

# Nuclear localization signals in the *Xenopus* FGF embryonic early response 1 protein

Janine N. Post, Laura L. Gillespie, Gary D. Paterno\*

Division of Basic Medical Sciences, Faculty of Medicine, Memorial University of Newfoundland, 300 Prince Phillip Drive, St. John's, NF, Canada

Received 30 May 2001; revised 22 June 2001; accepted 22 June 2001

First published online 9 July 2001

Edited by Felix Wieland

**Abstract** *Xenopus* early response 1 (XER1) is a fibroblast growth factor-inducible transcription factor whose developmentally regulated nuclear localization is thought to be important in the control of cell differentiation during embryonic development [Luchman et al., Mech. Dev. 80 (1999) 111–114]. Analysis of the XER1 amino acid sequence revealed four regions which contain potential nuclear localization sequences (NLSs). Using mutant XER1 proteins and portions of XER1 fused to green fluorescent protein (GFP) transfected into NIH 3T3 cells, we have determined that only one of these, NLS4, located near the carboxy-terminus of XER1, is necessary and sufficient for targeting exclusively to the nucleus. Of the other three predicted NLS sequences, only NLS1, consisting of the sequence <sup>138</sup>RPRRCK<sup>143</sup> was shown to function as a cryptic, weak NLS. NLS4 contains a core region consisting of the sequence <sup>463</sup>RPIKRQRM<sup>471</sup> which is similar to the core NLS directing the human c-MYC protein to the nucleus. The core sequence is flanked by a predicted cdc2/protein kinase A phosphorylation motif, however mutation of the serine<sup>472</sup> to alanine or aspartic acid had no detectable effect on accumulation of GFP-XER1 fusion proteins in the nucleus, demonstrating that this putative phosphorylation site plays no role in regulating nuclear transport. © 2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

**Key words:** Fibroblast growth factor; Nuclear localization; Early response 1; Green fluorescent protein; cdc2; Phosphorylation; *Xenopus*

## 1. Introduction

*Xenopus* early response 1 (XER1) is a fibroblast growth factor (FGF)-inducible early response gene activated during mesoderm induction in *Xenopus laevis* embryos [2]. XER1 protein was shown to be specifically localized to the nucleus and to possess potent transcription regulatory activity and thus is implicated in the regulation of differentiation-specific gene regulation. During *Xenopus* embryogenesis, the translocation of XER1 protein to the nucleus is precisely regulated [1]. XER1 protein is present only in the cytoplasm of the fertilized egg and persists in the cytoplasm throughout cleavage stages of early development. At mid-blastula (stage 8), movement of XER1 into the nucleus begins, but ubiquitous nuclear localization is not complete until several hours later,

at early gastrula stages. Following gastrulation, XER1 gradually disappears from the nuclei and by tadpole stages, there is no nuclear XER1 except in a few endodermal cells. The precise developmental regulation of this nuclear localization of XER1, together with its potential role as a regulator of transcription, suggests that the control of nuclear entry may be an important mechanism for regulating XER1 activity during embryonic development. We have also cloned and characterized the human homolog of XER1 and have shown that this protein is differentially expressed in breast tumors [3]. Thus, elucidation of the sequences responsible for directed nuclear localization will be important for understanding the function of XER1.

The nuclear pore complex is the checkpoint for protein transport into the nucleus [4]. In principle, proteins smaller than 60 kDa can enter the nucleus by diffusion through the nuclear pore. However, larger proteins require the importin transport mechanism and its recognition of intrinsic nuclear localization signals (NLSs) uniquely present in proteins which are targeted to the nucleus [4]. Analysis of the nuclear localization signals of many proteins has revealed four main classes of intrinsic nuclear localization signals. The first consists of a single cluster of basic amino acids, as exemplified by the NLS of SV40 large T antigen (PKKKRKV). In general, this NLS is a hexapeptide of which a minimum of four amino acids are positively charged. Moreover, this type of NLS contains no bulky, acidic, or hydrophobic residues in the core or flanking regions. The core cluster is usually flanked by proline or glycine that are thought to be important for secondary structure. The second class is a bipartite type, where two clusters of basic residues are separated by a 10–14 mutation-insensitive spacer, as first described in the *Xenopus* nucleoplasmin protein (KKPAATKKAGQAKKKK). The third class is exemplified by the c-MYC NLS in which only three of nine in the cluster are basic residues (PAAKRVKLD) and where the amino-terminal proline and carboxy-terminal acidic residue are thought to be important for function [5–7]. The fourth class includes various other NLSs, such as the ones associated with ribosomal proteins and hnRNPs [8,9].

In this report we describe the identification of a single, functional NLS that is similar in sequence to the c-MYC NLS and that is located near the carboxy-terminus of XER1. Although the XER1 protein possesses three additional putative NLS signals, only NLS1 may function as a cryptic, weak NLS for XER1. Furthermore, site-directed mutagenesis of a potential cdc2/protein kinase A (PKA) phosphorylation site adjacent to the core region of NLS4 had no effect on the targeting of XER1 constructs to the nucleus.

\*Corresponding author. Fax: (1)-709-777 7010.

E-mail address: gpaterno@morgan.ucs.mun.ca (Gary D. Paterno).

Fig. 1. Identification of potential NLS in XER1. A: Schematic of the XER1 protein showing all prolines and basic residues within the sequence, with the remaining residues as dots. Four clusters of basic amino acids (shaded) that we called NLS1-4 and that have the potential to function in targeting XER1 to the nucleus. B: Amino acid comparison of NLS4 of XER1 to the NLS of the c-MYC protein. A core region within the identified NLS4 shows significant homology to the core NLS of c-MYC. Identity and conservative changes are indicated by double and single dots, respectively.

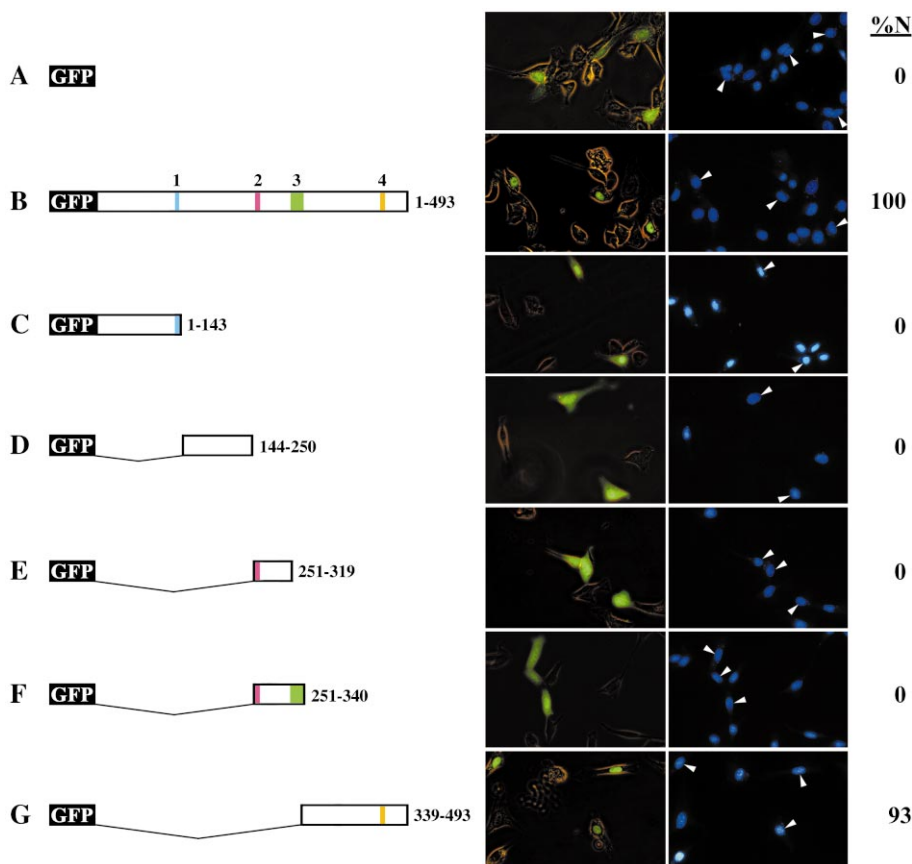


Fig. 2. Nuclear localization of GFP-ER1 fusion proteins. GFP-ER1 fusion proteins spanning the entire XER1 protein were constructed and transfected into NIH 3T3 cells. Subcellular localization was determined by fluorescence microscopy 48 h after transfection. A schematic illustrating each of the fusion constructs is shown on the left. Numbers on top of the GFP-ER1-493 construct indicate the position of the corresponding NLS described in Fig. 1. The numbers beside each construct indicate the XER1 amino acid numbers contained in the GFP fusion protein. Photographs illustrating the subcellular localization of each GFP fusion protein and DAPI staining of the same field to visualize nuclei (arrows) are shown for each construct. The percentage of cells showing exclusive nuclear localization (%N) is indicated in the column on the right hand side. A: GFP alone. B: GFP-ER1-493 (full-length). C: GFP-ER1-143. D: GFP-ER144-250. E: GFP-ER251-319. F: GFP-ER251-340. G: GFP-ER339-493.

in 4% paraformaldehyde in phosphate-buffered saline (PBS; pH 7.4) for 30 min. After fixation, the nuclei were stained by incubation in 20 ng/ml Hoechst 33342 (Sigma) in PBS for 10 min. After washing in PBS, the slides were mounted in 10% glycerol in PBS, viewed by combined fluorescence/brightfield microscopy on an Olympus BH2 microscope using DPlanApo/UV objectives and the images captured with a Coolsnap digital camera. Subcellular localization of the constructs was compared to GFP and GFP-ER1-493. All experiments were repeated at least three times and at least 200 cells were counted for each construct in each experiment. Cells were scored as either N (exclusively nuclear fluorescence), N>C (nuclear fluorescence was greater in intensity than cytoplasmic) or WC (whole cell fluorescence, in which nuclear fluorescence was equal in intensity to cytoplasmic fluorescence). XER1 possesses an estimated molecular weight of 56 000 Da, based upon amino acid composition, and thus the majority of the GFP-ER1 constructs will be below the approximately 60 kDa cut-off which prevents proteins from entering the nucleus by simple diffusion [10,11].

#### 2.4. Western blotting

At 48 h after transfection, cells grown in six well plates were lysed in sample buffer (0.125 M Tris pH 6.8, 2% SDS, 5%  $\beta$ -mercaptoethanol, 2% glycerol) and separated by SDS-PAGE. The proteins were electroblotted onto a Hybond membrane (Amersham Pharmacia Biotech) and fusion proteins were visualized using an anti-GFP antibody (1:1000) (Clontech) or an anti-XER1 polyclonal antiserum (1:5000) followed by chemiluminescence detection, using the Femtolucent system (Chemicon).

### 3. Results and discussion

#### 3.1. Identification of putative NLS sequences in XER1

We demonstrated previously that XER1 is localized exclu-

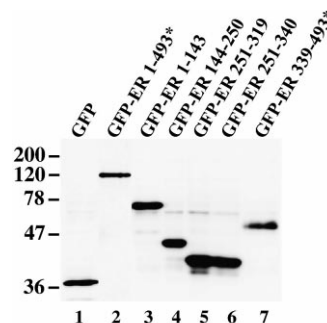


Fig. 3. Western blot analysis of GFP-ER1 fusion proteins expressed in NIH 3T3 cells. Cells transfected with the indicated GFP fusion construct were harvested 48 h after transfection and subjected to Western blot analysis using an anti-GFP antibody, as described in Section 2. The positions of molecular weight markers are indicated on the left. All constructs demonstrated similar levels of expression in NIH 3T3 cells; asterisks indicate the fusion proteins that localized to the nucleus.

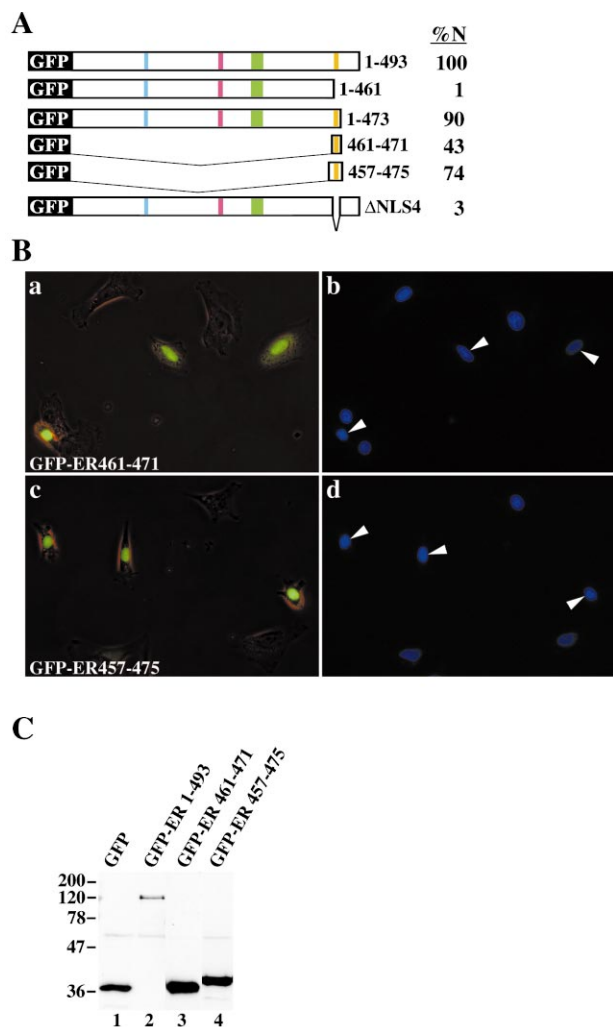


Fig. 4. Identification of NLS4 as a strong NLS sequence in XER1. GFP-ER1 fusion proteins containing various XER1 NLS4 constructs were transfected into NIH 3T3 cells and scored for subcellular localization after 48 h, in order to define the minimal sequence that could function as an NLS. A: Schematic illustrating the fusion constructs; the percentage of cells (%N) demonstrating exclusive nuclear localization is listed on the right of each construct. B: Photographs illustrating nuclear localization of the GFP-ER461–471 (panel a) and GFP-ER457–475 (panel c) constructs; DAPI staining of the same fields is shown in panels b and d. Arrows indicate the nuclei of cells expressing GFP fusion proteins. C: Western blot analysis of cells expressing GFP-GFP-ER461–471 (lane 3) or ER457–475 (lane 4), along with controls expressing either GFP (lane 1) or GFP-ER1–493 (lane 2). The positions of molecular weight markers are indicated on the left.

sively to the nucleus in transfected NIH 3T3 cells [2]. Examination of the amino acid sequence of XER1 and analysis by the PSORT computer program [12] revealed four clusters of positively charged amino acids which have the potential to serve as putative NLSs (Fig. 1A). In NLS1 (<sup>138</sup>RPRRCK<sup>143</sup>), the four positively charged residues are flanked by tyrosine and phenylalanine, both bulky residues which are predicted to interfere with recognition of the NLS by the import apparatus [13]. The NLS2 core cluster (<sup>261</sup>RRLR<sup>264</sup>) contains only three positively charged amino acids and is flanked by bulky residues. NLS3 (<sup>320</sup>KKSERDYFFAQQTRFGKKK<sup>338</sup>) was very similar to the described canonical NLS sequences of the bipartite type [4,14]. NLS4 (<sup>463</sup>RPIKRQRMDSPGK<sup>475</sup>)

was most similar to the c-MYC type of NLS (Fig. 1B) in which the cluster of positively charged residues is flanked by a proline and an acidic residue [5–7].

### 3.2. XER1 contains a functional NLS near the C-terminus

To determine which of the putative NLS sequences were functional, we fused portions of XER1 in frame to a heterologous cytoplasmic protein (GFP) in the expression plasmid EGFP2. These constructs were transiently transfected into NIH 3T3 cells and the subcellular localization of these fusion proteins monitored by fluorescence microscopy 48 h after transfection. In the first series of experiments, we generated six GFP fusion constructs containing different regions of XER1: GFP-ER1–493, GFP-ER1–143, GFP-ER144–250, GFP-ER251–319, GFP-ER251–340 and GFP-ER339–493 and transfected these constructs into NIH 3T3 cells (Fig. 2). We compared the subcellular localization of these fusion proteins to cells transfected with GFP alone or with GFP fused to full-length XER1 (GFP-ER1–493). The exclusive nuclear localization of GFP-ER1–493 in all experiments demonstrated that the GFP tag did not disturb the nuclear localization of the fusion proteins and that the chosen conditions did not result in a saturation of the import apparatus. In addition, cells transfected with each of the constructs had similar levels of fusion protein expression, as determined by Western blotting (Fig. 3). The only construct, other than the full-length XER1, that showed exclusively nuclear staining was GFP-ER339–493, containing NLS4 (Fig. 2G). In cells transfected with GFP-ER1–143, the intensity of nuclear fluorescence was noticeably greater than that in the cytoplasm (Fig. 2C, see Section 3.5).

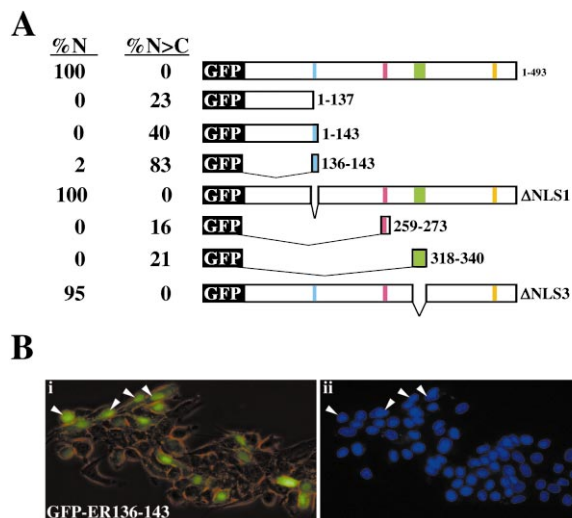


Fig. 5. Identification of NLS1 as a potential additional weak NLS sequence. A: Constructs expressing GFP-ER1 fusion proteins containing deletions of predicted XER1 NLS sequences and those containing putative core NLS sequences were transfected into NIH 3T3 cells and the subcellular localization scored 48 h after transfection. Cells were scored as having exclusively nuclear (%N) localization, nuclear localization greater than cytoplasmic (%N>C) and those where the whole cell fluoresced evenly (%N=C, not listed). The values for %N and %N>C are reported to the left of each construct. B: NIH 3T3 cells expressing GFP-ER136–143 (panel i), an example of N>C localization; corresponding DAPI staining for nuclei (arrows) is shown in panel ii.

### 3.3. NLS4 is necessary and sufficient for nuclear localization of XER1

In order to narrow down the region directing the localization of XER1 to the nucleus, we constructed GFP-ER1–461 which is truncated just before NLS4, and GFP-ER1–473 which includes the core NLS4. GFP-ER1–473 was almost exclusively nuclear (90%N), while GFP-ER1–461 was not (1%N; Fig. 4A), indicating that residues between 461 and 473 are important. In order to identify the minimal region of XER1 required for directing GFP to the nucleus, we generated two constructs: GFP-ER457–475 which contains the full NLS4 sequence and GFP-ER461–471 which contains the core region. GFP-ER457–475 (<sup>457</sup>PDDTNERPIKQR-MDSPGK<sup>475</sup>) was almost as efficient as the full-length sequence at directing exclusive nuclear targeting (74%N), while GFP-ER461–471 was much less efficient (43%N; Fig. 4A–C). To verify that NLS4 is a bona fide NLS, we constructed GFP-ERΔ464–467 which deletes residues **PIKR** from the core. When transfected into NIH 3T3 cells, this deletion virtually abolished nuclear localization (3%N, Fig. 4A). Thus, the NLS4 sequence is necessary and sufficient to direct XER1 to the nucleus.

GFP-ER461–471 contains the predicted core NLS, including the amino-terminal P<sup>464</sup>, the neutral Q<sup>468</sup> and the acidic D<sup>471</sup> residues, all of which have been shown to augment the efficacy of the core c-MYC NLS [5–7]. The reduced nuclear targeting efficiency of this construct may be due to the exclusion of flanking P<sup>457</sup> and <sup>473</sup>PG<sup>474</sup> residues that might be important for secondary structure and/or recognition by the import apparatus. Alternatively, differential phosphorylation of T<sup>460</sup> or S<sup>472</sup> may be critical.

### 3.4. Mutation of a cdc2/PKA site adjacent to NLS4 does not alter nuclear localization of XER1

It is well documented that phosphorylation of sites in and around NLSs can have a profound influence on the recognition of the NLS, the rate of nuclear transport and the level of nuclear accumulation of targeted proteins [4,15]. Examination and computer analysis of the GFP-ER457–475 sequence using pbase\_predict ([www.cbs.dtu.dk/databases/PhosphoBase/predict/predform.html](http://www.cbs.dtu.dk/databases/PhosphoBase/predict/predform.html)) and NetPhos ([www.cbs.dtu.dk/services/NetPhos](http://www.cbs.dtu.dk/services/NetPhos)) [16] revealed a potential cdc2/PKA phosphorylation site at S<sup>472</sup>, adjacent to the core. We investigated the potential role of the cdc2/PKA site by mutating S<sup>472</sup> → A or → D in GFP-ER339–493 and in the full-length XER1 fusion protein, GFP-ER1–493. When these mutants were transfected into NIH 3T3 cells and scored for nuclear localization after 24 and 48 h, no significant augmentation or reduction in nuclear localization was evident for any of the mutated constructs (data not shown). This suggests that phosphorylation of this cdc2/PKA site is not important for regulating nuclear targeting of XER1 in NIH 3T3 cells.

### 3.5. Additional weak NLSs in XER1

Several of the GFP-ER fusion constructs which did not

contain NLS4 and which did not demonstrate exclusive nuclear localization appeared to have a greater level of fusion protein in the nucleus than in the cytoplasm, not accountable for by simple diffusion. This suggested that there might be additional, albeit weak NLSs in XER1. We generated additional fusion constructs in an attempt to identify regions of XER1 which could function as weak NLSs. These included GFP-ER1–137, GFP-ER1–143, GFP-ER136–143, GFP-ERΔ136–143, GFP-ER259–273, GFP-ER318–340 and GFP-ERΔ318–340 (Fig. 5A). After transfection, only GFP-ER1–143 and GFP-ER136–143 demonstrated a significant number of cells (40% and 83%, respectively) in which nuclear fluorescence was greater than cytoplasmic (Fig. 5B). Thus, in absence of NLS4, the predicted NLS1 appears to function as a weak nuclear targeting signal for XER1. There are several nuclear proteins which contain multiple NLS sequences which cooperate to achieve nuclear localization, such as in the APC [17] and p53 [18] proteins. Alternatively, there are those in which the additional predicted NLSs function in a cryptic, less efficient manner, such as in c-MYC [5] and BRCA1 [19] proteins.

**Acknowledgements:** This work was supported by a grant from the Canadian Institutes for Health Research to L.L.G. and G.D.P.

### References

- [1] Luchman, H.A., Paterno, G.D., Kao, K.R. and Gillespie, L.L. (1999) *Mech. Dev.* 80, 111–114.
- [2] Paterno, G.D., Li, Y., Luchman, H.A., Ryan, P.J. and Gillespie, L.L. (1997) *J. Biol. Chem.* 272, 25591–25595.
- [3] Paterno, G.D., Mercer, F.C., Chayter, J.J., Yang, X., Robb, J.D. and Gillespie, L.L. (1998) *Gene* 222, 77–82.
- [4] Jans, D.A. and Hubner, S. (1996) *Physiol. Rev.* 76, 651–685.
- [5] Dang, C.V. and Lee, W.M. (1988) *Mol. Cell. Biol.* 8, 4048–4054.
- [6] Makkerh, J.P., Dingwall, C. and Laskey, R.A. (1996) *Curr. Biol.* 6, 1025–1027.
- [7] Hodel, M.R., Corbett, A.H. and Hodel, A.E. (2001) *J. Biol. Chem.* 276, 1317–1325.
- [8] Weighardt, F., Biamonti, G. and Riva, S. (1995) *J. Cell Sci.* 108, 545–555.
- [9] Gorlich, D. and Mattaj, I.W. (1996) *Science* 271, 1513–1518.
- [10] Paine, P.L., Moore, L.C. and Horowitz, S.B. (1975) *Nature* 254, 109–114.
- [11] Ribbeck, K. and Görlich, D. (2001) *EMBO J.* 20, 1320–1330.
- [12] Nakai, K. and Kanehisa, M. (1992) *Genomics* 14, 897–911.
- [13] Boulikas, T. (1997) *Anticancer Res.* 17, 843–863.
- [14] Robbins, J., Dilworth, S.M., Laskey, R.A. and Dingwall, C. (1991) *Cell* 64, 615–623.
- [15] Jans, D.A., Xiao, C.Y. and Lam, M.H. (2000) *BioEssays* 22, 532–544.
- [16] Blom, N., Gammeltoft, S. and Brunak, S. (1999) *J. Mol. Biol.* 294, 1351–1362.
- [17] Zhang, F., White, R.L. and Neufeld, K.L. (2000) *Proc. Natl. Acad. Sci. USA* 97, 12577–12582.
- [18] Shaulsky, G., Goldfinger, N., Ben-Ze'ev, A. and Rotter, V. (1990) *Mol. Cell. Biol.* 10, 6565–6577.
- [19] Thakur, S., Zhang, H.B., Peng, Y., Le, H., Carroll, B., Ward, T., Yao, J., Farid, L.M., Couch, F.J., Wilson, R.B. and Weber, B.L. (1997) *Mol. Cell. Biol.* 17, 444–452.