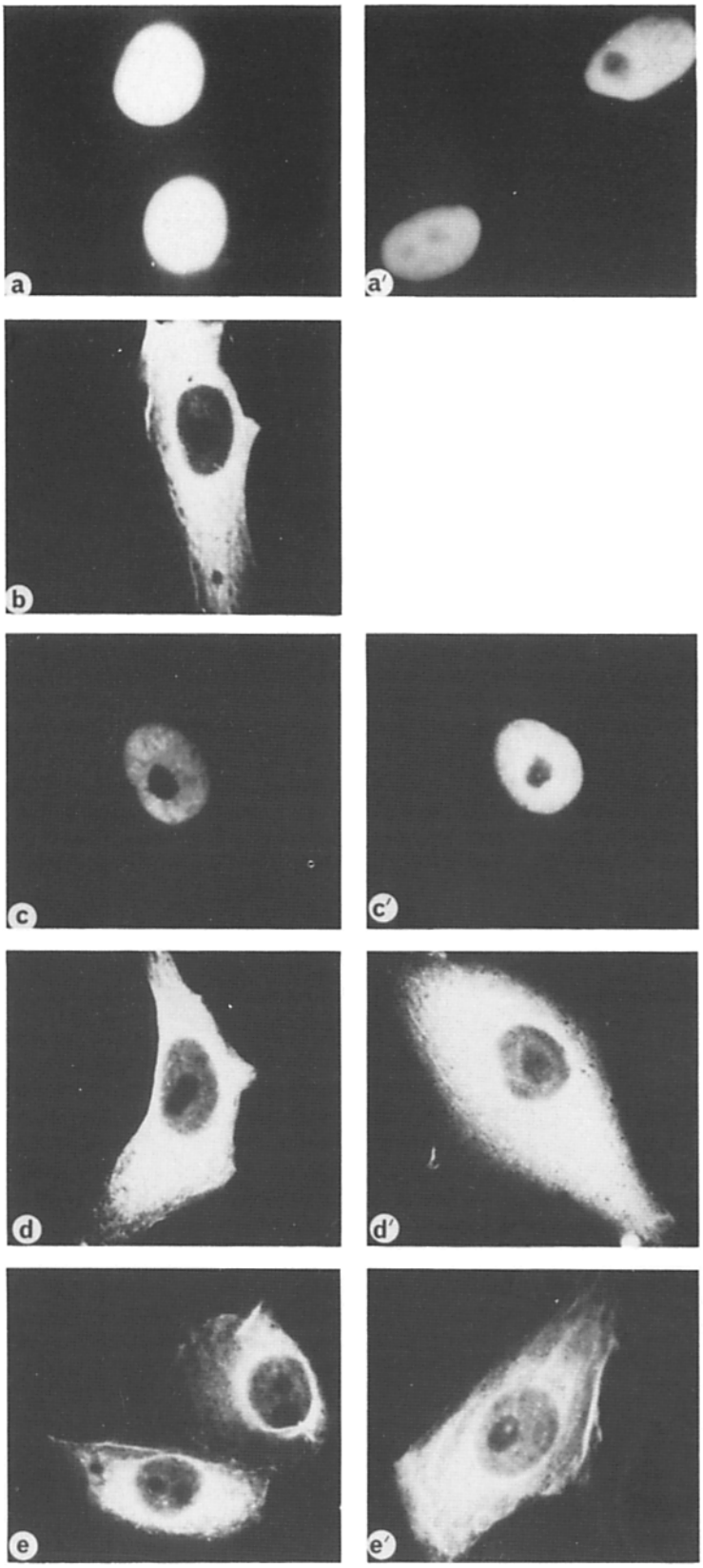


Figure 3. Subcellular locations of pyruvate kinase/nucleoplasmin fusion proteins. Plasmid DNAs encoding the fusion proteins diagrammed in Fig. 2 *b* were injected into the nucleus of Vero cells. Following injection cells were incubated overnight then fixed and stained for pyruvate kinase (*a–e*) or nucleoplasmin (*a'–e'*) as described in Materials and Methods. (*a*, *a'*) pPK-NP1; (*b*) pPK-NP2; (*c*, *c'*) pPK-NP3; (*d*, *d'*) pPK-NP4; (*e*, *e'*) pNP4-PK. No image is possible for b' because no monoclonal antibody is available that recognizes that fragment of nucleoplasmin.

demonstrate the ability of the SV-40 nuclear location sequence to locate pyruvate kinase to the cell nucleus. Therefore we have demonstrated conclusively that the sequence in nucleoplasmin that is closely similar to the SV-40 nuclear location sequence is unable to locate PK to the cell nucleus when it is fused to either the amino or carboxy terminus of the protein.

To establish that the sequence encoded in pPK-NP3 is the

only sequence in the tail region of nucleoplasmin that is capable of translocating pyruvate kinase to the cell nucleus we linked amino acids 165–200 of nucleoplasmin to the carboxy terminus of pyruvate kinase (plasmid pPK-NP6 in Fig. 2 *b*). The fusion protein expressed from this plasmid is located in the cell cytoplasm.

We therefore undertook a detailed analysis of the region of nucleoplasmin (amino acids 149–176) linked to pyruvate kinase in pPK-NP3 in order to determine precisely the sequence responsible for nuclear location.

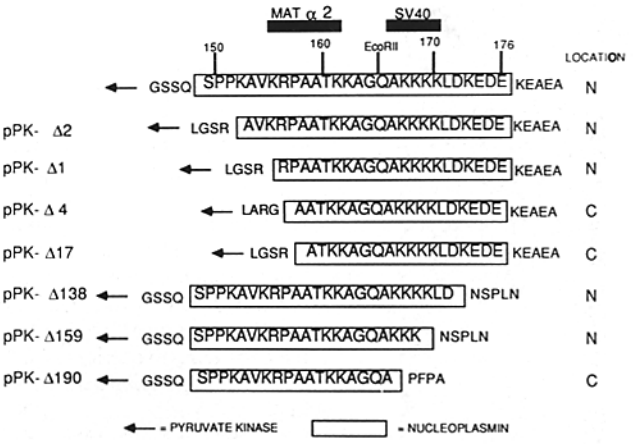
Deletion Analysis of the Nuclear Location Sequence

The DNA fragment which encodes amino acids 149–176 of nucleoplasmin was treated with Bal 31 nuclease (see Materials and Methods) to generate molecules with deletions from either the 5′ (amino) end or the 3′ (carboxyl) end of this fragment. The amino acid sequences of the deletions produced by this procedure and the subcellular location of the corresponding fusion protein are shown in Fig. 4 *top*. The immunofluorescence patterns observed for certain of the deletions are shown in Fig. 4 *bottom*.

Deletions from the Amino Terminus. Analysis of deletions from the amino terminus (amino acid 149) indicate a transition of the location of the fusion protein from nuclear to cytoplasmic when the deletion end point lies in the region of amino acid 157 (proline).

In PK-Δ1 the fusion protein is nuclear (Fig. 4 *bottom*, *a*). The end point of this deletion lies at arginine residue 156. However the lysine residue 155 present in the original fragment is replaced in the deletion by an arginine residue encoded by the linker DNA. Therefore the presence of two basic residues at this location is maintained; lys-arg-pro becomes arg-arg-pro.

In pPK-Δ4 the fusion protein is cytoplasmic (Fig. 4 *bottom*, *b*). The end point of this deletion lies at alanine residue 158. In this case the nucleoplasmin sequence lys-arg-pro is



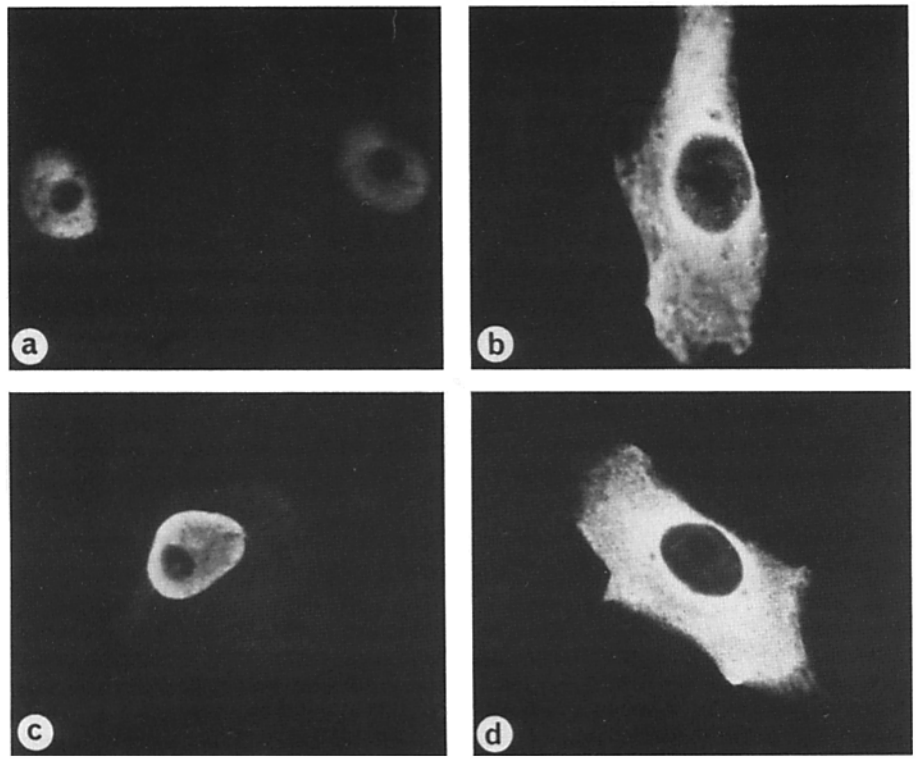


Figure 4. (*Top*) Deletion analysis of the minimal nuclear location sequence. The plasmid vectors listed on the left were made as described in Materials and Methods. Nucleoplasmin sequences (boxed) were linked to the COOH terminus of pyruvate kinase. Amino acid sequences encoded by linker DNA are shown outside the boxed sequences. Numbers indicate amino acids in the nucleoplasmin sequence. Subcellular location of the fusion proteins (*N*, nuclear; *C*, cytoplasmic) was determined by immunofluorescence assay after microinjection of plasmid DNAs into Vero cells. (*Bottom*) Subcellular location of pyruvate kinase/nucleoplasmin fusion proteins. Plasmid DNAs encoding the fusion proteins shown in the top were microinjected into the cell nucleus of Vero cells. After overnight incubation the cells were fixed and stained for pyruvate kinase as described in Materials and Methods. (*a*) pPK-Δ1; (*b*) pPK-Δ4; (*c*) pPK-Δ159; (*d*) pPK-Δ190.

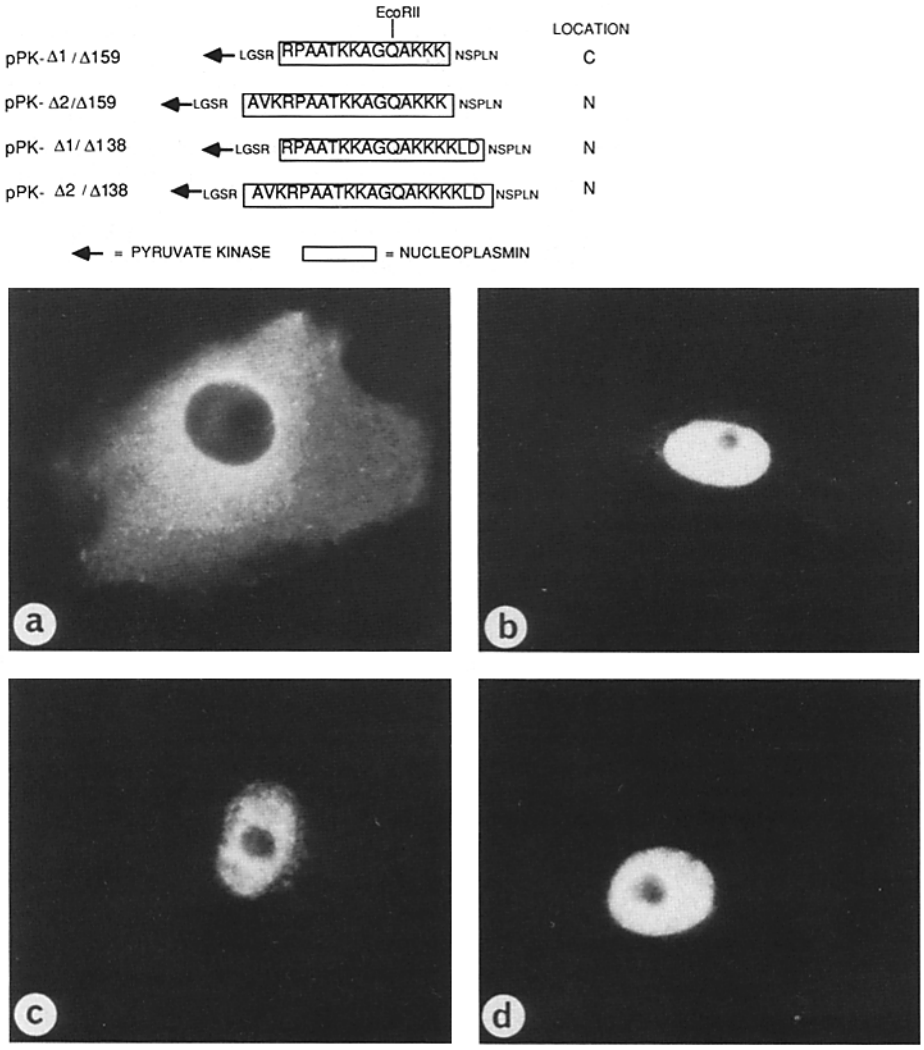


Figure 5. (*Top*) Minimal sequence for nuclear location of pyruvate kinase fusion proteins. The plasmid vectors listed on the left were made as described in Materials and Methods. Nucleoplasmin sequences (boxed) were linked to the carboxy terminus of pyruvate kinase. Amino acid sequences encoded by the linker DNA are shown outside the boxed sequences. (*Bottom*) Subcellular location of pyruvate kinase/nucleoplasmin fusion proteins carrying the proposed minimal nuclear location sequences. Plasmid DNA encoding the fusion protein shown in the top were microinjected into the nucleus of Vero cells. After overnight incubation the cells were fixed and stained for pyruvate kinase as described in Materials and Methods. (*a*) pPK-Δ1/Δ159; (*b*) pPK-Δ2/Δ159; (*c*) pPK-Δ1/Δ138; (*d*) pPK-Δ2/Δ138.

replaced by ala-arg-gly encoded by the linker DNA. Therefore this deletion effectively generates two amino acid changes, lysine[155] is replaced by an alanine residue and proline[157] is replaced by a glycine residue.

Therefore the difference between pPK-Δ1 (nuclear) and pPK-Δ4 (cytoplasmic) is the simultaneous replacement of a basic amino acid (lysine) by a neutral one (alanine) and of a bend-forming amino acid (proline) by glycine.

Deletion from the Carboxy Terminus. Deletion of amino acids from the carboxy terminus (amino acid 176) reveals that only part of the lysine tract (residues 167–170) is necessary for location to the cell nucleus. The fusion protein expressed from pPK-Δ159 bears amino acids 149–169 of nucleoplasmin (Fig. 4 *top*) and three of the four contiguous lysine residues remain. This fusion protein is located in the cell nucleus (Fig. 4 *bottom*, *c*), however, when these three remaining lysine residues are removed (pPK-Δ190) the fusion protein is located in the cell cytoplasm (Fig. 4 *bottom*, *d*).

This result clearly shows that within the context of this fragment, at most three of the four lysine residues are necessary for the location of pyruvate kinase to the cell nucleus, however, as described below the local sequence context of these deletions influences their ability to locate the fusion protein to the nucleus.

Minimal Sequence for Nuclear Location

The deletion analyses described above indicate that a minimal sequence of fourteen amino acids (amino acids 156–169) should be sufficient to locate pyruvate kinase to the cell nucleus. This minimal sequence was tested directly by linking a fragment of DNA from pPK-Δ1 (nuclear located fusion protein) to a fragment of DNA from pPK-Δ159 (nuclear located fusion protein) through the common Eco RII site (Fig. 5 *top*).

Surprisingly we found that the pyruvate kinase fusion protein bearing this proposal minimal sequence (pPK-Δ1/Δ159) was located in the cell cytoplasm (Fig. 5 *bottom*, *a*).

To determine which additional sequences were required to generate a sequence which could locate pyruvate kinase to the nucleus we restored nucleoplasmin sequence at the amino-terminal end (pPK-Δ2/Δ159), the carboxyl-terminal end (pPK-Δ1/Δ138), or at both ends (pPK-Δ2/Δ138). We found that the fusion proteins from all three plasmid con-

structs were located in the cell nucleus. (Fig. 5 *bottom*, *b–d*). We observed some differences in the extent of nuclear location. The fusion protein expressed from plasmid pPK-Δ2/Δ138 was exclusively nuclear while the fusion proteins expressed from plasmids pPKΔ1/Δ138 and pPKΔ2/Δ159 gave some cytoplasmic staining in a number of cells (data not shown).

Therefore this analysis shows that the proposed minimal sequence constitutes a necessary element in sequences that can locate the fusion protein to the cell nucleus. However this sequence element is not sufficient for transport and additional amino acids are required at either end for efficient nuclear location.

Discussion

Selective Entry Model for Nuclear Accumulation Is Applicable to Cultured Mammalian Cells

As a prerequisite for further studies we have established that nucleoplasmin accumulates in the nucleus of Vero cells by the same general mechanism as it accumulates in the germinal vesicle of the *Xenopus* oocyte despite the fact that the two types of cell are markedly different in terms of their nuclear architecture and composition (10).

The rapid rate of accumulation of nucleoplasmin that we have observed is greater than that reported for fluorescently labeled nucleoplasmin in hepatoma tissue culture heterokaryons (34). We feel that this difference is probably due to the chemical modification of lysine residues by their fluorochrome leading to reduced efficiency of the nuclear location sequence.

None of the Putative Nuclear Location Sequences in the Tail Region Can Locate Pyruvate Kinase to the Cell Nucleus

As mentioned above we chose the pyruvate kinase vector system to analyse the nuclear location sequence in nucleoplasmin because of the wealth of data that has been accumulated for the SV-40 and polyoma virus large T antigen nuclear location sequences using this system. The advantage of pyruvate kinase is its exclusively cytoplasmic distribution while the structural data available for the closely homologous cat muscle enzyme (26) indicates that the carboxy and amino termini are both exposed in the monomer and are ideal locations for the placement of a nuclear location sequence. *E. coli* β-galactosidase has been used as a marker protein for several studies of nuclear location sequences (1, 13, 25, 36). However, β-galactosidase enters the nucleus of cells to some extent (15) complicating the assay by producing a background level against which the activity of any putative nuclear location sequence must be assessed.

The investigations described here have led to the surprising finding that none of the four sequences in the nucleoplasmin tail region that resemble other nuclear location sequences are able to locate pyruvate kinase to the cell nucleus. In particular, one of these sequences differs from the SV-40 large T antigen nuclear location sequence at two amino acid positions, both of which have been mutated individually and shown not to influence nuclear accumulation of large T (37). However we find that this similar sequence from nucleoplasmin is unable to locate pyruvate kinase to the cell nucleus whether it is linked to the amino acid or carboxy terminus of the protein. Instead we have demonstrated that a single sequence able to locate pyruvate kinase to the cell nucleus lies between amino acids 149 and 176, in the amino-terminal 28 amino acids of the tail region.

Minimal Sequence of 14 Amino Acids Is Revealed by Deletion Analysis

Deletion analysis of the 28 amino acid segment able to locate pyruvate kinase to the nucleus allowed us to define the limits of a nuclear location sequence.

Deletion of amino acids from the amino terminus of this segment showed that a transition in the location of the fusion protein from nuclear to cytoplasmic is concomitant with simultaneous substitution of lysine^{155} by alanine, and proline^{157} by glycine. Both of these substitutions are nonconservative; a positively charged residue (lysine) is replaced by a hydrophobic amino acid (alanine) and proline, which fixes the conformation of the polypeptide backbone, is replaced by glycine which permits maximum flexibility. Analysis of single mutations in the sequence is required to determine the individual contribution of these two amino acid substitutions to the transport defect.

Deletion of amino acids from the carboxy terminus of this segment removes lysine residues which are part of the sequence showing greatest similarity to the SV-40 large T antigen nuclear location sequence. Removal of the most carboxy-terminal lysine residue does not abolish transport into the cell nucleus but removal of the next three does. This shows that at most three basic amino acids are needed at this end of the sequence for efficient nuclear location.

However, the minimal sequence defined by these end points (RPAATKKAGQAKKK) is unable to locate pyruvate kinase to the cell nucleus, indicating that the context within which this sequence is presented influences its activity. Thus, restoration of three amino acids of the original nucleoplasmin sequence at either the amino- or carboxy-terminal end of this minimal sequence (or both ends) produces sequences that can locate pyruvate kinase to the cell nucleus. Therefore the functional nuclear location sequences contain the proposed minimal sequence as a necessary central sequence element. This sequence has periodic structure in which two sequences of four hydrophobic residues separate three basic regions.

It is notable that the functional sequences contain a sequence showing similarity to the yeast MAT α2 nuclear location sequence and the sequence having the closest similarity to the SV-40 large T antigen nuclear location sequence.

The unusual finding that restoring sequences at one end or the other is sufficient to generate a functional nuclear location sequence suggests models for the interaction of this sequence with a putative receptor in the nuclear transport machinery. We will briefly consider three.

First, it could be envisaged that the termini of the minimal sequence defined by the deletions described above represent two discrete binding domains that interact with corresponding binding sites in the receptor. Disruption of one binding domain can be tolerated but only as long as the other domain is maintained intact. A sequence in which both domains are disrupted cannot locate the fusion protein to the cell nucleus due to the combinatorial effect of weak binding at both sites.

Secondly, the nucleoplasmin–receptor interaction may not be limited to distinct sites but could occur over an extended region which includes these sites. Weak interaction at either

end of the region can be tolerated, but weak interaction at both ends reduces the overall binding to a level that renders the sequence inactive. In this case amino acids in the central region of this sequence would be expected to play a significant role in the function of this sequence.

Third, a precise conformation is necessary for recognition and transport and only the longer sequences can assume this conformation.

These models can be tested by in vitro mutagenesis.

Comparison with Other Nuclear Location Sequences

The properties of the nucleoplasmin nuclear location sequence described here are unusual but there are indications that other nuclear location sequences may have similar features. In the case of N1, another abundant, acidic *Xenopus* nuclear protein also involved in chromatin assembly (4, 16, 18), there is a sequence which closely resembles the SV-40 nuclear location sequence (17). As in nucleoplasmin, this similar sequence is not sufficient to locate the protein to the nucleus. Further amino acids are required that provide additional positively charged residues in a manner analogous to nucleoplasmin (16a).

In bovine HMG 1 (29) there is a sequence that bears a striking resemblance to the nucleoplasmin nuclear location sequence identified here. This sequence is KKGVVKAEKSKKKK, compared with (AVK)RPAATKKAGQAKKK (KLD) in nucleoplasmin and lies on the amino-terminal side of the polyglutamic acid tract.

We would like to thank Ron Laskey for continued support during this work and helpful comments on the manuscript. We would also like to thank Charlie Hodgman and Tom Creighton for helpful discussions on protein structure.

We would like to thank The Cancer Research Campaign for financial support.

Received for publication 15 March 1988, and in revised form 26 May 1988.

References

1. Burglin, T. R., and E. M. De Robertis. 1987. The nuclear migration signal of Xenopus laevis nucleoplasmin. *EMBO (Eur. Mol. Biol. Organ.) J.* 6: 2617–2625.
2. Burglin, T. R., I. W. Mattaj, D. D. Newmeyer, R. Zeller, and E. M. De Robertis. 1987. Cloning of nucleoplasmin from Xenopus laevis oocytes and analysis of its developmental expression. *Genes & Dev.* 1:97–107.
3. Colledge, W. H., W. D. Richardson, M. D. Edge, and A. E. Smith. 1986. Extensive mutagenesis of the nuclear location signal of simian virus 40 large T antigen. *Mol. Cell. Biol.* 6:4136–4139.
4. Dilworth, S. M., S. J. Black, and R. A. Laskey. 1987. Two complexes that contain histones are required for nucleosome assembly in vitro; role of nucleoplasmin and N1 in Xenopus egg extracts. *Cell.* 51:1009–1018.
5. Dingwall, C., and R. A. Laskey. 1986. Protein import into the cell nucleus. *Annu. Rev. Cell Biol.* 2:367–390.
6. Dingwall, C., T. R. Burglin, S. E. Kearsey, S. M. Dilworth, and R. A. Laskey. 1986. Sequence features of the nucleoplasmin tail region and evidence for a selective entry mechanism for transport into the cell nucleus. *In* Nucleocytoplasmic Transport. R. Peters and M. Trendelenberg. Springer-Verlag, New York. 159–169.
7. Dingwall, C., S. M. Dilworth, S. J. Black, S. E. Kearsey, L. S. Cox, and R. A. Laskey. 1987. Nucleoplasmin cDNA sequence reveals polyglutamic acid tracts and a cluster of sequences homologous to putative nuclear location signals. *EMBO (Eur. Mol. Biol. Organ.) J.* 6:69–74.
8. Dingwall, C., S. V. Sharnick, and R. A. Laskey. 1982. A polypeptide domain that specifies migration of nucleoplasmin into the nucleus. *Cell.* 30: 449–458.
9. Feldherr, C. M., E. Kallenbach, and N. Schultz. 1984. Movement of a karyophilic protein through the nuclear pores of oocytes. *J. Cell Biol.* 99: 2216–2222.
10. Franke, W. W., U. Scheer, G. Krohne, and E. D. Jarasch. 1981. The nuclear envelope and the architecture of the nuclear periphery. *J. Cell Biol.* 91(3, Pt. 2):39s–50s.
11. Goldfarb, D. S., J. Gariepy, G. Schoolnick, and R. D. Kornberg. 1986. Synthetic peptides as nuclear localization signals. *Nature (Lond.).* 322: 641–644.
12. Graessmann, A., M. Graessmann, and C. Mueller. 1980. Microinjection of early SV40 DNA fragments and T antigen. *Methods Enzymol.* 65:816–825.
13. Hall, M. N., L. Hereford, and I. Herskowitz. 1984. Targeting of E. coli β-galactosidase to the nucleus in yeast. *Cell.* 36:1057–1065.
14. Kalderon, D., W. D. Richardson, A. T. Markham, and A. E. Smith. 1984. Sequence requirements for nuclear location of Simian Virus 40 large T antigen. *Nature (Lond.).* 311:33–38.
15. Kalderon, D., B. L. Roberts, W. D. Richardson, and A. E. Smith. 1984. A short amino acid sequence able to specify nuclear location. *Cell.* 39: 499–509.
16. Kleinschmidt, J. A., and W. W. Franke. 1982. Soluble acidic complexes containing histones H3 and H4 in nuclei of Xenopus laevis oocytes. *Cell.* 29:799–809.

16a. Kleinschmidt, J. A., and A. Seifer. 1988. Identification of domains involved in nuclear uptake and histone binding of protein N1 of Xenopus laevis. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:1605–1614.

17. Kleinschmidt, J. A., C. Dingwall, G. Maier, and W. W. Franke. 1986. Molecular characterisation of a karyophilic, histone binding protein: cDNA cloning, amino acid sequence and expression of nuclear protein N1/N2 of Xenopus laevis. *EMBO (Eur. Mol. Biol. Organ.). J.* 5:3547–3552.
18. Kleinschmidt, J. A., E. Fortkamp, G. Krohne, H. Zentgraff, and W. W. Franke. 1985. Co-existence of two different types of soluble histone complexes in nuclei of Xenopus laevis oocytes. *J. Biol. Chem.* 260:1166–1176.
19. Lanford, R. E., and J. Butel. 1984. Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. *Cell.* 37:801–813.
20. Lanford, R. E., P. Kanda, and R. C. Kennedy. 1986. Induction of nuclear transport with a synthetic peptide homologous to the SV40 T antigen transport signal. *Cell.* 46:575–582.
21. Laskey, R. A., B. M. Honda, A. D. Mills, and J. T. Finch. 1978. Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature (Lond.).* 275:416–420.
22. Lonberg, N., and W. Gilbert. 1983. Primary structure of chicken muscle pyruvate kinase mRNA. *Proc. Natl. Acad. Sci. USA.* 80:3661–3665.
23. Lyons, R. H., B. Q. Ferguson, and M. Rosenberg. 1987. Pentapeptide nuclear localisation signal in adenovirus E1a. *Mol. Cell. Biol.* 7:2451–2456.
24. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
25. Moreland, R. B., G. L. Langevin, R. H. Singer, R. L. Garcea, and L. M. Herford. 1987. Amino acid sequences that determine the nuclear localization of yeast histone H2B. *Mol. Cell. Biol.* 7:4048–4057.
26. Muirhead, H., D. A. Clayden, D. Barford, C. G. Lorimer, L. A. Fothergill-Gilmore, E. Schiltz, and W. Schmitt. 1986. The structure of cat muscle pyruvate kinase. *EMBO (Eur. Mol. Biol. Organ). J.* 5:475–481.
27. Newmeyer, D., and D. Forbes. 1988. Nuclear import can be separated into distinct steps in vitro: nuclear pore binding and translocation. *Cell.* 52: 641–653.
28. Newmeyer, D., J. Lucocq, T. R. Burglin, and E. M. De Robertis. 1986. Assembly in vitro of nuclei active in nuclear protein transport: ATP is required for nucleoplasmin accumulation. *EMBO (Eur. Mol. Biol. Organ.) J.* 5:501–510.
29. Pentecost, B., and G. H. Dixon. 1984. Isolation and partial sequence of bovine cDNA clones for the high-mobility-group protein (HMG-1). *Biosci. Rep.* 4:49–57.
30. Picard, D., and K. Yamamoto. 1987. Two signals mediate hormone dependent nuclear localization of the glucocorticoid receptor. *EMBO (Eur. Mol. Biol. Organ) J.* 6:3333–3340.
31. Richardson, W. D., and H. Westphal. 1983. Adenovirus early gene regulation and the adeno-associated virus helper effect. *Curr. Top. Microbiol. Immunol.* 109:147–165.
32. Richardson, W. D., B. L. Roberts, and A. E. Smith. 1986. Nuclear location signals in polyoma virus large T. *Cell.* 44:75–85.
33. Roberts, B. L., W. D. Richardson, and A. E. Smith. 1987. The effect of protein context on nuclear location signal function. *Cell.* 50:465–475.
34. Schultz, B., and R. Peters. 1986. Intracellular transport of a karyophilic protein. *In* Nucleocytoplasmic Transport. R. Peters and M. Trendelenberg, editors. Springer-Verlag, New York. 171–184.
35. Richardson, W. D., A. D. Mills, S. M. Dilworth, R. A. Laskey, and C. Dingwall. 1988. Nuclear protein migration involves two steps: rapid binding at the nuclear envelope followed by slower translocation through nuclear pores. *Cell.* 52:655–664.
36. Silver, P., P. L. Keegan, and M. Ptashne. 1984. Amino terminus of the yeast GAL4 gene product is sufficient for nuclear localization. *Proc. Natl. Acad. Sci. USA.* 81:5951–5955.
37. Smith, A. E., D. Kalderon, B. L. Roberts, W. H. Colledge, M. Edge, P. Gillett, A. Markham, E. Paucha, and W. D. Richardson. 1985. The nuclear location signal. *Proc. Royal Soc. London B. Biol. Sci.* 226:43–58.