The nuclear localization signal of zebrafish *terra* is located within the DM domain

Lixiang Zhang[1], Zhengchun Hua[1], Jianguo Ren, Anming Meng*

Institute of Cellular and Developmental Biology, Department of Biological Sciences and Biotechnology, National Laboratory of Protein Sciences of MOE, Tsinghua University, Beijing 100084, China

Received 25 June 2001; revised 5 July 2001; accepted 7 July 2001

First published online 23 July 2001

Edited by Ned Mantei

Abstract Zebrafish Terra is a member of the DM domain-containing transcription factor family and is involved in somitogenesis. The other known members of this family play a role in sex differentiation across species from *Caenorhabditis elegans* to human. Using the green fluorescence protein–Terra fusion constructs, we have identified the nuclear localization signal (NLS) of terra by transfecting human HeLa cells. The terra NLS is located between the two intertwined zinc-binding sites of the DNA-binding domain. However, the nuclear translocation of terra is independent of the structure required for DNA binding. Mutational analysis demonstrates that basic residues K77 and R78 within the DM domain are absolutely required for the translocation of Terra into the nuclei. Sequence comparison discloses that the NLS of Terra is also present in the other known members of the DM family, indicating the conservative nature of the NLS of this family during evolution. © 2001 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.

Key words: Zebrafish; Terra; DM domain; Nuclear localization signal

1. Introduction

The DM gene family, a novel transcription factor family, is characterized by sharing a zinc finger-like DNA-binding motif called the DM domain. The majority of the DM family members, including the *doublesex* (*dbx*) in *Drosophila* [1], the *mab-3* in *Caenorhabditis elegans* [2] and the *Dmrt1* in mouse [3], play important roles in sex determination in a variety of animal species from invertebrates to mammals. The fly *dbx* gene, the first member of this family to be identified, is specifically expressed in germ cells and required for the development of sexually dimorphic characteristics through alternative splicing [1]. Recent studies have shown that the *Mab-3* in the nematode and *Dmrt1* in vertebrates are essential for male development [2–4]. The human *DMRT1*, *DMRT2* and *DMRT3* genes are located on distal 9p, a region deleted in patients with 46, XY male-to-female sex reversal [5,6]. In chicken, *Dmrt1* is Z-linked and expressed in the genital ridge and Wolffian duct prior to sexual differentiation [4]. The rainbow trout *Dmrt1* is expressed in males during testicular differentiation [7]. Transcription factor Terra, a DM family member identified in zebrafish and mouse, is different in expression pattern from the other known members [8]. The *terra* gene is transiently expressed in the presomitic mesoderm and the newly formed somites, suggesting a role in somitogenesis during vertebrate embryogenesis [8]. It is obvious that the DM genes have divergent functions.

After synthesis in the cytoplasm, transcription factors are targeted to the nucleus by an active transport process. Efficient targeting requires nuclear localization signals (NLS) present in the import proteins. The NLS-binding proteins recognize and bind to the NLS of an import protein, and the formed complexes transport through nuclear pore complexes into the nucleus in an ATP-dependent manner [9]. A large number of NLSs have been identified, but their sequences are divergent. The NLSs identified so far are classified into two main types. The first one consists of a stretch of basic amino acids, which is represented by the NLS (PKKKRKV) present in the simian virus 40 large T-antigen [10]. The second type contains two stretches of basic amino acids and a spacer of 10–15 amino acids [11]. It appears that NLS of most DNA-binding proteins are located within or in the vicinity of their DNA-binding domains.

In this study, we have identified a NLS of zebrafish Terra by fluorescence microscopy using green fluorescence protein (GFP) marker. This signal is located in between the first zinc-binding sites of the DM domain.

2. Materials and methods

2.1. Plasmid construction

A sequence encoding the second to the last amino acids of zebrafish Terra was amplified with primers P5′*Sal*I (5′-ACGCGTCGACG-GATCTGTCCGGCACGG-3′) and P3′*Bam*HI (5′-CGCGGATCCT-TACTGAGATTTCCGATTTAAAG-3′). After digestion with *Sal*I and *Bam*HI, the amplified fragment was inserted into *Xho*I and *Bam*-HI sites of expression vector pEGFP-C2 (Clontech). The resulted recombinant plasmid pGFP–Terra can produce GFP–Terra fusion protein. The following constructs were designated according to the retained residues of Terra. Construct pK115–Q507 was generated by fusing a region positioning from 543 to 1724 of *terra* cDNA in frame to the GFP coding region in pEGFP-C2, which lacks the sequence encoding the putative DNA-binding domain of Terra. Constructs pT2–G148 and pT2–R350 were generated by deleting the 1030-bp *Eco*V/*Bam*HI fragment and the 605-bp *Xho*I/*Bam*HI fragment in the 3′ region of *terra* cDNA from the pGFP–Terra, respectively. The other deletion and mutation constructs were generated by inserting a required fragment, which was a *Sal*I/*Eco*RI-digested product of amplified fragment using a pair of primers, in frame into *Xho*I and *Eco*RI sites of pEGFP-C2. Primers P5′*Sal*I and P470 (5′-CG-GAATTCCCAGTTTGCGCACTGACAGTC-3′) were used for generating pT2–C90, P5′*Sal*I and P434 (5′-CGGGAATTCCCT-CTTGTGCCCCTTTAGACA-3′) for pT2–R78, P5′*Sal*I and P419

*Corresponding author. Fax: (86)-10-62772256.
E-mail address: mengam@mail.tsinghua.edu.cn (A. Meng).

[1] These authors contributed equally.

0014-5793/01/$20.00
PII: S0014-5793(01)02677-1

(5′-CGGGAATTCTAGACAAGACACAACGCCGTG-3′) for pT2–L73, P5′*Sal*I and P401 (5′-CGGGAATTCGTGGTTCCTACATCGGGCGCA-3′) for pT2–H67, P5′*Sal*I and P356 (5′-CGGAATTCTGGTCTCCAGAGCCGAG-3′) for pT2–Q52, P5′*Sal*I and PK77E (5′-CGGGAATTCCCTCTCGTGCCCCTTTAGACAAGA-3′, the underlined is a mutated base) for pT2–R78–K77E, P5′*Sal*I and PR78E (5′-CGGGAATTCCTCCTTGTGCCCCTTTAGACAAGA-3′) for pT2–R78–R78E, P354 (5′-ACGCGTCGACGCAGCGCAAACTGAGCCGCA-3′) and P716 (5′-GGAATTCCATGCGGTCGCTCAGAGGAG-3′) for pQ52–M173, P366 (5′-ACGCGTCGACAGCCGCACGCCCAAATGCGC-3′) and P716 for pS56-M173, and P420 (5′-ACGCGTCGACAAGGGGCACAAGAGGTTCTGC-3′) and P716 for pK74-M173. The sequences containing mutations were confirmed by sequencing.

To test whether KGFKR alone can confer the ability to be translocated, a region downstream of the multiple cloning sites of pEGFP-C2 was amplified with primers NLS5′ (5′-GGAATTC AAG GGG CAC AAG AGG ATCCACCGGATCTAGATA-3′; the underlined sequence encodes the KGFKR) and GFP3′ (5′-GTGCAATCCATCTTGTTCAATC-3′). The amplified product was digested with *Eco*RI and *Afl*III, and the resulted fragment was inserted into *Eco*RI/*Afl*III sites of pEGFP-C2 to generate construct pK74-R78 so that the putative NLS KGFKR was placed C-terminal to GFP. Two identified clones for pK74-R78 were sequenced to confirm the correct addition of the sequence encoding the NLS.

2.2. Cell culture and transfection

Human HeLa cells were grown in Dulbecco's modified Eagle's medium (Life Technologies, Inc.) supplemented with 10% fetal bovine serum at 37°C in a chamber supplied with 5% CO_2. When cell density in a culture plate reached 70% confluence, the cells were transfected with plasmid DNA using the Calcium Phosphate Transfection system (Life Technologies, Inc.) according to the manufacturer's recommendations. The old medium was replaced with fresh medium 12 h after transfection and then incubation continued until observation.

2.3. GFP observation

The transfected cells were observed 40–48 h after transfection for GFP expression by fluorescence microscopy under a GFP filter (Leica). Pictures were taken using a digital camera connected to the microscope. The presented pictures were produced by superimposing a fluorescent image on the parallel bright-field image using Adobe Photoshop software.

3. Results and discussion

3.1. NLS of Terra is located within the DNA-binding domain

We have previously shown that zebrafish *terra* encodes a putative transcription factor, since it contains a potential DNA-binding domain sharing a high identity to that of *Drosophila* Doublesex (DSX). Nuclear localization of Terra was evidenced by the fact that Terra–GFP fusion protein was completely translocated into the nuclei after the mRNA encoding the fusion protein was microinjected into zebrafish fertilized eggs [8]. To facilitate the identification of Terra's NLSs through deletion analysis, we generated construct pGFP–Terra by fusing Terra to the C-terminus of GFP in this study. Human HeLa cells were transfected with pGFP–Terra and transient expression of GFP was observed by fluorescence microscopy. Control transfection with pEGFP-C2 shows that GFP is distributed both in the cytoplasm and the nuclei of the transfected cells (Fig. 2A). By contrast, cells transfected with pGFP–Terra accumulate the GFP fusion protein exclusively in the nuclei (Fig. 2B), suggesting that Terra protein must contain NLS. Since the cytoplasmic localization of the GFP fusion protein is almost invisible, the retention of any new GFP–Terra fusion proteins in the cytoplasm will indicate loss of the NLSs in the retained Terra regions.

To identify the NLSs of Terra, a series of deletion constructs was generated for transfection (Fig. 1). The first two deletion constructs, pT2–R350 and pT2–G148, which both retain the putative DNA-binding domain (Q52–D114) of

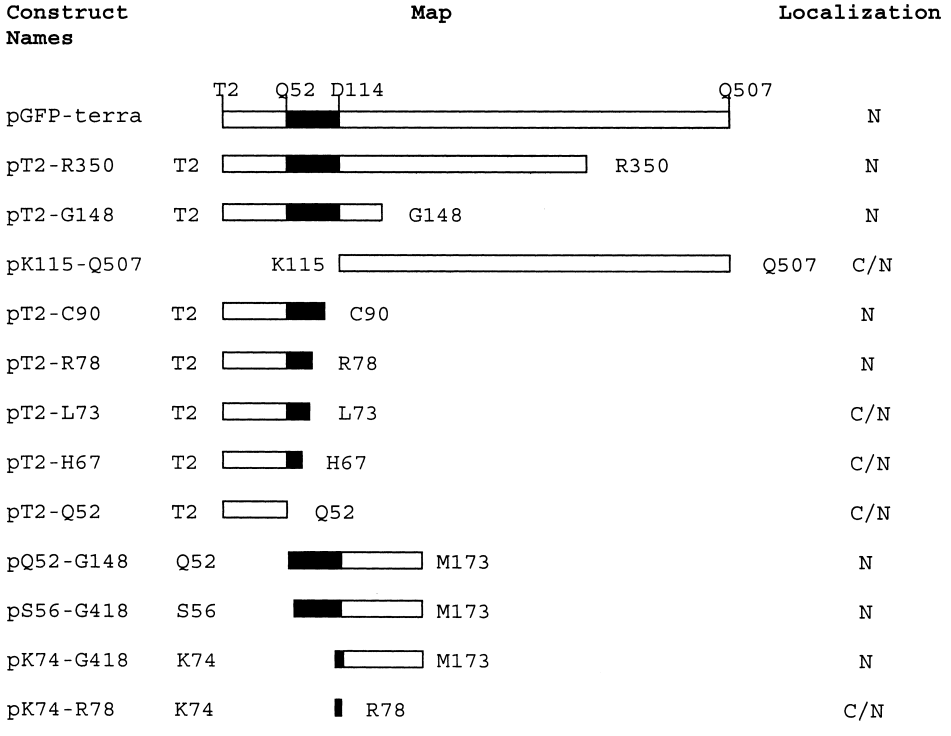

Fig. 1. Schematic drawing of Terra and deletion constructs. The filled box represented the putative DNA-binding domain. The N- and C-terminal residues of retained Terra peptides in each construct are indicated. Localization of the GFP–Terra fusion proteins after transfection is indicated on the right as N for the nuclei and C for the cytoplasm.

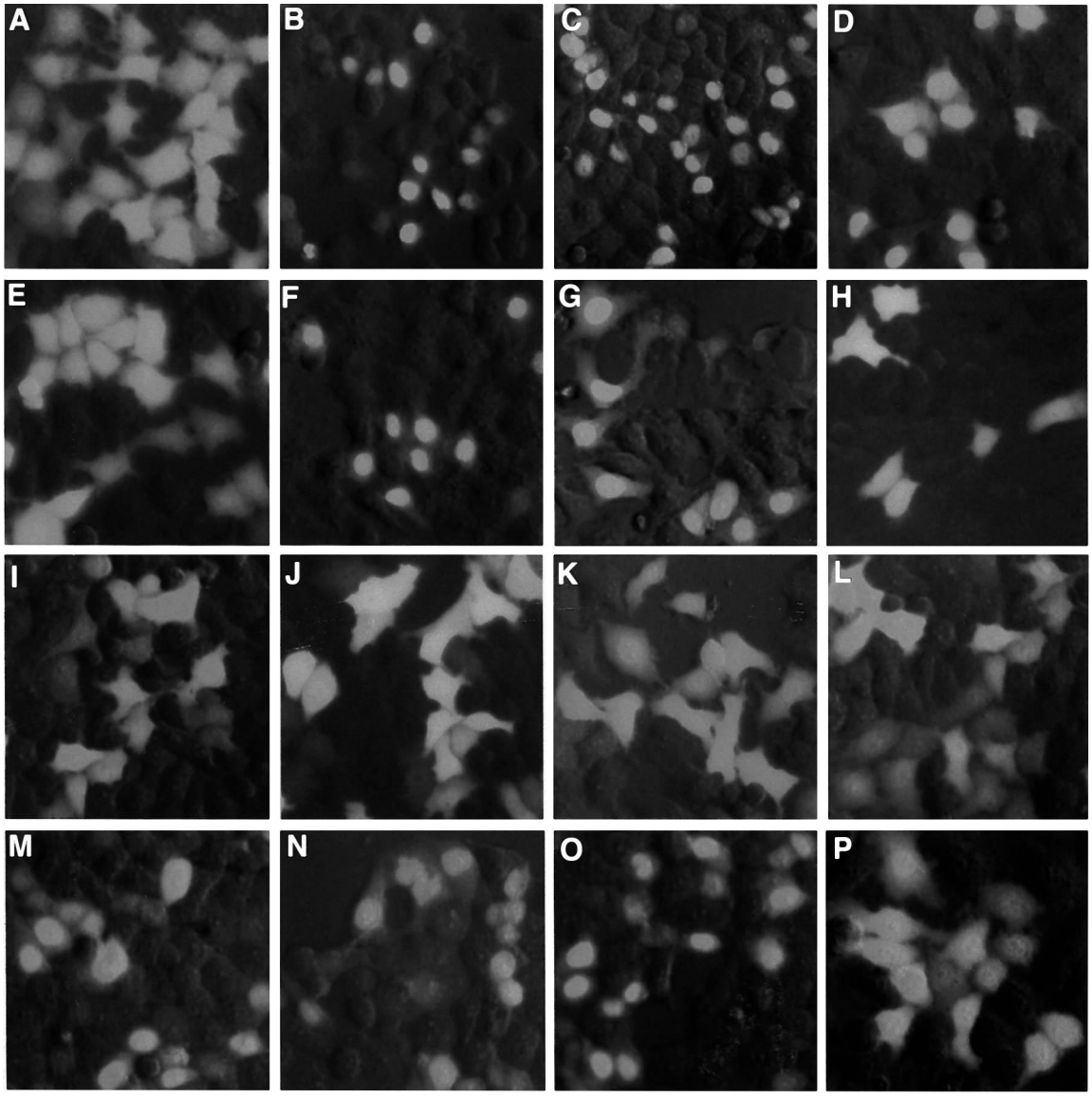

Fig. 2. A–L: Localization of GFP–Terra fusion proteins in HeLa cells transfected with various constructs. The constructs include pEGFP-C2 (A), pGFP–Terra (B), pT2–R350 (C), pT2–G148 (D), pK115–Q507 (E), pT2–C90 (F), pT2–R78 (G), pT2–L73 (H), pT2–H67 (I), pT2–Q52 (J), pT2–R78–K77E (K), pT2–R78–R78E (L), pQ52–M173 (M), pS56–M173 (N), pK74–G173 (O) and pK74–R78 (P).

Terra, can efficiently target the GFP to the nuclei (Fig. 2C,D). Construct pK115–Q507, which lacks the DNA-binding domain, gives rise to preservation of GFP in the cytoplasm (Fig. 2E) as does pEGFP-C2. Another construct, pT2–Q52, which has 51 amino acids at the N-terminus of Terra but excludes the entire DNA-binding domain, also results in the retention of GFP in the cytoplasm (Fig. 2J). These results suggest that the potential NLSs of Terra should be localized within the DNA-binding domain. In fact, concomitance of NLSs and DNA-binding domains has been seen in many types of transcription factors, including zinc finger factors [12], homeodomain factors [13], HMG domain factors [14], and basic helix–loop–helix factors [15]. This economical design should be conserved during evolution.

3.2. *Nuclear translocation of Terra is independent of DNA-binding structure*

To identity amino acid stretches controlling nuclear translocation of Terra, constructs pT2–C90, pT2–R78, pT2–L73 and pT2–H67 were generated, which removed different lengths of the DNA-binding domain from the C-terminus. Transfection experiment revealed that the T2–H67 and T2–L73 regions were unable to target the fusion proteins into the nuclei, while T2–R78 and T2–C90 could do so (Fig. 2F–I). This indicates that the K74–R78 domain, a motif KGHKR, is required for nuclear translocation of Terra. The T2–R78 region lacks four cystines that are zinc-binding residues required for forming the first and second zinc fingers as suggested by the structure of the DM domain of *Drosophila* DSX [16]. This strongly suggests that the nuclear targeting of Terra is independent of Zn^{2+} coordination and DNA recognition structure.

It is noted that the T2–R78 region allows a little more GFP to be retained in the cytoplasm compared to the T2–C90 region (Fig. 2F,G). This could be because the extra 12 amino acid residues C-terminal to the NLS can help to stabilize the

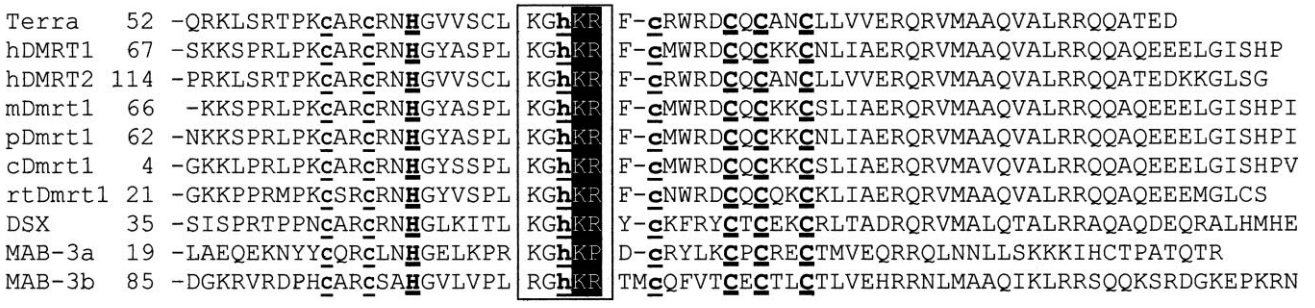

Fig. 3. Sequence alignment of DM domains of different DM proteins. Ligand-binding residues in the first zinc-binding site are shown in underlined bold lowercase and those in the second zinc-binding site are in underlined bold capital letters. The two strongly basic residues (K77 and R78 of Terra) required for nuclear translocation are shaded in black. These sequences were obtained from the GenBank database with accession numbers AF080622 for zebrafish Terra, AF130728 for human hDMRT1, AF130729 for human hDMRT2, AF202778 for murine mDmrt1, AF216651 for pig pDmrt1, AF123456 for chicken cDmrt1, AF209095 for rainbow trout rtDmrt1, M25292 for *Drosophila* DSX, and Z992788 for *C. elegans* MAB-3. MAB-3a and MAB-3b are the first and second DM domains of *C. elegans* MAB-3, respectively.

conformation of the NLS domain or the formation of the nuclear pore targeting complex.

The above results cannot exclude the possibility that Terra requires a bipartite NLS which has other components in the N-terminus of the DM domain. To test this possibility, we generated constructs pQ52–M173, pS56-M173 and pK74-G173 in which different N-terminal parts of Terra were deleted. After transfection with any of these constructs, the fusion protein was translocated into the nuclei (Fig. 2M–O). Along with the result from the C-terminal deletion analysis, this indicates that the ^{74}KGHKR^{78} is the only NLS.

We next asked whether KGHKR is sufficient for nuclear translocation. We generated another construct, pK74-R78, that encodes GFP coupled to KGHKR at its C-terminus. The cells transfected with this construct have little more GFP in the nuclei than those transfected with pEGFP-C2 (Fig. 2P,A). It is likely that the KGHKR domain, when located at the terminus of a protein, has an unstable conformation and thus can not serve as an effective NLS.

3.3. K77 and R78 are essential for nuclear targeting of Terra

Scrutiny of the ^{74}KGHKR^{78} motif identifies two consecutive strongly basic amino acids, i.e. K77 and R78. Since a common feature of NLSs is the presence of continuous basic amino acids, K77 and R78 of the Terra's NLS should play a key role in the nuclear translocation of Terra. To confirm their roles, K77 and R78 were mutated to glutamate (E), a strongly acidic residue, to generate constructs pT2–R78–K77E and pT2–R78–R78E. The cells transfected by either of these two constructs have cytoplasmic GFP as much as the cells transfected with pEGFP-C2 (Fig. 2K,L,A). This indicates that K77 and R78 are key components of the NLS of Terra.

3.4. The Terra's NLS is conserved among DM family members

Terra is a member of the expanding DM protein family that is characterized by the presence of the DM domain, a DNA-binding domain. The DM domain contains a novel zinc module consisting of intertwined CCHC and HCCC Zn^{2+}-binding sites [16]. Although Terra is the only member not known to be involved in sex determination, the NLS of the form KGHKR of Terra is also present in the other sex-related DM family members across species from invertebrates to vertebrates (Fig. 3), suggesting its conservative nature during evolutionary process. Residue H76 of the NLS, which corresponds to H59 in *Drosophila* DSX, is a ligand-binding residue for the first Zn^{2+}-binding site [12]. Residues K59 and R60 of DSX, corresponding to K77 and R78 of Terra, are in the chain connecting the two ligand-binding sites. In *C. elegans* Mab-3 contains two DM domains [2]. The first DM domain of Mab-3 is significantly divergent from its own second DM domain and the DM domains in other species in the C-terminal basic tail that is required for DNA recognition [16]. We note that the potential NLSs within the first and the second DM domains of Mab-3 are ^{41}KGHKP^{45} and ^{107}RGHKR^{111}, respectively, both of which have one residue different from the NLS of Terra. Because R107 is also a strongly basic residue, this substitution for the first lysine residue of the KGHKR motif should not affect the nature and function of the motif. On the other hand, P45 is a non-polar residue and thus the ^{41}KGHKP^{45} motif may not be able to act as an effective NLS. It is most likely that *C. elegans* Mab-3 uses ^{107}RGHKR^{111} within the second DM domain as a NLS for nuclear translocation. It has been found that several mutations in the DM domains of *Drosophila* DSX and *C. elegans* Mab-3 cause intersex development [2,17]. It remains unknown, but interesting, whether the mutations within the potential NLS in these species can also affect development of sex-specific phenotypes.

Acknowledgements: We thank other members of the Meng laboratory for discussions. This work was supported by the National Natural Science Foundation of China (#30025020 and #39970360), TRAPOYT of MOE, National Key Basic Research and Development Program (#G19990756), and the '985' Program of Tsinghua University.

References

[1] Burtis, K.C. and Baker, B.S. (1989) Cell 56, 997–1010.
[2] Raymond, C.S., Shamu, C.E., Shen, M.M., Seifert, K.J., Hirsch, B., Hodgkin, J. and Zarkower, D. (1998) Nature 391, 691–695.
[3] Raymond, C.S., Murphy, M.W., O'Sullivan, M.G., Bardwell, V.J. and Zarkower, D. (2000) Gene. Dev. 14, 2587–2595.
[4] Raymond, C.S., Kettlewell, J.R., Hirsch, B., Bardwell, V.J. and Zarkower, D. (1999) Dev. Biol. 215, 208–220.
[5] Ottolenghi, C., Veitia, R., Quintana-Murci, L., Torchard, D., Scapoli, L., Souleyreau-Therville, N., Beckmann, J., Fellous, M. and McElreavey, K. (2000) Genomics 64, 170–178.
[6] Stumm, M., Wieacker, P., Kessel-Weiner, E., Theile, U., Pascu, F. and Ottolenghi, C. (2000) Pediatr. Pathol. Mol. Med. 19, 415–423.
[7] Marchand, O., Govoroun, M., D'Cotta, H., McMeel, O., Lareyre, J.J., Bernot, A., Laudet, V. and Guiguen, Y. (2000) Biochim. Biophys. Acta Gene Struct. Expr. 1493, 180–187.

[8] Meng, A., Morre, B., Tang, H., Yuan, B. and Lin, S. (1999) Development 126, 1259–1268.
[9] Kaffman, A. and O'Shea, E.K. (1999) Annu. Rev. Cell Dev. Biol. 15, 291–339.
[10] Kalderon, D., Roberts, B.L., Richardson, W.D. and Smith, A.E. (1984) Cell 39, 499–509.
[11] Dingwall, C. and Laskey, R.A. (1991) Trends Biochem. Sci. 16, 478–481.
[12] Matheny, C., Day, M.L. and Milbrandt, J. (1994) J. Biol. Chem. 269, 8176–8181.
[13] Ghaffari, M., Zeng, X., Whitsett, J.A. and Yan, C. (1997) Biochem. J. 328, 757–761.
[14] Sudbeck, P. and Scherer, G. (1997) J. Biol. Chem. 272, 27848–27852.
[15] Vandromme, M., Cavadore, J.C., Bonnieu, A., Froeschle, A., Lamb, N. and Fernandez, A. (1995) Proc. Natl. Acad. Sci. USA 92, 4646–4650.
[16] Zhu, L.Y., Wilken, J., Phillips, N.B., Narendra, U., Chan, G., Stratton, S.M., Kent, S.B. and Weiss, M.A. (2000) Gene Dev. 14, 1750–1764.
[17] Erdman, S.E. and Burtis, K.C. (1993) EMBO J. 12, 527–535.