Journal of Cellular Biochemistry 55:32–58 (1994)

Putative Nuclear Localization Signals (NLS) in Protein Transcription Factors

Teni Boulikas

Institute of Molecular Medical Sciences, Palo Alto, California 94306

Abstract We have recognized about ten distinct forms of strongly basic hexapeptides, containing at least four arginines and lysines, characteristic of nuclear proteins among all eukaryotic species, including yeast, plants, flies and mammals. These basic hexapeptides are considered to be different versions of a core nuclear localization signal, NLS. Core NLSs are present in nearly all nuclear proteins and absent from nearly all "nonassociated" cytoplasmic proteins that have been investigated. We suggest that the few ($\sim$10%) protein factors lacking a typical NLS core peptide may enter the nucleus via their strong crosscomplexation with their protein factor partners that possess a core NLS. Those cytoplasmic proteins found to possess a NLS-like peptide are either tightly associated with cell membrane proteins or are integral components of large cytoplasmic protein complexes. On the other hand, some versions of core NLSs are found in many cell membrane proteins and secreted proteins. It is hypothesized that in these cases the N-terminal hydrophobic signal peptide of extracellular proteins and the internal hydrophobic domains of transmembrane proteins are stronger determinants for their subcellular localization. The position of core NLSs among homologous nuclear proteins may or may not be conserved; however, if lost from an homolgous site it appears elsewhere in the protein.

This search provides a set of rules to our understanding of the nature of core nuclear localization signals: (1) Core NLS are proposed to consist most frequently of an hexapeptide with 4 arginines and lysines; (2) aspartic and glutamic acid residues as well as bulky amino acids (F, Y, W) need not to be present in this hexapeptide; (3) acidic residues and proline or glycine that break the α-helix are frequently in the flanking region of this hexapeptide stretch; (4) hydrophobic residues ought not to be present in the core NLS flanking region allowing for the NLS to be exposed on the protein. In this study we attempt to classify putative core NLS from a wealth of nuclear protein transcription factors from diverse species into several categories, and we propose additional core NLS structures yet to be experimentally verified.

Key words: pore complex, transporter proteins, transcription factors, nuclear localization, nuclear proteins, basic region, cytoplasmic proteins, membrane proteins, c-Myc, c-Jun, NF-κB

The nuclear envelope effectively separates transcription of genes from translation of their mRNA into proteins and allows proteins destined to function in the nucleus to pass selectively through the lumen of the nuclear pores [reviewed by Newport and Forbes, 1987; Boulikas, 1987]. Pore complexes are anchored to the double nuclear membrane via the transmembrane glycoprotein gp210 [Wozniak et al., 1989] and with the aqueous channel that has an effective diameter of approximately 10 nm allow small proteins with a molecular weight inferior of 20 kD to pass slowly by diffusion [Paine et al., 1975]. However, both larger and smaller proteins destined to function in the nucleus are transported rapidly. Selective transport through pores creates a unique biochemical environment within the nucleus.

Transport across the pore complex is a two-step process involving binding to the pore complex and translocation across the lumen of the pore complex. The center of the pore is the actual site of protein translocation. Only the translocation step requires ATP [Newmeyer and Forbes, 1988; Richardson et al., 1988]. Macro- and micronuclei of *Tetrahymena* accumulate different subsets of proteins due to selectivity and discrimination in protein transport across the pore complexes [White et al., 1989]. Since both types of nuclei in this protozoan are in the same cytoplasmic compartment, subtle mechanisms

Abbreviations: ER, endoplasmic reticulum; aa, amino acid(s); NLS, nuclear localization signal.

Received June 23, 1993; accepted August 6, 1993.

Address reprint requests to Teni Boulikas, Institute of Molecular Medical Sciences, 460 Page Mill Road, Palo Alto, CA 94306.

linked to the transcriptional activity of nuclei and processing of primary transcripts occurring in the neighborhood of the nuclear pores may control protein transport across the annuli of pore complexes.

A very short sequence of seven amino acids (Pro-Lys-Lys-Lys-Arg-Lys-Val or PKKKRKV), first recognized in the SV40 large T antigen, is required for its normal nuclear localization [Kalderon et al., 1984a,b]. When the NLS of large T antigen was inserted within the pyruvate kinase or the β-galactosidase by genetic engineering at the gene level, the mosaic proteins containing this seven amino acid sequence patch were directed to the nucleus. Four putative NLS were identified in the C-terminal tail of *Xenopus* oocyte nucleoplasmin by sequence similarity to SV40 large T antigen and yeast MATα2 NLS [Dingwall et al., 1987]. However, subsequent studies have shown that only sequences containing the first two putative NLS were functional simultaneously in directing nuclear localization of fusion proteins [Bürglin and DeRobertis, 1987]. This finding was contradicted by the studies of Chelsky and co-workers [1989], who were able to direct chicken serum albumin crosslinked with only the second putative NLS of *Xenopus* nucleoplasmin to the nuclei of HeLa cells.

Yeast histone H2B has been shown to contain the seven amino acid residue NLS G**KKRSK**A that can direct β-galactosidase to the nucleus in vivo [Moreland et al., 1987]. Gilmore and Temin [1988] determined that the v-Rel oncoprotein has the NLS GN**KAKRQR**ST; addition of this peptide to the normally cytoplasmic protein β-galactosidase directed that protein to the nucleus. It is noteworthy that the hexapeptide KKKDLK occurs in chicken v-Rel oncoprotein (amino acid position 54) but does not appear to have NLS function [Hannik and Temin, 1989]. Yet the RKDRR motif of human estrogen receptor appears to be an efficient nuclear location signal [Hamy et al., 1992]. A systematic study of 85 peptides with NLS function [Boulikas, 1993] shows that D or E are not frequently placed between K and R. We can thus conclude that a negative charge within the core NLS hexapeptide diminishes NLS function. It is thus evident that not only the exact sequence of the particular NLS peptide, but also the 3D structure of the protein it belongs to and the species or cell type from which this protein is derived, are important parameters in defining NLS function.

Protein phosphorylation is a mechanism used by the cell to unmask a cryptic NLS or to increase its potency and thus to regulate the onset of nuclear import for specific proteins during development. The *pelle* gene in *Drosophila,* which encodes a protein with a protein kinase catalytic domain, is required for the nuclear import of dorsal protein. Dorsal establishes dorsoventral polarity in *Drosophila* embryos [Shelton and Wasserman, 1993]. The spatially regulated nuclear import of dorsal protein commences prior to gastrulation [reviewed by Govind and Steward, 1991]; on the ventral side of the embryo dorsal protein is translocated into nuclei, whereas on the dorsal side it remains in the cytoplasm. The gradient of nuclear localization of dorsal protein in wild-type embryos directs the formation of the dorsoventral axis by activating some downstream genes and repressing others [Ray et al., 1991]. Other transcription factors, including the glucocorticoid receptor, rel-related proteins, the *Xenopus* nuclear factor xnf7 and v-Jun, all appear to undergo regulated nuclear import [Picard et al., 1990; Gilmore, 1990; Blank et al., 1991; Miller et al., 1991; Chida and Vogt, 1992].

NF-κB, although ubiquitously expressed, is active in only a subset of cell types including B lymphocytes and mature macrophages. The regulation of NF-κB activity depends on its nuclear import. Two mechanisms for regulating the nuclear import of NF-κB have been unravelled: First, Henkel and coworkers [1992] have demonstrated using antibodies directed against NLS that the NLS of NF-κB (p50) is unavailable for binding to transporter proteins in the 110 kD precursor. Removal of only 191 aas from the C-terminus of the 110 kD precursor restored antibody accessibility as well as nuclear uptake by unmasking the NLS near its N-terminus. Thus, processing of the 100 kD precursor to mature p50 exposes the NLS. Second, in cells where NF-κB is inactive, an inhibitor, IκB, binds NF-κB in the cytoplasm. Upon activation by a variety of cytokines and mitogens, including IL-1 and TNF-α, NF-κB/IκB complex breaks down and NF-κB translocates to the nucleus [Baeuerle and Baltimore, 1988a,b]. This release of NF-κB from IκB appears to be mediated by phosphorylation of IκB [Shirakawa and Mizel, 1989].

A wealth of nuclear protein transcription factors from species as diverse as yeast, *Dro-*

sophila, and mammals is available. The vast majority of these molecules ought to possess NLS signals and a majority of cytoplasmic proteins ought not to possess such signals. In order to acquire a comprehensive view of the nature of NLS we have screened a representative number of protein transcription factors as well as of proteins known to be confined to the cytoplasmic, extracellular, and cell membrane compartments for NLS-like oligopeptides. The present study is an attempt to classify putative NLS into several categories.

MATERIALS AND METHODS

Several cDNA structures of transcription protein factors and their predicted protein sequences come from the current literature. Protein sequences, mainly from vertebrate transcription factors [e.g., Faisst and Meyer, 1991], were screened manually and all lysine and arginine residues were marked. All hexapeptide motifs containing 4 (K + R) and their flanking one to several amino acids were entered on the first column of our tables. If more such putative NLS are found, they are listed in the order in which they occur in the protein sequence from the N- to the C-terminus. NLS were classified into "highly basic NLS" (Table I), containing 5 or 6 (K + R) in a hexapeptide stretch; "typical NLS" (Table II), containing 4 (K + R) in a stretch of 6 amino acids; and "non-typical NLS," containing 3 or 2 (K + R) within a hexapeptide (Table III). "Typical NLS" motifs were further grouped into θθθθ motifs with four consecutive arginine or lysine residues, θθθxθθ, θθθxθ, θθxθθ, θxθθxθ, θθxθxθ, θθθxxθθ, and θθxxθθ (Table II). In this case θ is lysine or arginine, from the greek θ*etikós*, "positive."

As a control, other non-nuclear protein cDNA structures were taken randomly from the literature. Only those proteins found to contain a NLS-like peptide are listed (Table IV), indicating their subcellular location, whereas the total number of non-nuclear proteins not containing NLS-type peptides is scored (Table VI).

RESULTS

Protein transcription factors are nuclear proteins that interact with the regulatory sequences of genes (promoter or enhancer) to activate, and in fewer cases to repress, gene expression. The majority of these nuclear proteins are larger than ~20–40 kD and are thus expected to enter the nucleus via nuclear localization signals in their sequence. This prompted us to examine the primary structures of protein transcription factors for four (K + R) residues contained within an hexapeptide stretch, suggested from a previous study [Boulikas, 1993] to constitute a core NLS structure. Nuclear localization signals anchor the proteins to be imported into nuclei to specific transporter molecules [see Boulikas, 1993] and are thus essential components for the nuclear import of protein factors in eukaryotic cells.

A total number of 117 protein transcription factors were screened. Of those, 19 were found to include a "highly basic NLS," operationally defined as a stretch with 5 or more consectuve (K + R) residues (Table I). 87 were found to possess a typical NLS (Table II) and 11 were found not to contain an NLS-like peptide (Table III).

The protein factors that were found to possess a highly basic NLS core include the mammalian Oct-1, 2, 3, and 6, ATF-1 and 3, the ets-like Pu.1, the human thyroid hormone receptor α, the human GC factor, the *Drosophila* Suvar (3)7 gene product, a zinc-finger protein involved in position-effect variegation, as well as several proteins that contain HMG boxes (Table I). We notice that the contiguous 5 or 6 arginines and lysines occur with some preference for the motifs RRRKKR, RRKKK, RKRKR, and KKKKRKR. The GC factor possesses a contiguous stretch of 9 (K + R) in the motif RRRRQRTRKKKKKKKR (Table I). This density of positive charges in the GC protein from human cells is higher than the 8 consecutive K + R residues of starfish H2B (Boulikas, in preparation). All other proteins in Table I possess a 5-7 contiguous R + K.

The vast majority of protein transcription factors were found to possess a "typical NLS," defined as four (K + R) residues in a hexapeptide stretch (Table II). Protein transcription factors containing a typical NLS motif were subdivided into the 9 groups θθθθ, θθθxθθ, θθθxθ, θθxθθ, θxθθxθ, θθxθxθ, θθθxxθθ, θθθxxθ, and θθxxθθ, where θ is arginine or lysine (Tables II, V), depending on the position of the positive charges on the hexapeptide. The types θθθθ, θθxθθ, θθθxθ, and θθxθxθ are by far the most abundant type of motifs (Table V). Nuclear factors containing only the θθxxθθ type are rather rare; this type of motif (KKGSKK) from *Xenopus* H2B was a low potency NLS when tagged to BSA or IgG and the proteins were microinjected into *Xenopus* oocytes [Goldfarb et al., 1986].

TABLE I. Transcription Factors With "Highly Basic" Putative NLS*

NLS and flanks	Protein factor and features	References
H**R**$_4$**QRTRK**$_7$**R** L**RRKSR**P S**RRTKRRQ**	Human **GCF** (GC-factor)	Kageyama and Pastan [1989]
G**RKRKKR**T	**Oct-6** protein transcription factor from mouse cells	Meijer et al. [1990] Suzuki et al. [1990]
G**RRRKKR**T	Mouse **Oct-2** protein transcription factors (Oct-2.1 for Oct-2.6 isoforms)	Wirth et al. [1991]
A**RKRKR**T N**RRQKGKR**S	**Oct-3** from mouse P19 embryonal carcinoma cells	Okamoto et al. [1990]
EC**RRKKK**E	Human **ATF-1**. In basic region/leucine zipper.	Hai et al. [1989]
E**RKKRRR**E A**K**C**R**N**KKKE**KT	Human **ATF-3** (in basic region that binds DNA)	Hai et al. [1989]
S**KKKIR**L Q**K**GN**RKK**M V**KKVKKK**L	Mouse **Pu. 1** (Friend erythroleukemia cells). Related to *ets* oncogene	Klemsz et al. [1990] Paul et al. [1991]
V**KRKK**I C**R**N**RYRK**LE **IRKRRK**M**K** **PKKKR**L**R**L	Human **PRDII-BF1** that binds to IFN-β gene promoter. (The largest DNA-binding protein known, of 298 kD).	Fan and Maniatis [1990]
G**KKKKRKRE**K**L** (within the HMG-box)	Murine **LEF-1** (397 aa). Lymphoid-specific with an HMG1-like box. NLS is identical to that of human TCT-1α.	Travis et al. [1991]
G**KKKKRKRE**K**L** (within the HMG-box)	Human **TCF-1**α (399 aa) (T cell-specific transcription factor that activates the T cell receptor Cα). Contains an HMG box. NLS core is identical to that of murine LEF-1.	Waterman et al. [1991]
G**KKKRRSRE**K**H** (within the HMG-box) P**KK**C**RARF**	Human **TCF-1** (uniquely T cell-specific). HMG box containing.	van de Wetering et al. [1991]
F**K**Q**RRIK**L N**RRRKKR**T N**RRQKE**K**RI**	*Xenopus laevis* **Oct-1** (within POU-domain)	Smith and Old [1991]
D**KRSRKRKRSK** **RLRIDRKR**N A**KRSRR**S	*Drosophila* ***Suvar (3) 7*** gene product involved in position-effect variegation (932 aas); five widely spaced zinc-fingers could help condensation of the chromatin fiber	Reuter et al. [1990]
I**RKRRK**M**K**SVGD$_2$E$_2$ (not suggested as NLS by the authors; between the 1st and 2nd zinc finger) PP**KKKRLRL**AE (suggested as NLS by the authors; just before 2nd zinc finger) C**RNRYRK**LE (within 1st zinc finger)	Human **MBP-1** (class I MHC enhancer binding protein 1) mw 200 kD. Induced by phorbol esters and mitogens in Jurkat T cells	Baldwin et al. [1990]

(*Table I continued on next page.*)

TABLE I. Transcription Factors With "Highly Basic" Putative NLS* (*continued*)

NLS and flanks	Protein factor and features	References
P**RRKRR**V H**RYKMKRQ**	Rat **TTF-1** (thyroid nuclear factor that binds to the promoter of thyroid-specific genes); an homeodomain protein	Guazzi et al. [1990]
DG**KRKRK**N DDS**KRVAKRK**L N**RERRRKEE** W**KQRRK**F	Human **thyroid hormone receptor** α (*c-erbA-1* gene). Belongs to the family of cytoplasmic proteins that are receptors of hydrophobic ligands such as steroids, vitD, retinoic acid, and thyroid hormones; the ligand binding may expose the NLS for nuclear import of the receptor-ligand complex	Laud et al. [1991]
N**RRKRKR**S P**KKKK**L	*Drosophila* **gcl** (germ cell-less) gene product (569 aa, 65 kD), located in nuclei, required for germ line formation	Jongens et al. [1992]
A**RRKRRR**L L**KFKKVR**D **FKKFRK**F G**KQKRR**F ER**LKRDKEKREK**E T**RGRPKKVK**E S**KKRGRRRKK**T T**RRQKRAK**V S**RKSKKRLRA**	*C. elegans* **Sdc-3** protein (sex-determining protein) (2,150 aas). Zinc finger protein	Klein and Meyer [1933]
L**KKIRRKIKNKI** ES**RRKKK**E	*Drosophila* **BBF-2** (related to CREB/ATF)	Abel et al. [1992]

*Transcription factors with "highly basic" putative core NLS include those transcriptional regulators that contain 5 or more consecutive arginines or lysines in their sequence. Where more than one such signal occur, they are listed in the order they occur in the protein structure from the N- to C-terminus.

Our study suggests that a single protein transcription factor may possess more than one NLS (Tables I, II). This suggestion is in agreement with studies by others who have shown the presence of multiple nuclear localization signals on a single nuclear protein [Welsh et al., 1986; Richardson et al., 1986; Picard and Yamamoto, 1987]; presumably a multitude of NLS may increase the rate of protein factor-transporter protein interaction [Dworetzky et al., 1988].

Homologous proteins may possess the same type of NLS motif. For example, both human TFEB and TFE3 have the ERRRRF motif (Table II.1). Mouse Oct-2, 3, and 6 possess a 5 or 6 (K + R) flanked by G or A to the N-side and T to the C-side (Table I). The human mineralocorticoid, glucocorticoid, and progesterone receptors all possess the KIRRK motif that changes into KNRRK in estrogen receptor (Table II.3). The HNF-1 factor in rat, mouse, and human cells as well as the variant HNF-1 (vHNF-1) share the $KK^{M}_{G}RRNR$ motif (Table II.4).

Figures 1–4 show that putative core NLS peptides among homologous nuclear protein factors may or may not be conserved. However, when a putative core NLS disappears from a protein site within a group of homologous proteins, NLS appears at another site of the protein molecule; for example, the c-Myc and AP-4 putative core NLSs are lost from the basic region, normally present in their homologs Myn, Max, L-Myc, TFE-3, MyoD1, CBF-1, USF, and others, and appear in the leucine zipper region of the protein (Fig. 1). The more distant but homologous proteins da, E12, E2-2, E47, Pho4, Lyl-1, twist, EMC, and ID do not appear to have a typical NLS in their basic region; however, NLS may be found in other domains of these proteins. Evidently, the basic region of protein transcription factors is the most plausible site for the occurence of a putative NLS. Figure 2 shows putative NLS in the basic region of Jun, CREB, Fos, and related transcriptional regulators. Most of these molecules have two clusters of basic residues in their basic region with some conservation of the actual putative NLS (double underlined) between pairs of proteins.

Figure 3 shows the conservation, with some variations, of a putative NLS in the basic region of p50 and p65 subunits of NF-κB and their

TABLE II. Transcription Factors With "a Typical" Putative NLS*

NLS and flanks	Protein factor and features	Reference
1. Group θθθθ (with four consecutive arginines/lysines) (21 members)		
DRN**KKKK**E A**RRRR**P	*Xenopus* **RAR** (retinoic acid receptor)	Ellinger-Ziegelbauer and Dreyer [1991]
G**RRRR**A DE**KRRK**V C**RQKRK**V	Human **ATF-2** (the 2nd and 3rd NLS are in a basic region that binds DNA)	Hai et al. [1989]
E**RKRR**D S**RKK**L**R**ME	**Myn** (murine homolog of Max); forms a specific DNA-binding complex with c-Myc oncoprotein through a helix-loop-helix/leucine zipper	Prendergast et al. [1991]
EE**KRKR**TYE	Human **NFκB p65** (550 aa); not binding DNA; complexed with p50 that binds DNA; NFκB p50 also contains a NLS (Table IIIb)	Nolan et al. [1991]; Ruben et al. [1991]
G**RRRR**A DE**KRRK**F S**R**C**RQKRK**V	Human **HB16,** a cAMP response element-binding protein	Kara et al. [1990]
S**KKKK**TKV N**R**PD**KKK**I **QRRKK**P **QKKRR**FKT	Human TFIIE-β (general transcription initiation protein factor; forms tetramer $\alpha_2\beta_2$ with TFIIE-α)	Sumimoto et al. [1991]
S**RKRKM**	Human **kup** transcriptional activator (433 aas); two distantly spaced zinc fingers; expressed in hematopoietic cells and testis.	Chardin et al. [1991]
E**RKR**L**R**N**R**LA AT**K**C**RKRK**L (a 19 aa stretch)	Mouse **Jun-B** homologue to avian sarcoma virus 17 oncogene *v-jun* product; one region is similar to yeast GCN4 and to Fos	Ryder et al. [1988]
D**KR**x_6E**RKRR**D (N-terminus) Q**R**S**RKK**L**R**ME (C-terminus)	**Max** (specifically associates with c-Myc, N-Myc, L-Myc); the Max-Myc complex binds to DNA; neither Max nor Myc alone exhibit appreciable DNA binding	Blackwood and Eisenman [1991]
D**K**E**KK**I**K**LEEDE (within an acidic region) I**KK**A**KK**V T**RRKK**N	Chicken **VBP** (vitellogenin gene-binding protein); leucine zipper; related to rat DBP	Iyer et al. [1991]
T**R**D**D**K**RR**A EVE**RRRRDK**	*Xenopus borealis* **B1 factor;** closely related to the mammalian USF; binds to CACGTG in the TFIIIA promoter to developmentally regulate its expression	Kaulen et al. [1991]
T**R**D**E**K**RR**A EVE**RRRRDK**	Human **USF** (upstream stimulatory factor) activating the major late adenovirus promoter	Gregor et al. [1990]
Y**RR**YP**RRR**G **QRR**PY**RRRR**F Y**RP**R**FRR**G **QRRYRR**N Y**RRRR**P	**YB-1,** a protein that binds to the MHC class II Y box; YB-1 is a negative regulator	Didier et al. [1988]
A**K**E**R**Q**KK**D E**RRRR**F	Human **TFEB**; binds to IgH enhancer	Carr and Sharp [1990]

(*Table II continued on next page.*)

TABLE II. Transcription Factors With "a Typical" Putative NLS* (*continued*)

NLS and flanks	Protein factor and features	Reference
L**K**E**R**Q**K**KD IE**RRRR**FN YF**RRRR**LE**K**D	Human **TFE3** (536 aa); binds to μE3 enhancer of IgH genes	Beckmann et al. [1990]
KTVAL**KRRK**ASSRL	Human **Dr1** (176 aa, 19kD); interacts with TBP (TATA-binding protein), thus inhibiting association of TFIIA and/or TFIIB with TBP; TBP-Dr1 association is affected by Dr1 phosphorylation to repress activated and basal transcription	Inostroza et al. [1992]
1. L**RRRG**R**Q**TY 27. LT**RRRR**IEM 51. QN**RR**M**K**L**KK**EI	*Drosophila* **ultrabithorax** protein (from the conserved 61 amino acid homeodomain segment only); conserved in the antenappedia homeodomain protein	Chan and Mann [1993]
SN**RRR**PD**HR** VY**R**G**RRRV**R**RE** P_7AP_2**RRRR**SADN**K**D_2 P**KK**P**RH**QF	*C. elegans* sex-determining **Tra-1** protein; zinc finger; peaks in the second larval stage	Zarkower and Hodgkin [1992]
E**KRKK**E**R**N LL**RRLKK**EVE EPLG**RIRQKKRV**Y_2D_2 (EDAI**KKRR**EA**RERRRLRQ**) D**K**ETTAS**RSKRR**SS**RKKRT** ES**KKKK**P**K**L **KK**TAA**KKTKTK**S	Yeast **NPS1** transcription protein factor (1359 aa) involved in cell growth control at G2 phase; has a catalytic domain of protein kinases	Tsuchiya et al. [1992]
QRKRQKL **KAKKQK** **LRRKRQK**	Human **243** transcriptional activator (968 aas), induced by mitogens in T cells; N-terminal half is homologous to oncoprotein Rel and *Drosophila* Dorsal protein involved in development; the C-terminal half contains repeats found in proteins involved in cell-cycle control of yeast and tissue differentiation in *Drosophila*	Bours et al. [1990]
URDI**RRRGK**N**K**V QN**CRKRL**LE	Mouse **NF-E2** (45 kD), an erythroid transcription factor from mouse erythroleukemia (MEL) cells; involved in globin gene regulation. Binds to AP-1-like sites; homology to Jun B, GCN4, Fos, ATF1, and CREB in the basic region/leucine zipper (see Fig 2)	Andrews et al. [1993]
2. Group θθθxθθθ		
D**KIRRK**N A**RK**T**KKK**I	Human **glucocorticoid receptor**	Arriza et al. [1987]
473D**KIRRK**NCP EA**RK**T**KKKIK**GIQ	Mouse and human **GR** (glucocorticoid receptor)	Danielsen et al. [1986]
3. Group θθθxθ		
Y**R**V**RR**E**R**N V**RK**S**R**D**K**A D**RLRKR**VE	**C/EBP** (CCAAT/enhancer binding protein); functions in liver-specific gene expression	Landschulz et al. [1988]
D**KIRRK**N A**RKSKK**L	Human **mineralocorticoid receptor**	Arriza et al. [1987]
D**KIRRK**N G**RKFKK**F	Human **PR** (progesterone receptor)	Kastner et al. [1990]
EEVQ**RKRQK**LMP	Human and mouse **NKκB** 105 kD precursor of p50 (968 aas) (first R is at 361 position)	Ghosh et al. [1990]

(*Table II continued on next page.*)

TABLE II. Transcription Factors With "a Typical" Putative NLS* (*continued*)

NLS and flanks	Protein factor and features	Reference
EEVQ**RKRQK**L	Human **NK-κB p50** (DNA-binding subunit). Identical to protein KBF1, homologous to *rel* oncogene product. NF-κB p65 also contains a NLS (Table IIIa)	Kieran et al. [1990]; Blank et al. [1991]
G**K**T**R**T**RK**Q A**RRK**S**R**D	Human **TEF-1** (SV40 transcriptional enhancer factor 1); 426 aa	Xiao et al. [1991]
Q**R**K**E**R**K**SKS T**K**S**K**T**KRK**L	Rat, mouse, human **IRF-1** (interferon regulatory factor-1); induced in lymphoma T cells by the pituitary peptide hormone prolactin; regulates the growth-inhibitory interferon genes	Yu-Lee et al. [1990]
G**K**C**KKK**N	Ehrlich ascites **S-II** transcription factor; a general factor that acts at the elongation step	Hirashima et al. [1988]
E**RSKKRS**RE	Tobacco **TAF-1** transcriptional activator	Oeda et al. [1991]
E**R**E**L**K**R**E**K**R**K**Q A**RRS**R**LRK**Q		
Y**K**L**D**H**MRRR**IETDE	*Drosophila* **TFIIE**α (433 aa), a general transcription factor for RNA polymerase II. Composed of subunits α and β	Peterson et al. [1991]
D**K**N**RRK**S I**R**K**D**R**R**G I**KRSKK**N	Human **ER** (estrogen receptor); 595 aa	Green et al. [1986]
EQ**RRHR**IE TT**R**AE**KKR**LL ID**KKRS**KEAKE	Yeast **ADA2** (434 aa), a potential transcriptional adaptor required for the function of certain acidic activation domains	Berger et al. [1992]
EAAL**RRKIRTISK**	Yeast ***GCN5*** gene product (439 aa), required for the function of GCN4 transcriptional activator and for the activity of the HAP2-3-4 complex	Georgakopoulos and Thireos [1992]
4. Group θθxθθ		
N**KKMRRNRF** N**RRK**x_4**RQK**	Mouse **LFB3**	De Simone et al. [1991]
T**KKGRRNRF** N**RRK**x_4**RHK**	Mouse **LFB1**	De Simone et al. [1991]
N**KKMRRNRFK**	Rat **vHNF1-A**	Rey-Campos et al. [1991]
N**KKMRRNR**	Murine **HNF-1**β	Mendel et al. [1991]
T**KKGRRNRF**	Mouse **HNF-1**	Kuo et al. [1990]
N**KKMRRNRF**	Human **vHNF1**	Bach et al. [1991]
T**KKGRRNRF**	Rat liver **HNF1**	Baumhueter et al. [1990]
LRRQKRFK QQ**H**$_3$**S**H$_4$Q(?)	Rat **HNF-3**β	Lai et al. [1991]
LRRQKRFK	Rat **HNF-3**γ	Lai et al. [1991]
LRRQKRFK	Rat **HNF-3**α	Lai et al. [1991]
L**K**E**K**E**RK**A M**KKARK**V	Rat **DBP** a protein factor that binds to the D site of the albumin gene promoter	Mueller et al. [1990]

(*Table II continued on next page.*)

TABLE II. Transcription Factors With "a Typical" Putative NLS* (*continued*)

NLS and flanks	Protein factor and features	Reference
P**R**R**E**R**R**Y	Rat **AT-BP1**; highly acidic domain. Two zinc fingers, binds to the B-domain of α_1-antitrypsin gene promoter and to the NF-κB site in the MHC gene enhancer	Mitchelmore et al. [1991]
D**RR**V**RK**G**K**V	A 19 kD *Drosophila melanogaster* nonhistone associated with heterochromatin	James and Elgin [1986]
S**K**HG**RR**A**RR**LDP	Murine EBF (early B-cell factor) of 591 aa. Regulates the pre-B and B lymphocyte-specific *Mb-1* gene; expressed in pre-B and B-cell lines but not in plasmocytomas, T-cell and nonlymphoid cell lines	Hagman et al. [1993]
G**RR**T**RR**E	Human **Sp1**	Kadonaga et al. [1987]
DEQ**K**R**A**E**K**K**A**KE I**RR**I**HK**VIRP LL**RRLKK**DVE	Yeast SNF2, a transcriptional regulator of many genes	Tsuchiya et al. [1992]
5. Group θxθθxθ		
A**K**A**K**A**KK**A Y**K**M**R**R**E**R**N** V**R**K**S**R**D**K**A**	Mouse **AGP/EBP** (87% similarity to C/EBP), ubiquitously expressed	Chang et al. [1990]
A**K**A**K**A**KK**A Y**K**M**R**R**E**R**N** V**R**K**S**R**D**K**A**	Rat **LAP**, a 32-kD liver-enriched transcriptional activator, also present in lung, with 71% sequence similarity to C/EBP; leucine zipper; accumulates to maximal levels around birth	Descombes et al. [1990]
Y**R**Q**RR**E**R** V**KK**SRLK**SK**Q**K**	**Ig/EBP-1** (immunoglobulin gene enhancer-binding protein); forms heterodimers with C/EBP	Roman et al. [1990]
EDPE**K**E**K**R**I**K**E**LE M**RRK**V	Mouse **c-Myb**	Howe et al. [1990]
DYY**KV**K**RP**K**T**D G**R**AR**GRR**HQ F**R**YR**K**I**K**DIY	*Drosophila* ***eyes absent*** protein (760 aas), a nuclear protein that functions in early development to prevent programmed cell death and to allow the event that generate the eye to proceed; mutations cause programmed cell death of eye progenitor cells	Bonini et al. [1993]
6. Group θxθxθθ		
A**K**A**K**A**KK**A	Rat **IL-6DBP** interacting with interleukin-6 responsive elements; has a leucine zipper domain	Poli et al. [1990]
D**K**R**Q**R**N**R**C** F**k**r**t**i**rk**D	Mouse **H-2RIIBP** (MHC class I genes H-2 region II binding protein); member of the nuclear hormone receptor superfamily	Hamada et al. [1989]
F**k**r**t**i**rk**D D**K**R**Q**R**N**R**C**	Chicken **RXR**, related to RAR (retinoic acid receptor), a nuclear protein factor from the thyroid/steroid hormone receptor family	Rowe et al. [1991]
V**KS**K**AKK**T Y**K**I**R**R**E**R**N** V**R**K**S**R**D**K**A**	Human **NF-IL6** (345 aa); specifically binds to IL1-responsive element in the IL-6 gene; leucine zipper; homology to C/EBP	Akira et al. [1990]
Q**KK**N**R**N**K**C	Mouse **PPAR** (peroxisome proliferator activated receptor)	Issemann and Green [1990]
7. Group θθ0xxθθ		
EQI**RK**L**V KKH**G	Yeast **RAP1** it binds regulatory sites at yeast mating type silencers.	Shore and Nasmyth [1987]
F**RR**SM**K**R**K**A	Human **vitamin D receptor** (427 aa)	Baker et al. [1988]

(*Table II continued on next page.*)

TABLE II. Transcription Factors With "a Typical" Putative NLS* (*continued*)

NLS and flanks	Protein factor and features	Reference
8. Group θθxxθθ		
LKRHQRRH	Mouse **WT1** (the murine homolog of human Wilms' *tumor predisposition gene WT1*)	Buckler et al. [1991]
LKRHQRRH	Human **WT33** (Wilms' tumor predisposition)	Call et al. [1990]
9. Group θθθxxθ		
LKESKRKYDE	Yeast **SWI3** 99 kD, highly acidic protein; global transcription activator	Peterson and Herskowitz [1992]
EVLKVQKRRIYD	Human **RBAP-1** (retinoblastoma-associated protein 1) factor (412 aa); a protein that binds to the pocket *(functional domain) of the retinoblastoma (RB)* protein involved in suppression of cell growth (tumor suppressor); the transcription factor E2F, implicated in cell growth, binds to the same pocket of RB	Kaelin et al. [1992]; Helin et al. [1992]

*Transcription factors with "a typical" putative core NLS include nuclear proteins that interact with the regulatory regions of the various genes and contain four arginines and lysines within an hexapeptide segment. These protein factors are furthermore classified into several groups in order of decreasing charge concentration, θθθθ, θθθxθθ, θθθxθ, θθxθθ, θxθθxθ, θθxθxθ, θθθxxθθ, and θθxxθθ, where θ is arginine (R) or lysine (K) depending on the distribution of positive charges in the protein sequence. When more than one putative NLS are present in the protein sequence, the classification is according to the "strongest NLS" with the highest concentration of positive charges. Hexapeptide stretches containing one aspartic or glutamic acid in addition to the four arginines and lysines, and considered to be weak NLS [Hannik and Temin, 1989] are dotted-underlined. Hexapeptide motifs of the type θθxxθθ that bear resemblance to the KKGSKK motif from the *Xenopus* histone H2B that has a low potency in specifying nuclear localization [Goldfarb et al., 1986] are in lower case.

related proteins, dorsal and rel. A putative NLS is also conserved in a homologous region among thyroid hormone receptor, retinoic acid receptor, and related proteins (Fig. 4). These protein molecules are cytoplasmic and, upon binding of the hormone, are imported into the nucleus and bind to the regulatory regions of a distinct class of genes [see Forman and Samuels, 1990].

A class of protein transcription factors (11 out of 117 or 9.4%) did not possess a typical core NLS. These are listed in Table III. It is noteworthy that the NLS of the yeast MATα2 protein does not possess a high concentration of positively charged residues either [Hall et al., 1984]. It is suggested that such proteins may acquire a NLS during their folding by the juxtaposition of positively-charged residues. However, it is proposed that several of the nuclear proteins listed in Table III may enter the nucleus via strong cross-complexation in the cytoplasm with a protein transcription factor possessing a typical NLS. Formation of strong complexes is typical between several transcription protein factors, like TFIIE-α with TFIIE-β [Ohkuma et al., 1991; see Table III], and NF-κB p50 and p65 subunits [Kieran et al., 1990]. Protein-protein association plays an important role in the formation of protein-protein-DNA complexes in the regulatory regions of genes.

Our hypothesis that four arginine/lysine residues within a hexapeptide constitute a core NLS would be supported if the majority of cytoplasmic proteins do not have such motifs. We have searched 101 non-nuclear proteins randomly compiled from the literature. Of these, 32 were cytoplasmic, 41 were cell membrane, 16 were extracellular, one was a protein of the lumen of the ER, 3 were lysosomal and 8 were nucleus-encoded mitochondrial proteins (Table VI). Out of 101 non-nuclear proteins, 48 or 47.5% had a NLS-like peptide and are listed in Table IV. Of the 32 cytoplasmic proteins examined, about half had a NLS; however, the vast majority of these proteins were tightly associated with membrane proteins such as the chicken αN-catenin, a human tyrosine kinase, the bovine phosphatidylinositol 3-kinase, and the yeast *GPA2* gene product (Table IV). Other cytoplasmic proteins found to contain a NLS were associated with large cellular structures such as the meromyosin component of contractile fibers and the rabbit SMHC-29. We propose that the presence of a NLS in such cytoplasmic proteins is unable to target them to nuclei because of their strong commitment and association to other subcellular structures.

Other non-nuclear proteins that possess NLS-like core hexapeptides are either integral parts

Protein factor	Basic region	leucine zipper
Myn	DKRAHHNALERKRRDH	LKRQNALLEQQVRALEKAR
Max	DKRAHHNALERKRRDH	LKRQNALLEQQVRALEKAR
Mad	SSRSTHNEMEKNRRA	LQREQRHLKRQLEKLGIER
c-Myc	DKRRTHNVLERQRRNE	LRKRREQLKHKLEQLRNSC
N-Myc	ERRRNHNILERQRRND	LQARQQQLLKKIEHARTC
L-Myc	TKRKNHNFLERKRRND	LRCRQQQLQKRIAYLSGY
AP-4	IRREIANSNERRRMQS	LKRFIQELSGSSPKRRRAE
TFE-3	QKKDNHNLIERRRRFN	LEQANRSLQLRIQELELQ
TFE-B	QKKDNHNLIERRRRFN	LEMTNKQLWLRIQEL
Myo D1	DRRKAATMRERRRLSK	
Myogenin	DRRRAATLREKRRLK	
Lc	TGTKNHVMSERKRREK	
CBF-1	QRKDSHKEVERRRREN	
USF	KRRAQHNEVERRRRDK	
da	ERRQANNARERIRIRD	
E12	ERRVANNARERLRVRD	
E2-2	ERRMANNARERLRVR	
E47	ERRMANNARERVRVR	
Pho 4	DKRESHKHAEQARRN	
Lyl-1	ARRVFTNSRERWRQQN	
twist	NQRVMANVRERQRTQS	
EMC	RIQRHPTHRGDGENA	
ID	LPALLDEQQVNVLLY	

Fig. 1. Putative NLS in the basic and leucine zipper regions of Myc, Max, MyoD1, and other homologous proteins. The alignment of sequences is from Carr and Sharp [1990]; Beckmann et al. [1990]; Gregor et al. [1990]; Blackwood and Eisenman [1991]; Prendergast et al. [1991]; and Ferré-D'Amaré et al. [1993]. Double-underlined hexapeptides indicate, according to our proposal, NLS. The dotted-underlined motif indicates a "weak" NLS motif because of the presence of E (glutamic acid).

of organelles, such as mitochondria, or are tightly associated with a cell membrane or endomembrane (Table IV). We suggest that this association makes them unavailable for nuclear import even if the NLS-like core peptide is functional and correctly exposed and available for binding to transporter proteins. The signal peptide in the N-terminus will transfer extracellular or lysosomal proteins to the lumen of the endoplasmic reticulum. Similarly, the presence of trans-

Protein	Basic region
GCN4	DPAALKRARNTEAARRSRARKLQ
c-JUN	IKAERKRMRNRIAASKCRKRKLE
JUN B	IKVERKRLRNRLAATKCRKRKLE
v-JUN	IKAERKRMRNRIAASKSRKRKLE
BP1	EKRRKFLERNRAAASRCRQKRKV
CREB	RKREVRLMKNREAARECRRKKKE
C/EBP	NEYRVRRERNNIAVRKSRDKAKQ
c-FOS	EKRRIRRERNKMAAAKCRNRRRE
FRA1	RRRVRRERNKLAAAKCRNRRKE
SKN-1	RKIRRRGKNKVAARTCRQRRTD
CNC	RDIRRRGKNKVAAQNCRKRKLD
NF-E2 (p45)	RDIRRRGKNKVAAQNCRKRKLE
BZLF1	DSELEIKRYKNRVASRKCRAKFKQ
ATF-1	KREIRLMKNREA RECRRKKKE
dJRA	KLERKRQRNRVAASK CRKRKLE
dFRA	KRAVRRERNKQAAAR CRKRRVD
BBF-2	KKIRRKIKNKISA QESRRKKKE
DBP	EKYWSRRYKNNEAAKRSRDARRL

Fig. 2. Putative NLS in the basic domain of Fos, Jun, CREB, C/EBP, and related proteins. The alignment of sequences is from Mueller et al. [1990], Abel et al. [1992], and Andrews et al. [1993].

membrane domains on cell membrane proteins will direct such proteins into the cell membrane or intracellular membranes; the NLS-like core peptide in this case is either to the cytoplasmic side and even though a transporter protein can bind it, the protein will remain as an integral part of the membrane, or the NLS is exposed on the cell surface.

Few of the cytoplasmic proteins found to contain a NLS-like peptide have one or two negatively-charged amino acids within the putative NLS hexapeptide. It is noteworthy that a similar hexapeptide KKKDLK occurs in chicken v-Rel oncoprotein (amino acid position 54) but does not appear to have NLS function [Hannik and Temin, 1989]. We suggest that hexapeptides containing any negatively charged amino acids are weak NLS. However, the hexapeptide core of NLSs can be flanked by negatively charged amino acids [Boulikas, 1993].

DISCUSSION

Is an hexapeptide with four (K + R) a core NLS? A recent survey of the literature [Boulikas, 1993] led us to propose that a consensus core NLS is composed of 4 or more arginines/lysines within a hexapeptide motif. Assigning a peptide as a NLS depends on the method followed for NLS identification (gene fusion between a gene segment coding for NLS and the

Protein species	Basic region	Position 90-97
Drosophila **dorsal**	MDSDPAHL**RRKRQK**T	V**KKK**DIEA
NF-κB p50 (KBF1)	EI**K**D**K**EEVQ**RKRQK**L	VT**KKK**VFE
NF-κB p65	DTDD**R**H**R**IEE**KRKR**T	V**KKR**DLEQ
mouse **rel**	DE**K**DAYAN**KSKKQK**T	V**KKK**EVKG
human **c-rel-A**	DE**K**DTYGN**KAKKQK**T	V**KKK**EVKE
turkey **rel**	DEEDPSGN**KAKRQR**S	V**KKK**DLKE
v-Rel	DEEDPSGN**KAKRQR**S	

Fig. 3. Putative NLS in the basic region of the p50 and p65 subunits of NF-κB, dorsal and rel proteins. Sequences to the N-side of putative NLS are acidic. The alignment of homologous regions is from Kieran et al. [1990]; Nolan et al. [1991]; and Ruben et al. [1991].

E**RRRK**EE	hT3R-α1 (human thyroid hormone receptor-α1)
E**KRRR**EE	hT3R-β
E**RRRK**EE	human *erbA*-α2
E**RRRK**EE	v-*erb*A
V**RKRR**P	human RAR-α (retinoic acid receptor-α)
I**RKRR**P	hRAR-β
A**RRRR**P	hRAR-γ

Fig. 4. Conservation of putative NLS in an homologous basic region of nuclear hormone receptors. The alignment of sequences is from Forman and Samuels [1990].

gene coding for a cytoplasmic protein, cross-linking of a peptide to a cytoplasmic protein, mutagenesis of the NLS region of a normally nuclear protein). NLS identification is also dependent on the type of cytoplasmic protein molecule chosen to tag a NLS (pyruvate kinase, β-galactosidase, human serum albumin, immunoglobulins, etc.) and method of detection of nuclear import (microinjection of proteins into Xenopus oocytes followed by immunostaining, microinjection of DNA constructs, mutagenesis, etc.). It also greatly depends on the particular protein structure since a number of NLS are today-known that are reconstituted by protein folding from half or bipartite NLS [see Boulikas, 1993].

The primary structures of protein transcription factors, predicted from cDNA cloning and sequencing, were screened for the presence of 6 amino acid motifs with 4 or more positively charged amino acids (K + R). Such motifs could constitute core elements of signals for the transport of protein transcription factors across the nuclear pore complex to the nucleoplasm and are presumably recognized by specialized transporter protein molecules. This search demonstrates that the vast majority of protein factors examined possess such motifs often occuring in the basic region (Figs. 1–3), but also in other regions of nuclear proteins. Types θθθθ, θθθxθ, θθxθθ, and θθxθxθ of NLS-like peptides are over-represented in this random sample (Table V).

TABLE III. Protein Factors That Lack a Typical NLS*

NLS and flanks	Features	Reference
1. Group RRR		
DHM**RRR**IETDE**R** **K**FD**RK**QL**R** (but may enter nucleus via its tetramerization with TFIIE-β into an $\alpha_2\beta_2$ complex)	Human **TFIIE**-α (general transcription factor)	Ohkuma et al. [1991]
DLDDF**RRR**G **IKNKR**x_5**K**x_6**KK** **K**Ex_5**RKEKE**Ex_3**K**	Human **PRDI-BF1** (repressor of IFN-β gene expression)	Keller and Maniatis [1991]
2. Group θθxθ		
Rx_6**RTRR**x_3**R**x_3**R**	**RFX**	Reith et al. [1990]
Rx_5**KDKR** **RLKK**CF**R**	**HNF-4** (member of the steroid hormone superfamily)	Sladek et al. [1990]
V**RGRGRGK**Y I**R**C**KK**P L**K**K**H**I**R**T T**KHMKSK**A	Human **HIV-EP1** (HIV1 enhancer-binding protein 1)	Maekawa et al. [1989]
IRNEP**RRIK**I V**RGRGRGK**Y (Does it dimerize with AT-BP1 to enter the nucleus?)	Rat **AT-BP2;** highly acidic domain. Two zinc fingers. Binds to the B-domain of α_1-antitrypsin gene promoter and to the NF-κB site in the MHC gene enhancer	Mitchelmore et al. [1991]
YK**AKR**E **RNQK**x_7**KCRD** **KMSRALR** **K**x_2**R**x_2**R**x_3**R**x_4**RK**x_4**R**x_6 **R**x_5**K**x_2**R**x_5**R**	*Drosophila* ***yan*** gen product (688 aa) with an ETS DNA-binding domain; negative regulator of photoreceptor development. Accumulates in the nuclei of undifferentiated cells during the early stages of eye development	Lai and Rubin [1992]
V**KHIRDYK**	Mouse **Pax-1** (Homeotic protein)	Chalepakis et al. [1991]
KVPQKKVKL(D/E)$_{25}$ **KK**SV**R**DTPA**K** **K**x_2**K**x_4**RSK**x_5**KK**x_2**K**x_2**K** x_2**K**	Rat **B23**, a highly acidic nucleolar phosphoprotein (292 aa, 38 kD, pl = 5.1); associated with nucleolar structure, binds single-stranded RNA and DNA; does not possess the 90 aa RNA-binding domain or "RNP consensus" of other RNP proteins	Chang et al. [1988]
3. Group θθx_nθθ		
L**KDRFK**LSLK Y**KNK**DLG**K**I E**KK**VISI**KK**E	Mouse **IREBF-1** (interferon response element binding factor-1) 95% similarity to human YB-1; within a 79-aa range one side of the α-helical region contains a preponderance of hydrophobic aas and the other side contains hydrophilic aas; this type of structure provides a strong hydrophobic force for protein-protein interaction	Yan and Tamm [1991]
RARx_6**KR**x_6**R**I**R**EP**R** **RK**YARVVQ**K** **K**GF**RK**T	Mouse **TFIID**	Tamura et al. [1991]

*Protein factors that lack a typical NLS include protein transcriptional regulators which have a maximum concentration of arginines and lysines of 3 or 2 on a stretch of 6 amino acids, 4 (K+R) over a stretch of 7 or more amino acids, or a more weak distribution of positive charges over a longer stretch of the polypeptide. In spite of the possibility that a class of NLS are similar to those found in yeast MATα2 protein with only two (K+R) [see Boulikas, 1993 (Table 3)], we suggest that a fraction of these proteins acquire a stronger concentration of a positively charged domain by protein folding and that a second fraction are transported via their strong complexation in the cytoplasm with a protein transcription factor possessing a NLS.

TABLE IV. NLS-Like Peptides in Non-Nuclear Proteins*

Putative NLS	Protein	Location	Reference
a. Cytoplasmic			
G**KKK**G**RSKK**A DD**RRKR**I **DLRRQLRK**A	Chicken α**N-catenin,** a neural protein that regulates cadherin-mediated cell adhesion in the nervous system	Cytoplasmic but associated with the cytoplasmic domain of transmembrane N-cadherin and E-cadherin proteins	Hirano et al. [1992]
H**kryl**k**k**l (tyk2) E**krfl**k**r**l (JAK-1)	Human **tyrosine kinase** associated with the interferon α/β receptor	Cytoplasmic but associated with a cell membrane receptor	Velazquez et al. [1992]
I**RKK**T**RS** (position 225) H**KKKK**F (position 940)	Bovine **phosphatidylinositol 3-kinase** 110 kD catalytic subunit (1068 aa); functions in signal transducing pathways; associated with receptor protein kinases to form multiprotein complexes; involved along with phospholipase C in phosphoinositide metabolism	Cytoplasmic but associated with the cell membrane colony-stimulating factor 1 receptor	Hiles et al. [1992]
G**QRSERKK**WI 299-308	Yeast G protein ***GPA2*** gene product or guanine nucleotide-binding regulatory (G) protein (449aa, 50 kD); may be involved in regulation of cAMP levels and in signal transduction in yeast; possibly interacts with the mating factor receptors STE2 and 3.	Cytoplasmic but associated with a receptor molecule on cytoplasmic membrane	Nakafuku et al. [1988]
N**RRK**L**QR**ELDE 445-455	Rabbit light meromyosin of smooth muscle **myosin heavy chain** (484 aa), NLS is not conserved in four related polypeptides in rat, *Dictyostelium* and a nematode	Cytoplasmic; component of a large polymeric structure	Nagai et al. [1988]
S**ERAKKR**L**ESE** E**KDRKK**YE D**RSRKR**LE EL**KEYRKK**FG DY**KRAKK**E	*Dictyostelium discoideum* skeletal muscle **meromyosin of myosin heavy chain**	Cytoplasmic component of a large polymeric structure	Nagai et al. [1988]
L**RIKKK**LE	Nematode *unc-54;* skeletal muscle **meromyosin of myosin heavy chain**	Cytoplasmic component of a large polymeric structure	Nagai et al. [1988]
$E_2AE_3x_6$N**RRK** L**QR**ELDE	Rabbit **SMHC-29.** Smooth muscle myosin heavy chain (485 aas); extremely acidic, α-helical protein; involved in conversion of chemical energy into mechanical work	Cytoplasmic	Nagai et al. [1988]

(*Table IV continued on next page.*)

TABLE IV. NLS-Like Peptides in Non-Nuclear Proteins* (*continued*)

Putative NLS	Protein	Location	Reference
D**K**E**P**R**K**K (this is the C-terminus)	Rat 42**A** adrenal pheochromocytoma cell protein (101 aas) induced by NGF (nerve growth factor), related to calcium-*binding proteins*	Cytoplasmic	Masiakowski and Shooter [1988]
D**RR**SG**KK**LED	Human **EF-1**α (elongation factor-1α) in protein chain elongation. Induced by vitamin A in cultured human tracheobronchial epithelial cells; catalyzes the binding of aminoacyl-tRNA to the ribosome under hydrolysis of GTP	Cytoplasmic? ER membrane?	Ann et al. [1988]
H**RRKK**N T**K**D**KRR**L**K**Q I**K**V**RK**C**K**S E**KRKK**N	Murine **2-5A-dependent RNAase,** induced by interferon. In mRNA degradation; it requires 2-5A unusual oligoadenylates with 2′, 5′-phosphodiester linkages produced from ATP by a family of synthetases and requiring dsRNA	Cytoplasmic	Zhou et al. [1933]
I**R**R**K**E**T**K**D**	Human **UMP** synthase (480 aas, 52,199 D) In de novo pyrimidine biosynthesis	Cytoplasmic	Suttle et al. [1988]
L**RRK**T**K**ED DEE**K**H**KK**M**K**L**K**I	Human **2-5A-dependent RNAase,** induced by interferon. In mRNA degradation, it requires 2-5A unusual oligoadenylates with 2′, 5′-phosphodiester linkages produced from ATP by a family of synthetases and requiring dsRNA	Cytoplasmic	Zhou et al. [1993]
M**K**Q**K**G**KK** (this is the C-terminus)	Porcine p11 **calcium binding protein** (97 aas)	Cytoplasmic	Masiakowski and Shooter [1988]
T**kk**am**kk**A	Human **CFa** cystic fibrosis antigen (93 aas)	Cytoplasmic	Masiakowski and Shooter [1988]
V**KKRK**Q_2D	Chimeric oncoprotein product of ***DBL*** transforming gene recombined with an unrelated segment of human DNA	Equally distributed between cytoplasmic and crude membrane fractions	Eva et al. [1988]
V**R**I**E**K**R**K**Y** L**K**Q**H**R**RK**E E**RNKKK**Q	Human **5-lipoxygenase**	Cytoplasmic	Dixon et al. [1988]

(*Table IV continued on next page.*)

TABLE IV. NLS-Like Peptides in Non-Nuclear Proteins* (*continued*)

Putative NLS	Protein	Location	Reference
b. Transmembrane			
PI**RRKR**SIE	Human PDGF A-chain (platelet-derived growth factor) (30 kD); glycoprotein, potent mitogen for connective tissue cells	Cell surface receptor, transmembrane tyrosine kinase	Bonthron et al. [1988]
L**KSRRFKM** QLL**RKKARRM**	Human **RCNC2,** a subunit of the cyclic GMP-gated cation channel. Component of the retinal rod	Transmembrane	Chen et al. [1993]
(E**EKKKKKKEK KSK**) **EKKKKKK**DKE (**EKKKKEEKSK** DKKE$_3$)	Human **RCNC1,** a subunit of the retinal rod channel; rods respond to light with a membrane hyperpolarization by cGMP hydrolysis and closure of a cGMP-gated channel that is open in darkness	Transmembrane	Chen et al. [1993]
ARRHS**KR**L	*Drosophila* **raf** protein kinase	Transmembrane	Kieber et al. [1993]
DD**KKK**VKN	Human **A-Raf** protein kinase	Transmembrane	Kieber et al. [1993]
CL**KKKR**DERP	Human **B-Raf** protein kinase	Transmembrane	Kieber et al. [1993]
L$_3$C**RLRRLRAR**	Human **platelet glycoprotein Ib** β-chain	Transmembrane	Lopez et al. [1988]
T**KRKKQRSRRN**	Human **T11** (T-cell surface glycoprotein)	Transmembrane	Diamond et al. [1988]
F**KRKK**P C**KRRKR**N	Murine **T11** (T-cell surface glycoprotein)	Transmembrane	Diamond et al. [1988]
Q**R**D**E**R**R**K**W**	Human α-subunit of the guanine nucleotide binding protein Gs	Transmembrane	Kozasa et al. [1988]
AR**LLKRK**Q N**K**E**E**K**K**K**Y**	*Plasmodium* yoelii (malaria protozoan) **surface antigen** (230 kD)	Cell membrane	Burns et al. [1988]
C**RKKR**I	Human **IgG Fc** receptor; NLS is next to a hydrophobic segment.	Transmembrane	Hibbs et al. [1988]
V**RKKPK**L**K**E G**KKRKRKR**L**K**P	**PDGF A** chain (platelet-derived growth factor A chain); cell surface receptor membrane of transmembrane tyrosine kinase family; mitogen for connective tissue cells	Transmembrane	Bonthron et al. [1988]
Q**R**S**E**R**K**K**W**	Yeast **GPA2** guanine nucleotide-binding regulatory protein	Cell membrane	Nakafuku et al. [1988]

(*Table IV continued on next page.*)

TABLE IV. NLS-Like Peptides in Non-Nuclear Proteins* (*continued*)

Putative NLS	Protein	Location	Reference
NRYRFKKL (two bulky groups)	Arabidopsis **CHL1,** an electrogenic nitrate transporter; when mutated confers resistance to herbicide chlorate and results in decreased nitrate uptake	Transmembrane (12 membrane-spanning segments)	Tsay et al. [1993]
T**KKKK**A A**KR**MA**RK**N	Chicken α-subunit of the **(Na^+ + K^+)-ATPase**	Integral membrane protein	Takeyasu et al. [1988]
S**RRRK**L	Rat **P450 f**	Transmembrane	Nelson and Strobel [1988]
G**RSRR**PRL	Human **p450 P1**	Transmembrane	Nelson and Strobel [1988]
G**RARR**PR**F**	Rabbit **P450 6**	Transmembrane	Nelson and Strobel [1988]
G**RERR**PRL	Human **P450 4**	Transmembrane	Nelson and Strobel [1988]
G**R**G**RRER**G F**RKK**Y**RI**	Bovine **atrial natriuretic peptide (ANP) receptor precursor** (537 aas) with N-terminal membrane translocation signal; ANP is a hormone synthesized and secreted by the atria of the heart affecting cardiovascular homeostasis	Membrane	Fuller et al. [1988]
G**RR**N**RK**Q	*Xenopus* **IP_3 receptor**	Cell membrane	Kume et al. [1993]
QRKQKQRI	(2,693 aa)		
E**A**R**R**R**C**D**R**DS**R**	Tumor suppressor protein **Gas1;** Gas 1 inhibits the serum-induced $G_0 \rightarrow S$ phase transition; expression of Gas 1 is increased by growth arrest	Integral plasma membrane	Del Sal et al. [1992]
c. Extracellular			
EF**KRKK**PPF 69-77 IC**KRRKR**NDEE 233-243	Mouse **T11** T-cell surface glycoprotein (50 kD); contains a N-terminal leader peptide	Extracellular or cell surface	Diamond et al. [1988]
(IT**KRKKQRSR** **R**NDEE) 210-224	Human **T11** T-cell surface glycoprotein (327aa, 50 kD); contains a N-terminal leader peptide	Extracellular or cell surface	Diamond et al. [1988]
G**RK**P**RY**K**Y**	*Drosophila* **FMRF**amide polyprotein precursor (342 aas) with a 25 residue hydrophobic leader sequence giving rise to $FMRFNH_2$ at the skeletal neuromuscular junction; released into circulation of insects from neurohemal organs	Extracellular	Schneider and Taghert [1988]

(*Table IV continued on next page.*)

TABLE IV. NLS-Like Peptides in Non-Nuclear Proteins* (*continued*)

Putative NLS	Protein	Location	Reference
EKKKLKC	Human plasma **Zn-α_2-glycoprotein** (276 aas); unknown function, closely related to MHC antigens in amino acid sequence and domain structure; the EKKKLKC sequence is not conserved among MHC antigens	Extracellular	Araki et al. [1988]
MRRRRKRQ DDRRTRGRG TRRRDMKL	*Drosophila* **connectin,** a cell adhesion molecule expressed on motoneurons and the muscles they innervate	Extracellular (with a 24 aa signal peptide)	Nose et al. [1992]
IRTKRKRKKQ RVKI	Human **LACI** (lipoprotein-associated coagulation inhibitor), a protease inhibitor of plasma	Extracellular (blood plasma)	Wun et al. [1988]
d. Lumen of ER			
EDFKAKKKELE	**BiP** (immunoglobulin heavy chain-binding protein); NLS-like peptide is surrounded by acidic motifs	Lumen of ER	Haas and Meo [1988]
e. Nucleus-encoded mitochondrial proteins			
GKPKRPRS LRRTIKKQRKY KRKAMTKKK	**mt TF1** (mitochondrial transcription factor 1), a nucleus-encoded DNA binding protein (204 aas) HMG-box similar to hUBF	Nucleus-encoded mitochondrial protein	Parisi and Clayton [1991]

*Non-nuclear proteins with NLS-like peptides. These are classified into cytoplasmic, transmembrane embedded into the cell membrane, extracellular exported via the lumen of the endosplasmic reticulum, lumen of the endoplasmic reticulum, and nucleus-encoded mitochondrial proteins.

This finding makes the prediction that NLS-receptor molecules might possess the αααα, αααxα, ααxαα, and ααxαxα motifs where α is D, E, or phosphoserine [Meier and Blober, 1992]. More specifically, E or phosphoserine might complement K, and D might complement R. For example, the ERKKRRRE motif of human ATF-3 (Table I) may need the KDEEDDDK or the phosphorylated form of the KDSSDDDK motif on the receptor-transporter molecule. A detailed analysis of the type of acidic motifs in NLS-receptor molecules and their match to the putative NLS described here is in progress.

It is proposed here that most protein factors listed on Table III lacking a putative typical "core NLS" are imported to the nucleus via their strong crosscomplexation with other transcription factors. Indeed, Tsuneoka and co-workers [1986] have shown that a monoclonal antibody against HMG1 co-migrates with HMG1 into the nucleus.

Fusion of the peptide PAAKRVKLD containing only 3 positively charged amino acids to pyruvate kinase at the gene level induced complete nuclear localization [Dang and Lee, 1988]. Similarly, fusion of KIPIK, containing only two positively charged residues, to β-galactosidase specified nuclear import [Hall et al., 1984]. We suggest that either positively charged residues of pyruvate kinase or β-galactosidase are juxtaposed to these peptides to reconstitute a completely functional NLS or that acidic domains on these cytoplasmic proteins participate in anchoring them to a nuclear transporter protein. Thus, cytoplasmic proteins may contain potential elements for their anchoring to NLS-receptor proteins.

Location of NLS on Protein Transcription Factors

NLS may be unavailable for binding to transporter proteins; the SV40 large T protein NLS is nonfunctional when it is located in a region of pyruvate kinase predicted not to be exposed on the surface [Roberts et al., 1987]. Thus, the NLS needs to be exposed on the surface of the protein and to be available for binding to the nuclear transporter protein molecules. Indeed, the FKQRRIKL peptide motif on the DNA-binding POU-specific domain of Oct-1 predicted from this study to be a core NLS lies at the C-terminus of helix I just before a turn of 3 amino acids and helix II [Dekker et al., 1993]. In addition, nuclear proteins have been identified that are synthesized in the form of precursor molecules and that remain in the cytoplasm presumably because their NLS is hidden. Cleavage of such proteins by specific proteases to mature molecules exposes their NLS, and they are then rapidly transported to the nucleus. An example of this kind is the p50 subunit of the transcription factor NF-κB (50 kD) which is synthesized in the form of a 110 kD precursor with the NLS buried in the protein; proteolytic cleavage giving the 50 kD transcription factor exposes the NLS [Henkel et al., 1992].

Nuclear translocation of factors, presumably by exposure of their hidden NLS or by reconstitution of a functional NLS from two remote half NLS can be triggered by dephosphorylation [Moll et al., 1991], subunit association [Levy et al., 1989], or subunit dissociation [Baeuerle and Baltimore, 1988a,b]. In other cases, binding of hormone to a cytoplasmic protein such as glucocorticoid receptor induces a conformational change that exposes the NLS, or reconstitutes a functional NLS from remote half NLSs and the protein is rapidly transported to the nucleus [Picard and Yamamoto, 1987].

A fraction of the putative NLS peptides identified in transcription factors in this study might be masked by the interaction of the transcription factor with inhibitory polypeptides prohibiting the binding of the transporter protein molecule to the NLS.

We propose here that NLS have two different functions: 1) one in the cytoplasm for the identification of the protein by transporter cytoplasmic or pore proteins and its transportation to the nucleoplasm; 2) another in the nucleus in the sequence-specific binding to DNA or in protein-protein interaction on the regulatory regions of genes.

Loose evolutionary constraints with regard to the position of the NLS in the protein structure

TABLE V. Occurrence of Various Types of NLS Among a Random Sample of Protein Transcription Factors*

Putative NLS type	Frequency	%
θ_9	1	0.4
θ_8	0	0.0
θ_7	2	0.7
θ_6	6	2.2
θ_5	15	5.5
θθθθ	68	25.1
θθθxθθ	13	4.8
θθθxθ	45	16.6
θθxθθ	38	14.0
θxθθxθ	21	7.8
θθxθxθ	37	13.6
θθθxxθθ	3	1.1
θθθxxθ	14	5.2
θθxxθθ	8	3.0
Total	271	100.0

*Types of putative NLS among a sample of protein transcription factors. Out of the 117 proteins examined, the 106 of Tables I and II that were found to contain one or more NLS-like peptides (total 208 peptides) and those shown in Figures 1–4 (total 63 peptides) were used for Table V. Each NLS-like peptide (total 271) was scored for the type of distribution of K and R among the 14 groups shown.

TABLE VI. NLS-Like Peptides in Non-Nuclear Proteins*

Subcellular localization	Without NLS	With NLS	Total	% with NLS
Cytoplasmic	15	17	32	53
Endomembrane, cell membrane	18	23	41	56
Extracellular	10	6	16	37
Mitochondrial or chloroplast	7	1	8	12
Lysosomal	3	0	3	—
Lumen of ER	0	1	1	—
Total	53	48	101	47.5

*Distribution of NLS-like peptides among classes of non-nuclear proteins.

are expected to be imposed if the proposed dual function model for NLS is correct. Thus as Figures 1–4 indicate, putative core NLS peptides may or may not occur on homologous positions in proteins. This loose evolutionary conservation of the position of the core NLS within the polypeptide chain will serve the second function of NLS proposed here, i.e., in DNA binding or in the interaction with other regulatory proteins of the transcription apparatus. One argument in favor of the dual role of NLS is that, whereas the N terminal hydrophobic signal peptide of endoplasmic reticulum trafficking has one single function and is lost after protein is inserted into the lumen of ER, the NLS is retained.

NLS-Binding Proteins

Two proteins of about 60 and 70 kD from rat liver, distributed mainly in the cytoplasm but also in the nuclear interior and nuclear envelope, were shown to interact with the NLS of SV40 large T protein by chemical cross-linking [Adam et al., 1989]. Yoneda et al. [1988] have raised antibodies against the peptide DDDED supposed to be present in nuclear pore or cytoplasmic receptor (transporter) protein molecules that would interact electrostatically with the NLS KKKRK of SV40 large T protein. Affinity chromatography with anti-DDDED antibody or nucleoplasmin NLS has further identified a 69 kD protein in rat liver nuclear pore fraction as a candidate for a nuclear pore receptor-transporter [Imamoto-Sonobe et al., 1990].

Lee and coworkers [1991] have determined the primary structure from cDNA cloning of a yeast protein, NSR1 (nuclear signal receptor) of 384 amino acids, located at the nuclear periphery and involved in the specific recognition of nuclear localization signals. Meier and Blobel [1992] have determined the structure of Nopp140 from rat cells, also a NLS-receptor phosphoprotein that shuttles between the nucleolus and the cytoplasm. Both proteins are characterized by a number of acidic motifs in their sequence representing NLS-binding sites. In addition, positively charged domains in NSR1 and Nopp140 might interact with acidic domains on transcription factors for their nuclear import. Yeast NSR1 has two extensive RNA binding domains [Lee et al., 1991]. One exciting suggestion is that NSR1 might bind RNA and thus the same molecule might be involved in importing nuclear proteins to the nucleus and exporting RNA to the cytoplasm [Lee et al., 1991] thus regulating bidirectional traffic of both proteins and RNA though the pore.

Several other polypeptides have been identified in diverse species from their ability to bind to the NLS of SV40 large T, yeast histone H2B, yeast MATα2, and *Xenopus* nucleoplasmin. Some of these proteins appear to be cytoplasmic, others loosely bound to the pores and others nucleolar or nucleoplasmic [reviewed by Boulikas, 1993]. One evolutionarily conserved NLS-binding protein, immunologically indistinguishable, with a molecular weight of 70 kD was present in yeast, *Drosophila, Z. mays,* and human cells [Stochaj and Silver, 1992]. Some of these proteins have the ability to recognize heterologous core NLS such as the PKKKRKV (SV40 large T), the AKKKKLD (*Xenopus* nucleoplasmin), the GKKRSKA (yeast histone H2B), and the CRLKKLKC (yeast GAL4) [Silver et al., 1989]. It remains to be shown how many NLS-receptor proteins are present in a single cell, and what their specificity is for the heterologous classes of NLS suggested here or found before. Preliminary studies by Finlay and Forbes [1990] suggest the presence of 8–10 different possible NLS-receptor molecules associated with the nuclear pore. NLS-binding polypeptides, shuttling between the cytoplasm and the nucleoplasm or the nucleolus, may be present in addition to those associated with the pore [Borer et al., 1989; Yamasaki et al., 1989; Adam et al., 1989; Meier and Blobel, 1990, 1992; Lee et al., 1991; Li et al., 1992].

NLS in Non-Nuclear Proteins

An important finding of this sample-search for clusters of basic amino acids among proteins is that, whereas virtually all nuclear proteins have one to three or more such putative NLS motifs, about half of cytoplasmic proteins compiled also have such motifs. These cytoplasmic proteins containing a NLS are engaged in their cellular functions in association with membrane proteins or large cytoplasmic structures and are therefore unavailable for nuclear import. A previous attempt by Smith and coworkers [1985] revealed NLS-like peptides in non-nuclear proteins and led them to the conclusion that it is exceedingly difficult to predict a priori a functional nuclear localization signal by simple primary structure inspection.

A higher number of transmembrane, lumen of endoplasmic reticulum, and extracellular proteins were found to have typical NLS-like hexapeptides; we suggest that the N-terminal hydro-

phobic peptide of proteins transported through the endoplasmic reticulum [reviewed by Silver and Goodson, 1989] will direct them away from the nuclear and cytoplasmic compartment regardless of their possession of an NLS-like oligopeptide.

An unusual behavior of the secreted hepatitis B virus (HBV) precore P22 protein has been noted by Ou and coworkers [1989]: although this protein possesses a 19 aa signal sequence that directs it to the ER, upon cleavage of this segment by signal peptidase only 30% of P22 is secreted via the Golgi apparatus; 70% of this protein is released back into the cytoplasm, followed by its translocation to the nucleus. This nuclear targeting of P22 is due to the presence of a NLS. The sequence CLGWLWG was identified as an important constituent of the NLS. Similarly, Lee and coworkers [1987] have determined that the introduction of a charged amino acid within the hydrophobic N-terminal signal peptide of the simian sarcoma virus v-*sis* gene product aborted its normal transport across the lumen of the ER because of the destruction of its signal peptide and resulted in nuclear import of the mutated protein due to the presence of a NLS at a different region of the molecule. These studies support a major conclusion of this study, i.e., that many transmembrane proteins or proteins to be secreted may indeed possess functional NLS-like peptides but the hydrophobic signal N-terminal peptides or transmembrane domains are stronger determinants for their subcellular localization.

Is there a difference between putative core NLS peptides on transcription factors on one hand and NLS-like peptides on non-nuclear proteins on the other? Some NLS-like peptides in non-nuclear proteins contain two bulky amino acids in the hexapeptide such as tyrosine and phenylalanine in the RYRFKK motif of the *Arabidopsis* nitrate transporter CHL1 transmembrane protein [Tsay et al., 1993; Table IV]. None of the hexanucleotide motifs in nuclear putative NLS examined contain two bulky amino acids such as Y, F, W [Tables I–III]. Many other non-nuclear protein NLS-like peptides seem to possess one or two aspartic/glutamic acid residues within the hexapeptide (Table IV), something not favored in functional NLS from experimental data [Boulikas, 1993].

In future studies human disorders may be discovered which are caused by mutations leading to amino acid substitution at the core NLS site of nuclear protein factors, resulting in failure for nuclear localization of a particular nuclear protein.

ACKNOWLEDGMENTS

Special thanks to Emile Zuckerkandl, Bernhard Hirt, and Russ Doolittle for critically reading the manuscript and for their valuable input, to Dawn Brooks for her patience in typing the manuscript and to James Liu for bibliographical searches.

REFERENCES

Abel T, Bhatt R, Maniatis T (1992): A *Drosophila* CREB/ATF transcriptional activator binds to both fat body- and liver-specific regulatory elements. Genes Dev 6:466–480.

Adam SA, Lobl TJ, Mitchell MA, Gerace L (1989): Identification of specific binding proteins for a nuclear location sequence. Nature 337:276–279.

Akira S, Isshiki H, Sugita T, Tanabe O, Kinoshita S, Nishio Y, Nakajima T, Hirano T, Kishimoto T (1990): A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. EMBO J 9:1897–1906.

Andrews NC, Erdjument-Bromage H, Davidson MB, Tempst P, Orkin SH (1993): Erythroid transcription factor NF-E2 is a haematopoietic-specific basic-leucine zipper protein. Nature 362:722–728.

Ann DK, Wu MMJ, Huang T, Carlson DM, Wu R (1988): Retinol-regulated gene expression in human tracheobronchial epithelial cells. J Biol Chem 263:3546–3549.

Araki T, Gejyo F, Takagaki K, Haupt H, Schwick HG, Bürgi W, Marti T, Schaller J, Rickli E, Brossmer R, Atkinson PH, Putnam FW, Schmid K (1988): Complete amino acid sequence of human plasma Zn-α_2-glycoprotein and its homology to histocompatibility antigens. Proc Natl Acad Sci USA 85:679–683.

Arriza JL, Weinberger C, Cerelli G, Glaser TM, Handelin BL, Housman DE, Evans RM (1987): Cloning of human mineralocorticoid receptor complementary DNA: Structural and functional kinship with the glucocorticoid receptor. Science 237:268–274.

Bach I, Mattei M-G, Cereghini S, Yaniv M (1991): Two members of an HNF1 homeoprotein family are expressed in human liver. Nucleic Acids Res 19:3553–3559.

Baeuerle PA, Baltimore D (1988a): Activation of DNA-binding activity in an apparently cytoplasmic precursor of the NF-κB transcription factor. Cell 53:211–217.

Baeuerle PA, Baltimore D (1988b): IκB: a specific inhibitor of the NF-κB transcription factor. Science 242:540–546.

Baker AR, McDonnell DP, Hughes M, Crisp TM, Mangelsdorf DJ, Haussler MR, Pike JW, Shine J, O'Malley BW (1988): Cloning and expression of full-length cDNA encoding human vitamin D receptor. Proc Natl Acad Sci USA 85:3294–3298.

Baldwin AS Jr, LeClair KP, Singh H, Sharp PA (1990): A large protein containing zinc finger domains binds to related sequence elements in the enhancers of the class I major histocompatibility complex and kappa immunoglobulin genes. Mol Cell Biol 10:1406–1414.

Baumhueter S, Mendel DB, Conley PB, Kuo CJ, Turk C, Graves MK, Edwards CA, Courtois G, Crabtree GR (1990): HNF-1 shares three sequence motifs with the POU domain proteins and is identical to LF-B1 and APF. Genes Dev 4:372–379.

Beckmann H, Su L-K, Kadesch T (1990): TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes Dev 4:167–179.

Berger SL, Piña B, Silverman N, Marcus GA, Agapite J, Regier JL, Triezenberg SJ, Guarente L (1992): Genetic isolation of ADA2: A potential transcriptional adaptor required for function of certain acidic activation domains. Cell 70:251–265.

Blackwood EM, Eisenman RN (1991): Max: A helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. Science 251:1211–1217.

Blank V, Kourilsky P, Israël A (1991): Cytoplasmic retention, DNA binding and processing of the NF-κB precursor are controlled by a small region in its c-terminus. EMBO J *10:4159–4167.*

Blank V, Kourilsky P, Israel A (1991): NF-κB and related proteins: Rel/dorsal homologies meet ankyrin-like repeats. Trends Biochem Sci 17:135–140.

Bonini NM, Leiserson WM, Benzer S (1993): The eyes absent gene: Genetic control of cell survival and differentiation in the developing *Drosophila* eye. Cell 72:379–395.

Bonthron DT, Morton CC, Orkin SH, Collins T (1988): Platelet-derived growth factor A chain: Gene structure, chromosomal location, and basis for alternative mRNA splicing. Proc Natl Acad Sci USA 85:1492–1496.

Borer RA, Lehner CF, Eppenberger HM, Nigg EA (1989): Major nucleolar proteins shuttle between nucleus and cytoplasm. Cell 56:379–390.

Boulikas T (1987): Nuclear envelope and chromatin structure. Int Rev Cytol [Suppl] 17:493–571.

Boulikas T (1993): Nuclear localization signals CRC Crit Rev Euk Gene Expr 3:193–227.

Bours V, Villalobos J, Burd PR, Kelly K, Siebenlist U (1990): Cloning of a mitogen-inducible gene encoding a κB DNA-binding protein with homology to the *rel* oncogene and the cell-cycle motifs. Nature 348:76–80.

Buckler AJ, Pelletier J, Haber DA, Glaser T, Housman DE (1991): Isolation, characterization, and expression of the murine Wilms' tumor gene (WT1) during kidney development. Mol Cell Biol 11:1707–1712.

Bürglin TR, De Robertis EM (1987): The nuclear migration signal of *Xenopus laevis* nucleoplasmin. EMBO J 6:2617–2625.

Burns JM Jr, Daly TM, Vaidya AB, Long CA (1988): The 3′ portion of the gene for a *Plasmodium yoelii* merozoite surface antigen encodes the epitope recognized by a protective monoclonal antibody. Proc Natl Acad Sci USA 85:602–606.

Call KM, Glaser T, Ito CY, Buckler AJ, Pelletier J, Haber DA, Rose EA, Kral A, Yeger H, Lewis WJ, Jones C, Housman DE (1990): Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. Cell 60:509–520.

Carr CS, Sharp PA (1990): A helix-loop-helix protein related to the immunoglobulin E box-binding proteins. Mol Cell Biol 10:4384–4388.

Chalepakis G, Fritsch R, Fickenscher H, Deutsch U, Goulding M, Gruss P (1991): The molecular basis of the *undulated/Pax-1* mutation. Cell 66:873–884.

Chan S-K, Mann RS (1993): The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. Genes Dev 7:796–811.

Chang C-J, Chen T-T, Lei H-Y, Chen D-S, Lee S-C (1990): Molecular cloning of a transcription factor, AGP/EBP, that belongs to members of the C/EBP family. Mol Cell Biol 10:6642–6653.

Chang J-H, Dumbar TS, Olson MOJ (1988): cDNA and deduced primary structure of rat protein B23, a nuceolar protein containing highly conserved sequences. J Biol Chem 263:12824–12827.

Chardin P, Courtois G, Mattei M-G, Gisselbrecht S (1991): The KUP gene, located on human chromosome 14, encodes a protein with two distant zinc fingers. Nucleic Acids Res 19:1431–1436.

Chelsky D, Ralph R, Jonak G (1989): Sequence requirements for synthetic peptide-mediated translocation to the nucleus. Mol Cell Biol 9:2487–2492.

Chen T-Y, Peng Y-W, Dhallan RS, Ahamed B, Reed RR, Yau K-W (1993): A new subunit of the cyclic nucleotide-gated cation channel in retinal rods. Nature 362:764–767.

Chida K, Vogt PK (1992): Nuclear translocation of viral Jun but not of cellular Jun is cell cycle dependent. Proc Natl Acad Sci USA 89:4290–4294.

Dang CV, Lee WM (1988): Identification of the human c-*myc* protein nuclear translocation signal. Mol Cell Biol 8:4048–4054.

Danielsen M, Northrop JP, Ringold GM (1986): The mouse glucocorticoid receptor: mapping of functional domains by cloning, sequencing and expression of wild-type and mutant receptor proteins. EMBO J 5:2513–2522.

Dekker N, Cox M, Boelens R, Verrijzer CP, van der Vliet PC, Kaptein R (1993): Solution struture of the POU-specific DNA-bindin domain of Oct-1. Nature 362:852–855.

Del Sal G, Ruaro ME, Philipson L, Schneider C (1992): The growth arrest-specific gene, *gas1,* is involved in growth suppression. Cell 70:595–607.

Descombes P, Chojkier M, Lichtsteiner S, Falvey E, Schibler U (1990): LAP, a novel member of the C/EBP gene family, encodes a liver-enriched transcriptional activator protein. Genes Dev 4:1541–1551.

DeSimone V, De Magistris L, Lazzaro D, Gerstner J, Monaci P, Nicosia A, Cortese R (1991): LFB3, a heterodimer-forming homeoprotein of the LFB1 family, is expressed in specialized epithelia. EMBO J 10:1435–1443.

Diamond DJ, Clayton LK, Sayre PH, Reinherz EL (1988): Exon-intron organization and sequence comparison of human and murine T11 (CD2) genes. Proc Natl Acad Sci USA 85:1615–1619.

Didier DK, Schiffenbauer J, Woulfe SL, Zacheis M, Schwartz BD (1988): Characterization of the cDNA encoding a protein binding to the major histocompatibility complex class II Y box. Proc Natl Acad Sci USA 85:7322–7326.

Dingwall C, Dilworth SM, Black SJ, Kearsey SE, Cox LS, Laskey RA (1987): Nucleoplasmin cDNA sequence reveals polyglutamic acid tracts and a cluster of sequences homologous to putative nuclear localization signals. EMBO J 6:69–74.

Dixon RAF, Jones RE, Diehl RE, Bennett CD, Kargman S, Rouzer CA (1988): Cloning of the cDNA for human 5-lipoxygenase. Proc Natl Acad Sci USA 85:416–420.

Dworetzky SI, Lanford RE, Feldherr CM (1988): The effects of variations in the number and sequence of targeting signals on nuclear uptake. J Cell Biol 107:1279–1287.

Ellinger-Ziegelbauer H, Dreyer C (1991): A retinoic acid receptor expressed in the early development of *Xenopus laevis.* Genes Dev 5:94–104.

Eva A, Vecchio G, Rao CD, Tronick SR, Aaronson SA (1988): The predicted DBL oncogene product defines a distinct class of transforming proteins. Proc Natl Acad Sci USA 85:2061–2065.

Faisst S, Meyer S (1991): Compilation of vertebrate-encoded transcription factors. Nucleic Acids Res 20:3–26.

Fan C-M, Maniatis T (1990): A DNA-binding protein containing two widely separated zinc finger motifs that recognize the same DNA sequence. Genes Dev 4:29–42.

Ferré-D'Amaré AR, Prendergast GC, Ziff EB, Burley SK (1993): Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. Nature 363:38–45.

Finlay DR, Forbes DJ (1990): Reconstitution of biochemically altered nuclear pores: transport can be eliminated and restored. Cell 60:17–29.

Forman BM, Samuels HH (1990): Interactions among a subfamily of nuclear hormone receptors: the regulatory zipper model. Mol Endocrinol 4:1293–1301.

Fuller F, Porter JG, Arfsten AE, Miller J, Schilling JW, Scarborough RM, Lewicki JA, Schenk DB (1988): Atrial natriuretic peptide clearance receptor. J Biol Chem 263: 9395–9401.

Georgakopoulos T, Thireos G (1992): Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. EMBO J 11:4145–4152.

Ghosh S, Gifford AM, Riviere LR, Tempst P, Nolan GP, Baltimore D (1990): Cloning of the p50 DNA binding subunit of NF-κB: homology to rel and dorsal. Cell 62: 1019–1029.

Gilmore TD (1990): NF-κB, KBF1, *dorsal,* and *rel*ated matters. Cell 62:841–843.

Gilmore TD, Temin HM (1988): *v-rel* oncoproteins in the nucleus and in the cytoplasm transform chicken spleen cells. J Virol 62:703–714.

Goldfarb DS, Gariépy J, Schoolnik G, Kornberg RD (1986): Synthetic peptides as nuclear localization signals. Nature 322:641–644.

Govind S, Steward R (1991): Dorsoventral pattern formation in *Drosophila:* Signal transduction and nuclear targeting. Tends Genet 7:119–125.

Green S, Walter P, Kumar V, Krust A, Bornert J-M, Argos P, Chambon P (1986): Human oestrogen receptor cDNA: Sequence, expression and homology to *v-erb-A.* Nature 320:134–139.

Gregor PD, Sawadogo M, Roeder RG (1990): The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes Dev 4:1730–1740.

Guazzi S, Price M, De Felice M, Damante G, Mattei M-G, Di Lauro R (1990): Thyroid nuclear factor 1 (TTF-1) contains a homeodomain and displays a novel DNA binding specificity. EMBO J 9:3631–3639.

Haas IG, Meo T (1988): cDNA cloning of the immunoglobulin heavy chain binding protein. Proc Natl Acad Sci USA 85:2250–2254.

Hagman J, Belanger C, Travis A, Turck CW, Grosschedl R (1993): Cloning and functional characterization of early B-cell factor, a regulator of lymphocyte-specific gene expression. Genes Dev 7:760–773.

Hai T, Liu F, Coukos WJ, Green MR (1989): Transcription factor ATF cDNA clones: an extensive family of leucine zipper proteins able to selectively form DNA-binding heterodimers. Genes Dev 3:2083–2090.

Hall MN, Hereford L, Herskowitz I (1984): Targeting of E. coli β-galactosidase to the nucleus in yeast. Cell 36:1057–1065.

Hamada K, Gleason SL, Levi B-Z, Hirschfeld S, Appella E, Ozato K (1989): H-2RIIBP, a member of the nuclear hormone receptor superfamily that binds to both the regulatory element of major histocompatibility class I genes and the estrogen response element. Proc Natl Acad Sci USA 86:8289–8293.

Hamy F, Helbecque N, Henichart J-P (1992): Comparison between synthetic nuclear localization signal peptides from the steroid/thyroid hormone receptors superfamily. Biochem Biophys Res Commun 182:289–293.

Hannik M, Temin HM (1989): Transactivation of gene expression by nuclear and cytoplasmic rel proteins. Mol Cell Biol 9:4323–4336.

Helin K, Lees JA, Vidal M, Dyson N, Harlow E, Fattaey A (1992): A cDNA encoding a pRB-binding protein with properties of the transcription factor E2F. Cell 70:337–350.

Henkel T, Zabel U, van Zee K, Müller JM, Fanning E, Baeuerle PA (1992): Intramolecular masking of the nuclear location signal and dimerization domain in the precursor for the p50 NF-κB subunit. Cell 68:1121–1133.

Hibbs ML, Bonadonna L, Scott BM, McKenzie IFC, Hogarth PM (1988): Molecular cloning of a human immunoglobulin G Fc receptor. Proc Natl Acad Sci USA 85:2240–2244.

Hiles ID, Otsu M, Volinia S, Fry MJ, Gout I, Dhand R, Panayotou G, Ruiz-Larrea F, Thompson A, Totty NF, Hsuan JJ, Courtneidge SA, Parker PJ, Waterfield MD (1992): Phosphatidylinositol 3-kinase: Structure and expression of the 110 kd catalytic subunit. Cell 70:419–429.

Hirano S, Kimoto N, Shimoyama Y, Hirohashi S, Takeichi M (1992): Identification of a neural α-catenin as a key regulator of cadherin function and multicellar organization. Cell 70:293–301.

Hirashima S, Hirai H, Nakanishi Y, Natori S (1988): Molecular cloning and characterization of cDNA for eukaryotic transcription factor S-II. J Biol Chem 263:3858–3863.

Howe KM, Reakes CFL, Watson RJ (1990): Characterization of the sequence-specific interaction of mouse c-*myb* protein with DNA. EMBO J 9:161–169.

Imamoto-Sonobe N, Matsuoka Y, Semba T, Okada Y, Uchida T, Yoneda Y (1990): A protein recognized by antibodies to Asp-Asp-Asp-Glu-Asp shows specific binding activity to heterogeneous nuclear transport signals. J Biol Chem 265:16504–16508.

Inostroza JA, Mermelstein FH, Ha I, Lane WS, Reinberg D (1992): Dr1, a TATA-binding protein-associated phosphoprotein and inhibitor of class II gene transcription. Cell 70:477–489.

Issemann I, Green S (1990): Activation of a member of the steroid hormone receptor superfamily by peroxisome proliferators. Nature 347:645–650.

Iyer SV, Davis DL, Seal SN, Burch JBE (1991): Chicken vitellogenin gene-binding protein, a leucine zipper transcription factor that binds to an important control element in the chicken vitellogenin II promoter, is related to rat DBP. Mol Cell Biol 11:4863–4875.

James TC, Elgin SCR (1986): Identification of nonhistone chromosomal protein associated with heterochromatin in *Drosophila melanogaster* and its gene. Mol Cell Biol 6:3862–3872.

Jongens TA, Hay B, Jan LY, Jan YN (1992): The *germ cell-less* gene product: a posteriorly localized component necessary for germ cell development in *Drosophila.* Cell 70:569–584.

Kadonaga JT, Carner KR, Masiarz FR, Tijan R (1987): Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51: 1079–1090.

Kaelin WG, Jr, Krek W, Sellers WR, DeCaprio JA, Ajchenbaum F, Fuchs CS, Chittenden T, Li Y, Farnham PJ, Blanar MA, Livingston DM, Flemington EK (1992): Expression of a cDNA encoding a retinoblastoma-binding protein with E2F-like properties. Cell 70:351–364.

Kageyama R, Pastan I (1989): Molecular cloning and characterization of a human DNA binding factor that represses transcription. Cell 59:815–825.

Kalderon D, Richardson WD, Markham AF, Smith AE (1984a): Sequence requirements for nuclear location of simian virus 40 large-T antigen. Nature 311:33–38.

Kalderon D, Roberts BL, Richardson WD, Smith AE (1984b): A short amino acid sequence able to specify nuclear location. Cell 39:499–509.

Kara CJ, Liou H-C, Ivashkiv LB, Glimcher LH (1990): A cDNA for a human cyclic AMP response element-binding protein which is distinct from CREB and expressed preferentially in brain. Mol Cell Biol 10:1347–1357.

Kastner P, Krust A, Turcotte B, Stropp U, Tora L, Gronemeyer H, Chambon P (1990): Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. EMBO J 9:1603–1614.

Kaulen H, Pagnonec P, Gregor PD, Roeder RG (1991): The Xenopus B1 factor is closely related to the mammalian activator USF and is implicated in the developmental regulation of TFIIIA gene expression. Mol Cell Biol 11:412–424.

Keller AD, Maniatis T (1991): Identification and characterization of a novel repressor of β-interferon gene expression. Genes Dev 5:868–879.

Kieber JJ, Rothenberg M, Roman G, Feldmann KA, Ecker JR (1993): *CTR1*, a negative regulator of the ethylene response pathway in arabidopsis, encodes a member of the Raf family of protein kinases. Cell 72:427–441.

Kieran M, Blank V, Logeat F, Vandekerckhove J, Lottspeich F, Le Bail O, Urban MB, Kourilsky P, Baeuerle PA, Israël A (1990): The DNA binding subunit of NF-κB is identical to factor KBF1 and homolgous to the *rel* oncogene product. Cell 62:1007–1018.

Klein RD, Meyer BJ (1993): Independent domains of the Sdc-3 protein control sex determination and dosage compensation in C. elegans. Cell 72:349–364.

Klemsz MJ, McKercher SR, Celada A, Van Beveren C, Maki RA (1990): The macrophage and B cell-specific transcription factor PU.1 is related to the *ets* oncogene. Cell 61:113–124.

Kozasa T, Itoh H, Tsukamoto T, Kaziro Y (1988): Isolation and characterization of the human $G_s\alpha$ gene. Proc Natl Acad Sci USA 85:2081–2085.

Kume S, Muto A, Aruga J, Nakagawa T, Michikawa T, Furuichi T, Nakade S, Okano H, Mikoshiba K (1993): The Xenopus IP_3 receptor: Structure, function, and localization in oocytes and eggs. Cell 73:555–570.

Kuo CJ, Conley PB, Hsieh CL, Francke U, Crabtree GR (1990): Molecular cloning, functional expression, and chromosomal localization of mouse hepatocyte nuclear factor 1. Proc Natl Acad Sci USA 87:9838–9842.

Lai E, Prezioso VR, Tao W, Chen WS, Darnell JE, Jr (1991): Hepatocyte nuclear factor 3α belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene *fork head*. Genes Dev 5:416–427.

Lai Z-C, Rubin GM (1992): Negative control of photoreceptor development in *Drosophila* by the product of the yan gene, an ETS domain protein. Cell 70:609–620.

Landschulz WH, Johnson PF, McKnight SL (1988): The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science 240:1759–1764.

Landsman D, Srikantha T, Bustin M (1988): Single copy gene for the chicken non-histone chromosomal protein HMG-17. J Biol Chem 263:3917–3923.

Laudet V, Begue A, Henry-Duthoit C, Joubel A, Martin P, Stehelin D, Saule S (1991): Genomic organization of the human thyroid hormone receptor α (*c-erbA-1*) gene. Nucleic Acids Res 19:1105–1112.

Lee BA, Maher DW, Hannink M, Donoghue DJ (1987): Identification of a signal for nuclear targeting in platelet-derived-growth-factor-related molecules. Mol Cell Biol 7:3527–3537.

Lee W-C, Xue Z, Mélèse T (1991): The NSR1 gene encodes a protein that specifically binds nuclear localization sequences and has two RNA recognition motifs. J Cell Biol 113:1–12.

Levy DE, Kessler DS, Pine R, Darnell JE, Jr (1989): Cytoplasmic activation of ISGF3, the positive regulator of interferon-α-stimulated transcription, reconstituted in vitro. Genes Dev 3:1362–1371.

Li R, Shi Y, Thomas JO (1992): Intracellular distribution of a nuclear localization signal binding protein. Exp Cell Res 202:355–365.

Lopez JA, Chung DW, Fujikawa K, Hagen FS, Davie EW, Roth GJ (1988): The α and β chains of human platelet glycoprotein Ib are both transmembrane proteins containing a leucine-rich amino acid sequence. Proc Natl Acad Sci USA 85:2135–2139.

Maekawa T, Sakura H, Sudo T, Ishii S (1989): Putative metal finger structure of the human immunodeficiency virus type 1 enhancer binding protein HIV-EP1. J Biol Chem 264:14591–14593.

Masiakowski P, Shooter EM (1988): Nerve growth factor induces the genes for two proteins related to a family of calcium-binding proteins in PC12 cells. Proc Natl Acad Sci USA 85:1277–1281.

Meier UT, Blobel G (1990): A nuclear localization signal binding protein in the nucleolus. J Cell Biol 111:2235–2245.

Meier UT, Blobel G (1992): Nopp140 shuttles on tracks between nucleolus and cytoplasm. Cell 70:127–138.

Meijer D, Gruas A, Kraay R, Langeveld A, Mulder MP, Grosveld G (1990): The octamer binding factor Oct 6: cDNA cloning and expression in early embryonic cells. Nucleic Acids Res 18:7357–7363.

Mendel DB, Hansen LP, Graves MK, Conley PB, Crabtree GR (1991): HNF-1α and HNF-1β (vHNF-1) share dimerization and homeo domains, but not activation domains, and form heterodimers in vitro. Genes Dev 5:1042–1056.

Miller M, Reddy BA, Kloc M, Li XX, Dreyer C, Etkin LD (1991): The nuclear-cytoplasmic distribution of the *Xenopus* nuclear factor, xnf7, coincides with its state of phosphorylation during early development. Development 113: 569–575.

Mitchelmore C, Traboni C, Cortese R (1991): Isolation of two cDNAs encoding zinc finger proteins which bind to the α_1-antitrypsin promoter and to the major histocompatibility complex class I enhancer. Nucleic Acids Res 19:141–147.

Moll T, Tebb G, Surana U, Robitsch H, Nasmyth K (1991): The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor SWI5. Cell 66:743–758.

Moreland RB, Langevin GL, Singer RH, Garcea RL, Hereford LM (1987): Amino acid sequences that determine the nuclear localization of yeast histone 2B. Mol Cell Biol 7:4048–4057.

Mueller CR, Maire P, Schibler U (1990): DBP, a liver-enriched transcriptional activator, is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. Cell 61:279–291.

Nagai R, Larson DM, Periasamy M (1988): Characterization of a mammalian smooth muscle myosin heavy chain cDNA clone and its expression in various smooth muscle types. Proc Natl Acad Sci USA 85:1047–1051.

Nakafuku M, Obara T, Kaibuchi K, Miyajima I, Miyajima A, Itoh H, Nakamura S, Arai K-I, Matsumoto K, Kaziro Y (1988): Isolation of a second yeast *Saccaromyces cerevisiae* gene (GPA2) coding for guanine nucleotide-binding regulatory protein: Studies on its structure and possible functions. Proc Natl Acad Sci USA 85:1374–1378.

Nelson DR, Strobel HW (1988): On the membrane topology of vertebrate cytochrome P-450 proteins. J Biol Chem 263:6038–6050.

Newmeyer DD, Forbes DJ (1988): Nuclear import can be separated into distinct steps in vitro: Nuclear pore binding and translocation. Cell 52:641–653.

Newport JW, Forbes DJ (1987): The nucleus: structure, function, and dynamics. Annu Rev Biochem 56:535–565.

Nolan GP, Ghosh S, Liou H-C, Tempst P, Baltimore D (1991): DNA binding and IκB inhibition of the cloned p65 subunit of NF-κB, a *rel*-related polypeptide. Cell 64:961–969.

Nose A, Mahajan VB, Goodman CS (1992): Connectin: a homophilic cell adhesion molecule expressed on a subset of muscles and the motoneurons that innervate them in Drosophila. Cell 70:553–567.

Oeda K, Salinas J, Chua NH (1991): A tobacco bZip transcription activator (TAF-1) binds to a G-box-like motif conserved in plant genes. EMBO J 10:1793–1802.

Ohkuma Y, Sumimoto H, Hoffmann A, Shimasaki S, Horikoshi M, Roeder RG (1991): Structural motifs and potential α homologies in the large subunit of human general transcription factor TFIIE. Nature 354:398–401.

Okamoto K, Okazawa H, Okuda A, Sakai M, Muramatsu M, Hamada H (1990): A novel octamer binding transcription factor is differently expressed in mouse embryonic cells. Cell 60:461–472.

Ou J-H, Yeh C-T, Yen TSB (1989): Transport of hepatitis B virus precore protein into the nucleus after cleavage of its signal peptide. J Virol 63:5238–5243.

Paine PL, Moore LC, Horowitz SB (1975): Nuclear envelope permeability. Nature 254:109–114.

Parisi MA, Clayton DA (1991): Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. Science 252:965–969.

Paul R, Schuetze S, Kozak SL, Kozak CA, Kabat D (1991): The *Sfpi*-1 proviral integration site of Friend erythroleukemia encodes the *ets*-related transcription factor Pu.1. J Virol 65:464–467.

Peterson CL, Herskowitz I (1992): Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. Cell 68:573–583.

Peterson MG, Inostroza J, Maxon ME, Flores O, Admon A, Reinberg D, Tjian R (1991): Structure and functional properties of human general transcription factor IIE. Nature 354:369–373.

Picard D, Kumar V, Chambon P, Yamamoto KR (1990): Signal transduction by steroid hormones: nuclear localization is differentially regulated in estrogen and glucocorticoid receptors. Cell Regul 1:291–299.

Picard D, Yamamoto KR (1987): Two signals mediate hormone-dependent nuclear localization of the glucocorticoid receptor. EMBO J 6:3333–3340.

Poli V, Mancini FP, Cortese R (1990): IL-6DBP, a nuclear protein involved in interleukin-6 signal transduction, defines a new family of leucine zipper proteins related to C/EBP. Cell 63:643–653.

Prendergast GC, Lawe D, Ziff EB (1991): Association of Myn, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and Ras cotransformation. Cell 65:395–407.

Ray RP, Arora K, Nüsslein-Volhard C, Gelbart WM (1991): The control of cell fate along the dorsal-ventral axis of the *Drosophila* embryo. Development 113:35–54.

Reith W, Herrero-Sanchez C, Kobr M, Silacci P, Berte C, Barras E, Fey S, Mach B (1990): MHC class II regulatory factor RFX has a novel DNA-binding domain and a functionally independent dimerization domain. Genes Dev 4:1528–1540.

Reuter G, Giarre M, Farah J, Gausz J, Spierer A, Spierer P (1990): Dependence of position-effect variegation in *Drosophila* on dose of a gene encoding an unusual zinc-finger protein. Nature 344:219–223.

Rey-Campos J, Chouard T, Yaniv M, Cereghini S (1991): vHNF1 is a homeoprotein that activates transcription and forms heterodimers with HNF1. EMBO J 10:1445–1457.

Richardson WD, Mills AD, Dilworth SM, Laskey RA, Dingwall C (1988): Nuclear protein migration involves two steps: rapid binding at the nuclear envelope followed by slower translocation through nuclear pores. Cell 52: 655–664.

Richardson WD, Roberts BL, Smith AE (1986): Nuclear location signals in polyoma virus large-T. Cell 44:77–85.

Rihs H-P, Peters R (1989): Nuclear transport kinetics depend on phosphorylation-site-containing sequences flanking the karyophilic signal of the Simian virus 40 T-antigen. EMBO J 8:1479–1484.

Roberts BL, Richardson WD, Smith AE (1987): The effect of protein context on nuclear location signal function. Cell 50:465–475.

Roman C, Platero JS, Shuman J, Calame K (1990): Ig/EBP-1: a ubiquitously expressed immunoglobulin enhancer binding protein that is similar to C/EBP and heterodimerizes with C/EBP. Genes Dev 4:1404–1415.

Rowe A, Eager NSC, Brickell PM (1991): A member of the RXR nuclear receptor family is expressed in neural-crest-derived cells of the developing chick peripheral nervous system. Development 111:771–778.

Ruben SM, Dillon PJ, Schreck R, Henkel T, Chen C-H, Maher M, Baeuerle PA, Rosen CA (1991): Isolation of a *rel*-related human cDNA that potentially encodes the 65-kD subunit of NF-κB. Science 251:1490–1493.

Ryder K, Lau L, Nathans D (1988): A gene activated by growth factors is related to the oncogene v-*jun*. Proc Natl Acad Sci USA 85:1487–1491.

Schneider LE, Taghert PH (1988): Isolation and characterization of a *Drosophila* gene that encodes multiple neuropeptides related to Phe-Met-Arg-Phe-NH_2 (FMRFamide). Proc Natl Acad Sci USA 85:1993–1997.

Shelton CA, Wasserman SA (1993): *pelle* encodes a protein kinase required to establish dorsoventral polarity in the *Drosophila* embryo. Cell 72:515–525.

Shirakawa F, Mizel SB (1989): *In vitro* activation and nuclear translocation of NF-κB catalyzed by cyclic AMP-dependent protein kinase and protein kinase C. Mol Cell Biol 9:2424–2430.

Shore D, Nasmyth K (1987): Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. Cell 51:721–732.

Silver P, Goodson H (1989): Nuclear protein transport. CRC Crit Rev Biochem Mol Biol 24:419–435.

Silver P, Sadler I, Osborne MA (1989): Yeast proteins that recognize nuclear localization sequences. J Cell Biol 109: 983–989.

Sladek FM, Zhong W, Lai E, Darnell JE, Jr (1990): Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. Genes Dev 4:2353–2365.

Smith AE, Kalderon D, Roberts BL, Colledge WH, Edge M, Gillett P, Markham A, Paucha E, Richardson WD (1985): The nuclear location signal. Proc R Soc Lond [Biol] 226:43–58.

Smith DP, Old RW (1991): *Xenopus laevis* Oct-1 does not bind to certain histone H2B gene promoter octamer motifs for which a novel octamer-binding factor has high affinity. Nucleic Acids Res 19:815–821.

Stochaj U, Silver PA (1992): A conserved phosphoprotein that specifically binds nuclear localization sequences is involved in nuclear import. J Cell Biol 117:473–482.

Sumimoto H, Ohkuma Y, Sinn E, Kato H, Shimasaki S, Horikoshi M, Roeder RG (1991): Conserved sequence motifs in the small subunit of human general transcription factor TFIIE. Nature 354:401–404.

Suttle DP, Bugg BY, Winkler JK, Kanalas JJ (1988): Molecular cloning and nucleotide sequence for the complete coding region of human UMP synthase. Proc Natl Acad Sci USA 85:1754–1758.

Suzuki N, Rohdewohld M, Neuman T, Gruss P, Schöler HR (1990): Oct. 6: A POU transcription factor expressed in embryonal stem cells and in the developing brain. EMBO J 9:3723–3732.

Takeyasu K, Tamkun MM, Renaud KJ, Fambrough DM (1988): Ouabain-sensitive (Na^+ + K^+)-ATPase activity expressed in mouse L cells by transfection with DNA encoding the α-subunit of an avian sodium pump. J Biol Chem 263:4347–4354.

Tamura T-A, Sumita K, Fujino I, Aoyama A, Horikoshi M, Hoffmann A, Roeder RG, Muramatsu M, Mikoshiba K (1991): Striking homology of the 'variable' N-terminal as well as the 'conserved core' domains of the mouse and human TATA-factors (TFIID). Nucleic Acids Res 19:3861–3865.

Travis A, Amsterdam A, Belanger C, Grosschedl R (1991): LEF-1, a gene encoding a lymphoid-specific with protein, an HMG domain, regulates T-cell receptor α enhancer function. Genes Dev 5:880–894.

Tsay Y-F, Schroeder JI, Feldmann KA, Crawford NM (1993): The herbicide sensitivity gene *CHL1* of arabidopsis encodes a nitrate-inducible nitrate transporter. Cell 72:705–713.

Tsuchiya E, Uno M, Kiguchi A, Masuoka K, Kanemori Y, Okabe S, Miayawa T (1992): The *Saccharomyces cerevisiae* NPS1 gene, a novel CDC gene which encodes a 160 kDa nuclear protein involved in G2 phase control. EMBO J 11:4017–4026.

Tsuneoka M, Imamoto NS, Uchida T (1986): Monoclonal antibody against non-histone chromosomal protein high mobility group 1 co-migrates with high mobility group 1 into the nucleus. J Biol Chem 261:1829–1834.

van de Wetering M, Oosterwegel M, Dooijes D, Clevers H (1991): Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box. EMBO J 10:123–132.

Velazquez L, Fellous M, Stark GR, Pellegrini S (1992): A protein tyrosine kinase in the interferon α/β signaling pathway. Cell 70:313–322.

Waterman ML, Fischer WH, Jones KA (1991): A thymus-specific member of the HMG protein family regulates the human T cell receptor Cα enhancer. Genes Dev 5:656–669.

Welsh JD, Swimmer C, Cocke T, Shenk T (1986): A second domain of simian virus 40 T antigen in which mutations can alter the cellular localization of the antigen. Mol Cell Biol 6:2207–2212.

White EM, Allis CD, Goldfarb DS, Srivastva A, Weir JW, Gorovsky MA (1989): Nucleus-specific and temporally restricted localization of proteins in *Tetrahymena* macronuclei and micronuclei. J Cell Biol 109:1983–1992.

Wirth T, Priess A, Annweiler A, Zwilling S, Oeler B (1991): Multiple Oct2 isoforms are generated by alternative splicing. Nucleic Acids Res 19:43–51.

Wozniak RW, Bartnik E, Blobel G (1989): Primary structure analysis of an integral membrane glycoprotein of the nuclear pore. J Cell Biol 108:2083–2092.

Wun T-C, Kretzmer KK, Girard TJ, Miletich JP, Broze GJ, Jr (1988): Cloning and characterization of a cDNA coding for the lipoprotein-associated coagulation inhibitor shows that it consists of three tandem kunitz-type inhibitory domains. J Biol Chem 263:6001–6004.

Xiao JH, Davidson I, Matthes H, Garnier J-M, Chambon P (1991): Cloning, expression, and transcriptional properties of the human enhancer factor TEF-1. Cell 65:551–568.

Yamasaki L, Kanda P, Lanford RE (1989): Identification of four nuclear transport signal-binding proteins that interact with diverse transport signals. Mol Cell Biol 9:3028–3036.

Yan C, Tamm I (1991): Molecular cloning and characterization of interferon α/β response element binding factors of the murine (2′-5′)oligoadenylate. Proc Natl Acad Sci USA 88:144–148.

Yoneda Y, Imamoto-Sonobe N, Matsuoka Y, Iwamoto R, Kiho Y, Uchida T (1988): Antibodies to Asp-Asp-Glu-Asp can inhibit transport of nuclear proteins into the nucleus. Science 242:275–278.

Yu-Lee L-Y, Hrachovy JA, Stevens AM, Schwarz LA (1990): Interferon-regulatory factor 1 is an immediate-early gene under transcriptional regulation by prolactin in Nb2 T cells. Mol Cell Biol 10:3087–3094.

Zarkower D, Hodgkin J (1992): Molecular analysis of the C. elegans sex-determining gene tra-1: a gene encoding two zinc finger proteins. Cell 70:237–249.

Zhou A, Hassel BA, Silverman RH (1993): Expression cloning of 2-5A-dependent RNAse: a uniquely regulated mediator of interferon action. Cell 72:753–765.