Volume 15 Number 6 1987 Nucleic Acids Research

Molecular cloning and characterization of rat estrogen receptor cDNA

Satoshi Koike, Masaharu Sakai and Masami Muramatsu

Department of Biochemistry, Faculty of Medicine, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113, Japan

Received December 29, 1986; Revised and Accepted February 23, 1987 Accession no. Y00102

ABSTRACT

A cDNA clone of rat uterus estrogen receptor (ER) has been isolated and sequenced. This clone contains a complete open reading frame encoding 600 amino acid residues which is 5 and 11 amino acids larger than the corresponding molecules of human and chicken, respectively. The molecular weight of this protein is calculated to be 67,029. When this clone was ligated to the pSV2 vector and transfected into COS7 cells, a protein was produced that had the same affinity to estrogen as rat uterus ER. This sequence shows 88% homology with human ER; 528 amino acids are identical and 14 amino acids are conservative substitutions. The comparison of rat, human and chicken ER sequences indicate the presence of three highly conserved regions suggesting that these regions play important roles in ER function. The putative DNA-binding domain is completely identical in rat, human and chicken. The C-terminal half region which is thought to be the estrogen binding domain is also highly conserved and is rich in hydrophobic amino acid residues.

Southern blot analysis of genomic DNA with ER cDNA as a probe has shown that related sequences are present in the genome.

INTRODUCTION

Steroid hormone action is mediated through its specific receptor. There is also evidence that the receptor is a transcriptional regulatory protein which, after binding with hormone, alter the transcriptional efficiency of hormone-inducible genes (1). This regulation involves the binding of the steroid with the receptor molecule, conversion of the latter to an activated form and its interaction with DNA. Experimental results from this and other laboratories suggest that the receptor molecules consist of the separate functional domains (2-5).

Recently, cDNA clones of human and chicken ER (6,7,8),

human and rat glucocorticoid receptor (GR) (9, 10) and chicken progesterone receptor (PR) (11, 12) were isolated and sequenced. The domain structures of these steroid receptors were proposed from the primary structures deduced from the cDNA sequences. Sequence homologies between these steroid receptors and v-erbA protein of oncogenic avian erythroblastosis virus were also demonstrated. The presence of similar cystein-containing motifs in these proteins suggests that they belong to a new multigene family of transcriptional regulatory proteins.

It is crucial to study the structures of the proteins of this new gene family and the molecular mechanisms of activation of gene expression by these proteins. On the other hand, more information on ER structures from different species is required to assess more precisely the functional as well as evolutionary significance of each domain. We report here the isolation and sequencing of a rat uterus ER cDNA. The presence of highly conserved regions in human, rat and chicken ER was demonstrated, though there were some differences even between human and rat ER. The relationship between the highly conserved regions and the putative functional domains of ER are discussed.

MATERIALS AND METHODS

Preparation of rat uterus cDNA library

Rat (Wistar strain) uteri were immedately frozen after extirpation and powdered with dry ice using a Waring blender. Powdered uterine tissue was homogenized in 0.1M Tris-HCl (pH 9.0), 0.1M NaCl, 1mM EDTA, 13mM dithiothreitol and the homogenate was then made to 1% with respect to sodium dodecyl sulfate (SDS). Total nucleic acids were extracted five times with phenol-$CHCl_3$-isoamylalchol (50:50:2) and then RNA was precipitated with 2M LiCl at 0°C. Poly(A)$^+$ RNA was selected by oligo(dT)-cellulose column chromatography (13) and was fractionated by a 5-30% linear sucrose density gradient centrifugation. Fractions of about 30S were collected and used for cDNA synthesis. cDNA was synthesized by the method of Gubler and Hoffman (14). Oligo$(dT)_{10-12}$ or random hexamer (Pharmacia) was used as a primer. EcoRI sites in cDNA were methylated by EcoRI methylase (Boehringer) and the cDNA was ligated to the EcoRI linker (TAKARA). It was digested with

EcoRI and fractionated by gel filtration on a Sepharose CL-4B column. Fractions longer than 1kb were pooled, ligated to λgt10 vector and packaged in vitro.

Screening procedures

The phages were plated onto E.coli C600 hfl. DNA probes corresponding to the three regions of human ER (positions 9-27, 220-245 and 526-543 of human ER published in Ref. 6) were synthesized using a 381A DNA Synthesizer (Applied Biosystems). Duplicate filters were hybridized with the above probes in 1M NaCl, 50mM Tris-HCl (pH 7.5), 10mM EDTA, 0.1% sodium N-lauroyl sarcosinate, 0.2% bovine serum albumin, 0.2% Ficoll 400, 0.1% polyvinylpyrrolidone at 65°C overnight and then washed twice with 3xSSC (1xSSC: 0.15M NaCl, 0.015M Na-Citrate) containing 0.1% SDS at 65°C for 30 min. each.

Construction of pSV2RcER and its expression in COS7 cells

Rat ER expression plasmid, pSV2RcER, was constructed as follows. The β-globin cDNA fragment of pSV2-βG was removed with HindIII and BglII digestion (15). The pRcER6 insert fragment was ligated to this vector using HindIII and BamHI linkers. pSV2-βG plasmid without β-globin insert was used as a control plasmid. Plasmid DNAs were transfected into COS7 cells (9 plates, $1x10^6$ cells/plate) by Ca-phosphate precipitation method (16). After 48h, the cells were harvested and cytosol fraction was prepared. The estradiol binding activity was measured by the dextran-coated charcoal method (17). The dissociation constant was determined by the method of Scatchard (18).

Southern blot analysis of ER gene

Rat genomic DNA was digested with either EcoRI, HindIII or BamHI. Digested DNA fragments were resolved by 0.6% agarose gels, transferred to a nitrocellulose filter and hybridized to the probes (19). Probes used were a 2.1kb insert of pRcER6, a 221bp Sau3AI fragment containing 5'-untranslated region and the N-terminal region of ER (N-terminal probe), a 369bp Sau3AI fragment encoding DNA-binding domain (DNA-binding domain probe), and a 405bp Sau3AI-EcoRI fragment containing the C-terminal region including a part of the 3'-untranslated region (C-terminal probe), as indicated in Figure 6.

RESULTS

Cloning of rat uterus ER cDNA

Northern blot hybridization (46) of rat uterus ploy(A)$^+$ RNA with synthetic DNA probes revealed a single band of about 6kb. Therefore, we fractionated the total poly(A)$^+$ RNA by the sucrose density gradient centrifugation and the fractions centering this region were used as the template for cDNA synthesis.

0.9×10^5 plaques from oligo(dT)-primed cDNA library and 2.1×10^5 plaques from random-primed cDNA library were screened with the three probes described in **MATERIALS AND METHODS.** Three clones from the oligo(dT)-primed cDNA library and 12 from the random-primed cDNA library were positive by at least one of the three probes. One randomly primed clone designated λRcER6 was positive by all three probes. All positive clones hybridized with the insert of λRcER6 under the stringent conditions.

The insert of this clone was subcloned into EcoRI site of pBR322 (pRcER6). Restriction map was determined and the fragments of this insert was further subcloned into M13 mp18 or mp19 (20) and sequenced by the dideoxy method (21). The sequences which showed some difficulty in read-outs of bands, presumably due to the secondary structure caused by the high G-C content, were determined by using 7-deaza-dGTP (22).

The sequence of rat ER cDNA

The nucleotide sequence and the deduced amino acid sequence of pRcER6 insert are shown in Figure 1. This clone contained a long open reading frame encoding 600 amino acid residues, together with a 210 nucleotide 5'-untranslated region and a 74 nucleotide 3'-untranslated region. This open reading frame is the longest and the DNA sequence surrounding the ATG codon is in agreement with the consensus sequence proposed by Kozak for translation initiation region (23), while others do not comply with it. The molecular weight calculated from the deduced amino acid sequence is 67,029 daltons. This value is compatible with the molecular weight of rat ER purified by affinity chromatography (24, 25) or with that labeled with (^{3}H)-tamoxifen aziridine (26).

A

```
                                                                    gaattcggCCACACACGCTCTGCCTTGATCACACACCGCGCCACTCGATCATTCGAGCACATTCCTTCCTTCCGTCTTACTGTCTCAGCTCTTGACTT

CTACAAACCCATGGAACATTTCTGGAAAGACGCTTTTGAACCAGCAGGGTGGCTCATCCGCTGCTGAGCCCTCTGCGTGCGCGGGAAGCCAATCTGTACCTCGGCGGCTGCCACTGACC

ATG ACC ATG ACC CTT CAC ACC AAA GCC TCG GGA ATG GCC TTG TTG CAC CAG ATC CAA GGG AAC GAG CTG GAG CCC CTC AAT CGC CCG CAG
Met Thr Met Thr Leu His Thr Lys Ala Ser Gly Met Ala Leu Leu His Gln Ile Gln Gly Asn Glu Leu Glu Pro Leu Asn Arg Pro Gln  30

CTC AAG ATG CCC ATG GAG AGG GCT CTG GGC GAG GTG TAC GTG GAC AAC AGC AAG CCC GCC GTG TTC AAC TAC CCC GAG GGC GCC GCC TAC
Leu Lys Met Pro Met Glu Arg Ala Leu Gly Glu Val Tyr Val Asp Asn Ser Lys Pro Ala Val Phe Asn Tyr Pro Glu Gly Ala Ala Tyr  60

GAG TTC AAC GCC GCC GCC GCC GCC GCG GCC GCC GGG GCC TCG GCT CCG GTC TAT GGC CAG TCG AGC ATC ACT TAC GGT CCG GGG TCC GAG
Glu Phe Asn Ala Ala Ala Ala Ala Ala Ala Ala Gly Ala Ser Ala Pro Val Tyr Gly Gln Ser Ser Ile Thr Tyr Gly Pro Gly Ser Glu  90

GCG GCC GCC TTT GGT GCC AAT AGT CTG GGG GCT TTC CCC CAG CTC AAC AGC GTG TCG CCC AGT CCG CTG ATG CTG CTG CAC CCG CCG CCG
Ala Ala Ala Phe Gly Ala Asn Ser Leu Gly Ala Phe Pro Gln Leu Asn Ser Val Ser Pro Ser Pro Leu Met Leu Leu His Pro Pro Pro 120

CAC GTG TCG CCG TTC CTG CAC CCG CAT GGC CAC CAG GTG CCC TAC TAC CTG GAG AAC GAG CCC AGC GCC TAC GCT GTA CGC GAC ACC GGC
His Val Ser Pro Phe Leu His Pro His Gly His Gln Val Pro Tyr Tyr Leu Glu Asn Glu Pro Ser Ala Tyr Ala Val Arg Asp Thr Gly 150

CCT CCC GCC TTC TAC AGG TCC AAT TCT GAC AAT CGA CGC CAG AAT GGC CGA GAG AGA CTC TCC AGC AGC AGC GAG AAG GGA AAC ATG ATC
Pro Pro Ala Phe Tyr Arg Ser Asn Ser Asp Asn Arg Arg Gln Asn Gly Arg Glu Arg Leu Ser Ser Ser Ser Glu Lys Gly Asn Met Ile 180

ATG GAG TCT GCC AAG GAG ACT CGC TAC TGT GCT GTG TGT AAC GAC TAT GCC TCT GGC TAC CAT TAT GGG GTC TGG TCC TGT GAA GGC TGC
Met Glu Ser Ala Lys Glu Thr Arg Tyr Cys Ala Val Cys Asn Asp Tyr Ala Ser Gly Tyr His Tyr Gly Val Trp Ser Cys Glu Gly Cys 210

AAG GCT TTC TTT AAG AGA AGC ATT CAA GGA CAT AAT GAC TAC ATG TGT CCA GCT ACA AAC CAA TGC ACC ATC GAT AAG AAC CGG AGG AAG
Lys Ala Phe Phe Lys Arg Ser Ile Gln Gly His Asn Asp Tyr Met Cys Pro Ala Thr Asn Gln Cys Thr Ile Asp Lys Asn Arg Arg Lys 240

AGT TGC CAG GCC TGC CGG CTG CGC AAG TGT TAC GAA GTG GGC ATG ATG AAA GGC GGG ATA CGA AAA GAC CGC CGA GGA GGG AGA ATG TTG
Ser Cys Gln Ala Cys Arg Leu Arg Lys Cys Tyr Glu Val Gly Met Met Lys Gly Gly Ile Arg Lys Asp Arg Arg Gly Gly Arg Met Leu 270

AAG CAC AAG CGT CAG AGA GAT GAC TTG GAA GGC CGA AAT GAA ATG GGC ACT TCA GGA GAC ATG AGA GCT GCC AAC CTT TGG CCA AGT CCA
Lys His Lys Arg Gln Arg Asp Asp Leu Glu Gly Arg Asn Glu Met Gly Thr Ser Gly Asp Met Arg Ala Ala Asn Leu Trp Pro Ser Pro 300

CTT GTG ATC AAG CAC ACT AAG AAG AAT AGC CCC GCC CTG TCC TTG ACA GCT GAC CAG ATG GTC AGT GCC TTA TTG GAT GCT GAA CCA CCT
Leu Val Ile Lys His Thr Lys Lys Asn Ser Pro Ala Leu Ser Leu Thr Ala Asp Gln Met Val Ser Ala Leu Leu Asp Ala Glu Pro Pro 330

TTG ATC TAT TCT GAA TAT GAT CCT TCT AGA CCC TTC AGT GAA GCC TCA ATG ATG GGC TTA TTG ACC AAC CTG GCA GAC AGG GAG CTG GTT
Leu Ile Tyr Ser Glu Tyr Asp Pro Ser Arg Pro Phe Ser Glu Ala Ser Met Met Gly Leu Leu Thr Asn Leu Ala Asp Arg Glu Leu Val 360

CAT ATG ATC AAC TGG GCA AAG AGA GTG CCA GGC TTT GGG GAC TTG AAT CTC CAC GAT CAA GTT CAC CTT CTG GAG TGT GCC TGG TTG GAG
His Met Ile Asn Trp Ala Lys Arg Val Pro Gly Phe Gly Asp Leu Asn Leu His Asp Gln Val His Leu Leu Glu Cys Ala Trp Leu Glu 390

ATC CTG ATG ATT GGT CTG GTC TGG CGC TCC ATG GAA CAC CCG GGG AAG CTC CTG TTT GCT CCT AAC TTG CTC TTG GAC AGG AAT CAA GGT
Ile Leu Met Ile Gly Leu Val Trp Arg Ser Met Glu His Pro Gly Lys Leu Leu Phe Ala Pro Asn Leu Leu Asp Arg Asn Gln Gly 420

AAA TGT GTA GAA GGC ATG GTG GAG ATC TTT GAC ATG TTG CTG GCT ACG TCA AGT CGA TTC CGC ATG ATG AAC CTG CAG GGA GAA GAG TTT
Lys Cys Val Glu Gly Met Val Glu Ile Phe Asp Met Leu Leu Ala Thr Ser Ser Arg Phe Arg Met Met Asn Leu Gln Gly Glu Glu Phe 450

GTG TGC CTC AAA TCA ATC ATT TTG CTT AAT TCT GGA GTG TAC ACA TTT CTA TCC AGC ACC TTG AAG TCT CTG GAA GAG AAG GAC CAC ATC
Val Cys Leu Lys Ser Ile Ile Leu Leu Asn Ser Gly Val Tyr Thr Phe Leu Ser Ser Thr Leu Lys Ser Leu Glu Glu Lys Asp His Ile 480

CAC CGA GTC CTG GAC AAG ATC AAC GAC ACT TTG ATC CAC TTG ATG GCC AAA GCT GGC CTG ACT CTG CAG CAA CAG CAT CGC CGT CTG GCC
His Arg Val Leu Asp Lys Ile Asn Asp Thr Leu Ile His Leu Met Ala Lys Ala Gly Leu Thr Leu Gln Gln Gln His Arg Arg Leu Ala 510

CAG CTC CTC CTC ATC CTT TCC CAT ATC CGG CAC ATG AGT AAC AAA GGC ATG GAG CAT CTC TAC AAC ATG AAA TGC AAG AAT GTC GTG CCT
Gln Leu Leu Leu Ile Leu Ser His Ile Arg His Met Ser Asn Lys Gly Met Glu His Leu Tyr Asn Met Lys Cys Lys Asn Val Val Pro 540

CTC TAT GAC CTG CTG CTG GAG ATG CTG GAT GCT CAT CGT CTT CAT GCC CCC GCC AGT CGC ATG GGA GTG CCC CCG GAG GAG CCT AGC CAG
Leu Tyr Asp Leu Leu Leu Glu Met Leu Asp Ala His Arg Leu His Ala Pro Ala Ser Arg Met Gly Val Pro Pro Glu Glu Pro Ser Gln 570

AGC CAG CTG ACC ACC ACC AGC TCC ACT TCA GCA CAT TCC TTA CAA ACC TAC TAC ATC CCC CCG GAA GCA GAG GGC TTC CCC AAC ACC ATC
Ser Gln Leu Thr Thr Thr Ser Ser Thr Ser Ala His Ser Leu Gln Thr Tyr Tyr Ile Pro Pro Glu Ala Glu Gly Phe Pro Asn Thr Ile 600

TGA GAACTCCCAGGCTCCCCACAAGGTTCTGCGAATCCCTGAAACGTTTTACCCATGTCCTGTATGACTTTAGCAGAATTC
***
```

B

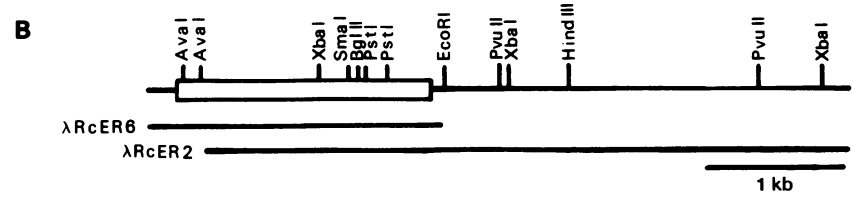

Fig. 1 (A) The nucleotide and deduced amino acid sequence of rat uterus estrogen receptor.
The insert of pRcER6 were further subcloned into appropriate sites of M13 mp18 or mp19 and sequenced (20). Small letters indicate the EcoRI linker sequence. (B) Restriction map of rat ER cDNA clones.

Fig. 2 Northern blot hybridization of rat ER mRNA.
Rat uterus poly(A)$^+$ RNA was electrophoresed on 1.0% agarose gel, transferred onto nitrocellulose filter and hybridized to the insert of pRcER6 (46).

We assume that the EcoRI site at the 3'-terminus of this cDNA is not the linker sequence, but is, in fact, present in the original ER mRNA sequence. Perhaps, it was not methylated during cDNA methylation and digested after linker ligation. We also obtained cDNA clones having the downstream region of this EcoRI site, one extending up to 3kb (Figure 1B). Northern blot analysis of rat uterus poly(A)$^+$ RNA showed a clear band of about 6kb (Figure 2). These results together indicate that rat ER mRNA has a very long 3'-untranslated region just like those of human ER and GR.

Sequence homology and domain structure of ER

The regions which are highly conserved are thought to have some important functions. Comparison of rat, human and chicken

A

```
human                                  I L  P     L S     Y                ----- N Q      75
rat     MTMTLHTKASGMALLHQIQGNELEPLNRPQLKMPMERALGEVYVDNSKPAVFNYPEGAAYEFNAAAAAAAAGASAPVYGQ  80
chicken            VT       T   T S     I L  S SDM  ESN TG        T D --------- TT     S  71

human   TGLP           S G  G  P                QL    Q   Q           G T  EA       P    155
rat     SSITYGPGSEAAAFGANSLGAFPQLNSVSPSPLMLLHPPPHVSPFLHPHGHQVPYYLENEPSAYAVRDTGPPAFYRSNSD 160
chicken TTLS A T  --S  SS  AG HS  N P   VVF QTA QL   I H SQ          QGSFGM EAA      PS   149

human      G     A TND  S A                                                            235
rat     NRRQNGRERLSSSSEKGNMIMESAKETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRK 240
chicken    HSI   M  TN   SLS                                                            229

human                                        G   G V SA             M  RS    L          315
rat     SCQACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDDLEGRNEMGTSGDMRAANLWPSPLVIKHTKKNSPALSLTADQM 320
chicken                                E M Q    EEQDS  GEAS TEL  PT  T    V  N           E   309

human            IL      T                                      V  T                      395
rat     VSALLDAEPPLIYSEYDPSRPFSEASMMGLLTNLADRELVHMINWAKRVPGFGDLNLHDQVHLLECAWLEILMIGLVWRS 400
chicken          IV      N   N     T                          V  T                      389

human       V                                                                            475
rat     MEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHI 480
chicken                                      AA                                    R Y  469

human          T                 Q                        S                        T   555
rat     HRVLDKINDTLIHLMAKAGLTLQQQHRRLAQLLLILSHIRHMSNKGMEHLYNMKCKNVVPLYDLLLEMLDAHRLHAPASR 560
chicken        T       LS  S                                                        A   549

human   G ASV  TD  H A AG   S    K   -TG      A V 595
rat     MGVPPEEPSQSQLTTTSSTSAHSLQTYYI-PPEAEGFPNTI 600
chicken SAA M  ENRN  -  APA S    SF  NSK E SMQ    589
```

B

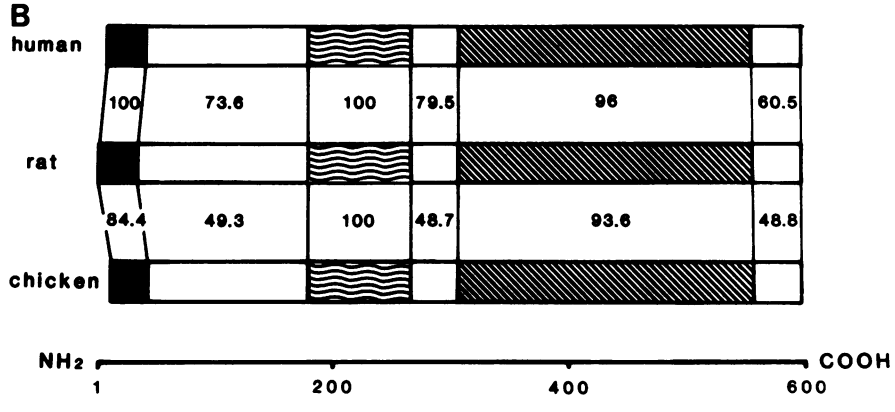

Fig. 3 (A) Comparison of amino acid sequence of rat, human (6, 7) and chicken (8) ER. For human and chicken sequences, only the amino acid residues different from those of rat are shown. (B) The domain structure of the conserved regions are shown schematically. Numbers among the boxes show the percentages of amino acid homology.
: putative DNA binding domain, : putative estrogen binding domain.

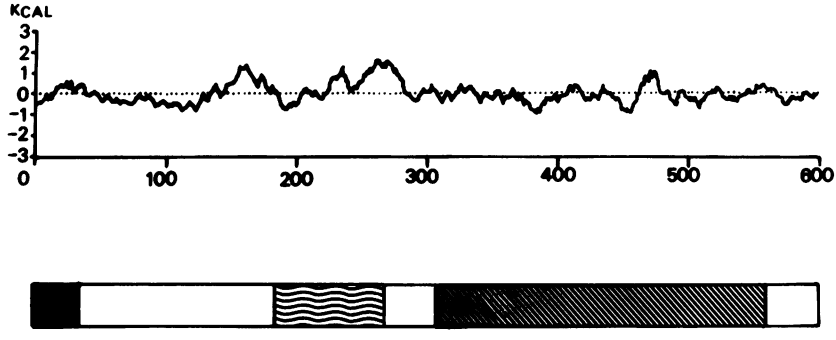

Fig. 4 Hydrophilicity value of rat ER.
Hydrophilicity value was calculated with the program of DNASIS system of Hitachi Software Engineering Co., Ltd. Abscissa represents the amino acid position from the N-terminus. Positive and negetive values on the ordinate indicate hydrophiclic and hydrophobic degrees, respectively.
: putative DNA binding domain, : putative estrogen binding domain.

ER sequences will help elucidate important functional domains of ER, as shown in Figure 3. Between rat and human ER, there are 528 identical and 67 different amino acids at corresponding positions, 14 of which are conserved substitutions when the chemical similarities of amino acids are grouped as follows: (S,T), (D,E), (N,Q), (K,R), (V,I,L), (F,Y,W). The overall homology is 88%. There is an insertion in rat ER of 5 amino acids which is considered to be caused by insertion of a GCC repeat followed by some base substitutions between positions 69 and 70 of human ER. This insertion makes rat ER 5 amino acids longer than human ER. Rat and chicken ER shows overall 77% amino acid homology. Chicken ER has 12 deletions and a single insertion compared with rat ER. Most of the changed amino acids are clustered in three restricted regions; 33-180, 268-306 and 558-600.

Counting of amino acid substitutions among the three species shows that the evolutionary distance of these species is well reflected in ER molecules. The numbers of substituted amino acids between human-rat, human-chicken and rat-chicken are 68, 126 and 130, respectively. The numbers of substitutions in chicken as compared with human and rat is consistent with the evolutionary distance between these species; i.e. 160

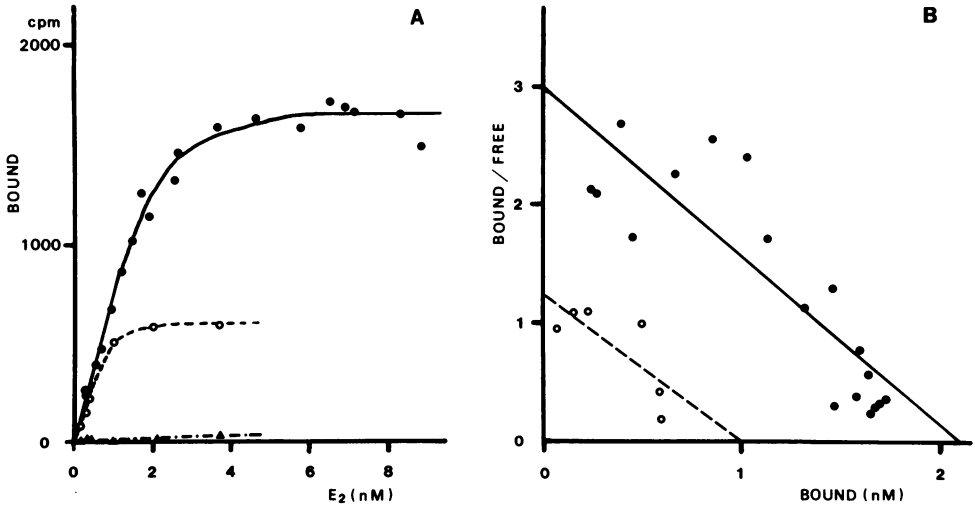

Fig. 5 Expression of rat ER cDNA in COS cells.
(A) Estrogen binding acitivities were determined for cytosols of rat uterus (●——●), COS cells transfected with pSV2RcER (O-----O) and those transfected with pSV2 vector (▲-·-·-▲) at various concentrations of (^{3}H)-estradiol (102Ci/mmol). The concentration of estradiol binding protein expressed in the COS cells was approximately 100fmol/mg. (B) Scatchard plot (18) of ER expressed in COS cells.

million years between human and rat and 600 million years between chicken and mammals.

The N-terminal region of rat ER, 1 to 32, are 100% and 84.4% homologous with human and chicken ER, respectively.

The putative DNA-binding domain (the region 181 to 267 of rat ER) is 100% identical in rat, human and chicken. This suggests that the conservation of this region is essential for ER function. This region is rich in basic amino acids and have characteristic cystein-containing motif sequences. This region is hydrophilic as shown in Figure 4 and presumably, is exposed outside of the protein and interacts with DNA.

The other highly conserved region is in the C-terminal half region (307 to 557). In this region, 96% amino acids are identical with human ER and 93.6% with chicken ER. This region is believed to be the estrogen binding domain, although the exact estrogen binding site is not known at present. The estrogen binding domain is hydrophobic throughout the region

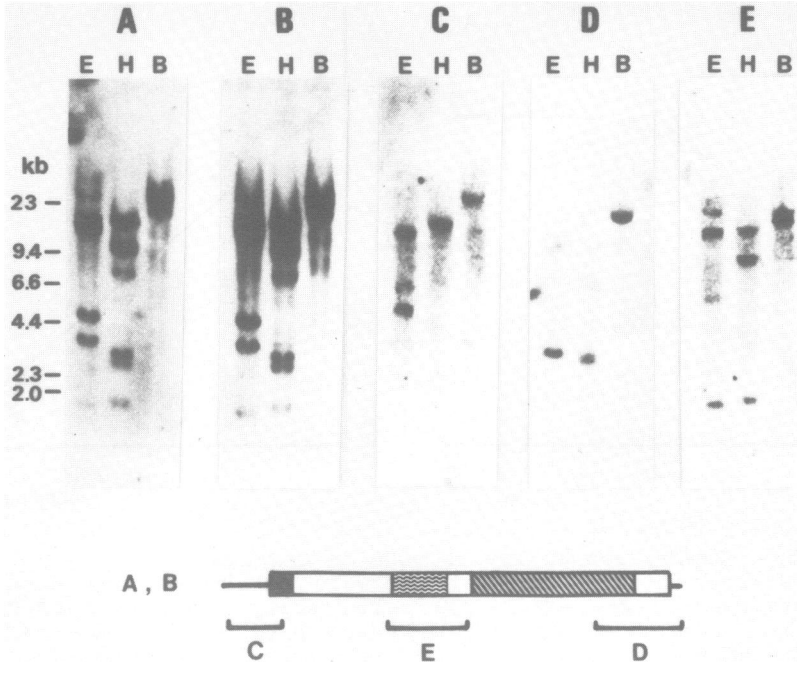

Fig. 6 Southern blot analysis of ER gene.
Rat genomic DNA was digested with restriction enzymes EcoRI (E), HindIII (H) or BamHI (B) and hybridized to the insert of pRcER6 (A), N-terminal probe (C), C-terminal probe (D) or DNA-binding domain probe (E) as described in **MATERIALS AND METHODS** and washed with 0.2xSSC containing 0.1% SDS at 65°C for 30 min twice. Panel (B) is identical to panel (A) except that the washing was done with 3xSSC.

(Figure 4) possibly to make a hydrophobic pocket (5, 6, 8) that facilitates the interaction with estrogen with such a high specificity and affinity.

Expression of rat ER cDNA in COS7 cells

The insert of pRcER6 was expressed in COS7 cells using pSV2 vector. Fourty-eight hours after transfection, the cytosol fraction contained a protein which had a high affinity to estrogen, whereas the COS cells which were transfected with control plasmid did not produce such a molecule (Figure 5). The dissociation constant of this protein (0.8nM) was almost the same as that of authentic rat uterus ER (0.7nM). This

confirms that the pRcER6 insert encodes the functional rat uterus ER.

Southern blot analysis of rat ER gene

Southern blot hybridization with the total insert of pRcER6 probe showed a very complex pattern (Figure 6A). We probed with shorter fragments containing the positions described in **MATERIALS AND METHODS.** The patterns were simple when probed with N- or C-terminal regions of ER (Figure 6C and D). This suggests that the ER gene is present probably as a single copy or at most a few copies in the genome. However, when the fragment of DNA-binding domain was used as probe, the pattern became complex (Figure 6E). We do not know whether this reflects the complex structure of ER gene in this region, possibly consisting of many exons and introns, or the presence of related sequences in the genome. Additional bands were detected when filters were washed under relaxed conditions (Figure 6B). This indicates that some related sequences are present in rat genome.

DISCUSSION

We have described here the isolation and characterization of a rat uterus ER cDNA. We have also shown that the cDNA can synthesize a protein which binds with estrogen at the same high affinity as rat ER, indicating that the message does produce a functional receptor. The comparison of this sequence with those of human and chicken ER disclosed several interesting aspects. Although the three sequences are derived from different species, they showed a high degree of homology with each other. The data of rat ER confirm the conclusion by Krust et al. (8) that the structures of ER are highly conserved during evolution. Indeed, Klein-Hitpass et al. (27) have shown that the human ER can activate the transcription of amphibian vitellogenin-CAT fusion gene. Three ER sequences are so far identified in different tissues, including one in a mammary tumor cell line, MCF-7. Since their structures are so similar that ERs present in different tissues may be identical. The very few number of ER gene, as shown by southern blot analysis, also argues for a very limited number of ER species.

Tissue-specific expression of estrogen-inducible genes, for example, the liver-specific expression of vitellogenin gene, may be due to some unknown factors rather than the difference of ER (28, 29).

The whole DNA-binding domain is the most conservative, 100% identical in these three species. This was anticipated when chicken ER was compared with that of humans (8). This suggests that the enormous selective pressure is imposed on this region during evolution. Presumably, a number of molecular criteria have to met with for the ER to interact correctly with cognate regulatory sequences. The DNA sequences of the binding sites for ER, GR and PR have been reported (30-35). Walker et al. (36) have found a consensus sequence in the 5'-flanking regions of estrogen regulated genes. This region of Xenopus vitellogenin A2 gene turned out to be essential for the estrogen dependent expression of vitellogenin-CAT fusion gene transfected in MCF-7 cells (27). The conserved region in the DNA-binding domain may play an important role in recognizing these specific sites on DNA. Kumar et al. (37) found that the ERs having a deletion in this conserved region could not be retained in the nuclear fraction even in the presence of estrogen.

We have previously shown the presence of a protease-resistant region in porcine ER which contains estrogen binding sites (5). We have also found that this protease-resistant region is required to maintain estrogen binding activity, arguing that the conformation of this entire region may be important (5). The size of the carboxy-side conserved region (250 amino acid residues) is almost the same as that of the protease-resistant region (27-30kd). Probably, the reason for the conservation of this region is to keep a conformation such that a hydrophobic pocket is formed for a selective and effecient binding of estrogen (5, 6, 8). Kumar et al. (37) recently proved by deletion mutants of ER that this entire region is required for ER to bind estrogen.

The N-terminal region (1-32) is also conserved, although the function of this region is totally unknown. Recently, Giguere et al. (38) have constructed a series of mutant GR

proteins having short amino acids insertions. They identified 4 domains that were needed for GR function, including steroid- and DNA-binding domains. The functions of the remaining two regions are not known. Although the N-terminal conserved region of ER might be the counterpart of the corresponding domain of the GR, no sequence homology was detected in these regions.

The DNA-binding domain is characterized by specific cystein-containing sequences and a high content of basic amino acids. This motif is found in other steroid hormone receptors (6-12) and also in v-erbA protein (39). This suggests that the steroid hormone receptors and erbA protein belong to a multigene family. Similar motifs were found in Xenopus transcription factor of 5S gene, TFIIIA (40), yeast regulatory protein GAL4 (41), PPRI (42), ADRI (43), Drosophila Kruppel (44) and Serendipity gene products (45). There are small differences among the motifs of these proteins: substitution of cystein by histidine, the distance between the cystein or histidine residues etc. The binding of metal ion (Zn^{2+}) has only been demonstrated for TFIIIA (40), but similar function is inferred from the amino acid sequences for other proteins. These proteins, altogether, may belong to a larger, so to speak, supermultigene family of regulatory proteins. From this point of view, the steroid hormone receptors may be categorized as one of the transcriptional regulatory proteins which acquired the steroid binding capacity during evolution, and the binding of steroid hormone becoming the 'switch' for the regulation.

When genomic DNA was probed with ER cDNA under relaxed conditions, additional bands appeared indicating that some ER-related sequences are present in the genome. These bands are probably detected because of the homology in the DNA-binding domain or estrogen binding domain. Indeed, we have cloned a cDNA which has only a partial homology in the estrogen binding domain from a human placental cDNA library (to be published). Further analysis of these related sequences will help understand the steroid hormone system and the evolution of steroid receptor gene family.

ACKNOWLEDGMENTS

We are grateful to our colleagues for valuable suggestions and discussion.

This research was supported in part Grants-in-Aid from the Ministry of Education, Science and Culture, Japan and from the Foundation for Promotion of Cancer Research backed by Japan Shipbuilding Industry Foundation.

REFERENCES

1. Yamamoto, K. R. (1985) Ann. Rev. Genet. 19, 209-252.
2. Wrange, O. and Gustafsson, J-A. (1978) J. Biol. Chem. 253, 856-865.
3. Carlstedt-Duke, J., Okret, S., Wrange, O. and Gustafsson, J-A. (1982) Proc. Natl. Acad. USA 79, 4260-4262.
4. Allegretto, E. A. and Pike, J. W. (1985) J. Biol. Chem. 260, 10139-10145.
5. Koike, S. Nii, A., Sakai, M. and Muramatsu, M. Biochemistry in press.
6. Green. S., Walter, P., Kumar, V., Bornert, J-M., Argos, P. and Chambon, P. (1986) Nature 320, 134-139.
7. Greene, G. L., Gilna, P., Waterfield, M., Baker, A., Hort, Y. and Shine, J. (1986) Science 231, 1150-1154.
8. Krust, A., Green, S., Argos, P., Kumar, V., Walter P., Bornert, J-M. and Chambon, P. (1986) EMBO J. 5, 891-897.
9. Hollenberg, S. M., Weinberger, C., Ong, E. S., Cerelli, G., Oro, A., Lebo, R., Thompson, E. B., Rosenfeld, M. G. and Evans, R. M. (1985) Nature 318, 635-641.
10. Miesfeld, R., Rusconi, S., Godowski, P. J., Maler, B. A., Okret, S., Wistrom. A-C., Gustafsson, J-A. and Yamamoto, K. R. (1986) Cell 46, 389-399.
11. Jeltsch, J. M., Krozowski, Z., Quirin-Stricker, C., Gronemeyer, H., Simpson, R. J., Garnier, J. M., Krust, A., Jacob, F. and Chambon, P. (1986) Proc. Natl. Acad. Sci. USA 83, 5424-5428.
12. Conneely, O. M., Sullivan, W. P., Toft, D. O., Birnbaumer, M., Cook, R. G., Maxwell, B. L., Zarucki-schulz, T., Greene, G. L., Schrader, W. T. and O'Malley, B. W. (1986) Science 233, 767-770.
13. Aviv. H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408-1412.
14. Gubler, U. and Hoffman, B. J. (1983) Gene 25, 263-269.
15. Gorman, C. M., Moffat, L. E. and Howard, B. H. (1982) Mol. Cell. Biol. 2, 1144-1151.
16. Graham, F. and van der Eb (1973) Virology 52, 456-457.
17. Schrader, W. T. and O'Malley, B. W. (1972) J. Biol. Chem. 247, 51-59.
18. Scatchard, G. (1949) Ann. N. Y. Acad. Sci. 511, 660-672.
19. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
20. Norrander, J., Kempe, T. and Messing, J. (1983) Gene 26, 101-106.
21. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.

22. Mizusawa, S., Nishimura, S and Seela, F. (1986) Nucleic Acids Res. 14, 1319-1324.
23. Kozak, M. (1986) Cell 44, 283-292.
24. Van Oosbree, T. R., Kim, U. H. and Mueller, G. C. (1984) Anal. Biochem. 136, 321-327.
25. Lubahn, D. B., McCarty, K. S., Jr. and McCarty, K. S., Sr. (1985) J. Biol. Chem. 260, 2515-2526.
26. Katzenellenbogen, J. A., Carlson, K. E., Heiman, D. F., Robertson, D. W., Wei, L. L. and Katzellenbogen, B. S. (1983) J. Biol. Chem. 258, 3487-3495.
27. Klein-Hitpass, L., Schropp, M., Wagner, U. and Ryffel, G. U. (1986) Cell 46, 1053-1061.
28. Jost, J-P., Moncharmont, B., Jiricny, J., Saluz, H. and Hertner, T. (1986) Proc. Natl. Acad. Sci. USA 83, 43-47.
29. Nakayama, T. (1985) Proc. Japan Acad. 61, (B) 52-55.
30. Jost, J-P., Seldran, M., Geiser, M. (1984) Proc. Natl. Acad. Sci. USA 81, 429-433.
31. Payver, F., DeFranco, D., Firestone, G. L., Edger, B., Wrange, O., Okret, S., Gustafsson, J-A. and Yamamoto, K. R. (1983) Cell 35, 381-392.
32. Moore, D. D., Marks, A. R., Buckley, D. I., Kapler, G., Payver, F. and Goodman, H. M. (1985) Proc. Natl. Acad. Sci. USA 82, 699-702.
33. Renkawitz, R., Schutz, G., von der Ahe, D., and Beato, M. (1984) Cell 37, 503-510.
34. Karin, M., Haslinger, A., Holtgreve, H., Richards, R. J., Krauter, P., Westphal, H. M. and Beato, M. (1984) Nature 308, 513-519.
35. Dean, D. C., Knoll, B. J., Riser, M. E. and O'Malley, B. W. (1983) Nature 305, 551-554.
36. Walker, P., Germond, J-E., Brown-Luedi, M., Givel, F. and Wahli, W. (1984) Nucleic Acids Res. 12, 8611-8626.
37. Kumar, V., Green, S., Staub, A. and Chambon, P. (1986) EMBO J. 5, 2231-2236.
38. Giguere, V., Hollenberg, S. M., Rosenfeld, M. G. and Evans, R. M. (1986) Cell 46, 645-652.
39. Debuire, B., Henry, C., Benaissa, M., Biserte, G., Cleverie, J. M., Saule, S., Martin, P. and Stehelin, D. (1984) Science 224, 1456-1459.
40. Miller, J., McLachlan, A. D. and Klug, A. (1985) EMBO J. 4, 1609-1914.
41. Laughan, A. and Gesteland, R. F. (1985) Mol. Cell. Biol. 4, 260-267,
42, Kammerer, B., Guyonvarch, A. and Hubert J. C. (1984) J. Mol. Biol. 180, 239-250.
43. Hartshorne, T. A., Blumberg, H. and Young, E. T. (1986) Nature 320, 283-287.
44. Rosenberg, U. B., Schroder, C., Preiss, A., Kienlin, A., Cote, S., Riede, I. and Jackle, H. (1986) Nature 319, 336-339.
45. Vincent, A., Colot, H. V. and Rosbach, M. (1985) J. Mol. Biol. 186, 149-166.
46. Thomas, P. S. (1980) Proc. Natl. Acad. Sci. USA 77, 5201-5205.