ELSEVIER

FEBS Letters 337 (1994) 175–178

FEBS 13520

FEBS LETTERS

Noncanonical Oct-sequences are targets for mouse Oct-2B transcription factor

A.G. Stepchenko*

Engelhardt Institute of Molecular Biology, The Russian Academy of Sciences, Vavilov str. 32, Moscow, 117984, Russian Federation

Received 24 November 1993

Abstract

We have suggested a random modification method for determining preferable binding sites of a DNA-binding protein and applied this method to the Oct-2B transcription factor. Our results indicate that the Oct-2B protein interacts with canonical oct sequence ATGC/TAAAT and degenerated sequences which contain TAAT motif in the binding site. We have determined nucleotides in the binding sites, involved in the DNA–protein interaction, and the equilibrium dissociation constants K_d for these sequences. These data show that a much greater number of potential targets for Oct proteins exist on DNA and changed our view on the gene expression regulation by this protein factor.

Key words: Mouse Oct-2B; DNA–protein interaction; POU domain; Octamer sequence

1. Introduction

The nuclear Oct proteins belong to a family of POU proteins, exhibiting some critical functions in cell differentiation and regulation of gene expression [1–3]. All these factors contain a POU-domain, recognizing a conservative oct sequence (ATGCAAAT) in the regulatory regions of target genes. The POU domain consists of 150–160 amino acid residues. Three parts can be distinguished in this domain: a conserved N-terminal POU-specific domain (POUs) and POU homeodomain (POUhd) connected by a variable spacer region [4–7]. The whole POU domain is involved in the site-specific recognition. Deletion of the POUhd, or POUs domain results in the elimination of the binding or a decrease in the affinity interaction [8,9]. The POUhd consists of a triple α-helical structure and a flexible N-terminal region. Two helices form a structure similar to helix-turn-helix motif. The third helix plays the leading role in DNA recognition since it creates many contacts in the major groove of the DNA.

Recently the ability of Oct-2 proteins [10], and Oct-1 proteins [11] to bind a set of noncanonical sequences have been shown. In this paper we have examined the interaction between an octamer-binding protein Oct-2B and a set of oct-related sequences. For this purpose we used the 'random' oligonucleotide to determine a preferable DNA binding site in the Oct-2B POU domain. Our results indicate that the Oct-2B POU domain affinities towards the octamer sequence ATGCAAAT and the octamer-related sequence ATGTAAAT are similar. The affinities to sequences containing the TAAT motif in the binding sites are low. We have investigated the reason for this reduction by analysis of the contact between DNA and the POU domain.

2. Materials and methods

2.1. Isolation of cDNA clones

The cDNA library from the mouse NS/O cell line was obtained in phage λgt11. The screening was based on the ability of recombinant proteins to bind the double-stranded oligonucleotide probe incorporating a binding site (ATGCAAAT) for POU proteins. Two clones were isolated and cDNA sequences were determined [10]. The sequence analysis of these clones has demonstrated their extensive similarity to human Oct-2B DNA [6]. The cDNA fragment coding for the POU domain was cloned in the pUR292 vector in the reading frame. The resulting fusion-protein, named rec7, consisting of the POU domain and β-galactosidase was purified and used for gel retardation assay with the DNA probes containing the oct-sequence.

To detect non-canonical sequences, recognized by Oct-2B protein, we have used 'random' oligonucleotide as a probe.

```
5′-CCGGGAAGCTGNNNNNNNNGTGCTGC
       CTTCGACNNNNNNNNCACGACGGGCC-5′
```

where N is any of four bases (A, C, G, T)

2.2 Gel retardation assay

Oligonucleotides for gel retardation assay were ^{32}P-end labeled with the large fragment of DNA polymerase I. The protein binding has been carried out under standard conditions. The binding buffer contained 50 mM KCl, 20 mM Tris-HCl, pH 7.8, 1 mM EDTA, 1 mM PMSF, 5% glycerol, 100 μg/ml BSA. 5–20 ng of purified rec7 protein and

*Corresponding author. Fax: (7) (095) 135 14 05.

SSDI 0014-5793(93)E1437-Q

0.5–0.75 μg of the ^{32}P-labeled random probe were added to the reaction mixture. To eliminate the non-specific binding 0.5–2 μg poly[d(G-C)]/poly[d(G-C)] DNA was used. The reaction mixture was incubated for 30 min at room temperature. After autoradiography, the band containing a specific DNA–protein complex was cut out from the gel. DNA fragments were purified, cloned in pUC19 vector and sequenced. We have used the gel retardation assay to compare the binding affinities of the POU domain with those of several octamer sequences. The equilibrium dissociation constant K_d values have been determined in the binding saturation experiments in which the protein concentration was increased in the presence of a fixed amount of the DNA probe [11]. Binding of the POU domain included in a fused protein was carried out under standard conditions described above in the presence of 0.25 μg poly[d(G-C)]/poly[d(G-C)] and $5 \cdot 10^{-10}$ M DNA probe.

2.3 DNA methylation interference experiments

The ^{32}P-labelled DNA fragments were methylated with DMS according to the method of Siebenlist and Gilbert [12].

3. Results and discussion

The oligonucleotide sequences, binding the Oct-2B POU domain are shown in Fig. 1. All these oligonucleotides bind the native Oct-1 and Oct-2B proteins. It is possible to distinguish two groups of sequences that interact with the Oct-2B POU domain. The first group contains a common tetranucleotide T/CAAA, the second one contains a tetranucleotide TAAT. Fig. 2 shows the rec7 protein binding with some cloned sequences. All probes differ by their affinity to rec7 protein. The stability of DNA–protein complexes depends on the structure of flanking sequences and for group 1, in particular, on the presence of the second half of the binding site (ATGC). It should be noted that MATα2 protein, which does not belong to the POU family of proteins, recognizes AATTTACAT sequence which also binds with high affinity to the Oct-2B POU domain [13]. Fig. 3 shows the results of DMS interference assay. This analysis indicates that the interaction of POU domain with octamer-related sequences is different. The modification

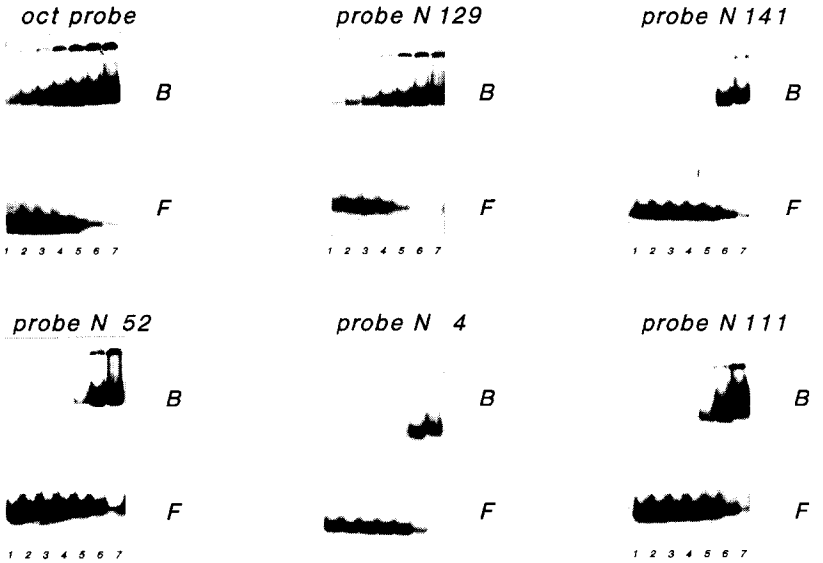
I group

67. GGGTCAGCACATGTAAATCAGCGTCCC
19. GGGTCAGCACATGCAAAACAGCTCCC
12. GGGTCAGCACAATGTAAATAGCTTCCC
129. GGGTCAGCATATGTAAATCAGCTTCCC
138. GGGTCAGCACAATGTAAATAGCCTTCCC
117. GGGAAGCTGATGCAAATGTGCTACCC
9. GGAAGGCTAATGCAAATCGTGCTGACCC
11. GGGAAACTTGTATGCAAAGTGCTGACCC
47. GGGTCAGCACATGCAAATTAGTTTCCC
1. GGGTCAGCACAATGTAAACAGCTTCCC
20. GGGTCAGCATGTAAATTGCAGCTTCCC
17. GGGAGCTGATATGCAAATGCTGACCC
38. GGGTCAGCACAATGTAAACAGCTTCCC
27. GGGAGCTGATATGCAAATGCTGACCC

II group

74. CTGTTTTAATTGTCTGACCC
70. GGGTCAGCACTTAATAAACTGCTTCCC
108. GGGTCAGCACATAATTTACAGCTTCCC
91. GGGAAGCTGAGATAATTGTGCTGCCC
5. GGGTGCTGCATAATCAGTGCTGACCC
86. GGGTCAGCACGTGGTAATTAGCTTCCC
124. GGGAAGTATTCTAATTGTGCTGACCC
141. GGGTCAGCACACGATAATCAGCTTCCC
24. GGGTCAGCACTTAATATGCAGCTTTCCC
41. GGGTCAGCACTCGATAATCAGCTTCCC
13. CCCAAACTGCATAATCGTGCTGACCC
122. GGGTCAGCACAGAATAATCAGGCTTCCC
14. GGGAAGCTGACCTAATAATGCTGA
15. GGGAAGCTGTATAATCTATGCTGACCC
16. GTCAGCATATATAATTCAGTTTCCC
98. GGGTCAGCACGAGTAATCAGCCC
52. GGGAAGCTGCATATCTAATGCTGACCC
2. GGGAGCTGATTTAATGTGCTGACCC
3. GGGAACTGTAATTTTAATGCTGACCC
5. GGGTCAGTATGGTAATGGCAGGCTTCCC
6. GGGTCAGCATAATGTAGTGAGCTTCCC
7. GGGGTCAGCATTTAATAATCAGCTTCCC
8. AGTATGTTAATGTAGCTTCCC
10. GGGAAGCTGATTTAATGTGCTGACCC
32. GGGAAACTGCATAATCGTGCTGACCC
26. GGGAAGCTGCTTATATAATGTTGACCC
4. GGGAAGCTGCATACCAAGGTGCTGACCC
111. GGGTCAGTACGGTATAAGTAGCTTCCC

Fig. 1. Aligned oligonucleotide sequences recognized Oct-2B POU domain. Sequences of the cloned oligonucleotides were selected as described in section 2.



region for the canonical sequence ATGCAAAT occupies 6–8 base pairs and our results on the DNA modification are in accordance with those obtained earlier [4].

If a DNA sequence has two POU domain binding sites TAATATTTGCAT (probe 47), then the POU domain preferably binds to the canonical site ATTTGCAT. The second group of sequences does not necessarily contain a target site ATGCA for POUs (probe 141, 41, 122) or contains a slightly modified sequence CTGCA (probe 13, 52, 24). Apparently, the presence of a sequence, recognized only by POUs, is enough for the DNA–protein interaction. For example, probe 4 binds the POU domain. In probes 24 and 111 POUhd apparently recognizes the TAAG tetramer.

From these data and the results obtained earlier [4,10,11,14,15] three possibilities of DNA–POU proteins interactions may be suggested:

(i) 'POU-specific' type interaction. There is a strong site for POUs domain (ATGCA) in the target sequence, while there is no such site for POUhd. Probably POUhd recognizes the DNA sequences close to the ATCGA site albeit with a low affinity and so stabilizes the DNA–protein complex. The most effective interaction is achieved in this case. Apparently, the POUs domain expels POUhd from its binding site and makes maximum possible contacts with DNA.

(ii) 'Homeo' type interaction. There is a strong site for POUhd RTAATNA (probe 41, 141, 122) in the target sequence, while a strong site for POUs is absent. POUhd makes all possible contacts with DNA and expels POUs which bind DNA with a very low affinity (apparently one G/C or C/G pair to the left from the POUhd site is sufficient for a weak POUs interaction).

(iii) Both parts of the binding site are suboptimal (probes 13, 52, 24) or degenerated (probes 4, 111) for DNA–POU domain interaction. In these case the binding strength depends on the affinity of each part of the binding site.

So, the POU proteins are able to interact in vitro with both the canonical oct-sequences and non-canonical ones. If this situation is reproduced in vivo (at least partially) the pattern of gene expression regulation by POU proteins should be essentially complicated, and the number of genes regulated by Oct proteins may be much higher than it is supposed at the present time.

←

Fig. 2. Comparison of the POU domain binding affinities for several probes determined by the gel retardation assay. The POU domain binding was carried out in the presence of $5 \cdot 10^{-10}$ M DNA probe. The amount of protein was gradually increased: 2 ng, 4 ng, 8 ng, 16 ng, 32 ng, 64 ng, 128 ng (lanes 1 to 7, respectively). Positions of the protein–DNA complex (B) and free DNA (F) are indicated with arrows.

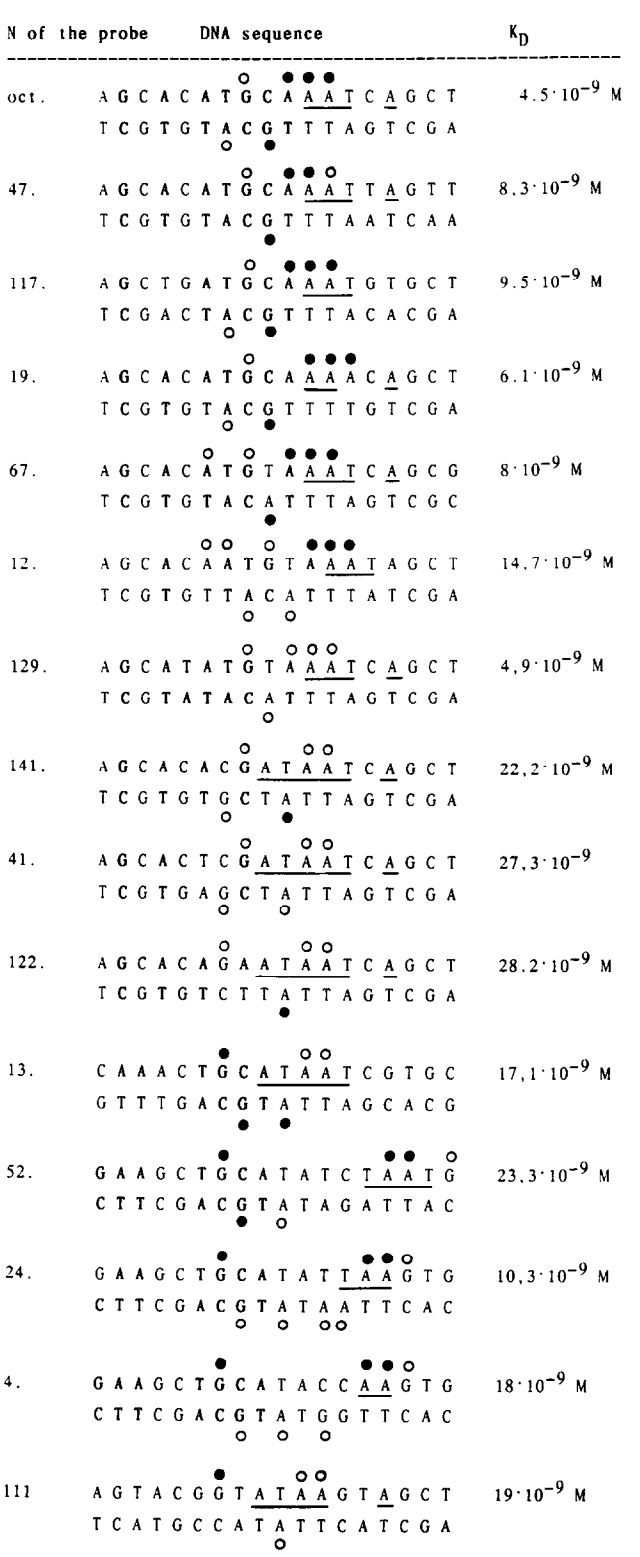


Fig. 3. Methylation interference analyses. The guanine and adenine residues which methylation interferes with the complex formation are marked by circles. The solid circles indicate a four-fold reduction in the band intensity while the open ones show about two-to-four-fold reduction in the band intensity. The consensus sequence for the POUs domain is indicated in boldface, the consensus sequence for POUhd is underlined. On the right, the K_d values for these sequences are cited.

Acknowledgements: I am grateful to Dr. E. V. Pankratova for advice and assistance in preparing this manuscript. This work was supported by 'The Human Genome' program and the Russian State program 'Frontiers in Genetics'.

References

[1] Gehring, W.J. (1987) Science 236, 1245–1252.
[2] Scott, M.P., Tamkun, J.W. and Hartzell, G.W. (1989) Biochim. Biophys. Acta 1989, 25–48.
[3] He, X, Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W. and Rosenfeld, M.G. (1989) Nature 340, 35–42.
[4] Sturm, R., Baumruker, T., Franza, B.R. and Herr, W. (1987) Genes Dev. 1, 1147–1160.
[5] Bodner, M., Castrillo, J.-L., Theill, L.T., Deerinck, T., Ellisman, M. and Karin, M. 91988) Cell 55, 505–518.
[6] Clerc, R.G., Corcoran, L.M., Le Bowitz, J.H., Baltimore, D. and Sharp, P.A. (1988) Genes Dev. 2, 1570–1581.
[7] Finney, M., Ruvkun, G. and Horvitz, H.R. (1988) Cell 55, 757–769.
[8] Sturm, R.A. and Herr, W. (1988) Nature 336, 601–604.
[9] Verrijzer, C.P., Kal, A.J. and van der Vliet, P.C. (1990) EMBO J. 9, 1883–1888.
[10] Stepchenko, A.G. (1992) Doklady Academii Nauk. Russia, 235, 175–179.
[11] Verrijzer, C.P., Alkema, M.J., van Weperen, W.W., van Lecuwen, H.C., Strating, M.J.J. and van der Vliet, P.C. (1992) EMBO J. 11, 4993–5003.
[12] Siebenlist, U. and Gilbert, W. (1980) Proc. Natl. Acad. Sci. USA 77, 122–126.
[13] Johnson, A.D. and Herskowitz, J. (1985) Cell 42, 237–247.
[14] Verrijzer, C.P., Kal, A.J. and van der Vliet, P.C. (1990) Genes. Dev. 2, 1964–1974.
[15] Kristic, T.M. and Sharp, P.A. (1990) Genes Dev. 4, 2383–2396.