Biochemistry (Moscow), Vol. 68, No. 4, 2003, pp. 377-384. Translated from Biokhimiya, Vol. 68, No. 4, 2003, pp. 467-476.

Consensus Sequence of Transcription Factor SF-1 Binding Site and Putative Binding Site in the 5′-Flanking Regions of Genes Encoding Mouse Steroidogenic Enzymes 3βHSDI and Cyp17

T. V. Busygina*, E. V. Ignatieva, and A. V. Osadchuk

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, ul. Lavrentieva 10, Novosibirsk 630090, Russia; fax: (383-2) 33-1278; E-mail: tbusig@bionet.nsc.ru

Received April 15, 2002
Revision received November 25, 2002

Abstract—Using 42 nucleotide sequences extracted from the Transcription Regulatory Regions Database (TRRD) containing SF-1 transcription factor binding site, we have determined the decanucleotide (GTCAAGGTCA) consensus sequence for SF-1 binding. In the frequency matrix of this sequence nucleotides between the 3rd and the 7th position had the highest frequency and guanine nucleotides at the 6th and the 7th positions were recognized in all nucleotide sequences. The latter suggests a crucial role of these guanines for the interaction of DNA with SF-1 protein. The determined consensus and frequency matrix were used for search of putative SF-1 binding sites in regulatory regions of two genes, encoding mouse Cyp17 (17α-hydroxylase/17-20-lyase) and 3βHSDI (3β-hydroxysteroid dehydrogenase/4Δ-5Δ-isomerase I), the microsomal enzymes involved in steroidogenesis. 5′-Flanking regions of genes encoding Cyp17 and 3βHSDI were shown to contain six and five such binding sites, respectively. The presence of the putative SF-1 binding sites in the regulatory regions of mouse *Cyp17* and *3βHSDI* suggests that gene *SF-1* could represent one of the putative genes which (as we predicted earlier) determine coordinated inheritable variability of hormonal activity in mouse Leydig cells.

Key words: SF-1 (steroidogenic factor 1), consensus of transcription factor binding site, steroidogenic enzymes, testicular steroidogenesis, Cyp17, 3βHSD, putative SF-1 binding site

Steroid hormones play an important role in control of various biological functions such as reproduction, stress adaptation, carbohydrate metabolism, and water and salt balance. According to biological functions steroid hormones are subdivided into five groups: progestins, estrogens, androgens, glucocorticoids and mineralocorticoids. Adrenal cortex and gonads are the main source of steroid hormones. Some species-specific expression of enzymes involved in biosynthesis of steroid hormones has also been found in placenta, uterus, brain, skin, and adipose tissue [1-7].

These classes of steroid hormones share structural similarity and common biosynthetic pathways. In all steroidogenic tissues cholesterol is the main precursor of steroid hormone synthesis *de novo* and steroidogenesis involves three mitochondrial and five microsomal enzymes operating in various tissues. The mitochondrial enzymes include cholesterol side chain cleaving cytochrome P450 (P450scc), 11β-hydroxylase (P450c11), and aldosterone synthase (P450aldo). The microsomal enzymes include 3β-hydroxysteroid dehydrogenase/4Δ-5Δ-isomerase (3βHSDI), 17α-hydroxylase/17-20-lyase (P450c17), 21-hydroxylase (P450c21), 17β-hydroxysteroid dehydrogenase/17-ketosteroid reductase (17KSR), and aromatase (P450arom).

Steroidogenic factor 1 (SF-1), a transcription factor of the nuclear receptor family, plays an essential role in regulation of gonadal and adrenocortical steroidogenic function. In the 1990s genes encoding steroidogenic enzymes were cloned and sequenced. Study of 5′-flanking regions of genes encoding steroid hydroxylases revealed one or more copies of nucleotide sequences containing motif AGGTCA responsible for binding of SF-1 monomer [8, 9]. Subsequent gene knockout employed in three independent laboratories [10-12] revealed that besides regulation of steroidogenesis in adrenals and gonads, SF-1 is essential for the development and functioning of all levels of the hypothalamo–pituitary–gonadal and adrenal axes. Mice with impaired expression of SF-1 were characterized by underdeveloped gonads

* To whom correspondence should be addressed.

0006-2979/03/6804-0377$25.00 ©2003 MAIK “Nauka/Interperiodica”

and adrenals, reduced level of genes encoding gonadotropins, and defects of the development of hypothalamus ventromedial nucleus involved in regulation of sex behavior. In such animals intrinsic female sex structures developed irrespectively of the genetically determined sex. Several hours after birth such mice died presumably due to endogenous steroid insufficiency [10-12]. SF-1 binding sites have now been found in regulatory regions of many genes involved in regulation of functioning of the hypothalamo–pituitary–gonadal and adrenal axes in various vertebrates. Steroidogenesis in adrenals and gonads is regulated by pituitary tropic hormones (adrenocorticotropic, luteinizing, and follicle stimulating hormones). Their effects in the target tissues are realized by cAMP. In some cases SF-1 is involved in cAMP-dependent regulation of transcription of genes encoding enzymes and other protein factors which regulate steroid hormone biosynthesis. These include transcription of bovine genes encoding CYP11B1 [13], CYP17 [14], and CYP11A [15], human genes encoding CYP11A [16], StAR [17-20], and CYP19 [21], and mouse gene encoding CYP11A [22].

Since 1993 the Institute of Cytology and Genetics (Siberian Branch of the Russian Academy of Sciences) has been developing the Transcription Regulatory Regions Database (TRRD) [23, 24]. Using computer technologies this database allows the accumulation of experimental data on eukaryotic gene regions involved in regulation of transcription. According to functions of genes described in TRRD (http://www.bionet.nsc.ru/trrd/) there are several specialized sections in this database. The Endocrine System–Transcription Regulatory Regions Database (ES-TRRD) is one of them [24, 25] where information on regulation of transcription of genes functionally related to the endocrine system is accumulated. We have developed the ES-TRRD subsection where data on regulation of gene transcription involved in control of steroidogenesis are accumulated. According to the ES-TRRD, transcription factor SF-1 is involved in the regulation of most genes listed in this subsection [25, 26]. Usually the consensus sequence of the SF-1 binding site is based on a rather limited set of its binding sites whereas ES-TRRD contains information on 61 SF-1 binding sites. This makes reasonable the possibility to obtain a more accurate consensus sequence of the SF-1 binding site based on the representative set of nucleotide sequences containing this site.

Studying genetic control of hormonal function of the male reproductive system in six inbred strain of mice, we found polymorphism of gonadal androgenic activity [27-32]. This polymorphism was mainly associated with peculiarities of steroidogenesis in Leydig cells [33, 34], the main testosterone producers in testes [35]. It was demonstrated that coordinated interstrain variability of activity of microsomal enzymes is involved in steroidogenesis [36].

Diallelic analysis of cAMP- and substrate-dependent testosterone production by mouse Leydig cells of inbred strains and their reciprocal hybrids of the first generation revealed that these parameters are under coordinate and polygenic control. Using these data we developed a 4-locus segregation model for coordinate inherited variability of hormonal activity of Leydig cells [37-39]. Using modern notions on the key role of SF-1 factor in the regulation of many genes involved in the hypothalamo–pituitary–gonadal system [2, 8, 40-43] we suggested that the gene encoding this factor can be considered as one of the putative candidates in the 4-locus model of inheritance of coordinate hormonal activity of Leydig cells. The existence of an SF-1 binding site in the mouse gene *Cyp11A* encoding P450scc was experimentally demonstrated [10, 44]. However, involvement of the SF-1 factor in the regulation of expression of mouse genes encoding 3βHSDI, Cyp17, and 17KSR has not been experimentally demonstrated. So, the search for putative SF-1 binding site(s) in 5′-flanking regions of these three genes is especially interesting.

Thus, in the present study we developed the consensus sequence for SF-1 binding site using the set of nucleotide sequences available in ES-TRRD and employed this consensus for the search of putative SF-1 binding in 5′-flanking regions of genes encoding enzymes involved into mouse testicular steroidogenesis.

MATERIALS AND METHODS

Nucleotide sequences (42) available at ES-TRRD (http://wwwmgs.bionet.nsc.ru/mgs/papers/ignatieva/es-trrd/) were used for determination of the consensus sequence of the SF-1 transcription factor binding site. According to literature data these sequences demonstrated SF-1 binding activity in the following tests: 1) gel retardation of the complex between DNA and purified SF-1 protein; 2) DNase I footprinting with purified SF-1; 2) gel retardation using nuclear extract and specific antibodies to SF-1.

Deviation of nucleotide occurrence in the frequency matrix of consensus from probable frequency was statistically evaluated by the χ^2 criterion using Microsoft Excel 97 software.

Putative SF-1 binding sites in 5′-flanking regions of genes encoding 3βHSDI and Cyp17 were sought using this consensus and the program for multiple local alignment of nucleotide sequences GIBBS Sampler available at http://wwwmgs.bionet.nsc.ru/mgs/programs/gibbs_nuc/ [45]. The list of putative SF-1 binding sites included 5′-regulatory regions of *Cyp17* and *3βHSDI* differing from the 10-letter consensus sequence by not more than three nucleotides. Nucleotide sequence of *3βHSDI* was taken from [46]. Nucleotide sequence of *Cyp17* (A00490) was taken from the EMBL database (S41708) and the position of the transcription start site was determined on the basis of information in TRRD on the positions of two ARE

(androgen responsive element) versus the start site. We did not find any information on 5′-flanking region for the mouse gene encoding 17KSR either in EMBL/GeneBank or the available literature.

RESULTS AND DISCUSSION

We developed a special section in TRRD, ES-TRRD, containing information on regulation of transcription of genes controlling steroidogenesis. This section, containing data on 60 genes, was used as the source of nucleotide sequences for consensus of SF-1 binding site. Table 1 lists 41 of 60 genes of ES-TRRD that have 61 SF-1 binding sites recognized using various experimental approaches. The first group consists of mouse (n = 12), human (10), rat (10), bovine (5), sheep (1), pig (1), horse (1), and salmon (1) genes encoding enzyme involved in steroid hormone biosynthesis, tropic hormones, receptors of tropic and releasing hormones, transcription factors, and other proteins.

Table 2 shows results of multiple alignments of 42 nucleotide sequences of at least 16 nucleotides in length. These sequences correspond to SF-1 binding sites listed in

Table 1. SF-1 regulated genes of ES-TRRD controlling steroid hormone biosynthesis

Genes	Number in TRRD[species]
Genes encoding enzymes and other factors of steroidogenesis	
Steroidogenic acute regulatory protein (StAR)	A00584[m], A00488[h], A01062[b], A01353[r], A01569[p]
Cytochrome P450 cholesterol side chain cleavage (Cyp11A)	A00496[r], A00497[b], A00498[h], A00561[m]
3β hydroxysteroid dehydrogenase/Δ-5–4 isomerase type II (3βHSDII)	A00858[h]
Cytochrome P-450 17α hydroxylase/ C17-20 lyase (Cyp17)	A00585[b], A00565[r], A00937[h]
Aromatase (Cyp19)	A00591[r], A00152[h]
Cytochrome P450 steroid 21-hydroxylase (Cyp21B)	A00790[m]
Cytochrome P450 11β hydroxylase (Cyp11B1)	A00583[m], A00587[r], A00789[b]
Aldosterone synthase (Cyp11B2)	A00586[h]
Aldose reductase-like protein 1 (ALD1)	A01206[m]
Genes encoding peptide hormones	
Luteinizing hormone beta subunit (LHB)	A00480[b], A00857[ho], A00628[r], A01058[s]
Beta subunit gonadotropin hormone (GTHIIB)	A00785[chs]
Glycoproteins alpha subunit (GHA)	A00791[m], A00056[h]
Genes encoding peptide hormone receptors	
Gonadotropin releasing hormone receptor (GRHR)	A01207[r], A00792[m]
Prolactin receptor (PRLR)	A00629[r]
Adrenocorticotropin receptor (ACTHR)	A00929[h], A00786[m]
Luteinizing hormone receptor (LHR)	A00783[r]
Follicule stimulating hormone receptor (FSHR)	A01208[m]
Genes encoding factors of autocrine and paracrine regulation of steroidogenesis	
Leydig insulin-like gene	A00784[m]
Mullerian inhibiting substance (MIS)	A00988[h], A00781[m]
Genes encoding transcription factors	
Steroidogenic factor 1 (SF1)	A00548[r]
Dosage-sensitive sex-reversal-adrenal hypoplasia congenita critical region on the X-chromosome (DAX1)	A00732[m], A01024[h]

Note: Superscripts designate: [h], human; [m], mouse; [r], rat; [b], bovine; [chs], salmon; [ho], horse; [s], sheep; [p], pig.

Table 2. Binding sites for SF-1 in regulatory regions of genes listed in ES-TRRD

Gene	Site number in TRRD	Nucleotide sequence	Positions versus gene transcription start site
StAR^{h}	S2638	gt**TCAAGGTCA**aa	–916/–928 R
StAR^{h}	S2639	ggg**GTCAAGG**at**A**ga	–93/–107 R
StAR^{b}	S4929	gtg**GTCAAGG**ca**A**ttt	–237/–252 R
StAR^{b}	S4928	tcac**CAAGG**ctgctga	–1191/–1176
Cyp11A^{h}	S2674	aggc**TCAAGGTCA**tca	–1642/–1627
Cyp11A^{b}	S2680	tcaccagc**TCAAGG**ct**A**agtgagaag	–32/–57 R
Cyp11A^{r}	S2682	agg**G**ggg**AGGTCA**acactcc	–83/–64
Cyp11A^{r}	S2683	agc**TCAAGG**ct**A**agagaggag	–38/–58 R
3βHSDII^{h}	S3992	gagt**TCAAGGT**a**A**taa	–68/–53
Cyp17^{h}	S5502	cAGGA**GTCAAGG**cttg	–274/–289 R
Cyp17^{h}	S5503	taaggg**AGGTCA**gttg	–197/–212 R
Cyp17^{h}	S5504	c**T**g**AAGG**c**C**tcctcaa	–137/–152 R
Cyp17^{b}	S3074	aaagtcaag**G**ag**AAGGTCA**ggg	–62/–41
Cyp17^{r}	S2649	ac**GTCAAGGT**g**A**caat	–69/–54
Cyp17^{r}	S3970	gagaaagagatcttca**AAGGT**t**A**gtcggcata	–411/–442 R
Cyp11B2^{h}	S3045	cg**AAGGTCA**aggctggag	–112/–129 R
Cyp19^{h}	S3313	tac**CAAGGTCA**gaaat	–135/–120
Cyp19^{r}	S3065	ctcc**CAAGGTCA**tcct	–85/–70
LHB^{b}	S2791	agaggca**G**a**CAAGGTCA**gggagagg	–109/–133 R
LHB^{r}	S3050	ca**G**a**CAAGGTCA**gaaa	–116/–131 R
LHB^{r}	S4810	t**G**g**CAAGG**c**CA**cta	–49/–62 R
LHB^{ho}	S3986	cg**G**a**CAAGGTCA**ggga	–113/–128 R
LHB^{ho}	S3988	ga**G**g**CAAGG**c**CA**ctgg	–44/–59 R
LHB^{s}	S4940	ca**G**a**CAAGGTCA**ggga	–115/–130 R
GHA^{h}	S3307	ac**G**a**CAAGGTCA**gccc	–208/–223 R
GHA^{m}	S4803	aggtagtggtgacc**TCAAGG**a**CA**gcttatg	–193/–222 R
GHIIB^{chs}	S3696	aaa**GT**ag**AGGTCA**gga	–175/–160
GHIIB^{chs}	S3694	tta**TCAAGGTC**caagc	–351/–366 R
PRLR^{r}	S3063	cag**G**c**CAAGGTCA**aac	–680/–665
GNRHR^{m}	S3754	cc**T**g**AAGG**c**CA**agtgt	–235/–250 R
ACTHR^{h}	S4367	ttat**TCAAGGT**a**A**tga	–102/–87
ACTHR^{h}	S4368	cggcc**CAAGGTC**cact	–39/–24
ACTHR^{h}	S4366	ta**GTCAAGGT**t**A**cttc	–198/–213 R
LeyIL^{m}	S3697	gact**TCAAGGTC**ccaa	–144/–129
LeyIL^{m}	S3698	cga**GTCA**c**GGTCA**ggg	–99/–114 R
LeyIL^{m}	S3699	ccc**G**c**CAAGG**c**C**catg	–65/–50
MIS^{h}	S4799	ggcactgtcccc**CAAGGTC**gcg	–103/–82
MIS^{m}	S3702	cccc**CAAGGTCA**cctt	–95/–80
DAX1^{m}	S4304	agt**TC**g**AGGTCA**gagt	–337/–322
DAX1^{m}	S4305	ttgga**CAAGG**g**C**gcag	–68/–83 R
DAX1^{m}	S3292	tt**TC**g**AGGTCA**tggcca	–131/–115
DAX1^{h}	S4796	accac**C**g**AGGTCA**tgg	–177/–162
Consensus		**GTCAAGGTCA**	

Note: Superscripts designate: ^{h}, human; ^{m}, mouse; ^{r}, rat; ^{b}, bovine; ^{chs}, salmon. Nucleotides of the consensus sequence are marked in bold. R indicates that the sequence is located on the complementary strand.

ES-TRRD. (Their positions in the regulatory regions of genes are given in the last column.) For optimization of alignment more than half of the sequences of Table 2 correspond to the complementary strand of DNA (marked with R).

Using this set of SF-1 binding sites we developed the 10-letter consensus sequence GTCAAGGTCA (Table 2) and its frequency matrix (Table 3). According to our data, nucleotides from the 3rd to the 7th positions have the highest frequencies. Nucleotides from the 8th to the 10th positions have lower frequencies in the set of nucleotide sequences. For all nucleotides from the 3rd to the 10th positions the hypothesis of equal frequency probability for nucleotide occurrence is rejected by χ^2 criterion with very high statistical significance ($p < 0.00001$). Nucleotides located at the first and the second positions of the consensus sequences are characterized by lower frequency in the set of nucleotide sequences analyzed. For nucleotides at these positions the null hypothesis is rejected at $p < 0.01$. A pair of guanine nucleotides at the 6th and the 7th positions of the consensus are ultimately required for SF-1 binding. Usually the consensus sequence for the SF-1 binding site is given as an eight or nine letter fragment: PyCAAGGPyC, PuPuAGGTCA [41], CCAAGGTCA [47], (C/T)CAAGGT(C/T)A [8]. Our analysis of a representative set of SF-1 binding sites revealed additional guanine nucleotide in the first position of the consensus sequence. According to the frequency matrix (Table 3) this nucleotide is essential for SF-1 interaction with DNA. In the nucleotide sequence, positions 0 and 11 (numerated on the basis position of the consensus sequence) can be occupied by any nucleotide and the distribution of number of nucleotides in these positions (Table 2) insignificantly differed from equally probable (Table 3). This suggests that SF-1 binding requires a 10-nucleotide sequence.

The consensus sequence contains the six nucleotide motif AGGTCA, which represents the binding site for monomers of several nuclear receptors: NGFI-B, ROR, ERR1, ERR2 [48]. An element that is the binding site for homo- or heterodimers of transcription factors of families of steroid/thyroid hormone receptors is represented by a double repeat of six- or five-nucleotide sequence in the right or reversed palindrome orientation separated by several nucleotides. A half-site for binding of such factors of this family as estrogen, thyroid, X-retinoic receptors, and COUP-TF factors contains the AGGTCA motif. The existence of the common motif in nucleotide sequences responsible for binding of transcription factors of nuclear receptor family may cause competition between SF-1 and other nuclear receptors. In some cases such competition may determine the regulatory mechanism for gene transcription. For example, orphan receptors COUP-TF (COUP-TFI and COUP-TFII) known as negative regulators of transcription may compete for binding sites with certain factors of the nuclear receptor family (estrogen receptors, SF-1) or form heterodimers with thyroid receptors, vitamin D receptors, and retinoic X-receptors. In these complexes the orphan receptors may act as transcription inhibitors [49]. Within the database ES-TRRD such competition for a binding site between COUP-TF and SF-1 was recognized for human *Cyp11B2* (A00586), *Cyp19* (A00152), bovine (A00585) and rat (A00565) *Cyp17*, and also for mouse *DAX1* (A00732).

Using alignment of 5′-flanking regions of mouse *Cyp17* and *3βHSDI* with the consensus sequence of SF-1 binding sites, we found 11 putative SF-1 binding sites. The nucleotide sequence of *Cyp17* contains six putative binding sites for SF-1 (figure). One of them (−53/−44) is located in the right orientation, others (−505/−496, −435/−426, −337/−328, −285/−276, −279/−270) are in reversed orientation. In the position from −285 to −270 we found two overlapping SF-1 binding sites (figure). Binding sites for SF-1 were experimentally recognized for human, bovine, and rat *Cyp17* (and included into ES-TRRD) (Table 2). In bovine and rat genes functional sites (−62/−41 and −69/−54, respectively) are located in the

Table 3. Frequency matrix for consensus sequence for SF-1 binding

Position in the consensus sequence	0	1	2	3	4	5	6	7	8	9	10	11
A	0.25	0.12	0.17	0.05	**0.86**	**0.98**	0.00	0.00	0.05	0.07	**0.79**	0.24
T	0.15	0.17	**0.48**	0.00	0.00	0.00	0.00	0.00	**0.69**	0.17	0.05	0.17
G	0.35	**0.46**	0.09	0.14	0.14	0.00	**1.00**	**1.00**	0.02	0.02	0.07	0.35
C	0.25	0.25	0.26	**0.81**	0.00	0.02	0.00	0.00	0.24	**0.74**	0.09	0.24
Consensus	N	**G***	**T***	**C****	**A****	**A****	**G****	**G****	**T****	**C****	**A****	N

Note: The table shows frequencies for nucleotide occurrence in the corresponding position of nucleotide sequences listed in Table 2. Asterisks show significance of differences (* $p < 0.01$, ** $p < 0.00001$) of deviation of the null hypothesis (equal frequency probability for nucleotide occurrence) by the χ^2 criterion. Bold script shows frequencies of nucleotide in the consensus sequence. N is any nucleotide.

5'-GTCAAGGTCA-3'

Consensus sequence for SF-1-binding site in right orientation

5'-TGACCTTGAC-3'

Consensus sequence for SF-1-binding site in reversed orientation

Cyp17

```
-1018 tgagtattggcattgcgtccctaaaatgtacctttaaaaatgattaaagaggcagagaga
 -958 tggctcaaatgtaataaatttttaaaatgacaattaaaatgttactctgtaatttttgtg
 -898 tttaagatactagttagatattttatatgatttttaactagtgttttattatttatagg
 -838 catgagtgttttacctgcatgtatgtatgcagaccgtgtgtgttgggggtgctttcaggg
 -778 gccagaaggtgctttcctggaaccagaggtacagaaagttgtgagctaccatgtcagtgc
 -718 tgggaatctaacctgggtcttctggaagagcaacgaggcatcttgactgctgagccatct
 -658 ctccagctcttagttaaagtattttaaatgatgcattttacgttttatacatagtaacat
 -598 gacaaaatcatttatttccagaaaatattaaatagcccattcaagagaatgtacctagaa
                    -520                -500                -480
 -538 agtaaccaagctatttgaacagagacctgcaatgGcCCTTGAgctgatgttctgtgtacc
                    -460                -440                -420
 -478 tgaaagcctatgacttctggattcttatcatcactgccttgtgTGACCTTatgcaaacta
                    -400                -380                -360
 -418 accctaaaagacctctctctcctcaactatcagataataagactgaagtctctttgacag
                    -340                -320                -300
 -358 ctttggctagctgcaacctgaTGACaTTaAttattaactgtgcagcacttttgacattac
                    -280                -260                -240
 -298 agcacgcactctgaaACCTTGAtCTTaAtctgatagcatttgcctctgggaggatccata
                    -220                -200                -180
 -238 gcgcagaagccccagggagatgaccaactctctgctctttagccagctcctcctccagga
                    -160                -140                -120
 -178 gttcgtccttaatctgagccagcccttgtggaggccaaattccctctcctctccctttc
                    -100                 -80                 -60
 -118 tgggtgaggactcaggcttggagacactccaggataaacttctgcaggccaagagataac
                     -40                 -20                  +1
  -58 acgtcttCAAGGTgAcaatcagaaacgccttttaaaagttctccttctctgctgaaattg

   +3 ctgtagcttctccactccacagctggccatctgcctacacctggctgccatgtgggaact
```

3βHSDI

```
 -337 atcctccaataaatgcttacctggggttgcattgtaggtattccataaaataaattccac
 -277 catcagtaaatcaaagatatctagtcatctagtaagtattctattcttttgttgttgttt
 -217 gtttgtttctttagtatgttgtatctattgcttctttatacattgaaaataatagtgtaa
                    -140                -120                -100
 -157 gaagtgcacatgtccaaacaaaacagtctttatcacagtgTaACCTTGAagctggctttt
                     -80                 -60                 -40
  -97 catcctgtggcaggaaTaAAGGaCAtaaggtttatcaatcactGggAAGGaCAgaccaca
                     -20                  +1                  +20
  -37 gaaaagaataaatatggtcacgaagcaggcatgcGTCAAGaTatgataagtcagcagact
                     +40                 +60                 +80
  +24 agagattaaaacactagtcctgatctgagggctgaggagatcagcatccagacactctca
                    +100                +120                +140
  +84 tcTGACtTTtAacaatttaacaggttagattttatttattttagctcctcttgctgctg

 +144 ctgtg
```

Putative binding sites for the transcription factor SF-1 in the regulatory regions of mouse genes *Cyp17* and *3βHSDI*, recognized by homology with the consensus sequence. Putative sites differing from the consensus sequence by not more than three nucleotides are shown. Capital letters show nucleotides coinciding with the consensus sequence. Arrows show the direction of the transcription factor binding site (5′→3′). The curved arrow shows the position of transcription start site

right orientation in the proximal part of the promoter. In the mouse *Cyp17* gene a putative SF-1-binding site is located at −53/−44. The existence of SF-1 binding in the reversed orientation was experimentally recognized in the distal part of the promoter (−442/−411) of the rat gene (Table 2). In the mouse *Cyp17* the site positioned from −435 to −426 can be considered as homologous to that found in the rat gene. Three sites (−289/−274, −212/−197, and −152/−137) are essential for human *Cyp17* functioning (Table 2). In the mouse *Cyp17* gene two overlapping potential SF-1 binding sites −285/−276 and −279/−270 have position similar to the position of SF-1 binding site −289/−274 of human *Cyp17*.

The nucleotide sequence of mouse 3βHSDI contains five putative sites for SF-1 binding (figure). Three of them (−82/−73, −54/−45, −3/+7) are in the right orientation whereas others (−117/−108, +86/+95) are in the reversed orientation. Three of five putative binding sites are located at the 5′-flanking region, the fourth overlaps with start site of transcription, whereas the last one is positioned on the site corresponding to the first exon. Three putative SF-1 binding sites (−321/−313, −117/−109, −26/−18) were also found in the nucleotide sequence of *3βHSDI* [46]. However, since nucleotide sequences of the sites at −321/−313 and −26/−18 differ from our consensus sequence by more than three nucleotides we do not consider them as putative SF-1 binding sites. This difference from our results may be explained by employment of a rather distinct consensus sequences for the search for potential SF-1 binding sites by homology with them: CCAAGGTCA [47] and (C/T)CAAGGT(C/T)A [9]. In human *3βHSDII* and mouse *3βHSDI*, which are preferentially expressed in adrenals and gonads, the experimentally recognized functional site for SF-1 binding (S3992) is located at −68/−53 (Table 2). Putative SF-1 binding site (−54/−45) of mouse *3βHSDI* can be considered as its homolog.

According to the frequency matrix (Table 3), the pair of guanine nucleotides at the 6th and the 7th positions of the consensus sequence in the right orientations (which corresponds to the pair of cytosine nucleotides at the 4th and the 5th positions of the consensus sequence in the reversed orientation) are essential for SF-1 binding. Among putative SF-1 binding sites four sites of *Cyp17* (−505/−496, −435/−426, −285/−276, −53/−44) and three sites of *3βHSDI* (−117/−108, −82/−73, −54/−45) meet these criteria (figure).

Thus, positions of putative SF-1 binding sites in the analyzed nucleotide sequences of two genes encoding enzymes involved in biosynthesis of steroid hormones coincide with experimentally recognized positions of SF-1 binding sites in homologous genes of other mammalian species. Consequently, these putative SF-1 binding sites are promising for subsequent experimental identification.

As shown in Table 1, SF-1 binding sites were experimentally recognized in the regulatory regions of genes encoding all key enzymes involved in steroidogenesis in various vertebrate species (Cyp11A, 3βHSD, Cyp11B1, Cyp11B2). A binding site for SF-1 was also recognized in the regulatory regions of bovine, human, rat, mouse, and pig genes encoding so-called steroidogenic acute regulatory protein (StAR). This StAR protein is involved in cholesterol transfer from the outer mitochondrial membrane to the inner mitochondrial membrane; this is the rate-limiting stage in steroid hormone biosynthesis in gonads and adrenals [50]. The existence of experimentally demonstrated SF-1 binding sites in mouse genes encoding StAR and Cyp11A (Table 1) and the presence of putative binding sites for this factor in regulatory regions of mouse *Cyp17* and *3βHSDI* recognized in the present study can be considered as additional evidence for the key role of SF-1 in the regulation of genes controlling steroidogenesis. It is possible that the gene encoding SF-1 represent one of the predicted loci determining coordinate inherited variability of hormonal activity of Leydig cells in mice [37-39].

Thus, the consensus sequence of the SF-1 binding site, specified in the present study, and its use for the search of putative SF-1 binding sites represent a promising basis for subsequent experimental studies on identification of SF-1 binding sites in regulatory regions of the above mentioned and other genes, representing actual (and potential) target for transcriptional control by this factor.

The authors are grateful to N. A. Kolchanov and T. I. Merkulova for constructive criticism and useful discussion.

This work was supported by the Russian Foundation for Basic Research (grants No. 01-07-90376, 02-07-90359, 01-04-49523).

REFERENCES

1. Miller, W. L. (1988) *Endocr. Rev.*, **9**, 295-318.
2. Parker, K. L., and Schimmer, B. P. (1995) *Vitam. Horm.*, **51**, 339-370.
3. Simpson, E. R., Mahendroo, M. S., Means, G. D., Kilgore, M. W., Hinshelwood, M. M., Graham-Lorence, S., Amarneh, B., Ito, Y., Fisher, C. R., Michael, M. D., Mendelson, C. R., and Bulun, S. E. (1994) *Endocr. Rev.*, **15**, 342-355.
4. Simpson, E. R., Michael, M. D., Agarwal, V. R., Hinshelwood, M. M., Bulun, S. E., and Zhao, Y. (1997) *FASEB J.*, **11**, 29-36.
5. Slominski, A., and Wortsman, J. (2000) *Endocr. Rev.*, **21**, 457-487.
6. Stromstedt, M., and Waterman, M. R. (1995) *Mol. Brain Res.*, **34**, 75-88.
7. Walsh, S. W., Resko, J. I., Grumbach, M. M., and Novy, M. J. (1980) *Biol. Reprod.*, **23**, 264-270.
8. Omura, T., and Morohashi, K. J. (1995) *Steroid Biochem. Mol. Biol.*, **53**, 19-25.
9. Morohashi, K., Honda, S., Inomata, Y., Handa, H., and Omura, T. (1992) *J. Biol. Chem.*, **267**, 17913-17919.

10. Luo, X., Ikeda, Y., and Parker, K. L. (1994) *Cell*, **77**, 481-490.
11. Sadovsky, Y., Crawford, P. A., Woodson, K. G., Polish, J. A., Clements, M. A., Tourtellotte, L. M., Simburger, K., and Milbrandt, J. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 10939-11043.
12. Shinoda, K., Lei, H., Yoshii, H., Nomura, M., Nagano, M., Shiba, H., Sasaki, H., Osawa, Y., Ninomiya, Y., and Niwa, O. (1995) *Dev. Dyn.*, **204**, 22-29.
13. Honda, S., Morohashi, K., and Omura, T. (1990) *J. Biochem.*, **108**, 1042-1049.
14. Bakke, M., and Lund, J. (1995) *Mol. Endocrinol.*, **9**, 327-339.
15. Liu, Z., and Simpson, E. R. (1997) *Mol. Endocrinol.*, **11**, 127-137.
16. Watanabe, N., Inoue, H., and Fujii-Kuriyama, Y. (1994) *Eur. J. Biochem.*, **222**, 825-834.
17. Sugawara, T., Holt, J. A., Kiriakidou, M., and Strauss, J. F. (1996) *Biochemistry*, **35**, 9052-9059.
18. Sugawara, T., Kiriakidou, M., McAllister, J. M., Holt, J. A., Arakane, F., and Strauss, J. F. (1997) *Steroids*, **62**, 5-9.
19. Sugawara, T., Kiriakidou, M., McAllister, J. M., Kallen, C. B., and Strauss, J. F. (1997) *Biochemistry*, **36**, 7249-7255.
20. Sugawara, T., Saito, M., and Fujimoto, S. (2000) *Endocrinol.*, **141**, 2895-2903.
21. Michael, M. D., Kilgore, M. W., Morohashi, K., and Simpson, E. R. (1995) *J. Biol. Chem.*, **270**, 13561-13566.
22. Rice, D. A., Kirkman, M. S., Aitken, L. D., Mouw, A. R., Schimmer, B. P., and Parker, K. L. (1990) *J. Biol. Chem.*, **265**, 11713-11720.
23. Kolchanov, N. A., Podkolodnaya, O. A., Ananko, E. A., Ignatieva, E. V., Podkolodny, N. L., Merkulov, V. M., Stepanenko, I. L., Pozdnyakov, M. A., Belova, O. E., Grigorovich, D. A., and Naumochkin, A. N. (2001) *Mol. Biol. (Moscow)*, **35**, 934-942.
24. Kolchanov, N. A., Ignatieva, E. V., Ananko, E. A., Podkolodnaya, O. A., Stepanenko, I. L., Merkulova, T. I., Pozdnyakov, M. A., Podkolodny, N. L., Naumochkin, A. N., and Romashchenko, A. G. (2002) *Nucleic Acids Res.*, **30**, 312-317.
25. Ignatieva, E. V., Busygina, T. V., Ananko, E. A., Podkolodnaya, O. A., Merkulova, T. I., Suslov, V. V., and Pozdnyakov, M. (2000) *Proc. II Int. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2000)*, Novosibirsk, Russia, Vol. 1, pp. 22-25.
26. Busygina, T. V., Ignatieva, E. V., and Osadchuk, A. V. (2000) *Proc. II Int. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2000)*, Novosibirsk, Russia, Vol. 1, pp. 41-44.
27. Osadchuk, A. V., and Naumenko, E. V. (1981) *Dokl. Akad. Nauk SSSR*, **258**, 746-749.
28. Osadchuk, A. V., and Naumenko, E. V. (1983) *Dokl. Akad. Nauk SSSR*, **268**, 983-987.
29. Osadchuk, A. V. (1990) *Ontogenetic and Genetic-Evolutional Aspects of Neuroendocrinal Regulation of Stress* [in Russian], Nauka, Novosibirsk.
30. Naumenko, E. V., Osadchuk, A. V., Serova, L. I., and Shishkina, G. T. (1983) *Genetic-Physiological Mechanisms of Regulation of Testicular Functions* [in Russian], Nauka, Novosibirsk.
31. Busygina, T. V., and Osadchuk, A. V. (2001) *Genetika*, **37**, 649-656.
32. Busygina, T. V., and Osadchuk, A. V. (2001) *Genetika*, **37**, 97-106.
33. Osadchuk, A. V., and Svechnikov, K. V. (1994) *Byull. Eksp. Biol. Med.*, **118**, 177-180.
34. Osadchuk, A. V., and Svechnikov, K. V. (1995) *Dokl. Ros. Akad. Nauk*, **343**, 281-283.
35. Saez, J. M. (1994) *Endocr. Rev.*, **15**, 574-626.
36. Osadchuk, A. V., and Svechnikov, K. V. (1998) *Genetika*, **34**, 1277-1285.
37. Osadchuk, A. V., Svechnikov, K. V., and Ahmerova, L. G. (1999) *13th Int. Mouse Genome Conf.*, October 31-November 3, Philadelphia, PA, USA, Abstract E24.
38. Osadchuk, A. V., Svechnikov, K. V., Ahmerova, L. G., Kozlova, O. N., and Huhtaniemi, I. (2001) *15th Int. Mouse Genome Conf.*, Edinburgh, Scotland, 79.
39. Osadchuk, A. V., Svechnikov, K. V., Ahmerova, L. G., Kozlova, O. N., and Huhtaniemi, I. (2002) *12th Europ. Workshop on Molecular and Cellular Endocrinology of the Testis*, Doorwerth, Nehterlands, 4a-6.
40. Ingraham, H. A., Lala, D. S., Ikeda, Y., Luo, X., Shen, W. H., Nachtigal, M. W., Abbud, R., Nilson, J. H., and Parker, K. L. (1994) *Genes Dev.*, **89**, 2302-2312.
41. Parker, K. L., and Schimmer, B. P. (1997) *Endocr. Rev.*, **18**, 361-377.
42. Caron, K. M., Clark, B. J., Ikeda, Y., and Parker, K. L. (1997) *Steroids*, **62**, 53-56.
43. Wong, M., Ikeda, Y., Luo, X., Caron, K. M., Weber, T. J., Swain, A., Schimmer, B. P., and Parker, K. L. (1997) *Recent Prog. Horm. Res.*, **52**, 167-182.
44. Rice, D. A., Mouw, A. R., Bogerd, A. M., and Parker, K. L. (1991) *Mol. Endocrinol.*, **5**, 1552-1562.
45. Lawrence, Ch. E., Altshul, S. F., Boguski, M. S., Liu, S. L., Neuwald, A. F., and Wootton, J. C. (1993) *Science*, **262**, 208-214.
46. Clarke, T. R., Bain, P. A., Burmeister, M., and Payne, A. H. (1996) *DNA Cell. Biol.*, **15**, 387-399.
47. Shen, W. H., Moore, C. C., Ikeda, Y., Parker, K. L., and Ingraham, H. A. (1994) *Cell*, **77**, 651-661.
48. Mangelsdorf, D. J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P., et al. (1995) *Cell*, **83**, 835-839.
49. Tsai, S. Y., and Tsai, M. J. (1997) *Endocr. Rev.*, **18**, 229-240.
50. Strauss, J. F., Kallen, C. B., Christenson, L. K., Watari, H., Devoto, L., Arakane, F., Kiriakidou, M., and Sugawara, T. (1999) *Prog. Horm. Res.*, **54**, 369-394.