The EMBO Journal vol.6 no.1 pp.69–74, 1987

Nucleoplasmin cDNA sequence reveals polyglutamic acid tracts and a cluster of sequences homologous to putative nuclear localization signals

Colin Dingwall, Stephen M.Dilworth, Susan J.Black, Stephen E.Kearsey[1], Lynne S.Cox and Ronald A.Laskey

CRC Molecular Embryology Research Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

[1]Present address: Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Communicated by R.A.Laskey

Nucleoplasmin is the most abundant protein in the *Xenopus* oocyte nucleus. It is involved in histone storage and chromatin assembly and it has been used extensively to study the transport of proteins into the cell nucleus. We have isolated λ gt11 phage containing nucleoplasmin cDNA and have determined the sequence of the entire protein coding region of 200 amino acids for one of the two genes. The translation product of the sp6 transcript of this cDNA has the same electrophoretic mobility as nucleoplasmin and is able to form pentamers. The protein sequence shows remarkable clusters of charged residues including a long polyglutamic acid tract which presumably constitutes the histone binding site. The short C-terminal domain which specifies nuclear entry contains four regions which are homologous to putative nuclear localization signals including two regions of homology to the nuclear migration signal of SV40 large T antigen.

Key words: Xenopus/nucleoplasmin/sequence

Introduction

Nucleoplasmin is an acidic, pentameric, thermostable protein which is able to assemble nucleosomes by binding histones and transferring them to DNA (Laskey *et al.*, 1978; Earnshaw *et al.*, 1980). It is the most abundant protein in the *Xenopus* oocyte nucleus (Mills *et al.*, 1980; Krohne and Franke, 1980a) and an apparently equivalent protein is present in the eggs and oocytes of several amphibian species (Krohne and Franke, 1980b).

Recently a role for nucleoplasmin in nucleosome assembly *in vivo* has been supported by the demonstration of nucleoplasmin–histone complex in *Xenopus* oocyte nuclei (Kleinschmidt *et al.*, 1985). These same investigations have demonstrated the existence of complexes between histones H3 and H4 and the acidic nuclear proteins N1–N2, the sequence of which is presented elswhere (Kleinschmidt *et al.*, 1986).

There are two further reasons for determining the sequence of nucleoplasmin. Firstly, recent evidence indicates that nucleoplasmin may also have a role in transcription, as it is highly concentrated in active lampbrush chromosomes and nucleoli (Moreau *et al.*, 1986). Secondly nucleoplasmin has been used widely to study the mechanism of protein transport into the cell nucleus. Nucleoplasmin provided the first evidence that information for accumulation in the nucleus resides in a local protein domain and that this domain specifies accumulation by selective entry through the nuclear envelope, rather than by selective binding after entry by diffusion (Dingwall *et al.*, 1982). In addition nucleoplasmin was used to demonstrate that the nuclear pore complex is a route of protein entry into the nucleus (Feldherr *et al.*, 1984) and that ATP is required for nucleoplasmin entry into reconstituted nuclei *in vitro* (Newmeyer *et al.*, 1986).

Information on the mechanism and route of nuclear entry for nucleoplasmin is exceptional, yet nuclear localization signal sequences have been determined for other proteins, notably SV40 large T antigen (Kalderon *et al.*, 1984a,b). The sequence of nucleoplasmin is necessary to integrate these types of information on protein entry into the nucleus. It should also elucidate the functions of nucleoplasmin in histone binding, nucleosome assembly and possibly transcription.

Results

Production and characterization of monoclonal antibodies against nucleoplasmin

Monoclonal antibodies were raised against nucleoplasmin purified from *Xenopus* eggs and their specificities were determined by

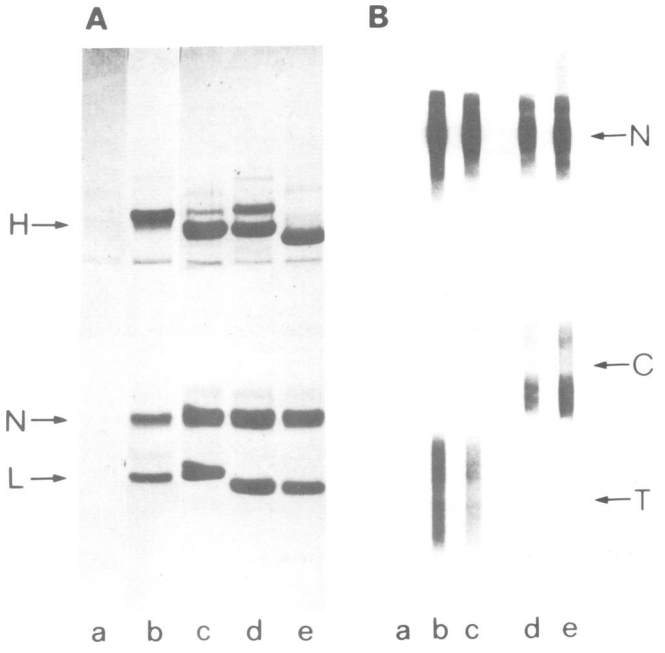

Fig. 1. (A) A Coomassie Blue-stained 13% SDS–polyacrylamide gel analysis of the proteins immunoprecipitated from oocyte lysates with four anti-nucleoplasmin monoclonal antibodies. 100 μl of a lysate of soluble proteins from *Xenopus* oocyte were immunoprecipitated with 100 μl of TCF containing monoclonal antibody as described in Materials and methods. H indicates the position of immunoglobulin heavy chain, N nucleoplasmin monomer and L immunoglobulin light chain. **Lane a** was immunoprecipitated with TCF from the parental myeloma line; **b** the monoclonal antibody PA3C5; **c** PB2D7; **d** PA1C2; **e** PB4D3. **(B)** An autoradiograph of an immunoblot of *Xenopus* oocyte nucleoplasmin cleaved with pepsin and detected with anti-nucleoplasmin monoclonal antibodies. 10 μg of nucleoplasmin in 100 mM HCl was digested with 0.1 μg of pepsin for 45 min at 37°C then electrophoresed on a 15% SDS–polyacrylamide gel and immunoblotted as described in Materials and methods. N indicates the position of intact nucleoplasmin monomer, C the position of the core polypeptide and T the tail peptides. The monoclonal antibody used in each lane was as in **(A)**.

immunoprecipitation of total soluble proteins from *Xenopus* oocytes (Figure 1A). The location of the antigenic site recognized by the monoclonal antibodies was determined by immunoblotting against pepsin-cleaved nucleoplasmin (Figure 1B).

Pepsin cleavage of nucleoplasmin occurs at specific sites to produce 'tail' fragments that contain the nuclear localization signal and a 'core' region that remains pentameric. None of the monoclonal antibodies produced reacted with both the 'core' and 'tail' domains. We have provisionally assigned the tail region of the molecule to the carboxy terminus of the polypeptide (Dingwall *et al.*, 1982). The 'core' or 'tail' specificity of the monoclonal antibodies has enabled us to confirm this assignment (see below).

```
                                    ***                 ***  ***
CTTTTAATACCGGATGTGCATGTGCCTGAAAGGAGCTGTTAGTTAGTGGAGTGCAATTGT
        10          20          30          40          50          60

                                                                1
                                                              MetAlaSer
TCCTCCGCCTGCAGAGTGTGCCGAAGCCAGTTACTTTGGGTTATCTACGTGACATGGCCT
        70          80          90          100         110         120

                        10                                  20
  ThrValSerAsnThrSerLysLeuGluLysProValSerLeuIleTrpGlyCysGluLeu
CTACAGTGAGCAACACCAGCAAACTTGAGAAACCTGTGTCCCTTATATGGGGGTGTGAAC
        130         140         150         160         170         180

                        30                                  40
  AsnGluGlnAspLysThrPheGluPheLysValGluAspAspGluGluLysCysGluHis
TGAATGAGCAAGACAAGACGTTTGAGTTTAAGGTAGAAGATGATGAGGAAAAATGTGAGC
        190         200         210         220         230         240

                        50                                  60
  GlnLeuAlaLeuArgThrValCysLeuGlyAspLysAlaLysAspGluPheAsnIleVal
ATCAGTTGGCGTTGCGCACGGTGTGTCTGGGGGACAAGGCAAAGGATGAGTTCAACATTG
        250         260         270         280         290         300

                        70                                  80
  GluIleValThrGlnGluGluGlyAlaGluLysSerValProIleAlaThrLeuLysPro
TAGAAATCGTTACACAAGAGGAGGGAGCGGAAAAATCTGTTCCAATTGCCACTCTAAAGC
        310         320         330         340         350         360

                        90                                  100
  SerIleLeuProMetAlaThrMetValGlyIleGluLeuThrProProValThrPheArg
CTTCTATTCTACCCATGGCAACTATGGTGGGCATTGAGCTGACTCCTCCAGTTACTTTCC
        370         380         390         400         410         420

                        110                                 120
  LeuLysAlaGlySerGlyProLeuTyrIleSerGlyGlnHisValAlaMetGluGluAsp
GGTTAAAAGCTGGTTCCGGCCCACTGTACATCAGTGGTCAACACGTAGCGATGGAGGAAG
        430         440         450         460         470         480

                        130                                 140
  TyrSerTrpAlaGluGluGluAspGluGlyGluAlaGluGlyGluGluGluGluGluGlu
ATTACTCATGGGCAGAAGAGGAAGATGAAGGGGAAGCTGAAGGAGAAGAGGAGGAAGAAG
        490         500         510         520         530         540

                        150                                 160
  GluGluAspGlnGluSerProProLysAlaValLysArgProAlaAlaThrLysLysAla
AGGAAGAAGATCAAGAATCTCCACCCAAAGCTGTAAAGAGGCCTGCGGCTACCAAAAAAG
        550         560         570         580         590         600

                        170                                 180
  GlyGlnAlaLysLysLysLysLeuAspLysGluAspGluSerSerGluGluAspSerPro
CAGGCCAGGCAAAGAAGAAGAAACTTGACAAAGAGGACGAGAGCTCCGAGGAAGACAGTC
        610         620         630         640         650         660

                        190                                 200
  ThrLysLysGlyLysGlyAlaGlyArgGlyArgLysProAlaAlaLysLys***
CAACCAAAAAAGGGCAAAGGAGCCGGAAGAGGAAGAAAGCCGGCTGCTAAGAAGTGAACCT
        670         680         690         700         710         720

CTTGGAGGCTGTTCCTTGCCCCTATTCCCTCCTGTGTAATTTTTTTTGTTGTTTAAAATA
        730         740         750         760         770         780

AAATATAAATAAGCTTAAACAAAAAAAAAAAAAAAA
        790         800         810
```

Fig. 2. Nucleotide sequence of a cDNA clone encoding nucleoplasmin and the deduced amino acid sequence. Stop codons are asterisked.

Cloning of nucleoplasmin cDNA

Total oocyte RNA was prepared from *Xenopus* ovaries and the poly(A)$^+$ fraction was prepared by chromatography on poly(U) Sepharose. Complementary DNA was synthesized using published procedures (see Materials and methods) and ligated into the expression vector λ gt11 (Young and Davis, 1983a,b). The library so produced was amplified and screened for nucleoplasmin cDNA clones with a mixture of 'core'- and 'tail'-specific monoclonal antibodies. These recombinants were plaque purified and then screened with individual monoclonal antibodies. DNA was then prepared from a number of recombinant phage for sequencing.

One cDNA clone that reacted with the tail-specific monoclonal antibody PA3C5 but not with core-specific monoclonal antibodies contained DNA that encoded only the carboxy terminal 48 amino acids, confirming the assignment of the tail to the carboxy terminus. The majority of phage that reacted wtih both classes of antibody contained cDNA that was excisable by digestion with the restriction endonuclease *Eco*RI and was ~620 bp in length as judged by mobility in agarose gels.

Northern blot analysis of *Xenopus* oocyte RNA indicated a single hybridizing species of ~980 nucleotides (data not shown). The largest cDNA clone had an insert with a mobility on agarose gels corresponding to 760 bp. DNA sequence analysis (see Materials and methods) indicated that the actual length was 817 bp. The DNA sequence of this clone and the amino acid sequence deduced from it are shown in Figure 2. This cDNA encodes a polypeptide of 200 amino acids having a mol. wt of 22 024.

We have also isolated a clone which differs from the sequence presented here in which the sequence Glu-Gly-Glu-Ala (residues 132−135) is deleted indicating that there are possibly two genes encoding nucleoplasmin. This is supported by Southern blot analysis of *Xenopus* genomic DNA and by the isolation of two genomic clones for nucleoplasmin which have different restriction maps (data not shown).

Confirmation that the cDNA clone contains the entire protein coding sequence

That the sequence shown in Figure 2 is the entire protein sequence of nucleoplasmin is indicated by the presence of a stop codon at position 26 in the cDNA sequence 86 nucleotides from the putative initiator methionine residue. In fact there are stop codons in all three reading frames in this upstream region as shown in Figure 2.

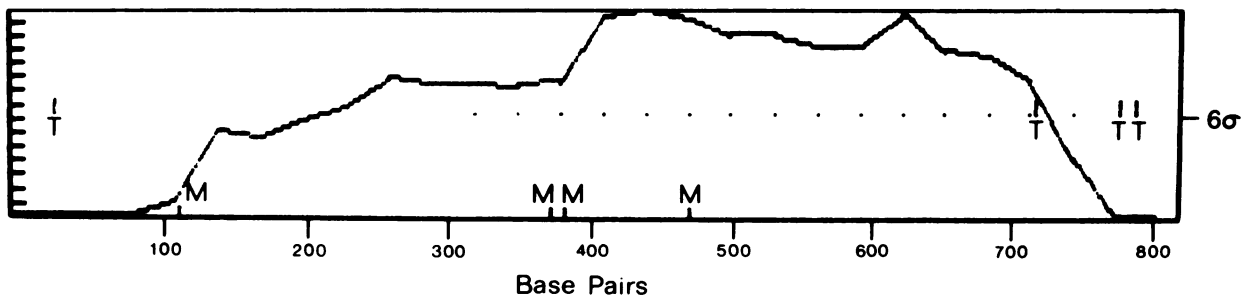

Fig. 3. Probability of coding in the nucleoplasmin cDNA sequence. The probability of coding is plotted (vertical axis) against nucleotide position for the sequence shown in Figure 2. 6 represents the level of probability 6 SD above the expected level. The top of the plot corresponds to 12 SD (McLachlan *et al.*, 1984). T = termination codons; M = methionine residues.

A computer analysis of the cDNA sequence was undertaken which looks for uneven usage of codons within a sequence: only coding regions will give codon biases above those expected from the base composition (McLachlan *et al.*, 1984). The result of this analysis is shown in Figure 3 and it supports the assignment of the boundaries of the coding sequence.

The most compelling evidence that the sequence presented is the entire protein sequence comes from translation *in vitro* of mRNA produced by transcription of the cDNA in an sp6 vector system (Krieg and Melton, 1984). The protein product produced by translating this RNA has the same mobility on SDS−polyacrylamide gels as nucleoplasmin isolated from *Xenopus* oocytes (Figure 4). In addition is it immunoprecipitated by monoclonal antibodies to nucleoplasmin and has the ability to form pentamers which are resistant to denaturation by SDS (Figure 4). The pentamer is resolved into two major bands. The reason for this is not clear but nucleoplasmin isolated from *Xenopus* eggs or oocytes behaves identically on SDS−polyacrylamide gels (Dingwall *et al.*, 1982).

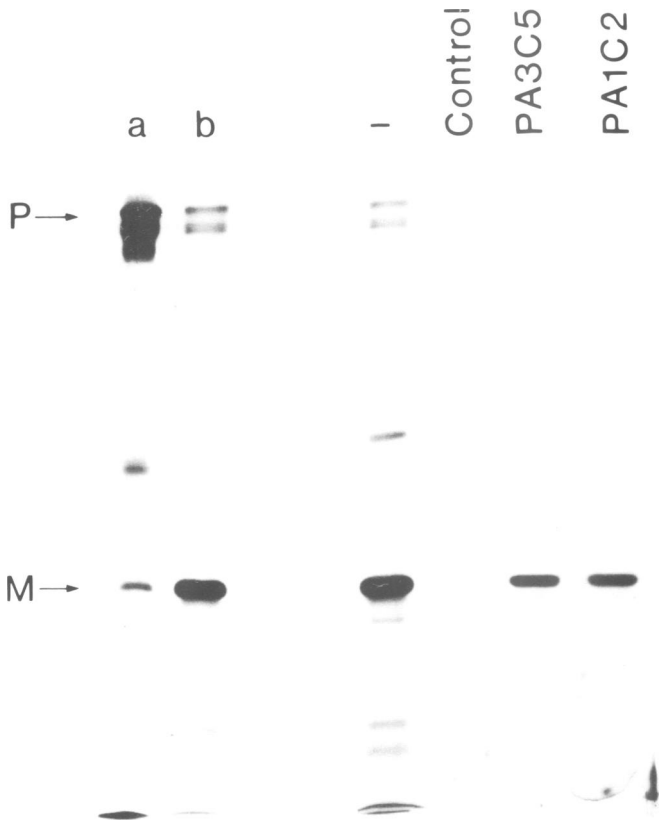

Fig. 4. *In vitro* translation of mRNA transcribed from nucleoplasmin cDNA in an sp6 vector. 1 μg of an sp6 plasmid with the nucleoplasmin cDNA inserted at the *Eco*RI site was linearized and transcribed with sp6 polymerase. The RNA product was translated in 40 μl of rabbit reticulocyte lysate. **Left hand panel:** autoradiograph of a 13% SDS−polyacrylamide gel analysis of transcription translation products. **Lane a:** 4 μl of the *in vitro* translation product analysed in the presence of 0.1% SDS without boiling. P = pentamer. **Lane b:** as **lane a** but the sample was boiled in 1% SDS for 5 min. M = monomer. **Right hand panel:** immunoprecipitations of transcription−translation products. **Lane (−):** as **lane b**, left hand panel. **Lane 'Control':** 4 μl of the *in vitro* translation reaction mixture immunoprecipitated with 50 μl of tissue culture fluid from mouse myeloma cells. **Lane 'PA3C5':** 4 μl of the *in vitro* translation reaction mixture immunoprecipitated with 50 μl of tissue culture fluid from hybridoma cells producing monoclonal antibody PA3C5 specific for the nucleoplasmin tail region. **Lane 'PA1C2':** as above but with antibody PA1C2 which is specific for the core region of nucleoplasmin.

These experiments show that this protein of 200 amino acids with a mol. wt of 22 024 migrates at a position in SDS−polyacrylamide gels corresponding to a mol. wt of ~33 000 (Dingwall *et al.*, 1982). They also demonstrate that the protein produced from this cDNA clone has all the properties of the native protein that enable it to form pentamers and confer on it its unusual stability in SDS and anomalous mobility on SDS−polyacrylamide gels.

The data described above therefore allow us to state confidently that the sequence presented here represents the entire protein coding sequence of one of the nucleoplasmin polypeptides.

Location of protease cleavage sites and the domain which specifies entry into the cell nucleus

As stated previously (Dingwall *et al.*, 1982), nucleoplasmin has a protease-sensitive C-terminal tail domain and a relatively protease-resistant N-terminal core that remains pentameric. The C-terminal tail specifies entry into the nucleus.

In order to focus the search for nuclear localization sequences it is necessary to know the limits of the core and tail domains in the sequence presented in Figure 2. We have attempted to map the known protease cleavage sites by reference to cyanogen bromide cleavage sites and by correlation of these data with the amino acid compositions of proteolytic fragments.

The amino acid composition of the core molecule produced by trypsin cleavage (Table I) shows that it contains at least part of the major polyglutamic acid tract. However there are no trypsin cleavage sites between residues 105 and 152, so the trypsin-resistant core must extend to at least residue 152. In this case the amino acid composition agrees well with that predicted from the cDNA sequence (Table I). This means that the trypsin-sensitive tail region is no more than 48 amino acids long and has a mol. wt of 5220 but its removal from the nucleoplasmin polypeptide causes a change in mobility on SDS gels corresponding to ~10 kd.

This is supported by the observation that papain cleaves nucleoplasmin to produce a core molecule with the same electrophoretic mobility as the core produced by trypsin digestion and an intact 'tail' fragment with a relative mol. wt of ~10 kd (Dingwall *et al.*, 1982; Dingwall, 1985).

Table I. Amino acid compositions (mol %) of the trypsin-resistant core of nucleoplasmin and the core predicted from the cDNA sequence

	Core (acid hydrolysis)	cDNA (residues 1−152)
Asp	7	6.6
Thr	5	5.9
Ser	7	6.6
Glu	21	23.6
Pro	10	5.9
Gly	8	5.9
Ala	7	6.6
Val	6	6.6
Cys	–	2.0
Met	2	2.6
Ile	4	4.6
Leu	7	7.2
Tyr	1	1.3
Phe	2	2.6
His	2	2.0
Lys	8	7.2
Arg	2	1.3
Trp	–	1.3

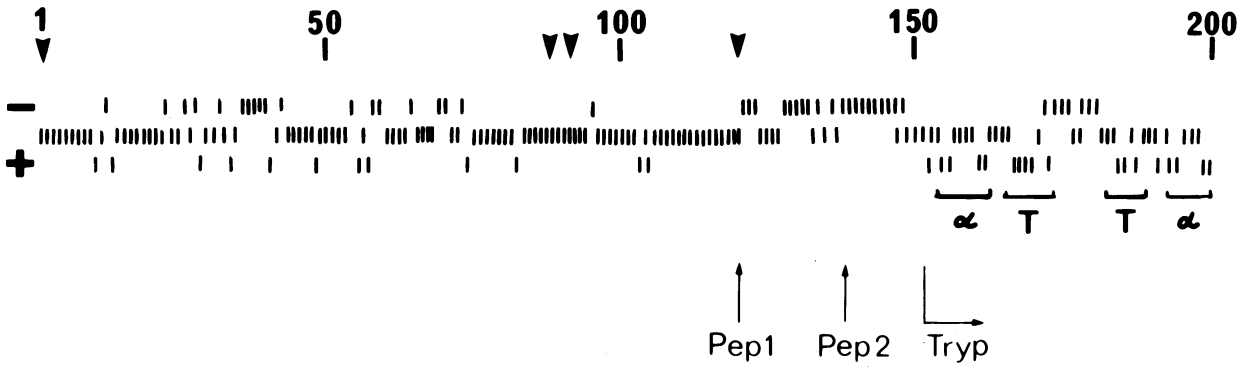

Fig. 5. The distribution of charged and neutral amino acids in the nucleoplasmin sequence. The individual amino acids in the nucleoplasmin sequence are represented by vertical bars. Their charge is shown by the allocation of the bars to rows representing acidic amino acids (−); neutral amino acids and basic amino acids (excluding histidine) (+). α — the location of sequences showing homology to the MAT α2 consensus sequence for nuclear localization. T — the location of sequences showing homology to the SV40 large T antigen nuclear localization sequence. Arrowheads mark the position of methionine residues. Approximate sites of pepsin cleavage (Pep1 and Pep2) and the limit of the trypsin (Tryp)-resistant core are marked.

Table II. Amino acid sequences in the tail region showing homology to the SV40 nuclear localization sequence

				128				
SV40 T antigen		Pro	Lys	Lys	Lys	Arg	Lys	Val
Substitutions not affecting transport (Smith *et al.*, 1985)		Ser Leu Ala	Thr Ile		Thr Arg Met	Thr Ile Lys	Thr Met Arg	Ile
	166–172	Ala*	Lys*	Lys*	Lys*	Lys*	Leu	Asp
Nucleoplasmin sequences								
	183–189	Pro*	Thr*	Lys*	Lys*	Gly	Lys*	Gly

Two sequences from the tail region are compared with the wild-type SV40 nuclear localization sequence and with mutations in that sequence which do not abolish transport into the nucleus.
Asterisks indicate identity.

It is interesting to note that the amino acid composition of the trypsin-sensitive tail region resembles that of the C terminus of the histone H1 (Dingwall and Allan, 1984). This fragment also electrophoreses anomalously slowly on SDS–polyacrylamide gels (J.Allan, personal communication).

The primary pepsin cleavage that produces the 16-kd tail fragment must occur extremely close to a methionine residue since cyanogen bromide cleavage produces a fragment which is the same size and which is cleaved internally by pepsin to produce an identical '12-kd tail'. Therefore if we assume that all the methionine residues are substrates for cyanogen bromide we can assign the primary pepsin cleavage to the region of residue 120 (Figure 5). This indicates that the 16-kd fragment contains only 80 amino acids, a point we discuss further below.

Our previous mapping studies (Dingwall *et al.*, 1982) showed that the secondary pepsin site lies between the primary pepsin site and the papain site which is also the boundary of the trypsin-resistant core. This means that the secondary pepsin site must lie within the major polyglutamic acid tract. This is supported by the known high glutamic acid content of the 12-kd tail fragment produced by prolonged pepsin digestion (Dingwall and Allan, 1984) and preliminary protein sequence data that indicated a run of several contiguous glutamic acid residues at the amino terminus of this fragment (data not shown). While this may be surprising, it is entirely consistent with early studies of pepsin specificity which showed that long stretches of polyglutamic acid (but not short stretches) are efficiently cleaved by pepsin (Simons *et al.*, 1961; Neumann *et al.*, 1962). Therefore we propose that the secondary cleavage site in nucleoplasmin occurs close to residue 138. The polyglutamic acid tract is unlikely to represent the nuclear localization sequence since it is present in the trypsin-resistant core which fails to enter the nucleus (Dingwall *et al.*, 1982). However the possibilty that the polyglutamic acid tract is the nuclear localization signal is not absolutely excluded. The structure of the tract and therefore its function might be changed by removal of the C-terminal tail. However we feel that this is unlikely since a similar change could be expected to occur in the production of the 16-kd and 12-kd 'tail' fragments, both of which accumulate rapidly in the nucleus.

Instead the sum of the evidence indicates clearly that the information which specifies entry into the nucleus is located within the carboxy terminal 48 amino acids. It is striking that this region contains over half the lysine residues of the protein, and all of these, except one, lie in four regions which are homologous to postulated nuclear localization signal sequences (Figure 5 and Table II).

Discussion

Four sequences in the carboxy terminal 50 amino acids show homology to putative nuclear localization signals

The carboxy terminal 50 amino acids contain two regions of close homology to the nuclear localization signal of SV40 T antigen (Smith *et al.*, 1985; Table II). Neither of these sequences is identical to the wild-type SV40 sequence, but the homology is strong when neutral mutations in the SV40 sequence are considered (Table II). However, each of the nucleoplasmin sequences shown differ from the wild-type SV40 sequence at two positions and the transport properties of such double mutations have not been tested directly. Despite this we consider that the sequence lying between residues 166 and 172 is the strongest candidate as it contains four consecutive basic residues which is the predominant general feature of the SV40 nuclear localization sequence (Smith *et al.*, 1985).

There are also two regions (amino acids 156–162 and 195–199) which show weaker homology to the putative nuclear localization consensus for yeast MAT α2 deduced by Hall *et al.* (1984). We have found no sequences in this region which show homology to the influenza NP accumulation signal (Davey *et al.*, 1985) or the polyoma large T signal (Richardson *et al.*, 1986), and there are no sequences in the 'core' region that show homology to any nuclear localization signals identified to date. (For

a review of nuclear localization signals see Dingwall and Laskey, 1986.)

If it emerges that all of these homologies to putative localization signals are functional, this would have implications for the transport properties of the protein since the number of signal sequences on a protein appears to influence the rate of accumulation within the nucleus. This has been shown by Lanford *et al.* (1986) who have increased the number of synthetic peptides having the SV40 nuclear localization signal linked to another protein and thereby increased the rate of entry of the peptide−protein conjugate into the nucleus of cells in culture. Multiple signals in one polypeptide have been observed in polyoma large T antigen (Richardson *et al.*, 1986). Dreyer *et al.* (1981, 1982) have observed differences in the rate of uptake of *Xenopus* oocyte nuclear proteins in the nuclei of the developing embryo. The 'rapidly migrating' class of nuclear proteins includes nucleoplasmin (and N1−N2). These authors suggest that proteins of this group may contain the most potent signal or multiple signals for nuclear localization (Dreyer *et al.*, 1986). This is consistent with our earlier findings that as tail regions are removed proteolytically from nucleoplasmin so the rate of entry of these partially cleaved pentamers into the oocyte nucleus is reduced relative to the rate of intact nucleoplasmin (Dingwall *et al.*, 1982). The results described here indicate that each tail region could possibly contain four potential nuclear localization signals suggesting that the nucleoplasmin pentamer could have as many as 20 signals.

We do not yet have direct evidence of the involvement of these homologous sequences in nuclear transport of nucleoplasmin. However the close resemblance to tolerated substitutions in the SV40 T antigen sequence make the two sequences shown in Table II conspicuously good candidates. Therefore it will be interesting to determine which amino acids are essential for nuclear entry.

Anomalous mobility of nucleoplasmin and the tail peptide on SDS−polyacrylamide gels

From the cDNA sequence the nucleoplasmin subunit is now known to have a mol. wt of 22 024 giving a pentamer of 110 120 but the subunit migrates on SDS−polyacrylamide gels at a position corresponding to a mol. wt of ~33 000. This anomalous mobility is similar to that noted for the acidic nuclear proteins N1−N2 (Kleinschmidt *et al.*, 1986), which are functionally related to nucleoplasmin by their involvement in histone binding and nucleosome assembly and also to that noted for Chromogranin A which has repeated clusters of polyglutamic acid residues (Benedum *et al.*, 1986).

In these cases the protein in question is highly negatively charged and this may reduce the binding of SDS and hence the electrophoretic mobility of the protein. This would be consistent with known resistance of the nucleoplasmin pentamer to denaturation by SDS.

The C-terminal 50 amino acids of nucleoplasmin can be removed as an intact fragment by papain and remarkably this basic fragment has an electrophoretic mobility corresponding to an apparent mol. wt of 10 kd. The removal of the C-terminal domain from the nucleoplasmin subunit causes a corresponding large decrease in its apparent mol. wt (33 kd to 23 kd). In this case an abundance of basic residues may decrease the overall negative charge on the SDS−protein complex and thereby reduce mobility by a different mechanism from that which brings about the anomalous mobility of the acidic polypeptides.

The fact that the pentameric core molecule has a true mol. wt of 84 110 rather than 115 000 as previously thought places it above, but close to, the size limit of protein that is able to enter the nucleus by diffusion (reviewed by Dingwall and Laskey, 1986). The fact that the core molecule does not enter the oocyte nucleus to any detectable extent reinforces our earlier assertion that the tail region specifies selective entry into the nucleus (Dingwall *et al.*, 1982).

The polyglutamic acid tract and histone binding

In the nucleoplasmin sequence there are a number of clusters of acidic amino acids (Figures 2 and 5). There is an almost uninterrupted run of acidic amino acids from residue 128 to 148. Of these 20 amino acids 15 are glutamic acid and two are aspartic acid. On the amino terminal side of this tract the core region is fairly acidic (21 acidic and 18 basic residues out of 127) while the tail region has only seven acidic and 18 basic residues out of 51. However in both regions the acidic residues are clustered, notably in the tail region where a short acidic tract lies between two basic regions containing the putative nuclear localization signals (Figure 5).

As stated in the Introduction, nucleoplasmin facilitates nucleosome assembly at physiological ionic strength by binding histones and transferring them to DNA. The polyacidic tracts are the obvious candidates for the histone binding site in nucleoplasmin since it is known that both polyglutamic acid (Stein *et al.*, 1979) and HMG1 (Bonne-Andrea *et al.*, 1984) facilitate the assembly of nucleosome core particles at physiological ionic strength.

In HMG1 there is a sequence of 41 consecutive acidic residues at the C-terminal end of the protein (Walker *et al.*, 1980). This long polyacidic tract appears to be able to enable a single HMG molecule to neutralize the basic charge on a histone tetramer (Bonne-Andrea *et al.*, 1984).

However, in nucleoplasmin there are reasons for thinking that each subunit does not provide a single binding site. Firstly the polyacidic tracts are much shorter than that present in HMG1. Secondly it is difficult to reconcile histone assembly into dimers (H2A, H2B), tetramers (H3, H4) and octamers (H2A, H2B, H3, H4) with binding sites in clusters of five. Thirdly the stoichiometry of binding *in vivo* is thought to be one or possibly two histone molecules to one nucleoplasmin pentamer (Kleinschmidt *et al.*, 1985). Therefore an alternative which we favour would be that the polyacicid regions of each subunit act together to produce a large negatively charged region that functions as a single binding site for a single histone or pair of histones.

Hence the nucleoplasmin molecule would be asymmetric with the polyacidic region on one side. In this way the molecule can be envisaged to act like a hand or five-fingered grab which can pick up histones by virtue of a large acidically lined region produced by a contribution from each subunit.

The existence of at least two genes for nucleoplasmin raises interesting questions with respect to histone binding. The difference between the two genes is a small deletion in the longest polyglutamic acid tract. The translation product from sp6 transcripts of the cDNA clone indicate that pentamers can form with five identical subunits.

This raises the possibility that *in vivo* different types of homo pentamer may exist with corresponding differences in histone binding specificity. It will be interesting to determine the histone binding properties of such homo pentamers produced by expression of the cloned nucleoplasmin genes.

Materials and methods

Monoclonal antibody production

Monoclonal antibodies directed against nucleoplasmin were generated by conventional procedures. Briefly, BALB/c × B10BR F_1 generation mice were immunized with purified nucleoplasmin in Freund's adjuvant at 2-weekly intervals. After three s.c. injections the mice were rested for 2 months then boosted by an i.p.

injection of nucleoplasmin alone. Four days later the animals were killed, the spleens removed and splenocytes fused with the myeloma Sp2/0-Ag14 (Shulman *et al.*, 1978) by PEG 1500 treatment. Selection was achieved by azaserine treatment at 1 μg/ml (Foung *et al.*, 1982). Clones secreting antibodies directed against nucleoplasmin were located by EIA and single cell cloned at least three times. The cells were allowed to grow to confluency then the tissue culture fluid (TCF) containing the antibody was clarified by centrifugation.

Immunoprecipitation

Centrifuged lysates of soluble proteins were reacted for 2 h at 0°C with TCF containing antibody. The immune complexes were then collected on protein A-bearing *Staphylococcus aureus* Cowan I (SAC) (Kessler, 1975), washed three times with NET buffer (50 mM Tris–HCl pH 7.5, 5 mM EDTA, 150 mM NaCl, 0.05% NP-40), eluted with SDS sample buffer and electrophoresed on SDS–polyacrylamide gels as in Laemmli (1970). Proteins were then detected by staining with Coomassie Blue.

Immunoblotting

Proteins electrophoresed on SDS–polyacrylamide gels were transferred to nitrocellulose (Amersham) by the method of Towbin *et al.* (1979). The nitrocellulose was then cut into strips and the unoccupied binding sites saturated with fetal calf serum. The proteins reacting with antibody were then detected by TCF containing monoclonal antibody, followed by biotinylated goat anti-mouse immunoglobulin, ^{35}S-labelled streptavidin (Amersham), and autoradiography.

RNA preparations and in vitro translation

Xenopus oocyte RNA was prepared as described by Bienz and Gurdon (1982) from which poly(A)$^+$ RNA was prepared by the method of Adesnik *et al.* (1972). The RNA was translated *in vitro* according to the method of Jackson and Hunt (1983) using nuclease-treated rabbit reticulocyte lysate that was the generous gift of Dr T.Hunt.

cDNA synthesis and cloning

cDNA was synthesized using methods described by Maniatis *et al.* (1982) or by the method of Gubler and Hoffmann (1984). Briefly, 5 μg of poly(A)$^+$ RNA was used to synthesize 1 μg of complementary DNA; 250 ng of this DNA were ligated into the expression vector λ gt11 as described by Young and Davis (1983a, b). This DNA produced a library containing ~4 × 10^5 phage. This library was amplified to give a phage stock containing 3 × 10^{11} phage/ml.

Screening for nucleoplasmin cDNA clones and DNA sequence analysis

The amplified library was screened with a TCF containing a mixture of monoclonal antibodies to nucleoplasmin exactly as described by Young and Davis (1983b). Approximately 4 × 10^5 phage were screened initially from which a total of 75 positive phage were selected. A number of these were plaque purified by plating at higher dilution and DNA was prepared from these phage for DNA sequencing. The DNA to be sequenced was sonicated to produce random fragments that were subcloned into an M13 vector and sequenced exactly as described by Messing (1983). The DNA sequence data generated were analysed using the standard computer programs available on the University of Cambridge 'Phoenix' computer.

sp6 transcription

The *Eco*RI fragment containing the nucleoplasmin cDNA was ligated into the vector pSP65 (Promega Biotec) and RNA was produced from the plasmid DNA exactly as described by Krieg and Melton (1984). The RNA was purified by phenol extraction and ethanol precipitation and translated *in vitro* as described above.

Acknowledgements

We would like to thank Julian Blow for helpful comments on the manuscript and Tim Mohun and David Judge for their help with computing. We also thank Sandy Carne for amino acid compositions and Ian Fearnley for protein sequencing. We gratefully acknowledge the generous support of the Cancer Research Campaign.

References

Adesnik,M., Salditt,M., Thomas,W. and Darnell,J.E. (1972) *J. Mol. Biol.*, **17**, 21–30.

Benedum,U.M., Baluerle,P.A., Konecki,D.S., Frank,R., Powell,J., Mallet,J. and Huttner,W.B. (1986) *EMBO J.*, **5**, 1495–1502.

Bienz,M. and Gurdon,J.B. (1982) *Cell*, **29**, 811–819.

Bonne-Andrea,C., Harper,F., Sobczak,J. and DeRecondo,A.-M. (1984) *EMBO J.*, **3**, 1193–1199.

Davey,J., Dimmock,N.J. and Colman,A. (1985) *Cell*, **40**, 667–675.

Dingwall,C. (1985) Ph.D. Thesis, Council for National Academic Awards.

Dingwall,C. and Allan,J. (1984) *EMBO J.*, **3**, 1944–1937.

Dingwall,C. and Laskey,R.A. (1986) *Annu. Rev. Cell Biol.*, **2**, 365–388.

Dingwall,C., Sharnick,S.V. and Laskey,R.A. (1982) *Cell*, **30**, 449–458.

Dreyer,C., Singar,H. and Hausen,P. (1981) *Wilhelm Roux Arch. Entwicklungsmech. Org.*, **190**, 197–207.

Dreyer,C., Scholz,E. and Hausen,P. (1982) *Wilhelm Roux Arch. Entwicklungsmech. Org.*, **191**, 228–233.

Dreyer,C., Stick,R. and Hausen,P. (1986) In Peters,R. and Trendelenburg,M.F. (eds), *Nucleocytoplasmic Transport*. Springer-Verlag, Berlin, pp. 143–157.

Earnshaw,W.C., Honda,B.M., Thomas,J.O. and Laskey,R.A. (1980) *Cell*, **21**, 373–383.

Feldherr,C.M., Kallenbach,E. and Schultz,N. (1984) *J. Cell Biol.*, **99**, 2216–2222.

Foung,S.K.H., Sasaki,D.T., Grunet,F.C. and Engleman,E.G. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7484–7488.

Gubler,U. and Hoffmann,B.J. (1983) *Gene*, **25**, 263–269.

Hall,M.N., Hereford,L. and Herskowitz,I. (1984) *Cell*, **36**, 1057–1065.

Jackson,R.J. and Hunt,T. (1983) *Methods Enzymol.*, **96**, 50–74.

Kalderon,D., Richardson,W.D., Markham,A.T. and Smith,A.E. (1984a) *Nature*, **311**, 33–38.

Kalderon,D., Roberts,B.L., Richardson,W.D. and Smith,A.E. (1984b) *Cell*, **39**, 499–509.

Kessler,S.W. (1975) *J. Immunol*, **115**, 1617–1624.

Kleinschmidt,J.A., Fortkamp,E., Krohne,G., Zentgraf,H. and Franke,W. (1985) *J. Biol. Chem.*, **260**, 1166–1176.

Kleinschmidt,J.A., Dingwall,C., Maier,G. and Franke,W.W. (1986) *EMBO J.*, **5**, 3547–3552.

Krieg,P.A. and Melton,D.A. (1984) *Nucleic Acids Res.*, **12**, 7057–7070.

Krohne,G. and Franke,W.W. (1980a) *Proc. Natl. Acad. Sci. USA*, **77**, 1034–1038.

Krohne,G. and Franke,W.W. (1980b) *Exp. Cell Res.*, **129**, 167–189.

Laemmli,U.K. (1970) *Nature*, **227**, 680–685.

Lanford,R.E., Kanda,P. and Kennedy,R.C. (1986) *Cell*, **46**, 575–582.

Laskey,R.A., Honda,B.M., Mills,A.D. and Finch,J.T. (1978) *Nature*, **275**, 416–420.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

McLachlan,A.D., Staden,R. and Boswell,D.R. (1984) *Nucleic Acids Res.*, **12**, 9567–9575.

Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.

Mills,A.D., Laskey,R.A., Black,P. and De Robertis,E.M. (1980) *J. Mol. Biol.*, **139**, 561–568.

Moreau,N., Angelier,N., Bonnanfant-Jais,M.-L., Gounon,P. and Kubisz,P. (1986) *J. Cell Biol.*, **103**, 683–690.

Neumann,H., Sharon,N. and Katchalski,E. (1962) *Nature*, **195**, 1002.

Newmeyer,D.D., Lucoca,J.M., Burglin,T.R. and De Robertis,E.M. (1986) *EMBO J.*, **5**, 501–510.

Richardson,W.D., Roberts,B.L. and Smith,A.E. (1986) *Cell*, **44**, 77–85.

Shulman,M., Wilde,C.D. and Kohler,G. (1978) *Nature*, **276**, 269–270.

Simons,E.R., Fasman,G.D. and Blout,E.R. (1961) *J. Biol. Chem.*, **236**, Preliminary Communication 64.

Smith,A.E., Kalderon,D., Roberts,B.L., Colledge,W.H., Edge,M., Gillett,P., Markham,A., Paucha,E. and Richardson,W.D. (1985) *Proc. R. Soc. London, Ser. B*, **221**, 43–53.

Stein,A., Whitlock,J.P. and Bina,M. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5000–5004.

Towbin,H., Staehelin,T. and Gordon,J. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 4350–4354.

Walker,J.M., Gooderham,K., Hastings,J.R.B., Mays,E. and Johns,E.W. (1980) *FEBS Lett.*, **122**, 264–270.

Young,R.A. and Davis,R.W. (1983a) *Proc. Natl. Acad. Sci. USA*, **80**, 1194–1198.

Young,R.A. and Davis,R.W. (1983b) *Science*, **222**, 778–782.

Received on 6 October 1986; revised on 10 November 1986