Nuclear localization signals overlap DNA- or RNA-binding domains in nucleic acid-binding proteins

Eric C. LaCasse⁺ and Yvonne A. Lefebvre^{*}

Departments of Medicine and Biochemistry, University of Ottawa, Loeb Institute for Medical Research, Ottawa Civic Hospital, Ottawa, Canada

Received February 9, 1995; Revised and Accepted March 30, 1995

INTRODUCTION

Many proteins are composed of modular domains with each domain being derived from a sequence of amino acid residues that forms a distinctive structure. Moreover, many of the domains perform well-defined functions that contribute to the overall activity of a protein (1). The domains are often able to move about within proteins with no obvious adverse functional consequence (see below), which suggests that sequences specifying different functions need not be in proximity to each other. Therefore, when the proximity of two functions is conserved through evolution a consideration of the underlying advantage is in order.

Nuclear receptors are part of a large superfamily of transcription factors which all share a similar modular design (2). The domains were identified by sequence homology and by functional analyses although a structural basis is assumed and has been proven only for the DNA-binding domain (3–5). Sequence comparisons of the chick and human estrogen receptors resulted in the identification of six distinct regions of sequence similarity, termed A to F (6). All regions except the small N-terminal A region and the small C-terminal F region are conserved for all members of the nuclear receptor superfamily (2). The modular design obtained from sequence data agrees with previous biochemical proteolytic studies of steroid receptors that separated the receptors into four regions; an N-terminal domain (A/B domain), a central DNA-binding domain (C domain) and a C-terminal ligand-binding domain (E/F domain) that is separated from the DNA-binding domain by a hinge region (D region) (7,8). The separate functions of each modular domain of these receptors is clearly demonstrated by experiments in which the whole DNA-binding domain of one receptor has been replaced with that of another receptor to change the DNA-binding specificity to the original receptor (9), or in which the ligand-binding domain has been fused to many other transcription factors and shown to confer hormonal regulation of transcription for the resulting chimera (10,11).

Nuclear localization signals (NLSs) are sequences within nuclear proteins that target these proteins to the cell nucleus (12–17). Nuclear localization signals are short stretches of amino acids (4–30 amino acids) too short to form an independent

structure and therefore, may be part of a larger sequence that forms a structural or globular domain. For example, the NLS of NGFI-A is proposed to be dependent on the overall structure of the DNA-binding domain (18) and the bipartite NLS of the androgen receptor is dependent on the structure of the second ‘zinc finger’ of the DNA-binding domain (19). Although the position of NLSs within a protein is restricted to regions of the protein that are exposed (20), presumably so as to be accessible to regulatory molecules or to the nuclear import machinery, NLSs occur anywhere in the primary structure of a protein (17). However, we observed that within the nuclear receptor superfamily, i.e. in all the steroid, thyroid hormone and retinoic acid receptors, the NLS either overlaps directly with the DNA-binding function or is separated from the C-terminal end of the DNA-binding domain by only a few amino acids (21). This observation prompted us to investigate whether this close spatial relationship between the NLS and nucleic acid-binding domain was a general property of other nucleic acid-binding proteins as this would strengthen the basis for further investigation of the relative contributions to each other’s function.

Classification of NLSs in nucleic acid binding proteins

Consensus sequences for NLSs have been difficult to establish. Most NLSs however share the property of being composed primarily of basic residues. Early work identified a basic heptamer of SV40 large T-antigen as the typical NLS motif (12–17). Later, a bipartite basic structure in nucleoplasmin was put forward as the predominant NLS motif (22). Although the majority of NLSs in nuclear proteins fall into one of these two classes, some reported NLSs in nucleic acid binding proteins do not conform to either of these types of signals. Accordingly, we summarize below the two major classes together with five additional classes into which the remainder of proven NLSs found in nucleic acid-binding proteins fall.

NLSs have been identified experimentally either by deletion of a sequence (negative approach) which results in the loss of nuclear localization capability and/or by the ability of a short peptide sequence to target a non-nuclear protein to the nucleus (additive approach). Both approaches are required to prove

^{*} To whom correspondence should be addressed at: Division of Endocrinology and Metabolism, Ottawa Civic Hospital, 1053 Carling Avenue, Ottawa, Ontario, Canada K1Y 4E9

⁺ Present address: Ontario Cancer Institute, 500 Sherbourne Street, Toronto, Ontario M4X 1K9, Canada

absolutely the role of an NLS in the nuclear targeting of a specific protein. The deletional approach alone cannot be used to identify the precise location of a specific signal, as it might disrupt the structure of an upstream or downstream signal. On the other hand, the additive approach alone may identify a signal which is not operable in the context of the protein and as there may be more than one NLS in a protein (17,23), it will not identify a potential second signal. However, as the NLSs of only a few proteins have been identified by both of these approaches, we have considered all NLS which have been identified by either one of the two approaches.

SV40-like

The first NLS identified was that of the SV40 large T-antigen, PKKKRKV (24,25). As similar sequences were found to function as NLSs in other proteins (12–17), this sequence was thought to be the archetypical signal. The SV40-like NLSs are small sequences of seven or eight amino acids containing a single cluster of basic residues often flanked by a prolyl or acidic residue. Chelsky *et al.* (26) further refined this sequence to a tetrameric consensus sequence, K(R/K)x(R/K), in which x is limited to a few preferred amino acids, K, R, P, V or A, but not N. Two extensive reviews by Boulikas have presented a set of rules for what he has defined as a core NLS (27,28). First, a core NLS is proposed to consist of a hexapeptide with four R and K residues. Secondly, D and E residues as well as bulky amino acids (F, Y, W) are not frequently placed between K and R residues. Thirdly, acidic residues and P or G residues that break the α -helix are frequently in the flanking region of the hexapeptide stretch. Fourthly, hydrophobic residues ought not to be present in the core NLS flanking region so as to allow for the exposure of the NLS on the protein surface. Boulikas (27) classifies core NLSs into three main categories which were further subdivided. The categories are 'highly basic NLS' which contains five or six (K/R) residues in a hexapeptide stretch, 'typical NLS' containing four (K/R) residues in a stretch of six amino acids and 'non-typical NLS', containing two or three (K/R) residues within a hexapeptide. The 'typical NLS' motifs were further subdivided into nine motifs; (K/R)₄, (K/R)₃X(K/R)₃, (K/R)₃X(K/R), (K/R)₂X(K/R)₂, (K/R)X(K/R)₂X(K/R), (K/R)₂X(K/R)X(K/R), (K/R)₃X(K/R)₂, (K/R)₂XX(K/R)₂ and (K/R)₃XX(K/R). The 'non-typical NLS' motifs were subdivided into three motifs; (R)₃, (R/K)₂X(R/K) and (K/R)₂X_n(R/K)₂. X represents any residue but is usually not D, E, F, Y or W. (K/R)₄ is a stretch of four consecutive K or R residues and is the most predominant motif with a frequency of 25%.

Bipartite

The bipartite NLS was first identified in the non-nucleic acid binding protein, nucleoplasmin. It took many years to define the nucleoplasmin NLS because it was more complex than that of SV40 large T-antigen (29,30). The nucleoplasmin NLS consists of two basic clusters of amino acids separated by a spacer region of 10 amino acids, KRPAATKKAGQAKKKK. The C-terminal basic sequence is similar to the NLS of SV40 large T-antigen. However, in nucleoplasmin, both basic clusters are required for efficient nuclear import (30,31). This bipartite NLS may be the true archetype of an NLS as this consensus sequence (two adjacent basic residues separated by ~10 amino acids from a

second domain that contains at least three out of five basic amino acids) has been identified in sequence searches of the SWISS-PROT protein database in 56% of the nuclear proteins and in only 4% of non-nuclear proteins (22).

Mato2-like

This type of NLS, found in the yeast mating-type switch factor, consists of a short hydrophobic region that contains one or more basic amino acids (32,33). Two of these unusual signals exist in Mato2, NKIPKD at amino acid residues 1–13 (32) and VRILESWFAKNI at residues 141 to 159 (33). Each of these two NLSs is a competent nuclear targeting signal on its own. Similar sequences have been identified in a number of other nuclear proteins (17) and at least one other protein with a proven NLS similar to that of Mato2 is known, the maize R transcription factor (34).

Hydrophobic

There is only one demonstrated example of a hydrophobic NLS sequence, FV(X)_{7–20}MXSLXYM(X)₄MF. This signal is noteworthy because of the lack of any charged residues. It targets the adenovirus E1a transcription factor (also an oncoprotein) to the *Xenopus* oocyte nucleus in a developmentally regulated manner (35). A similar but as yet unproved sequence exists in the glucocorticoid receptor in a region that does contain a nuclear localization function (23,35), but which has yet to be further delimited.

Methylated arginyl residues or consecutive arginyl residues

The role of methylated arginyl residues in nuclear import has been suggested from studies which show that the only difference between forms of basic fibroblast growth factor (bFGF) that get into the nucleus from those which do not is that the former have methylated arginyl residues (36,37). The different forms of bFGF result from alternate splicing and the use of alternate initiation codons (38). Burgess *et al.* note that a common feature of methylated arginyl-containing proteins, such as histones, is their nuclear localization (36).

Although consecutive arginyl residues constitute a basic stretch of amino acids, these differ from the usual NLSs in that most NLSs of a basic nature appear to be a mix of R and K residues with a preference for K residues. The maize 'transcription factor R' contains in addition to an NLS of the SV40-type such an unusual NLS, containing three consecutive R residues and another of the Mato-2 type (34). Similar NLSs have been found in other viral proteins (34), such as RRNRRRR in HIV Rev (39) or RLPV(R)₆VP in the adenovirus preterminal protein, (40).

Trimethylguanosine cap structure

The U snRNPs are uracil-rich small nuclear ribonucleoprotein particles that play a role in spliceosome function. After export from the nucleus, the U snRNAs are complexed with Sm and other proteins in the cytoplasm and the complex is then imported into the nucleus. We include in a separate classification the trimethylguanosine cap structure of U snRNPs because import of U1 snRNA specific binding proteins, such as U1-70K (41) and U1-A (42) are dependent on NLS sequences within the protein constituents but their distribution will also be influenced by the

5'-terminal trimethylguanosine cap structure of the pol II U snRNAs (U1–U5) when they are part of an snRNP complex (42–44). It has been proven by competition and interference experiments that the import pathways for these 'cap structure NLSs' are distinct from the pathway for the basic sequence NLSs (45,46). However, this cap structure 'NLS' is limited in its functionality, as the cap is essential for U1 and U2 snRNP nuclear import in *Xenopus* oocytes but not in somatic cells (47). The differential cap dependence for nuclear import is mediated by soluble cytosolic factors (47). Although the cap structure is not essential for nuclear import in somatic cells, it does enhance the transport kinetics of U1 snRNPs (48).

Others

There are other examples of NLSs that show no similarity to the above classes, such as the signal found at residues 2–59 of the HTLV-1 Tax transactivator (49), residues 1–29 (MK(X)₁₁CRLKKLKCSKEKPKCA) of the yeast Gal4 transactivator (50,51) or AAFEDLRVLS of the influenza viral nucleoprotein (52). This latter sequence may not be an NLS, but rather a sequence that specifies nuclear accumulation (retention) (52). In addition, other signals show some similarity to one of the above classes but display some unusual features. For example, a serum-regulated pentapeptide NLS of adenovirus E1a, KRPRP, confers a serum dependent nuclear entry to non-nuclear proteins

(53), which suggests that in addition to this basic sequence a serum factor is required.

Finally, a sugar dependence has been demonstrated for the nuclear import of glycoconjugates that show ATP-dependent nuclear import (54). Glucose, fucose and mannose can target human serum albumin-conjugates to the nucleus while other sugars cannot. It is not known whether the sugar-dependent signal is physiologically relevant or whether the nuclear localization is mediated by some other protein, such as an NLS-bearing lectin. Although we have not found an example which might fit this category among the NLSs of nucleic acid binding proteins, we include it for completeness.

Proximity of NLSs and DNA/RNA-binding motifs

A review of the literature identified those nuclear proteins with known DNA- (Table 1) and RNA-binding domains (Table 2) and for which an NLS has been proven by an additive or negative approach. For the purposes of this survey, we accepted the published boundaries for the DNA- and RNA-binding domains, usually defined by deletion analysis or homology and made no attempt to refine the limits of the reported domains. More than 40 different nuclear proteins from non-homologous genes have been reported for which the positions of the DNA-binding motif and the NLS(s) are known, while only a small number of proteins with defined RNA-binding motifs and NLS(s) have been reported.

Table 1. DNA-binding proteins with proven nuclear localization signals

DNA-binding motif	Nuclear protein	#aa	Minimal DNA-binding domain	Nuclear location signal(s)	Class ^a	Ref.
HTH homeobox	mouse Pit-1 (GHF-1)	291	214–273	209–252	O	105
	yeast Mat α 2	210	136–188	1–13		32
Zinc finger nuclear receptor				141–159	O	33
	human ER	595	179–250	256–303	F	21,118
	chicken PR	786	410–478	490–506		21
				477–487	F	
				second finger	O	
	rabbit PR	930	568–633	638–642/645	F	119,120
				614–618	O	
				624–627	O	
	rat GR	795	438–505	497–524	O/F	23
				540–795		
	human GR	777	420–488	478–505	O/F	121
				521–777		
	human AR	910	550–615	608–625		
				628–634	O/F	89,122
		919	559–624	580–661	O/F	19,123
Zinc finger poly(ADP-ribose)pol Zinc finger C2H2	rat TR α 1 (c-erbA)	410	53–121	134–136	F	124
	chicken c-erbA	408	51–118	127–135	F	59
	AEV v-erbA	639	293–361	370–376	F	117,125
	human PARP	1014	12–66	207–226	D?	126
			116–166			
	mouse Egr-1 (NGF1-A, Krox24, Zif268, TIS8)	533	331–419	315–330		90
				361–419	O	
				328–433	O	18
	yeast Swi5	709	538–635	635–655	F	127
	<i>Xenopus</i> nuclear factor 7	609	145–273	106–120	P	128
Zinc finger B box (RING)						

Table 1. continued

Zinc finger (GAL4)	yeast Gal4	(881)	11–38/52	1–29	O	51
Zinc finger (other)	adenovirus E1a	289	139–185	282/285–289		53
	human XPAC	273	105–129	140–185	O	35
	human GATA-3	444	303–347	30–42	D	129
bHLH/bZIP	ASV v-Jun	296	223–273	249–311	O/F	130
Fos/Jun	chicken c-Jun	310	223–239	223–239	O	131
	human c-Jun	340	236–287	245–253/259	O	131
	mouse c-Fos	380	259–317	267–283	O	91
	FBJ v-Fos	379	131–193	139–161	O	132
bHLH/bZIP	AMCV v-Myc	416	131–193	139–161	O	132
Myc	human c-Myc	439	330–410	294–309	P	133
			354–439	320–328		26
				364–374	O	134
	human N-Myc	456	381–456	337–344	P	59
	mouse Max (Myn)	160	24–103	149–157	D	135
	human Max	151	15–99	140–148	D	132
bHLH/bZIP (other)	EBV ZEBRA (Zta, EB1)	245	167–202	178–183	O	
				187–194	O	91
	rat CREB	327	269–327	287–295	O	136
	tobacco TGA-1A	359	77–114	70–139	O	96
	maize Opaque-2	437	136–284	101–135		
				223–254	O	103
Helix–loop–helix	maize R protein	610	190–330?	100–109	D	34
				419–428		
				598–610		
	human c-myb	640	34–86	521–528	D	59
Ets (HTH)	chicken c-ets-1	485	375–459	421–427	O	137
		441	331–415	369–388	O	138
				388–441	O/F	
HSF	human HSF2 (heat shock factor)	536	5–113	108–122	F	113
				195–210		
HMG box	mouse UBF	795	6 boxes	449–480	O	80
			401–480 (#4)			
			480–560 (#5)			
	xenopus UBF-2	701	5 boxes	566–592?	F	139
			?–563 (#5)			
NRD or RH domain	v-rel	503	(rel 16–304)	274–317	F?	140,141
	human p65	551	33–226	300–304	F?	142
			(rel 21–291)			
Other	yeast Apr1 DNA repair enzyme	367	1–287?	334–367	D?	69
	SV40 large T-antigen	708	137–246	126–132	F	24,25
	SV40 Vp3	234	221–234?	198–206	F?	57,58,143
	<i>Agrobacterium</i> VirE2	556	?unknown	212–252	O/F?	144
	<i>Agrobacterium</i> VirD2	447	29–?	1–37	O/F?	145,146
				417–434		
	polyomavirus VP1	384	1–5	1–11	O	147,148
	BPV E1 replication protein	605	1–299	85–108	O?	149
	hamster RCC1	421	8–?	8–24/29	O	150
	HSV ICP8	1196	564–849	1169–1196	D	151,152
	murine p53	390	307/343–390	312–318	O/F	61
			280–390	365–370		
			80–290	375–380		
	human p53	393	319–393	316–325	O/F	59,60
			102–292	312–323		
	<i>Drosophila</i> HP1	206	95–206	152–206	O	71
	Adenovirus DNA-binding protein	529	?–529	23–105	P/D	70,153

Abbreviations: **O**, overlapping; **F**, flanking; **P**, proximal; **D**, distal; **HTH**, helix–turn–helix; **bHLH/bZIP**, basic region helix–loop–helix/basic region leucine zipper; **HSF**, heat shock factor; **HMG**, high mobility group; **NRD**, NF- κ B-Rel-Dorsal; **RH**, Rel homology; **ER**, estrogen receptor; **PR**, progesterone receptor; **GR**, glucocorticoid receptor; **AR**, androgen receptor; **TR**, thyroid hormone receptor; **?**, refers to uncertainties with exact borders of domains, or other uncertainties (see text).

^aWhen a protein contains multiple NLSs only the closest NLS to the DNA-binding domain is classified.

Table 2. RNA-binding proteins with proven nuclear localization signals

RNA-binding motif	Nuclear protein	#aa	Minimal RNA-binding domain	Nuclear location signal(s)	Class ^a	Ref.
RNP/RRM	chicken nucleolin	694	278–682	256–273	F	78
	<i>Xenopus</i> nucleolin90	651	236–639	271–290	F	154
				209–225	O	
	yeast NSR1	414	169–187 (RRM1)	139–142 159–162	O/F	79
	human U1-70K	437	92–199	92–199	O	
	human U1A	282	11–96 204–282	183–437 94–204	O/F O/F	41 42
Zinc finger	<i>Xenopus</i> XFG5-1	420	7 fingers (209–404)	in finger 3 (265–292)	O	155
Other	HDV p27 delta antigen	214	88–163? (79–107) (2–27)	35–44 69–89	O/F	156–158
	influenza PB1 (NS1)	237	19–38	34–38 216–221	O	159,160
	yeast L25	142	62–126	11–17 18–28	P/D	88
	HTLV-1 Rex (pX)	189	1–19/77	1–19	O	161
	HIV-1 Tat	86	37–72?	48–52/61	O	59,162
	HIV-1 Rev	116	35–50	38–45	O	39,162

Abbreviations: O, overlapping; F, flanking; P, proximal; D, distal; RNP, ribonucleoprotein; RRM, RNA recognition motif.

^aWhen a protein contains multiple NLSs only the NLS closest to the RNA-binding domain is classified.

We determined the relative positions of the DNA- or RNA-binding domain to the NLS in each of the proteins listed in the Tables. The proteins are grouped, based on the type of DNA- or RNA-binding motif, similar to the grouping in the Prosite library (e.g., nuclear receptor zinc finger) for ease of analysis (55). The positions of the DNA- and RNA-binding domains and of the NLS are indicated. Each of the nuclear proteins was then placed into one of four categories depending on whether the NLS and the DNA- or RNA-binding motifs are overlapping (**O**), immediately flanking each other (**F**, less than 10 amino acid separation), separated by <30 amino acids (proximal or **P**) or separated by >30 amino acids (distal or **D**). The first major finding from Table 1 and 2 is that >79% of the proteins contain overlapping or flanking NLSs and nucleic acid-binding motifs (42 out of 53) or 76% (32 out of 42) of proteins with DNA-binding motifs and 91% (10 out of 11) of proteins with RNA-binding motifs. Although homologous proteins are listed, these percentages were calculated from the non-homologous proteins so as not to give any weighted advantage to a protein with a proven NLS function which has many homologues, for example rat TR α 1, chicken c-erbA and avian erythroblastosis virus v-erbA are counted as one DNA-binding protein. Inclusion of the homologues into the calculations would increase the reported percentages. The second major finding from the Tables is that the proximity of NLSs and DNA- or RNA-binding motifs was observed for several types of DNA- and RNA-binding motifs in proteins of unrelated evolutionary origins. This suggests a functional basis for the association of NLS and DNA- or RNA-binding motifs. Third, similarly, the NLSs found close to DNA- and RNA-binding motifs are also representative of various classes of NLSs (e.g. SV40-like, bipartite, hydrophobic). For example, the SV40 large T-antigen

NLS is of course SV40-like and the NLSs of the androgen receptor (19,89), Egr1 (90) and ZEBRA (91) are bipartite. Although the NLSs for many of the nucleic acid binding proteins cannot be classified exactly into one or other of these two classes as they have not been defined precisely enough, these two classes represent by far the majority of NLSs in nucleic acid binding proteins. However, other NLSs are found: the mat α 2-like signal is present in yeast mat α 2; the hydrophobic signal is found in adenovirus E1a; the consecutive arginyl signal is found in HIV Rev as well as other signals which do not fit well into any of the six classes were present, such as that in yeast Gal4 (see first section for references).

Forty-three non-homologous proteins fell readily into one of the four categories chosen to designate the distance of the NLS from the nucleic-acid binding region. However, explanations are required for the classification of 10 other proteins:

(i) *Rel* family members were classified as proteins with flanking NLS and DNA-binding regions even though according to the criteria described above they could be classed as distal because the two motifs are separated by >100 amino acids. The *Rel* homology domain, also called the NRD (NF- κ B/*Rel*/dorsal) motif, is roughly a 300 amino acid region that is conserved from *Drosophila* to man (56). The *Rel* homology domain contains the DNA-binding domain at its N-terminus next to a dimerization domain with an NLS at the C-terminal end. All three functional domains are conserved. The dimerization domain between the NLS and DNA-binding domain may be required for high affinity DNA binding and therefore can be considered part of the DNA-binding domain which forms with the NLS a distinct well-conserved modulatory cassette known as the *Rel* homology domain.

(ii) Similarly, the NLS of SV40 Vp3 protein (amino acids 198–206) (57) is classified as flanking or overlapping the DNA-binding domain (amino acids 221–234) because the C-terminal amino acids which constitute the major DNA-binding domain require the C-terminal 35 amino acids (which also contain the NLS) for full DNA binding activity (58).

(iii) The murine and human p53 tumor suppressor proteins have three NLSs (59–61) that overlap with a region involved in non-specific DNA binding (amino acids 319–393) which is next to the region involved in sequence specific DNA binding (amino acids 102–292) (see refs 62–65). The NLSs also overlap with a region involved in the regulation of specific DNA binding. The NLSs in the basic C-terminus may actually interact and mask the region involved in sequence-specific DNA binding (see refs 62, 63 and references therein). The murine and human p53 proteins have been classified here as containing an overlapping/flanking NLS even though the NLS does not overlap a region of sequence specific DNA binding. However, the overlapping region does appear to be important for DNA binding.

(iv) Because the NLS of the poly(ADP-ribose)polymerase (PARP) bipartite NLS is functionally separate from the DNA-binding activity and 40 amino acids away from the second zinc finger, we have classified it in the D (distal) group (see Table 1; 126). However, it is part of a 46 kDa proteolytic fragment that forms the DNA-binding domain which suggests that the two functions share a common protease-resistant globular domain. Therefore, it could be argued that the NLS and DNA-binding motifs should be classified as flanking (66).

(v) *Agrobacterium spp.* VirE2 and VirD2 proteins are required for the nuclear import of the *Agrobacterium* T-DNA complex into plant nuclei (67). Although the VirE2 and VirD2 NLSs have been mapped, their DNA-binding domains have not been precisely mapped. The VirE2 DNA binding domain lies somewhere in the C-terminal two thirds of the protein (67). A possible overlap with the NLS has been inferred because a point mutation in the NLS also disrupts DNA binding (67). Likewise, the N-terminal NLS sequence of VirD2 is very close to a Y residue proposed to be involved in the covalent linkage of VirD2 to the T-DNA (68).

(vi) Although the DNA binding domain of the yeast Apn1 has not been precisely mapped it has been classified as distal by inference. It is assumed that its DNA binding function must correspond to a homologous domain of the *E.coli* endonuclease IV which excludes the region containing the NLS (69). However, this does not preclude the possible contribution of the NLS region to DNA binding of Apn1.

(vii) The adenovirus DNA-binding protein NLS is classified as distal because the DNA-binding domain, although not precisely mapped, lies in a C-terminal chymotryptic fragment of 44–46 kDa while the NLS lies in the first 105 N-terminal amino acids (70).

(viii) The heterochromatin-associated protein HP1 does not contain sequences homologous to previously identified NLS or DNA binding motifs (71). Nonetheless, regions specifying these functions have been identified and these have been included in the analysis. However, it is possible that the overlapping domains of HP1 which confer nuclear localization and heterochromatin binding properties may function by protein–protein interactions rather than HP1 binding directly to DNA or to an NLS binding protein.

(ix) We have considered the so-called nucleolar localization signal of Rex to be a nuclear localization signal. What has been called a ‘nucleolar localization signal’ is probably an RNA-bind-

ing function, which is overlapping a nuclear localization signal. Thus the RNA binding and not the NLS may be responsible for the sub-nuclear localization, by binding RNA (72–76) or by binding the nucleolar protein B-23 (76). It has clearly been demonstrated that the nucleolar localization of nucleolin (77,78), NSR1 (79) and UBF (80) is due to nucleolar component-binding sequences.

The only homeodomain proteins listed in Table 1 are Pit-1 and Mat α 2 because the study of NLS function in other homeodomain proteins has not received much attention as it is generally assumed that the small size of these proteins would allow their diffusion into the nucleus. However, it may be argued that all nuclear proteins no matter how small may possibly require an NLS: functional NLSs have recently been identified in numerous small viral proteins (see tables) and small mammalian proteins, such as histone H1 (81). Therefore, it is possible that the majority of homeodomain proteins may require an NLS for nuclear import. An argument against this possibility is that a putative basic NLS in homeodomains which is definitely part of the DNA-binding domain, is deleted or mutated in some endogenous homeodomain inhibitors, e.g., I-POU (82), Id (83,84), extramacrochaetae (85–87). Nonetheless, these inhibitors dimerize with specific homeodomain proteins and inhibit DNA binding of the heterodimers *in vitro*. One would have to postulate that if the basic stretch is truly an NLS and the sole NLS for homeodomain proteins then these inhibitors must piggyback into the nucleus either with their dimerization partner or with another protein.

Some of the proteins listed in the tables are conserved in prokaryotes. However, as might be expected the prokaryotic homologues lack an NLS. Yeast ribosomal binding protein L25 which binds rRNA in the nucleus (88) and the yeast DNA repair enzyme Apn1 (69) possess extensions with NLSs while their prokaryote homologues do not possess NLSs.

We have restricted our analysis to DNA- and RNA-binding motifs, although NLSs may overlap with other important functional motifs of nuclear proteins. For example, the C-terminal 28 amino acid NLS of herpes simplex virus protein ICP8 appears to overlap with a viral replication function (92). With respect to NLSs overlapping other functional motifs, it is interesting to note that the only nuclear proteins known to affect RNA export share the property that their NLSs and DNA- or RNA-binding motifs are overlapping, i.e. RCC1 (RCC1 is also a guanine nucleotide release factor for a GTPase, Ran/TC4, involved in nuclear import) (93,94), the influenza virus NS1 protein, the HIV Rev protein and its functional homologue HTLV Rex (95). This overlap may indicate that export, NLSs and RNA-binding are functionally related.

Modular, divergent and convergent evolution mechanisms

Modular, divergent or convergent mechanisms of evolution may be invoked to explain the association between NLSs and DNA- or RNA-binding motifs in one domain. Support for the modular theory of evolution in the context of this report can be understood in the following example: if a module coding for a zinc finger and an NLS were fused to another gene's coding region then it is more likely that this protein, originally cytoplasmic, would gain an additional function as the new protein now is also able to translocate to the nucleus. Hence, a module possessing both a nucleic acid-binding domain and an NLS is more likely to be

vertically transmitted and to be represented in greater proportion in the examples of DNA- and RNA-binding proteins in eukaryotic cells. The overlap of function would impose a considerable constraint on the evolution of such bifunctional domains. Such constraints have been proposed to account for the high degree of similarity of the bifunctional basic domain of bZIP proteins (96) and for the close proximity of RNA binding domains to other auxiliary domains, such as serine-arginine-rich (SR-rich) domains (97). Our findings (Table 1) are consistent with the proposal of Gashler *et al.* that NLSs and C2H2 zinc fingers (both typified by basic residues) have co-evolved (90).

An alternative evolutionary mechanism is divergence. It is possible that NLS and DNA-binding motifs were once identical and that this original motif has diverged to give two proximal specialized (and independent?) functions that further increase both DNA-binding and nuclear import efficiency.

A final possible mechanism for the overlap of NLS and DNA-binding domains is convergence. In this case, partial sequences of totally unrelated nucleic acid-binding domains would have evolved into a similar basic NLS sequence, to provide an evolutionary advantage. In fact, multiple NLSs have been shown to result in more efficient nuclear import in some systems (98). It is important to note that sequence convergence has never been proven (99).

Hypothesis

We have determined that NLSs and DNA/RNA-binding domains of the majority of nuclear proteins are either overlapping or flanking. The proximity of the motifs suggests that their association confers some evolutionary advantage. The obvious reason that DNA- or RNA-binding motifs have evolved with an overlapping or flanking NLS is the requirement of the DNA/RNA-binding proteins for the NLSs directly for their localization to the nucleus and thus indirectly for their function. However, we suggest that the proximity of the two functions may also sub-serve one or more of the following roles: the requirement of nucleic acid binding in retention of proteins in the nucleus, the additional function of the NLS in nucleic acid binding and/or the need to mask both functions simultaneously.

DNA binding, retention and nuclear localization

The localization of proteins in the nucleus is dependent on import into and export from the nucleus as well as anchorage in both cytosolic and nuclear compartments via binding sequences (retention signals). RNA- and DNA-binding domains may be considered nuclear retention signals. The import of large nuclear proteins (>70 kDa) requires an NLS. Although the role of retention in nuclear localization has been controversial, recent reports indicate that if the protein does not bind to intranuclear components it is exported out of the nucleus by a default pathway (77,100), even though the protein contains an NLS. While the protein is still competent for nuclear entry it is not retained in the nucleus. Thus, continued nuclear residency requires a nuclear-binding function such as DNA- or RNA-binding. However, it must be noted that nucleic acid binding proteins may have several nuclear retention signals mediating binding to other nuclear components which will also determine their nuclear occupancy. For example, nuclear receptors are known to bind to non-histone proteins (101) as well as to the nuclear matrix (102). It seems

likely that DNA- or RNA-binding will be but one determinant of nuclear retention of nucleic acid binding proteins.

Additional function of the NLS in DNA/RNA binding

In some of the DNA-binding proteins with overlapping NLSs it is possible to introduce point mutations that disrupt one function but not the other (18,103,104). An R to K mutation in the basic domain of the maize regulatory protein opaque-2 will no longer bind DNA but this mutant is nonetheless localized in the nucleus (104) (as noted above, another retention site is probably contributing to nuclear retention). Also, DNA binding of NGF1-A (Egr-1, Zif268, Krox-24) which has an overlapping NLS, is not required for nuclear localization as a mutant which does not bind DNA is still directed to the nucleus. On the other hand, another mutation which abolished nuclear localization had no effect on DNA binding (18). These results clearly demonstrate the overlap of two distinct and distinguishable motifs. However, possible co-operation in the action of these motifs is not ruled out. NLSs because of their basic nature may also increase DNA- and RNA-binding affinities. Hence, the NLS may have a dual function in the DNA- and RNA-binding proteins. This indeed appears to be the case for the POU homeodomain transcription factor, Pit-1, in which a deletion of part of the DNA-binding domain destroys an NLS function (105). The basic stretch of amino acids within this deletion that is believed to be the NLS of Pit-1 is known to make base specific contacts and phosphate backbone contacts for Pit-1 and other homeodomain proteins (106–108).

Co-ordinate regulation: masking of both NLS and DNA-binding functions and accessibility

NLSs have been shown to be present in regions of the protein that are accessible to antibodies, e.g., in the glucocorticoid receptor (109) and E1a protein (110). Furthermore, when an NLS was inserted into various regions of the cytosolic protein, pyruvate kinase, its function was dependent on the flanking sequences. Flanking sequences determine ultimately whether or not the sequence is exposed, distorted or folded into the protein (20). It is believed that to be functional, NLSs must be exposed so that they can interact with the molecular chaperones involved in nuclear import. The DNA-binding domain also must be exposed so that it can interact with DNA. It would be reasonable for both motifs to be placed close to each other in a protein sequence that is exposed and not buried inside the protein as part of the structural scaffold of a domain.

Regulation of the function of a protein is sometimes achieved through masking of an NLS or of a DNA-binding domain (56,111,112). Hence, another reason for the close association of NLS and DNA-binding motifs may be to facilitate mutual or simultaneous masking of both functions that occurs in some inactive proteins. The NLS and DNA-binding functions are masked either by the tertiary or quaternary structure of the protein itself, or another factor. The reason for the masking of both functions is not at first obvious. One would think that it would be necessary to inactivate only one function in order to inactivate a transcription factor. However, there may be reasons for masking both functions at once. The DNA-binding domain if exposed in the cytoplasm may bind RNA or may be inappropriately altered by post-translational modifications that will negatively affect its function at a later time.

Transcription factors such as the members of the Rel family (56,111) and some members of the nuclear receptor superfamily become 'anchored' in an inactive complex in the cytoplasm (112) in which the NLS and the DNA-binding domain are masked. Recently, this type of cytoplasmic anchoring also has been described for human heat shock factor 2 (HSF2). HSF2 is believed to reside in the cytoplasm of non-stressed cells because it is folded over onto itself through hydrophobic coiled-coil interactions or through interaction with some negative regulator (113) which results in the masking of the NLS and DNA-binding domains. Heat-shock disrupts this interaction and allows HSF2 to trimerize exposing its NLS and DNA-binding domains (113–115). Heat-shock inducible translocation of the transcription factor to the nucleus occurs followed by binding to heat-shock response elements of heat shock genes. HSF1 also displays stress-induced DNA-binding activity, oligomerization and nuclear localization (114). In support of a cytoplasmic effect which is limiting in these processes, HSF1 and HSF2 over-expressed in transfected 3T3 cells display constitutive DNA-binding activity, oligomerization and transcriptional activity and therefore by deduction, nuclear localization.

CONCLUSION

NLSs are required for the import of a large number of, and perhaps all, DNA- and nuclear RNA-binding proteins into the nucleus. A survey of members from each of the major DNA- and nuclear RNA-binding families with proven NLS sequences was conducted. We have determined that NLSs and DNA- or RNA-binding motifs directly overlap or are immediately adjacent in >79% of these nuclear proteins. The proximity of the two motifs appeared to be independent of the class of nucleic acid-binding protein and of the type of NLS. Either modular, divergent or convergent evolution may be invoked to explain the mechanism for this association. We postulate that the apparent linkage of the two motifs may be governed by one or more of the following physiological requirements in addition to the obvious requirement of a nucleic acid binding protein to be imported to the nucleus: one, continued nuclear occupancy of a protein is a consequence of both nuclear import (NLS-mediated process) and retention (DNA binding); two, the basic nature of a flanking or overlapping NLS may enhance DNA binding; three, the closeness of the two motifs may allow co-ordinate regulation.

The physical closeness of the two functions points to an association of two distinct motifs that are both required for the proper localization and functioning of DNA- and RNA-binding proteins. Our finding has implications for the mechanism of nuclear import of proteins. It suggests an additional role of DNA- and RNA-binding domains in the retention of proteins in the nucleus. Not only does the NLS target a protein to the nucleus but it is associated with a domain that retains the protein in the nucleus by binding DNA or RNA. This is further support for the increasing prominence that the role of retention is being accorded in the localization of proteins to the nucleus or nucleolus (77,79). A reasonable extension of our survey is to determine whether other binding functions in non DNA- or RNA-binding nuclear proteins are near the NLS.

Establishing the mechanism whereby NLSs and DNA/RNA-binding motifs became closely associated with each other is not testable as modular theories of evolution and other theories of evolution are not amenable to direct experimentation (116). On

the other hand, the proposed functional consequences of the proximity of the two motifs suggested are amenable to investigation. It would be of interest for example to test if the separation of the DNA-binding motif from the NLS would have detrimental consequences on the function of the nuclear protein. Although several laboratories have shown that mutations in an overlapping NLS destroy DNA-binding (67,105,117), it is not known to what extent the functions overlap, or how the specific amino acid residues contribute to the kinetics of nuclear import and/or affinity for DNA.

ACKNOWLEDGMENTS

YAL is supported by the MRC and NCI of Canada. ECLC is the recipient of an MRC studentship. We would like to thank Drs Donal Hickey, Mark Ekker and Rob Haché for their critical reading of the manuscript. We also wish to thank Dr Michael Zuker for his helpful discussions, Drs Art Szabo and Martin Tenniswood for providing computer access and Dr Heather Lochnan, Tanya Dahms and Jonathan Lakins for their help.

REFERENCES

- 1 Doolittle, R.F. and P. Bork (1993) *Scient. Amer.* **October**, 50–56.
- 2 Evans, R.M. (1988) *Science* **240**, 889–895.
- 3 Schwabe, J.W.R., D. Neuhaus and D. Rhodes (1990) *Nature* **348**, 458–461.
- 4 Härd, T., E. Kellenbach, R. Boelens, B.A. Maler, K. Dahlman, L.P. Freedman, J. Carlstedt-Duke, K.R. Yamamoto, J.-Å. Gustafsson and R. Kaptein (1990) *Science* **249**, 157–160.
- 5 Luisi, B.F., W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto and P.B. Sigler (1991) *Nature* **352**, 497–505.
- 6 Krust, A., S. Green, P. Argos, V. Kumar, P. Walter, J.M. Bornert and P. Chambon (1986) *EMBO J.* **5**, 891–897.
- 7 Gustafsson, J.-Å., J. Carlstedt-Duke, L. Poellinger, S. Okret, A.-C. Wikstrom, M. Bronnegard, M. Gillner, Y. Dong, K. Fuxe, A. Cintra, A. Harstrand and L. Agnati (1987) *Endocrinol. Rev.* **8**, 185–234.
- 8 Sherman, M.R., M.C. Moran, F.B. Tuazon and Y.-W. Stevens (1983) *J. Biol. Chem.* **258**, 10366–10377.
- 9 Green, S. and P. Chambon (1987) *Nature* **325**, 75–78.
- 10 Scherrer, L.C., D. Picard, E. Massa, J.M. Harmon, S.S. Simons Jr., K.R. Yamamoto and W.B. Pratt (1993) *Biochemistry* **32**, 5381–5386.
- 11 Jackson, P., D. Baltimore and D. Picard (1993) *EMBO J.* **12**, 2809–2819.
- 12 Newmeyer, D.D. (1993) *Curr. Opin. Cell Biol.* **5**, 395–407.
- 13 Osborne, M.A. and P. Silver (1993) *Annu. Rev. Biochem.* **62**, 219–254.
- 14 Nigg, E.A., P.A. Baueerle and R. Lührmann (1991) *Cell* **66**, 15–22.
- 15 Stochaj, U. and P. Silver (1992) *Eur. J. Cell Biol.* **59**, 1–11.
- 16 Hanover, J.A. (1992) *FASEB J.* **6**, 2288–2295.
- 17 Garcia-Bustos, J., J. Heitman and M.N. Hall (1991) *Biochim. Biophys. Acta* **1071**, 83–101.
- 18 Matheny, C., M.L. Day and J. Milbrandt (1994) *J. Biol. Chem.* **269**, 8176–8181.
- 19 Zhou, Z.-x., M. Sar, J. Simental, M.V. Lane and E.M. Wilson (1994) *J. Biol. Chem.* **269**, 13115–13123.
- 20 Roberts, B.L., W.D. Richardson and A.E. Smith (1987) *Cell* **50**, 465–475.
- 21 Ylikomi, T., M.T. Bocquel, M. Berry, H. Gronemeyer and P. Chambon (1992) *EMBO J.* **11**, 3681–3694.
- 22 Dingwall, C. and R.A. Laskey (1991) *Trends Biochem. Sci.* **16**, 478–481.
- 23 Picard, D. and K.R. Yamamoto (1987) *EMBO J.* **6**, 3333–3340.
- 24 Kalderon, D., B.L. Roberts, W.D. Richardson and A.E. Smith (1984) *Cell* **39**, 499–509.
- 25 Kalderon, D., W.D. Richardson, A.F. Markham and A.E. Smith (1984) *Nature* **311**, 33–38.
- 26 Chelsky, D., R. Ralph and G. Jonak (1989) *Mol. Cell. Biol.* **9**, 2487–2492.
- 27 Boulikas, T. (1994) *J. Cell. Biochem.* **55**, 32–58.
- 28 Boulikas, T. (1993) *CRC Crit. Rev. Euk. Gene Expr.* **3**, 193–227.
- 29 Dingwall, C., J. Robbins, S.M. Dilworth, B. Roberts and W.D. Richardson (1988) *J. Cell Biol.* **7**, 841–849.
- 30 Robbins, J., S.M. Dilworth, R.A. Laskey and C. Dingwall (1991) *Cell* **64**, 615–623.

- 31 Lanford, R.E., C.M. Feldherr, R.G. White, R.G. Dunham and P. Kanda (1990) *Exp. Cell Res.* **186**, 32–38.
- 32 Hall, M.N., L. Hereford and I. Herskowitz (1984) *Cell* **36**, 1057–1065.
- 33 Hall, N.M., C. Craik and Y. Hiraoka (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6954–6958.
- 34 Shieh, M.W., S.R. Wessler and N.V. Raikhel (1993) *Plant Physiol.* **101**, 353–361.
- 35 Standiford, D.M. and J.D. Richter (1992) *J. Cell Biol.* **118**, 991–1002.
- 36 Burgess, W.H., J. Bizik, T. Mehlman, N. Quarto and D.B. Rifkin (1991) *Cell Reg.* **2**, 87–93.
- 37 Cao, Y., M. Ekström and R.F. Pettersson (1993) *J. Cell Sci.* **104**, 77–87.
- 38 Bugler, B., F. Amalric and H. Prats (1991) *Mol. Cell. Biol.* **11**, 573–577.
- 39 Malim, H.H., S. Böhnlein, J. Hauber and B.R. Cullen (1989) *Cell* **58**, 205–214.
- 40 Zhao, L.-j. and R. Padmanabhan (1988) *Cell* **55**, 1005–1015.
- 41 Romac, J.M.-J., D.H. Graff and J.D. Keene (1994) *Mol. Cell. Biol.* **14**, 4662–4670.
- 42 Kambach, C. and I.W. Mattaj (1992) *J. Cell Biol.* **118**, 11–21.
- 43 Hamm, J., E. Darzynkiewicz, S.M. Tahara and I.W. Mattaj (1990) *Cell* **62**, 569–577.
- 44 Fischer, U., V. Sumpter, M. Sekine, T. Satoh and R. Lührmann (1993) *EMBO J.* **12**, 573–583.
- 45 Michaud, N. and Goldfarb, D. (1992) *J. Cell Biol.* **116**, 851–861.
- 46 van Zee, K., A. Dickmanns, U. Fischer, R. Lührmann and E. Fanning (1993) *J. Cell Biol.* **121**, 229–240.
- 47 Marshallsay, C. and R. Lührmann (1994) *EMBO J.* **13**, 222–231.
- 48 Fischer, U., J. Heinrich, K. van Zee, E. Fanning and R. Lührmann (1994) *J. Cell Biol.* **125**, 971–980.
- 49 Smith, M.R. and W.C. Greene (1992) *Virology* **187**, 316–320.
- 50 Silver, P.A., L.P. Keegan and M. Ptashne (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5951–5955.
- 51 Nelson, M. and P. Silver (1989) *Mol. Cell. Biol.* **9**, 384–389.
- 52 Davey, J., N.J. Dimmock and A. Colman (1985) *Cell* **40**, 667–675.
- 53 Lyons, R.H. (1991) *Mol. Endocrinol.* **5**, 1897–1902.
- 54 Duverger, E., V. Carpentier, A.-C. Roche and M. Monsigny (1993) *Exp. Cell Res.* **207**, 197–201.
- 55 Bairoch, A. (1992) *Nucleic Acids Res.* **20**, 2013–2018.
- 56 Gilmore, T.D. (1990) *Cell* **62**, 841–843.
- 57 Clever, J. and H. Kasamatsu (1991) *Virology* **181**, 78–90.
- 58 Clever, J., D.A. Dean and H. Kasamatsu (1993) *J. Biol. Chem.* **268**, 20877–20883.
- 59 Dang, C.V. and W.M.F. Lee (1989) *J. Biol. Chem.* **264**, 18019–18023.
- 60 Addison, C., J.R. Jenkins and H.-W. Sturzbecher (1990) *Oncogene* **5**, 423–426.
- 61 Shaulsky, G., N. Goldfinger, M.S. Tosky, A. Levine and V. Rotter (1991) *Oncogene* **6**, 2055–2065.
- 62 Clore, G.M., J.G. Omichinski, K. Sakaguchi, N. Zambrano, H. Sakamoto, E. Appella and A.M. Gronenborn (1994) *Science* **265**, 386–391.
- 63 Prives, C. and J.J. Manfredi (1993) *Genes Dev.* **7**, 529–534.
- 64 Foord, O.S., P. Bhattacharya, Z. Reich and V. Rotter (1991) *Nucleic Acids Res.* **19**, 5191–5198.
- 65 Wang, Y., M. Reed, P. Wang, J.E. Stenger, G. Mayr, M.E. Anderson, J.F. Schwedes and P. Tegtmeyer (1993) *Genes Dev.* **7**, 2575–2586.
- 66 Molinete, M., W. Vermeulen, A. Bürkle, J. Ménissier-de Murcia, J.H. Küpper, J.H. Hoeijmakers and G. de Murcia (1993) *EMBO J.* **12**, 2109–2117.
- 67 Citovsky, V. and P. Zambryski (1993) *Annu. Rev. Microbiol.* **47**, 167–197.
- 68 Vogel, A.M. and A. Das (1992) *J. Bacteriol.* **174**, 303–312.
- 69 Ramotar, D., C. Kim, R. Lillis and B. Demple (1993) *J. Biol. Chem.* **268**, 20533–20539.
- 70 Cleghorn, V., K. Voelkerding, N. Morin, C. Delsert and D.F. Klessig (1989) *J. Virol.* **63**, 2289–2299.
- 71 Powers, J.A. and J.C. Eissenberg (1993) *J. Cell Biol.* **120**, 291–299.
- 72 Böhnlein, E., J. Berger and J. Hauber (1991) *J. Virol.* **65**, 7051–7055.
- 73 Grassmann, R., S. Berchtold, C. Aepinus, C. Ballaun, E. Boehnlein and B. Fleckenstein (1991) *J. Virol.* **65**, 3721–3727.
- 74 McDonald, D., T.J. Hope and T.G. Parslow (1992) *J. Virol.* **66**, 7232–7238.
- 75 White, K.N., T. Nosaka, H. Kanamori, M. Hatanaka and T. Honjo (1991) *Biochem. Biophys. Res. Commun.* **175**, 98–103.
- 76 Adachi, Y., T.D. Copeland, M. Hatanaka and S. Oroszlan (1993) *J. Biol. Chem.* **268**, 13930–13934.
- 77 Schmidt-Zachmann, M.S., C. Dargemont, L.C. Kühn and E.A. Nigg (1993) *Cell* **74**, 493–504.
- 78 Schmidt-Zachmann, M.S. and E.A. Nigg (1993) *J. Cell Sci.* **105**, 799–806.
- 79 Yan, C. and T. Mélése (1993) *J. Cell Biol.* **123**, 1081–1091.
- 80 Maeda, Y., K. Hisatake, T. Kondo, K.-i. Hanada, C.-Z. Song, T. Nishimura and M. Muramatsu (1992) *EMBO J.* **11**, 3695–3704.
- 81 Breeuwer, M. and D.S. Goldfarb (1990) *Cell* **60**, 999–1008.
- 82 Treacy, M.N., L.I. Neilson, E.E. Turner, X. He and M.G. Rosenfeld (1992) *Cell* **68**, 491–505.
- 83 Benezra, R., R.L. Davis, D. Lockshon, D.L. Turner and H. Weintraub (1990) *Cell* **61**, 49–59.
- 84 Fairman, R., R.K. Beran-Steed, S.J. Anthony-Cahill, J.D. Lear, W.F. Stafford III, W.F. DeGrado, P.A. Benfield and S.L. Brenner (1993) *Proc. Natl. Acad. Sci. USA* **90**, 10429–10433.
- 85 Ellis, H.M., D.R. Spann and J.W. Posakony (1990) *Cell* **61**, 27–38.
- 86 Garrell, J. and J. Modolell (1990) *Cell* **61**, 39–48.
- 87 Jones, N. (1990) *Cell* **61**, 9–11.
- 88 Schaap, P.J., J. van't Riet, C.L. Woldringh and H.A. Raué (1991) *J. Mol. Biol.* **221**, 225–237.
- 89 Jenster, G., J. Trapman and A.O. Brinkmann (1993) *Biochem. J.* **293**, 761–768.
- 90 Gashler, A.L., S. Swamanathan and V.P. Sukhatme (1993) *Mol. Cell. Biol.* **13**, 4556–4571.
- 91 Mikaëlían, I., D. Emmanuel, V. Marechal, G. Denoyel, J.-C. Nicolas and A. Sergeant (1993) *J. Virol.* **67**, 734–742.
- 92 Gao, M. and D.M. Knipe (1993) *J. Virol.* **67**, 876–885.
- 93 Moore, M.S. and G. Blobel (1993) *Nature* **365**, 661–663.
- 94 Melchior, F., B. Paschal, J. Evans and L. Gerace (1993) *J. Cell Biol.* **123**, 1649–1659.
- 95 Krug, R.M. (1993) *Curr. Opin. Cell Biol.* **5**, 944–949.
- 96 van der Krol, A.R. and N.-H. Chua (1991) *Plant Cell* **3**, 667–675.
- 97 Fukami-Kobayashi, K., S. Tomoda and M. Go (1993) *FEBS Lett.* **335**, 289–293.
- 98 Dworetzky, S.I., R.E. Lanford and C.M. Feldherr (1988) *J. Cell Biol.* **107**, 1279–1287.
- 99 Doolittle, R.F. (1994) *Trends Biochem. Sci.* **19**, 15–18.
- 100 Laskey, R.A. and C. Dingwall (1993) *Cell* **74**, 585–586.
- 101 Spelsberg, T.C., B.J. Gosse, B.A. Littlefield, H. Toyoda and R. Seelke (1984) *Biochemistry* **23**, 5103–5113.
- 102 Barrack, E.R. and D.S. Coffey (1983) *Biochem Actions Horm.* **10**, 23–90.
- 103 Varagona, M.J., R.J. Schmidt and N.V. Raikhel (1992) *Plant Cell* **4**, 1213–1227.
- 104 Varagona, M.J. and N.V. Raikhel (1994) *Plant J.* **5**, 207–214.
- 105 Theill, L.E., J.-L. Castrillo, D. Wu and M. Karin (1989) *Nature* **342**, 945–948.
- 106 Kissinger, C.R., B. Liu, E. Martin-Blanco, T.B. Kornberg and C.O. Pabo (1990) *Cell* **63**, 579–590.
- 107 Laughon, A. (1991) *Biochemistry* **30**, 11357–11367.
- 108 Li, P., X. He, M.R. Gerrero, M. Mok, A. Aggarwal and M.G. Rosenfeld (1993) *Genes Dev.* **7**, 2483–2496.
- 109 Urda, L.A., P.M. Yen, S.S. Simons Jr. and J.M. Harmon (1989) *Mol. Endocrinol.* **2**, 251–260.
- 110 Arsenaault, H. and J.M. Weber (1993) *FEMS Microbiol. Lett.* **114**, 37–40.
- 111 Schmitz, M.L., T. Henkel and P.A. Baeuerle (1991) *Trends Cell Biol.* **1**, 130–137.
- 112 Hunt, T. (1989) *Cell* **59**, 949–951.
- 113 Sheldon, L.A. and R.E. Kingston (1993) *Genes Dev.* **7**, 1549–1558.
- 114 Baler, R., G. Dahl and R. Voellmy (1993) *Mol. Cell. Biol.* **13**, 2486–2496.
- 115 Sarge, K.D., S.P. Murphy and R.I. Morimoto (1993) *Mol. Cell. Biol.* **13**, 1392–1407.
- 116 DuBose, R.F. and D.L. Hartl (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9966–9970.
- 117 Boucher, P. and M.L. Privalsky (1990) *Oncogene* **5**, 1303–1311.
- 118 Picard, D., V. Kumar, P. Chambon and K.R. Yamamoto (1990) *Cell Regul.* **1**, 291–299.
- 119 Guiochon-Mantel, A., H. Loosfelt, P. Lescop, S. Sar, M. Atger, M. Perrot-Applanant and E. Milgrom (1989) *Cell* **57**, 1147–1154.
- 120 Guiochon-Mantel, A., P. Lescop, S. Christin-Maitre, H. Loosfelt, M. Perrot-Applanant and E. Milgrom (1991) *EMBO J.* **10**, 3851–3859.
- 121 Cadepond, F., J.-M. Gasc, F. Delahaye, N. Jibard, G. Schweizer-Groyer, I. Segard-Masurel, R. Evans and E.-E. Baulieu (1992) *Exp. Cell Res.* **201**, 99–108.
- 122 Jenster, G., H.A.G.M. van der Korput, C. van Vroonhoven, T. van der Kwast, J. Trapman and A.O. Brinkmann (1991) *Mol. Endocrinol.* **5**, 1396–1404.
- 123 Simental, J.A., M. Sar, M.V. Lane, F.S. French and E.M. Wilson (1991) *J. Biol. Chem.* **266**, 510–518.
- 124 Lee, Y. and V. Mahdavi (1993) *J. Biol. Chem.* **268**, 2021–2028.

- 125 Boucher, P., A. Koning and M.L. Privalsky (1988) *J. Virol.* **62**, 534–544.
- 126 Schreiber, V., M. Molinete, H. Boeuf, G. de Murcia and J. Mnissier-de Murcia (1992) *EMBO J.* **11**, 3263–3269.
- 127 Moll, T., G. Tebb, U. Surana, H. Robitsch and K. Nasmyth (1991) *Cell* **66**, 743–758.
- 128 Li, X. and L.D. Etkin (1993) *J. Cell Sci.* **105**, 389–395.
- 129 Miyamoto, I., N. Miura, H. Niwa, J. Miyazaki and K. Tanaka (1992) *J. Biol. Chem.* **267**, 12182–12187.
- 130 Yang, Z., L. Gu, P.-H. Romeo, D. Bories, H. Motohashi, M. Yamamoto and J.D. Engel (1994) *Mol. Cell. Biol.* **14**, 2201–2212.
- 131 Chida, K. and P.K. Vogt (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4290–4294.
- 132 Roux, P., J.-M. Blanchard, A. Fernandez, N. Lamb, P. Jeanteur and M. Piechaczyk (1990) *Cell* **63**, 341–351.
- 133 Min, S., N.T. Mascarenhas and E.J. Taparowsky (1993) *Oncogene* **8**, 2691.
- 134 Dang, C.V. and W.M.F. Lee (1988) *Mol. Cell. Biol.* **8**, 4048–4054.
- 135 Prendergast, G.C., R. Hopewell, B.J. Gorham and E.B. Ziff (1992) *Genes Dev.* **6**, 2429–2439.
- 136 Waeber, G. and J.F. Habener (1991) *Mol. Endocrinol.* **5**, 1431–1438.
- 137 Macleod, K., D. Leprince and D. Stehelin (1992) *Trends Biochem. Sci.* **17**, 251–256.
- 138 Boulukos, K.E., P. Pognonec, B. Rabault, A. Begue and J. Ghysdael (1989) *Mol. Cell. Biol.* **9**, 5718–5721.
- 139 Dimitrov, S.I., D. Bachvarov and T. Moss (1993) *DNA Cell Biol.* **12**, 275–281.
- 140 Gilmore, T.D. and H.M. Temin (1988) *J. Virol.* **62**, 703–714.
- 141 Hannink, M. and H.M. Temin (1989) *Mol. Cell. Biol.* **9**, 4323–4336.
- 142 Ganchi, P.A., S.-C. Sun, W.C. Greene and D.W. Ballard (1993) *Mol. Cell. Biol.* **13**, 7826–7835.
- 143 Gharakhanian, E., J. Takahashi and H. Kasamatsu (1987) *Virology* **157**, 440–448.
- 144 Citovsky, V., J. Zupan, D. Warnick and P. Zambryski (1992) *Science* **256**, 1802–1805.
- 145 Howard, A.H., J.R. Zupan, V. Citovsky and P.C. Zambryski (1992) *Cell* **68**, 109–118.
- 146 Tinland, B., Z. Koukolkov-Nicola, M.N. Hall and B. Hohn (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7442–7446.
- 147 Chang, D., J.I. Haynes II, J.N. Brady and R.A. Consigli (1992) *Virology* **191**, 978–983.
- 148 Chang, D., X. Cai and R.A. Consigli (1993) *J. Virol.* **67**, 6327–6331.
- 149 Lentz, M.R., D. Pak, I. Mohr and M.R. Botchan (1993) *J. Virol.* **67**, 1414–1423.
- 150 Seino, H., N. Hisamoto, S. Uzawa, T. Sekiguchi and T. Nishimoto (1992) *J. Cell Sci.* **102**, 393–400.
- 151 Gao, M. and D.M. Knipe (1992) *Mol. Cell. Biol.* **12**, 1330–1339.
- 152 Toh, Y., Y. Liu, S. Tanaka and R. Mori (1993) *Arch. Virol.* **129**, 183–196.
- 153 Morin, N., C. Delsert and D.F. Klessig (1989) *Mol. Cell. Biol.* **9**, 4372–4380.
- 154 Memer, B. and C. Dreyer (1993) *Eur. J. Cell Biol.* **61**, 369–382.
- 155 van Wijk, I., J. Burfeind and T. Pieler (1992) *Mech. Dev.* **39**, 63–72.
- 156 Poisson, F., P. Roingeard, A. Baillou, F. Dubois, F. Bonelli, R.A. Calogero and A. Goudeau (1993) *J. Gen. Virol.* **74**, 2473–2477.
- 157 Xia, Y.-P., C.-T. Yeh, J.-H. Ou and M.M.C. Lai (1992) *J. Virol.* **66**, 914–921.
- 158 Chang, M.-F., S.C. Chang, C.-I. Chang, K. Wu and H.-Y. Kang (1992) *J. Virol.* **66**, 6019–6027.
- 159 Nath, S.T. and D.P. Nayak (1990) *Mol. Cell. Biol.* **10**, 4139–4145.
- 160 Qian, X.-Y., F. Alonso-Caplen and R.M. Krug (1994) *J. Virol.* **68**, 2433–2441.
- 161 Siomi, H., H. Shida, S.H. Nam, T. Nosaka, M. Maki and M. Hatanaka (1988) *Cell* **55**, 197–209.
- 162 Churcher, M.J., C. Lamont, F. Hamy, C. Dingwall, S.M. Green, A.D. Lowe, P.J.G. Butler, M.J. Gait and J. Karn (1993) *J. Mol. Biol.* **230**, 90–110.