Journal of Sciences, Islamic Republic of Iran 18(1): 13-18 (2007) http://jsciences.ut.ac.ir
University of Tehran, ISSN 1016-1104

Expression and Purification of the Human Homeodomain TGIFLX

R. Ghassemi, F. Mansouri, and M. Heidari*

Department of Medical Genetics, Tehran University of Medical Sciences, Tehran, Islamic Republic of Iran

Abstract

The homeobox genes are known to play a crucial role in controlling the development of multicellular organisms. The majority of these genes have been determined to express regulatory proteins act as a regulatory protein. These trans-acting factors regulate the expression of proteins that are necessary during the developmental processes throughout the body. *TGIFLX/Y* is a homeobox gene and it contains two genes: *TGIFLX* (X-linked) and *TGIFLY* (Y-linked) are specifically expressed in human adult testes. *TGIFLX* has originated from retrotransposition of *TGIF2*, located on long arm of chromosome 20 (q11.2-12), onto the X chromosome. To date, the role of *TGIFLX/Y* are unknown, therefore, in order to get insight into the potential roles of *TGIFLX*, we cloned and produced the GST-TGIFLX fusion protein. The purified protein was recognized by anti-GST antibodies. Through a single purification procedure using MagneGST beads, approximately 8 mg of the recombinant protein was obtained per liter from bacterial culture. The production of this recombinant protein will now permit the investigation of *TGIFLX* target genes as well as identification of co-factors or partner proteins involved in *TGIFLX* function in normal and abnormal development.

Keywords: Homeobox gene; Recombinant TGIFLX; Room temperature; SDS-PAGE

Introduction

Homeobox-containing genes consist of a family of regulatory genes encoding transcription factors that are implicated in the control of developmental processes. The homeobox genes encode a 60 amino acid helix-turn-helix type of DNA-binding motif, termed the homeodomain (HD) [1]. Homeobox genes were initially identified as genes controlling *Drosophila* development [2] and subsequently were isolated from other organisms such as nematodes, plants and vertebrates [3,4]. In *Drosophila* and other invertebrates, these homeotic genes are organised as a complex (*HOM-C*), which shares significant homology with the four mammalian *HOX* gene clusters (*HOXA*-*HOXD*) [5]. In humans, thirty-nine homeobox genes belonging to these clusters have thus far been identified [6]. Many studies have revealed that changes in the expression patterns, target genes and functions of homeodomain proteins are also thought to be a driving force in animal evolution [7,8]. It is estimated that homeobox genes account for more than 0.1% of the vertebrate genome [9].

*E-mail: mheidari@sina.tums.ac.ir

Homeodomain proteins are categorised into distinctive groups. One group of homeodomain proteins, called the TALE group, has a 3-amino-acid insertion between helices 1 and 2 of the homeodomain [10,11], resulting in a 63, instead of the typical 60, amino acid homeodomain. TALE proteins have been discovered in numerous species and involved in regulation of gene expression [12-14].

*TGIF*s belong to the TALE superclass of homeobox-containing genes known to play critical roles in developmental process such as cell proliferation, differentiation and cell fate. *TGIFLX/Y* as a homeobox gene consists of two genes; *TGIFLX* (X-linked) and *TGIFLY* (Y-linked) are specifically expressed in human adult testes [15]. *TGIFLX* originated from retrotransposition of *TGIF2*, located on long arm of chromosome 20 (q11.2-12), onto the X chromosome. The *TGIFLX* gene has 2808 bps containing two exons and one intron, the whole exon 1 and 21 bps of exon 2 create a 5'UTR and *TGIFLX* coding sequences are located within exon 2 [15]. The *TGIFLX* contains a 726 bp open reading frame that encodes a 26.675 KDa protein belonging to a distinct subclass of HOX proteins. TGIFLX is suggested to be a transcriptional regulatory factor that shares a helix-turn-helix motif with other members of the homeodomain proteins. The *Hox* genes encode a network of evolutionarily conserved transcription factors that are involved in the specification of segmental identity along the anteroposterior body axis of animals as diverse as insects and vertebrates. The expression of *Hox* genes in the mammal starts at the stage of gastrulation [16]. Recently using analysis of 64 orthologous homeobox genes from humans and mice indicated higher rates of amino acid substitution in *TGIFLX* genes than in other genes. These findings suggest a general pattern of rapid evolution of mammalian *TGIFLX* homeobox genes [17].

It is quite clear that the majority of homeodomain proteins function as transcription factors which bind to DNA containing AT-rich sequences [1]. In order to study the role of this transcription factor we have recently studied TGIFLX/Y gene expression in 101 infertile male by the nested RT-PCR technique. The findings of this study showed a significant correlation between *TGIFLX/Y* and male infertility (Heidari *et al.*, unpublished), however, the precise mechanism in which TGIFLX acts in normal and abnormal developments is not clear.

Studies have shown that cellular target genes of transcription factors are best understood in the context of interacting partner proteins which form transcriptional complexes assembled on gene promoters and/or enhancers. As studies with other transcription factors have showed, approaches are available which utilise recombinant protein to functionally dissect their role by identifying/confirming target genes and/or interacting partner proteins [18-21]. Such studies require significant amounts of protein of high purity. The major aim of this study was to optimise the production of intact, soluble GST-TGIFLX by systematically varying various expression and purification parameters. We therefore cloned human TGIFLX coding region into a pGEX-6p-1 expression vector. A high level of purified GST-fusion protein was obtained. The production of this recombinant protein will now allow the investigation of DNA target sequence of TGIFLX as well as identification of co-factors involved in TGIFLX function in normal and abnormal development.

Materials and Methods

TGIFLX Expression

For construction of pGEX-6P-1/TGIFLX, the 726 bp human TGIFLX DNA encoding full-length TGIFLX was amplified by PCR using *Pfu* DNA polymerase (Promega, Madison, WI, USA) and the primers 5′-CGGGATCCATGGAGGCCGCTGCGGA-3′ and 5′-GCGAATTCATCATGGATTAGGCTCTTGC-3′. The underlined sequences represent the introduced *Bam* HI and *Eco* RI restriction sites to facilitate cloning.

Amplification was performed on a Corbet thermal cycler (Corbet, model: CG1-96, AUS) 3 min at 97°C for initial denaturation step, 30 sec at 97°C, 45 sec at 64°C with a 1°C decrease every following cycle down to 55°C then 55°C for the 14 cycles, 1 min at 72°C for extension, and finally 10 min at 72°C. After digestion with *Bam* HI and *Eco* RI (Promega, Madison, WI, USA), the purified PCR product was unidirectionally inserted into the *Bam* HI and *Eco* RI sites of the pGEX-6P-1 vector (Amersham Pharmacia, USA). The correct DNA sequence and reading frame between the GST and TGIFLX DNA was confirmed by automated DNA sequencing using an ABI Prism dye terminator cycle sequencing kit (Gene Fanavaran, IRI).

Expression and Purification of Recombinant TGIFLX

Expression of recombinant *TGIFLX* was carried out as previously described [22]. Briefly, *E. coli* strain ER2566 transformed with recombinant expression vectors pGEX-6P-1/TGIFLX was cultured overnight in 2x YT medium containing 100 mg/ml ampicillin. The cultures were then diluted 1:100 with fresh pre-warmed 2x YT and incubated at 37°C with vigorous shaking.

When the OD_{600} had reached a value of 0.5 (~2 h), fusion gene expression was induced by adding IPTG to a final concentration of 0.1 mM. Uninduced cultures were also included to compare bacterial protein expression. The culture was then incubated at 37°C for an additional 1.5 h and the cells harvested by centrifugation (14,000 rpm, 5 sec, 4°C). The pellet was resuspended in 50 µl of phosphate-buffered saline (PBS) per ml of bacterial culture. The total bacterial proteins were analysed by SDS-PAGE as described by Sambrook [23]. The purification of the GST-TGIFLX fusion protein was carried out, using MageneGene kit (Promega, USA) based on the instructions. Briefly, 1 ml of overnight cultures of bacteria containing the parental pGEX-6P-1 (Amersham Pharmacia, USA) and recombinant pGEX-6P-1/TGIFLX vectors was diluted 1:100 in fresh and pre-warmed 2x YT and further incubation was performed at 37°C for ~2 h with shaking. When the cell densities were reached to an OD_{600} = 0.5 induction carried out using IPTG to a final concentration of 0.1 mM. Cultures were then incubated a further 1.5 h at 37°C with vigorous shaking (250 rpm). One milliliter of culture was transferred into a labelled Eppendorf microtube, and the cells were centrifuged for 5 sec at 14,000 rpm. The pellet was resuspended in 200 µl lysis buffer (supplied with the kit) and lysed by freezing and thawing repeated four times and subsequently incubated at room temperature for 20-30 min on the shaker. The cell lysates were then added to 100 µl of equilibrated Magnet GST particles (supplied with the kit) and incubated with gentle mixing on a shaker for 30 min at RT. Subsequently, the Magnet GST particles were placed in a Magnet stand and washed three times with 250 µl of MagneGST cold binding/wash buffer (supplied with the kit) and the recombinant protein eluted by gentle mixing (15 min, RT) using 200 µl of glutathione elution buffer (50 mM Tris-HCl, pH 8.1, 50 mM reduced glutathione; supplied with the kit). The yield of the TGIFLX fusion proteins was estimated by measuring their absorbance at 280 nm.

The total protein concentration was estimated based on the following formula:

Protein Concentration (mg/ml) = $(1.55.\ A_{280}) - 0.76.\ A_{260}$ (lab FAQS, Roche Applied Science, Germany).

Western Blotting

Western blot analysis was performed as previously explained [22]. Briefly, 10 µl bacterial supernatant or purified proteins were electrophoresed through 12.5% SDS-PAGE gels and transferred to a nitrocellulose membrane (Amersham Pharmacia, USA). The membranes were blocked for 1 h with 3% Bovine Serum Albumin (BSA) at 37°C and then incubated with a 1/1000 dilution of a polyclonal rabbit anti-GST antiserum raised against the C-terminus (sc-459, Santa Cruz Biotechnology, Santa Cruz, CA, USA) (1 h, at room temperature). The membranes were incubated with 1:10,000 dilution of a goat HRP-conjugated anti-rabbit IgG secondary antibody (1 h) prior to development with 4-chloro-1-naphthol (Immun-Blot, Bio-Rad Laboratories, USA).

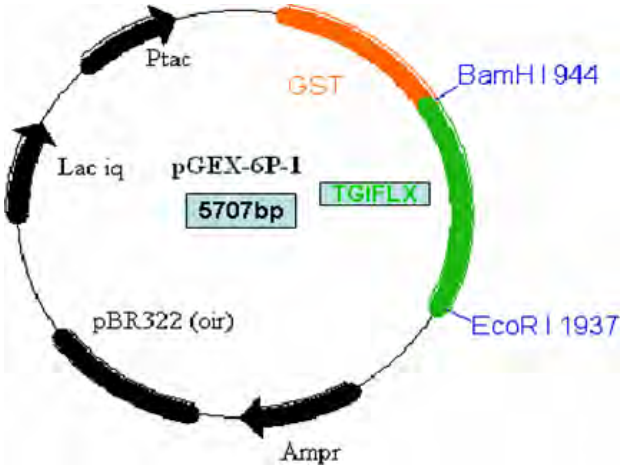
Results and Discussion

In this study, PCR products of TGIFLX was successfully cloned as N-terminal fusion into the EcoR I and BamH I sites of the pGEX-6P-1 expression vector (Fig. 1). The human DNA-binding protein TGIFLX was heterologously expressed in *E. coli*. For production of this rTGIFLX protein, overnight cultures of lon/OmpT protease-deficient ER2566 cells transformed with pGEX-6P-1/TGIFLX, were used to inoculate fresh media the next morning. Fusion protein production was carried out under conditions whereby degradation of the interested proteins was minimized. This entailed growing the bacterial broths to an OD_{600} of 0.5 before adding 0.1 mM IPTG, and culturing the cells for an additional 2 h only. Incubation for longer periods did not result in a higher yield of protein expression (data not shown). To establish the GST-TGIFLX fusion protein as a soluble form in *E. coli*, expressing cells were lysed and separated into supernatant and pellet fractions. Figure 2 shows supernatant fractions analysed by SDS-PAGE after IPTG induction. Our observations revealed that the recombinant proteins were present mainly in the supernatant (data not shown). Thus, GST-TGIFLX protein which has apparent molecular weights of 52.6 KDa was expressed as soluble proteins in *E. coli* (Fig. 2).

The expression of TGIFLX protein as GST fusion allowed its convenient purification from other soluble bacterial proteins using MagneGene GST-purification kit (Promega, USA). Supernatants of samples were purified using MagneGST particles at room temperature and the eluted recombinant TGIFLX protein was checked on a SDS-PAGE gel stained with Coomassie blue (Fig. 2). Purified rTGIFLX produced multiple bands when purification was performed in high cell densities (data not shown). The recombinant protein concentration was assayed as explained (materials and methods).

This result indicated that induction at high cell densities of bacterial cell containing pGEX-6p-1/TGIFLX lead to production of extremely unstable recombinant protein. This result was consistent with the

A)



B)

pGEX-6P-1-TGIFLX

PreScissionTM Protease ↓ *TGIFLX*

Leu Glu Val Leu Phe Gln Gly Pro Leu Gly Ser Met Glu Ala Ala
CTG GAA GTT CTG TTC CAG GGG CCC CTG GGA TCC ATG GAG GCC GCT

*Bam*H I

Figure 1. Schematic representation of pGEX-6P-1/TGIFLX expression construct which encodes the GST-TGIFLX fusion protein. A) Features of the expression vector and restriction sites used are shown; B) Sequence of pGEX-6P-1/5′-TGIFLX junction showing the reading frame.

nature of some homeodomain fusion proteins [24]. It is because some studies have indicted that the stability of the recombinant HOX proteins differs from one HOX protein to another [22]. In spite of the fact that we previously reported that recombinant homeodomain GST-HOX11 protein was extremely stable in induction of high cell densities [22], GST-TGIFLX is unstable under these conditions. Recombinant protein purification was done at low cell densities which resulted in a single undegraded band of recombinant protein, although overall yield was reduced (Fig. 2). The intact GST-TGIFLX could be produced with high level of purity, at a final yield of approximately 8 mg/L of bacterial culture. When the GST-TGIFLX was tested with an anti-GST antibody the expected bands of approximately 26 and 52.6 KDa gave a strongly positive result (Fig. 3). This confirmed the identity of the soluble protein as being the expected TGIFLX fusion protein.

High-level production of recombinant protein as a prerequisite for subsequent purification has become a standard method. Significant applications of recombinant proteins are: (1) immunization, (2) protein-protein interaction, (3) three-dimensional analysis of the protein, and (4) protein-DNA interactions and therapeutic use. Therefore, the successful production of recombinant TGIFLX (rTGIFLX) protein will now allow a range of investigations including the WG-PCR technique to study target DNA sequences that directly bind to TGIFLX in order to dissect its role in gene regulation. Recombinant protein can also be utilised in assays to help identify interacting protein partners. Such studies should ultimately help to define the role of this transcription factor in both normal and probably abnormal developments.

Even though during this study efforts were made to obtain an intact GST-TGIFLX protein using bacterial expression system with high yield and minimized autodegradation, the biological function of this recombinant protein remains to be determined. For analysing the biological function of recombinant protein which produced in bacterial system various studies such as Electrophoretic Mobility Shift Assay (EMSA) and Whole Genome PCR has been employed (Harris, 1999, Heidari, 2000, Heidari, 2006).

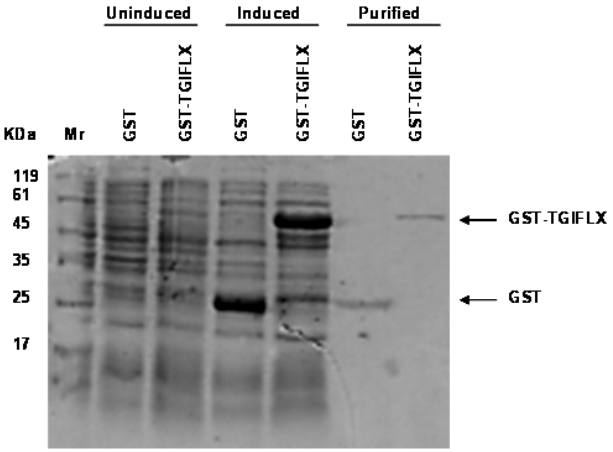


Figure 2. Expression and purification of soluble recombinant GST-TGIFLX and GST proteins under optimized conditions. *E. coli* strain ER2566 was transformed with either pGEX-6P-1 or pGEX-6P-1/TGIFLX. Purification of GST-TGIFLX fusion protein or GST alone were purified at room temperature (RT) by Megne Gene kit (materials and methods). After purification at RT, rTGIFLX and GST were separated onto a 12.5% SDS-PAGE gel and stained with Coomassie brilliant blue R-250. Lane 1, high-range protein molecular weight marker; Lanes 2 and 3 show uninduced *E. coli* culture lysate expressing GST-TGIFLX protein. Lanes 4 and 5 indicate the induced cell lysate. Lanes 6 and 7 shows purified GST and GST-TGIFLX protein. Arrows indicate the position of GST-TGIFLX and GST proteins, 52.6 KD and 26 KD, respectively.

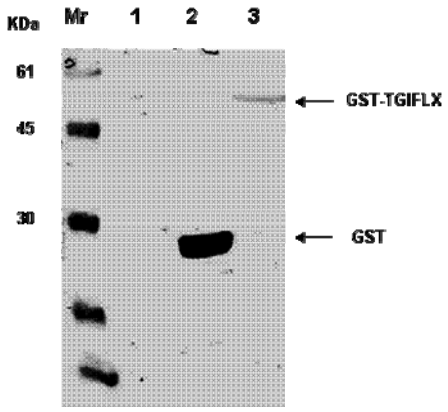


Figure 3. Western blot analysis of uninduced GST and induced GST or recombinant GST-TGIFLX proteins. Supernatant (10 µl) representing 1 ml uninduced or induced bacterial cell cultures containing parental vector of pGEX-6p-1 or pGEX-6p-1/TGIFLX were separated by 12.5% SDS-PAGE and electrophoretically transferred to nitrocellulose membrane. The membrane was probed with an antibody directed against GST, which revealed the presence of GST-TGIFLX and GST proteins (lanes 2 and 3) and uninduced GST-TGIFLX (lane 1). Mr shows Prestained protein molecular weight marker. Arrows show the position of recombinant GST-TGIFLX and GST proteins, 52.6 KD and 26 KD, respectively.

Whole Genome PCR (WG-PCR) is a useful method for investigation of transcription factor DNA target sequences [19, 21]. We have recently employed a whole genome PCR-based screening method using recombinant HOX11 expressed as a GST fusion protein. GST-HOX11 immobilised with glutathione-coated Sepharose was incubated with adaptor-ligated human genomic DNA cleaved with *Sau*3AI. Bound DNA was eluted, amplified by PCR, cloned into pCR 2.1 and sequenced. WG-PCR using GST-HOX11 revealed that HOX11 physically interacts with Satellite 2 DNA. GST-pulldown assay is another valuable method which is currently used in protein-protein interaction [21].

In summary, we have described a method to produce large amounts of an intact recombinant TGIFLX protein. The features that were particularly critical to the success of our procedure were: 1) induction of TGIFLX expression at low cell densities and 2) purification of GST-TGIFLX fusion proteins at RT. Using this protocol, soluble TGIFLX protein can be produced at a high level of purity and consequently used in protein-protein interaction or protein-DNA interactions which may provide significant understanding of the TGIFLX partner proteins and target sequences involved in gene regulation by TGIFLX. This in turn will help to define the role of this transcription factor in both normal and abnormal developments.

Acknowledgements

This work was supported by a grant from the Tehran University of Medical Sciences (TUMS) grant no 85-01-30-3322.

References

1. Gehring *et al.* Homeodomain-DNA recognition. *Cell*, **78**(2): 211-23. Review, No abstract available (1994).
2. Castelli G. Implications of the spatial and temporal regulation of Hox genes on development and evolution. *Int. J. Dev. Biol.*, **42**(3): 437-44 (1998).
3. Manak J.R., Mathies L.D., and Scott M.P. Regulation of a decapentaplegic midgut enhancer by homeotic proteins. *Development*, **120**(12): 3605-19 (1994).
4. Boyl *et al.* Otx genes in the development and evolution of the vertebrate brain. *Int. J. Dev. Neurosci.*, **19**(4): 353-63 (2001).
5. Magli M.C. *et al.* Effects of HOX homeobox genes in blood cell differentiation. *J. Cell Physiol.*, **173**(2): 168-77 (1997).
6. Krumlauf R. and Gould A. Homeobox cooperativity. *Trends Genet.*, **8**(9): 297-300 (1992).
7. Mann R.S. and Carroll S.B. Molecular mechanisms of selector gene function and evolution. *Curr. Opin. Genet. Dev.*, **12**(5): 592-600 (2002).

8. Vervoort M. Functional evolution of Hox proteins in arthropods. *Bioessays*, **24**(9): 775-9 (2002).
9. Stein *et al.* Differential activation of the clustered homeobox genes CNOT2 and CNOT1 during notogenesis in the chick. *Dev. Biol.*, **180**(2): 519-33 (1996).
10. Bertolino E., Wildt S., Richards G., and Clerc R.G. Expression of a novel murine homeobox gene in the developing cerebellar external granular layer during its proliferation. *Dev. Dyn.*, **205**: 410-420 (1996).
11. Burglin T.R. Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Res.*, **25**: 4173-4180 (1997).
12. Berthelsen J., Zappavigna V., Ferretti E., Mavilio F., and Blasi F. The novel homeoprotein Prep1 modulates Pbx-Hox protein cooperativity. *EMBO J.*, **17**: 1434-1445 (1998).
13. Jacobs Y., Schnabel C.A., and Clehindbraary M.L. Trimeric association of Hox and TALE homeodomain proteins mediates *Hoxb2* in enhancer activity. *Mol. Cell. Biol.*, **19**: 5134-5142 (1999).
14. Kamps M.P., Murre C., Sun X.H., and Baltimore D. A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL. *Cell*, **60**: 547-555 (1990).
15. Blanco-Arias P., Sargent C.A., and Affara N.A. The human-specific Yp11.2/Xq21.3 homology block encodes a potentially functional testis-specific TGIF-like retroposon. *Mamm. Genome*, **13**: 463-468 (2002).
16. Graham *et al.* The murine and Drosophila homeobox gene complexes have common features of organization and expression. Cell, **57**(3): 367-78 (1989).
17. Wang X. and Zhang J. Rapid evolution of mammalian X-linked testis-expressed homeobox genes. Genetics, **167**(2): 879-88 (2004).
18. Robinson L., Panayiotakis A., Papas T.S., Kola I., and Seth A. ETS target genes: identification of egr1 as a target by RNA differential display and whole genome PCR Techniques. *Proc. Natl. Acad. Sci. USA*, **94**: 7170-7175 (1997).
19. Harris *et al.* Functional analysis of the homeodomain protein SIX5. *Nucleic Acids Res.*, **28**(9): 1871-8 (2000).
20. Linderson Y., French N.S., Neurath M.F., and Pettersson S. Context-dependent Pax-5 repression of a PU.1/NF-kappaB regulated reporter gene in B lineage cells. *Gene*, **262**: 107-114 (2001).
21. Heidari M., Rice K.L., Phillips J.K., Kees U.R., and Greene W.K. The nuclear oncoprotein TLX1/HOX11 associates with pericentromeric satellite 2 DNA in leukemic T-cells. *Leukemia*, **20**(2): 304-12 (2006).
22. Heidari M., Rice K.L., Kees U.R., and Greene W.K. Expression and purification of the human homeodomain oncoprotein HOX11. *Protein Expr. Purif.*, **25**(2): 313-8 (2002).
23. Sambrook J., Fritsch E.F., and Maniatis T. *Molecular Cloning: A Laboratory Manual*. 2nd Ed., Cold Spring Harbour Laboratory Press, New York (2001).
24. Tang *et al.* The optimal binding sequence of the Hox11 protein contains a predicted recognition core motif. *Nucleic Acid Res.*, **23**: 1928-1935 (1995).