Biochemistry (Moscow), Vol. 70, No. 8, 2005, pp. 920-925. Translated from Biokhimiya, Vol. 70, No. 8, 2005, pp. 1112-1118.

Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM04-244, February 6, 2005.

Rapid Identification of Quox-1 Homeodomain DNA-Binding Sequence Using SAAB

Fan Zhu[1#], Zhong-Chun Liu[2#], Wen-Xin Li[1]*, and Xiu-Ling Xu[1]

[1]College of Life Sciences, Wuhan University, Wuhan 430072, People's Republic of China; fax: 86-027-87669560; E-mail: fanzhu@my3ia.sina.com
[2]Ren-Ming Hospital, Wuhan University, Wuhan 430060, People's Republic of China

Received August 30, 2004
Revision received September 30, 2004

Abstract—*Quox-1* is the only gene in the *hox* family whose expression occurs throughout the development of the central nervous system. Using the Quox-1 homeodomain produced in a bacterial expression system, we were able to identify DNA-binding targets of the Quox-1 protein from a library of randomly generated oligonucleotides by the selection and amplification binding (SAAB) technique. The results indicated that the Quox-1 protein recognizes a new consensus sequence, 5′-CAATC-3′, which has not been reported for any other Hox family homeoprotein. In addition, electromobility shift assay further confirmed that the Quox-1 homeoprotein preferentially binds to the 5′-CAATC-3′ sequence, but not to the binding sites for other Hox class homeoprotein (TAAT) or NKX class homeoprotein (CAAG). Based on mutation analyses of the DNA sequences, we found that the 5′-CAATC-3′ core sequences are required for high affinity binding by the Quox-1 protein. Furthermore, mutation analyses of the Quox-1 homeodomain showed that one of the major determinants participating in recognition of a minor groove is the Gln6 and Thr7 in the N-terminal arm of the homeodomain.

Key words: Quox-1, DNA–protein interaction, DNA-binding site, homeodomain protein

Homeobox genes are involved in the commitment of cells to specific developmental pathways and play an important role in pattern formation. They encode proteins that contain highly conserved DNA-binding regions called homeodomains [1]. Mutation studies and expression patterns of several members of the homeodomain gene family have indicated a role in controlling specification of cranial structures, including development of neurons and sensory organs. Homeodomain proteins can bind to specific DNA sequences and then function as transcription factors that regulate eukaryotic development with spatial and temporal specificity [2].

Quox-1 is a novel homeobox gene isolated from cDNA library of five-week quail embryo by Xue in 1990 [3]. It is the only gene in the *hox* family that has been found to express in both prosencephalon and mesencephalon involved in the differentiation of the central and peripheral nerve cells [4]. The over-expression of Quox-1 was associated with the development of tumors [5-8]. Sequence analysis of homeodomain proteins binding can reveal the carcinogenesis.

The optimal DNA binding site for Quox-1 or its mammalian homologs has not yet been defined. In this study, we aimed to identify potential DNA-binding sequences for the Quox-1 protein. We applied an approach that has been used successfully to identify optimal DNA binding sites for a variety of transcription factors [9]. This strategy relies on the ability of a protein factor to interact with specific DNA sequences selected from a population of randomly generated oligonucleotides. Electromobility gel shift assays (EMSAs) along with mutation studies were used to analyze binding affinity and the consensus sequences that are important for Quox-1 binding.

MATERIALS AND METHODS

Expression and purification of Quox-1 fusion proteins. *Quox-1* homeobox sequence was obtained by PCR amplification from a human embryo cDNA library [10]. The amplified DNA fragment was digested with *Eco*RV and *Xho*I and directly cloned into the *Sma*I and *Xho*I restriction site of the expression vector pGEMEX-xBalI and verified by DNA sequencing. *Quox-1* mutations were constructed by PCR [11] and cloned as well as verified as described above.

The recombinant plasmids were transformed into competent *Escherichia coli* strain BL21 (DE3) cells using

[#] These authors contributed equally to the work.
* To whom correspondence should be addressed.

0006-2979/05/7008-0920

standard procedure [12]. Freshly transformed cells were used to inoculate 10 ml of Luria–Bertani (LB) medium and the cultures were incubated overnight with shaking at 200 rpm at 30°C. Overnight culture was then grown in 100 ml of LB medium containing 0.2% glucose and 200 μg of ampicillin at 37°C with shaking at 300 rpm until the optical density at 600 nm reached 0.6. Isopropyl-β-D-thiogalactopyranoside was added to the culture at a final concentration of 0.3 mM and the culture was incubated for an additional 4 h at 37°C. Cells were harvested by centrifugation and frozen at –80°C. Approximately 1 g of frozen cells was thawed on ice and resuspended in 10 ml of buffer A (20 mM Tris-HCl, 200 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol (DTT)). Suspended cells were lysed by sonication and a cell-free extract was prepared by centrifugation. The cell-free extract was loaded onto an amylose column (New England Biolabs, USA) and washed with buffer A. Quox-1 fusion proteins were eluted from the column with buffer A containing 10 mM maltose. Protein purity was demonstrated by 15% SDS-PAGE and purified protein concentration was determined by the Bradford assay [13].

Selection and amplification binding assay. A selection and amplification-binding (SAAB) assay was performed essentially as described previously [9, 14, 15] with modifications. Briefly, 0.5 pmol of a 15 bp random sequence flanked by 20 bp regions of non-random sequence was radiolabeled with 10 μCi of [γ-^{32}P]CTP using PCR amplification. The radiolabeled probe was incubated with 25 pmol of purified Quox-1 homeodomain fusion protein in binding buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 7.5 mM $MgCl_2$, 1 mM EDTA, 5% glycerol, 5% sucrose, 0.1% Nonidet P-40, 0.1 μg salmon sperm DNA, and 5 mM DTT). Unbound probe was separated from protein-bound DNA by native 8% PAGE in 0.5× TBE buffer (89 mM Tris, 89 mM borate, 2 mM EDTA). Protein-bound DNA was detected by autoradiography, and the bands representing protein–DNA complexes were excised from the gel. The DNA was eluted from the gel slice overnight at 37°C in elution buffer (0.5 M ammonium acetate, 1 mM EDTA, and 0.1% SDS). Eluted DNA was amplified by PCR using primers complementary to the 20 bp non-random flanking sequences and purified by EZNA™ DNA/RNA Isolation Systems (Omega Bio-tek, Inc, USA). The purified DNA was radiolabeled and used as a probe for another round of SAAB with a total of six rounds. The amount of protein was reduced to 5 pmol for the fifth and sixth rounds of SAAB. Following the final PCR amplification and purification, the 55 bp fragment was cloned into pMD 18-T vector using a pMD 18-T cloning kit (Takara Biotechnology Corporation, Japan). The inserts were sequenced on an ABI377 Nucleotide sequencer (Perkin Elmer, USA).

Electromobility gel shift assays. As described, double-stranded DNA containing a Bicoid site (5′-TAATCC-3′), with flanking partial *Bam*HI ends, was annealed and filled with Klenow polymerase to generate a ^{32}P-labeled probe for EMSAs [14, 15]. For standard binding assays, the oligonucleotide (10 nM) was incubated in a 20 μl reaction containing binding buffer, 120 ng purified Quox-1 homeodomain fusion protein on ice for 15 min. Competition binding experiments were performed essentially as described by Amendt et al. [15] by including unlabeled competitor DNA concentrations of 62.5, 125, 250, and 500 nM. Unlabeled double-stranded competitor DNAs were preincubated with the protein for 15 min on ice prior to addition of the probe. The samples were electrophoresed for 2 h at 250 V in an 8% polyacrylamide gel in 0.25× TBE buffer at 4°C following pre-electrophoresis of gels for 1 h at 200 V. The dried gels were visualized by exposure to autoradiographic film. For quantitative analyses to establish binding constants and relative competitions, the amount of bound and free radioactive probe was measured from dried gels using an InstantImager (Packard, USA). For determination of the amount of binding competitor, the ratio of bound to free probe was normalized to the absence of competitor DNA, which was set at 100%.

RESULTS

Identification of DNA-binding site of the Quox-1 protein. Previous studies with other homeodomain proteins showed that the information necessary for DNA binding is located within the homeodomain region [16, 17]. We used the SAAB assay originally described by Blackwell and Weintraub [9] with modifications by Chen and Schwarz [14] to determine consensus DNA binding sequences for Quox-1 homeodomain fusion protein, which was expressed in bacteria. Purified proteins were used to identify preferred binding sequences from a random pool of double-stranded DNA (Fig. 1a). Six cycles of binding selection were performed to isolate optimal binding sequences. Enriched double-stranded DNA was cloned and sequenced. Alignment of the 18 selected Quox-1 binding fragments revealed an identical consensus sequence (CAATCT) with an individual nucleotide frequency of occurrence ranging from 83.3 to 100% (Fig. 1b). In particular, the 5′-CAATC-3′ core was predominant, which is present in all of the selected oligonucleotides. These results indicate that the 5′-CAATC-3′ is a consensus sequence to which the Quox-1 protein preferentially binds.

Comparison of the binding activities of Quox-1 to the CAATC and the CAATT (G/A) sequences. The results of the SAAB assay indicated that Quox-1 preferentially binds to a CAATC sequence. Using a competitive binding assay, we compared the binding abilities of Quox-1 to the CAATC site with that to the CAATT (G/A), a motif-binding site for other Hox proteins. Mutations of the binding sequences were generated via PCR by replacing C

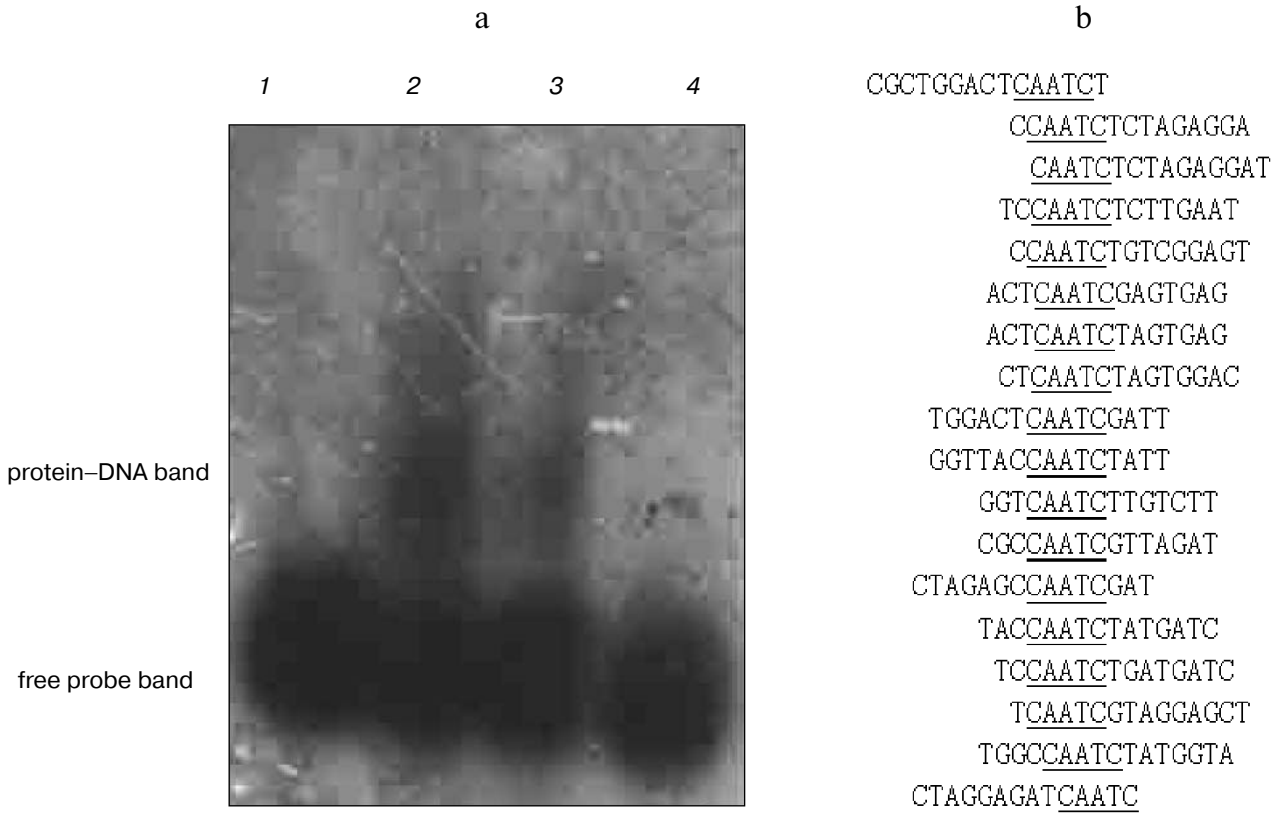

Fig. 1. Identification of the Quox-1 DNA-binding site using SAAB. a) The first round of SAAB: *1*, *4*) free probe; *2*, *3*) protein–DNA binding assay; b) the result of sequences.

with T (G/A) at position 5 (CAATC→CAATT (G/A)). The CAATT (G/A) motif-binding site is similar to the Quox-1 consensus sequence (CAATC) with only one nucleotide differing at position 5, which was proposed to be important for optimal binding [14, 18]. Our results, showed in Fig. 2a, suggest that Quox-1 cannot bind to the CAATT (G/A) motif-binding site. It also implied that the fifth nucleotide is essential for optimal binding of the Quox-1 protein.

Quox-1 DNA binding activity was measured by EMSA using the CAATC sequence as a probe in the presence or absence of competitor DNA fragments at 1-, 2-, 4-, and 8-fold molar excess (Fig. 2b).

Effect of mutations in the CAAT consensus sequence on Quox-1 binding. Quox-1 belongs to the Hox class of homeodomain proteins, but preferentially binds to the novel CAAT site rather than the TAAT core sequence recognized by most homeodomain proteins. An EMSA experiment was used to identify the nucleotide(s) within the CAAT site, which is important for the preferential binding of Quox-1. Sequences were generated to replace either C with T at position 1 (CAAT→TAAT) or T with G at position 4 (CAAT→CAAG) of the Quox-1 consensus binding site. The binding affinities of Quox-1 to the TAAT and CAAG were measured by EMSAs (Fig. 3). The results indicate that after replacing C with T at position 1 of the consensus site, Quox-1 cannot bind to the TAAT sequence, suggesting that C at the first position of the consensus sequence is preferred for optimal Quox-1 binding. Similar result was also observed when T was replaced with G at position 4—Quox-1 no longer binds to the altered sequence CAAG. Removal of the entire CAAT core completely eliminated DNA binding ability of Quox-1. These results demonstrated that the 5′-CAAT-3′ core is an essential recognition site for Quox-1 binding.

Mutations at residues 6 and 7 of the Quox-1 protein affect DNA-binding specificity. Structural and mutational studies further indicate that amino acid residues of helix β and the N-terminal arm of the homeodomain primarily determine DNA binding specificity [18]. However, the relative importance of each residue appears to be different, depending on the specific homeodomain. Residues 6 and 7 are important for DNA binding, since change of Gln to Thr at position 6 eliminated the binding ability of the mutant protein to CAAT, while replacing of Gln6 and Thr7 with Thr and Ala resulted in the change of the binding site: from CAAT to TAAT (Fig. 4).

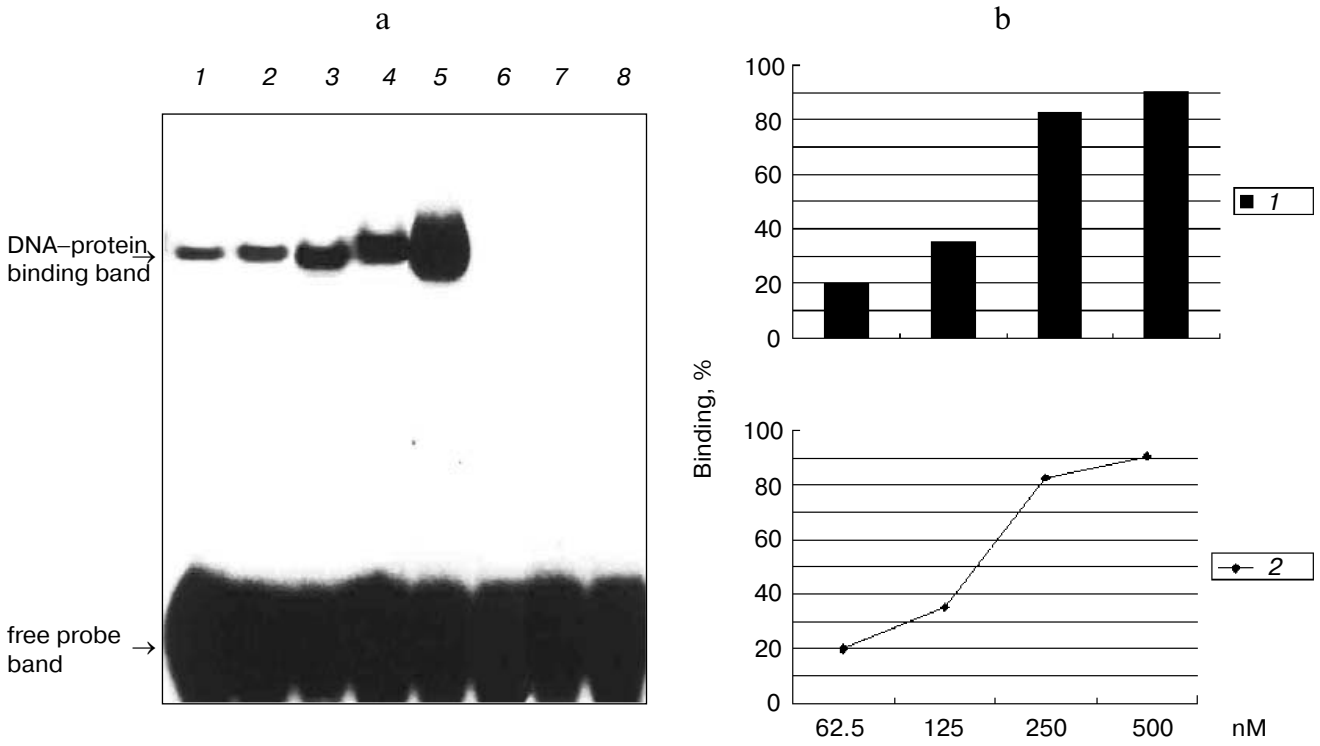

Fig. 2. a) Binding affinity of Quox-1 to CAATC was confirmed by a competitive gel shift assay: *1-4*) competitors were included at concentrations of 500, 250, 125, and 62.5 nM, respectively; *5*) no competitors; *6*) binding affinity of Quox-1 to CAATT; *7*) binding affinity of Quox-1 to CAATG; *8*) binding affinity of Quox-1 to CAATG-free probe. b) The data were quantified (*1*) and normalized to Quox-1 binding to probe with no competitor (*2*).

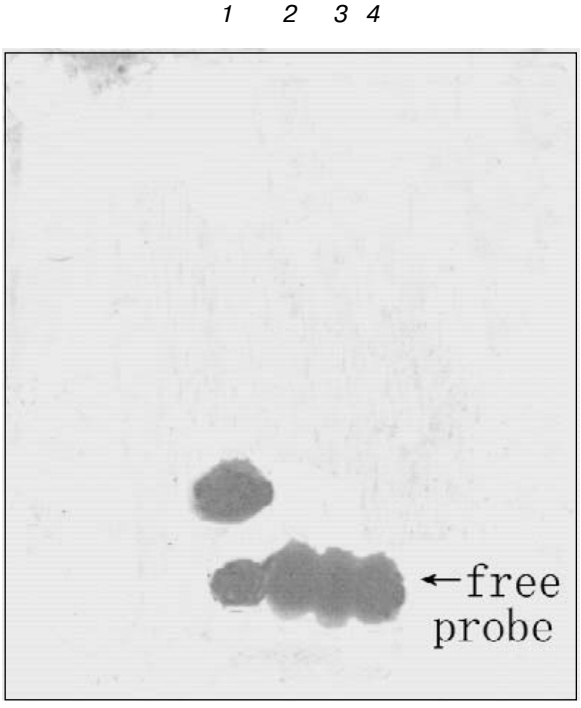

Fig. 3. Effect of mutations in the CAAT consensus site on Quox-1 binding: *1*) probe contains CAAT; *2*) probe contains TAAT; *3*) probe contains CAAG; *4*) the reaction medium has no protein.

DISCUSSION

By using purified fusion proteins in a random DNA selection assay, a CAATC consensus binding sequence for Quox-1 homeodomain was identified. This consensus sequence has not been recognized as a binding site for any other Hox class homeodomain proteins. In contrast, almost all other Hox class homeodomain proteins preferentially bind to the DNA-binding consensus sequence 5′-TAAT-3′. The usual bias among homeodomain proteins for the 5′-TAAT-3′ core sequence was recently confirmed by Wilson et al. [19] using the SAAB selection strategy similar to the one used in this study. For Quox-1 homeodomain, the 5′-CAAT-3′ core sequence is essential for its binding specificity and efficiency. Interestingly, this binding site is similar, but not identical, to the consensus sequence recognized by most Hox class homeodomain proteins, since the first nucleotide is a C rather than a T.

The homeodomain (HD) is one of the most common DNA binding motifs in eukaryotic transcriptional regulators [2]. Structural studies have shown that the HD is composed of three α-helices and a flexible N-terminal arm [19]. NMR showed that the third helix of the homeodomain is extended by two turns of a more flexible helix, termed helix χ [16]. The third helix (β/χ) of the HD

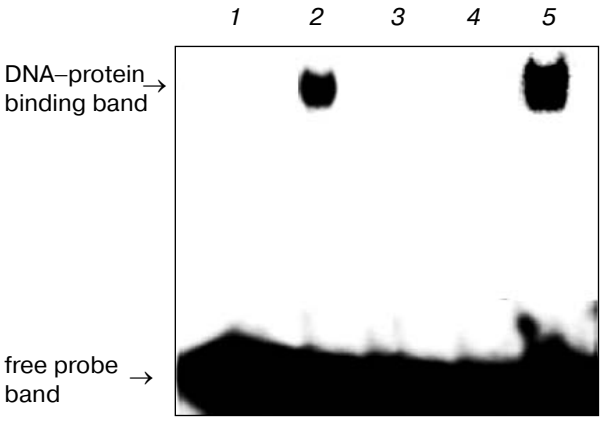

Fig. 4. Mutations in residues 6 and 7 change DNA binding site: *1*) mutations of Gln6, Thr7 to Thr, Ala (probe contains CAAT); *2*) mutations of Gln6, Thr7 to Thr, Ala (probe contains TAAT); *3*) mutation of Gln6 to Thr (probe contains CAAT); *4*) mutation of Gln6 to Thr (probe contains TAAT); *5*) protein carrying no mutation (probe contains CAAT).

inserts itself into the major groove of the DNA and is, therefore, called the recognition helix [19]. In addition, the N-terminal arm of the HD inserts into the minor groove of the DNA, making several contacts with the bases [19, 20]. Recently a set of detailed experiments was performed to determine the amino acids of the homeodomains required for recognizing DNA sequences of the core bind site. It has been demonstrated that the amino acid residues 6-8 of the N-terminal arm of the homeodomains are required for the recognition of the preferred nucleotides at the 5′-end of the core-binding site [21]. The amino acids of the homeodomain of Quox-1 at positions 6, 7, and 8 are Gln, Thr, and Tyr, respectively, while the amino acids of most other Hox class homeodomains at positions 6, 7, and 8 are Thr, Ala, and Tyr, respectively. The variation in side-chain compositions of residues 6 and 7 may result in preferential binding by Quox-1 of C instead of T at the first position of the CAATC binding site. Our findings and previous studies have consistently demonstrated that change of Glu6 to Thr6 led to the loss of binding ability to CAAT while replacement of Glu6 and Thr7 with Thr6 and Ala7 resulted in the change of the binding specificity (from CAAT to TAAT). These data suggest that residues at positions 6 and 7 of the homeodomain are important for DNA binding. Steadman et al. [22] also found that when Leu at position 7 of the NK-2 homeodomain was changed to Ala, the binding affinity for the CAAGTG site decreased by one order of magnitude, further suggesting that side-chain volume of residues 6 and 7 is an important determinant of homeodomain binding activity. Therefore, it is likely that both residue Gln6 and residue Thr7 of the Quox-1 homeodomain are required for preferential binding of C instead of T in the CAATC sequence. Thus, we think that the Gln6 and Thr7 in N-terminal arm of the homeodomain are important for participating in recognition of a minor groove.

NMR spectroscopy showed that the recognition helix (β/χ) lies in the major groove of the DNA [2]. Kissinger et al. [20, 23, 24] reported that amino acid replacements at positions 50 in helix β and 54 in helix χ might influence DNA binding specificity. Amendt et al. [15] demonstrated that the amino acid in position 54 of the homeodomain is involved in the recognition of the 3′-end of the core sequence. The amino acids at position 50 and 54 of Quox-1 homeodomain are Gln and Met, respectively, which are identical to the typical DNA binding sites of most homeodomains that recognize and bind to the canonical TAAT binding site. So it is not surprising that the fourth nucleotide of the core-binding site is T, which is the same for the typical binding sequence.

Our findings suggested that the Quox-1 homeodomain binding sequence is CAATC, not TAAT. Preliminary data generated in this study indicated that polymorphism occurs at positions 6 and 7 of the N-terminal arm of the homeodomains. These polymorphisms may affect DNA binding affinity as well specificity. Further analyses may provide insights into the effects of the polymorphisms on the function of the homeodomains.

This research was funded by the National Natural Sciences Foundation of China to Dr. W. X. Li from the Academy of Sciences, P. R. China (No. 39770373) and the Zhiqiang Foundation of Wuhan University to Dr. F. Zhu from Wuhan University, P. R. China (No. 273106).

Many thanks to Dr. Ma for providing us expression vector pGEMEX-xBal I.

REFERENCES

1. Scott, M. P., Tamkun, J. W., and Hartzell, G. W., 3rd. (1989) *Biochim. Biophys. Acta*, **989**, 25-48.
2. Gehring, W. J., Affolter, M., and Burglin, T. (1994) *Annu. Rev. Biochem.*, **63**, 487-526.
3. Xue, Z. G., Gehring, W. J., and le Douarin, N. M. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 2427-2431.
4. Xue, Z. G., Xue, J., and le Douarin, N. M. (1993) *Mech. Devel.*, **43**, 149-158.
5. Zhu, F., Li, W. X., Jiang, D. H., and Gou, D. M. (2002) *Cell Tissue Res.*, **308**, 333-337.
6. Zhu, F., Mao, X., and Li, W. X. (2001) *Acta Zool. Sinica*, **47**, 329-332.
7. Zhu, F., Li, J., Li, W. X., Liu, Z. C., and Long, X. (2004) *J. Biochem. Mol. Biol.*, **37**, 671-675.
8. Zhu, F., Mao, X., Liu, Z. C., Li, J., Long, X., and Li, W. X. (2004) *Chin. J. Stomatol.*, **39**, 544.
9. Blackwell, K. T., and Weintraub, H. (1990) *Science*, **250**, 1104-1110.

10. Gou, D. M., Sun, Y., Gao, L., Chow, L. M. C., Huang, J., Feng, Y. D., Jiang, D. H., and Li, W. X. (2001) *Biochim. Biophys. Acta*, **1518**, 306-310.
11. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene*, **77**, 51-59.
12. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
13. Ausubel, F. M. (ed.) (1995) *Short Protocols in Molecular Biology*, 3rd Ed., John Wiley & Sons, Inc.
14. Chen, C. Y., and Schwartz, R. J. (1995) *J. Biol. Chem.*, **270**, 15628-15633.
15. Amendt, B. A., Sutherland, L. B., and Russo, A. F. (1999) *J. Biol. Chem.*, **274**, 11635-11642.
16. Gehring, W. J., Qian, Y., Billeter, M., Furukubo-Tokunaga, K., Schier, A., Resendez-Perez, D., Affolter, M., Otting, G., and Wuthrich, K. (1994) *Cell*, **78**, 211-223.
17. Damante, G., Fabbro, D., Pellizzanri, L., Civitareale, D., Guazzi, S., Polycarpon-Schwartz, M., Cauci, S., Quadrifoglio, F., Formisano, S., and Lauro, R. D. (1994) *Nucleic Acids Res.*, **22**, 3075-3083.
18. Percival-Smith, A., Muller, M., Affolter, M., and Gehring, W. J. (1990) *EMBO J.*, **9**, 3967-3974.
19. Wilson, D. S., Sheng, G. L., Jun, S., and Desplan, C. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 6886-6891.
20. Otting, G., Qian, Y. Q., Billeter, M., Muller, M., Affolter, M., Gehring, W. J., and Wuthrich, K. (1990) *EMBO J.*, **9**, 3085-3092.
21. Damante, G., Pellizzari, L., Esposito, G., Fogolari, F., Viglino, P., Fabbro, D., Tell, G., Formisano, S., and Lauro, R. D. (1996) *EMBO J.*, **15**, 4992-5000.
22. Steadman, D. J., Giuffrida, D., and Gelmann, E. P. (2000) *Nucleic Acids Res.*, **28**, 2389-2395.
23. Kissinger, C. R., Liu, B., Martin-Blanco, E., Kornberg, T. B., and Pabo, C. O. (1990) *Cell*, **63**, 579-590.
24. Wolberger, C., Vershon, A., Liu, B., Johnson, A., and Pabo, C. (1991) *Cell*, **67**, 517-528.