Aust. J. Biol. Sci., 1988, **41**, 507–16

Oestrogen Sulfotransferase: Molecular Cloning and Sequencing of cDNA for the Bovine Placental Enzyme

Allan R. Nash[AB], *Wendy K. Glenn*[A], *Stephen S. Moore*[AC], *Judith Kerr*[A], *Adrienne R. Thompson*[A] *and E. O. P. Thompson*[A]

[A] Oncology Research Centre, Prince of Wales Hospital, Randwick, N.S.W., 2031, and School of Biochemistry, University of N.S.W., Kensington, N.S.W. 2033.
[B] To whom correspondence should be addressed.
[C] Present address: CSIRO Division of Tropical Animal Production, Rockhampton, Qld 4701.

Abstract

The female sex hormone, oestrogen, plays a central role in breast cell proliferation in both the normal and malignant state. It controls transcription from several genes, including that for the progesterone receptor, and in endometrial tissue, via this receptor, it controls the gene for the enzyme oestrogen sulfotransferase. This enzyme may control the level of the oestrogen receptor by sulfurylating free oestradiol. To study the mode of transcriptional control exercised by oestrogen, bovine oestrogen sulfotransferase cDNA has been cloned and the nucleotide sequence determined. The message, of which 1812 bases have been sequenced, contains an open reading frame of 885 bases which encode a protein of 295 amino acids and a maximum apparent molecular weight of 34 600. The deduced protein sequence is supported by existing peptide sequence data and appears to contain a steroid-binding region. Some physico-chemical characteristics of the enzyme appear to differ markedly from those previously reported.

Introduction

Oestrogen sulfotransferase (OST, EC 2.8.2.4) is an intracellular enzyme present in most human foetal tissues (Wengle 1966), but it has a more limited distribution in adult tissues (Bostrom and Wengle 1967; Pack *et al.* 1979; Tseng *et al.* 1985). It has been detected in some breast carcinomas (Adams 1964) and in these tumours the activity of the enzyme appears to correlate with the concentration of oestrogen receptors. It has been proposed that the level of enzyme activity may constitute an independent prognostic variable in breast cancer (Adams *et al.* 1979; Tseng *et al.* 1983). A role for the enzyme in this disease has yet to be described.

Similarly, nothing is known of the transcriptional control of the OST gene in breast cancer. However, in human endometrium progesterone appears to induce transcription of the gene (Pewnin *et al.* 1980), while transcription from the progesterone receptor gene is controlled by the oestrogen receptor (Leavitt *et al.* 1978) with oestrogen sulfotransferase contributing to the control of the level of oestrogen (reviewed by Hobkirk 1985). While it is reasonable to expect that this system may operate in hormone-dependent breast cancers, i.e. those with an absolute requirement for oestrogen for growth, it would appear unlikely that the same system could exist in those cancers which are hormone-independent and which constitute the majority of human breast cancers (65%).

To define a role for OST in human breast cancer, and as part of a larger program of study of the control of expression of genes for steroid-binding proteins, OST was isolated from bovine placental tissue (Moore *et al.* 1988) using modifications of the method of

0004-9417/88/040507$03.00

Adams and Low (1974). The enzyme appeared to have a monomeric molecular weight of around 36 000 although there was some evidence that the active form of the enzyme could be a dimer with a molecular weight of 72 000. Peptides from the OST preparation were sequenced and the sequences used to design oligonucleotide probes. In the study reported the base sequence of the entire message for the enzyme was determined.

Materials and Methods

All common chemicals and solvents were AR grade. All isotopes were obtained from Amersham (Australia) Pty Ltd. Unless otherwise indicated enzymes were from Promega Pty Ltd (Madison, WI., U.S.A.) or Pharmacia (Australia) Pty Ltd, all electrophoresis reagents and equipment were from Bio Rad (Sydney, Australia) and all common chemicals were from Sigma Chemical Company (St Louis, MO., U.S.A.).

Construction of cDNA Library

Bovine placentas were collected fresh from the abattoirs and stored at −70°C. Frozen tissue was homogenised in 3–4 volumes of 7 M guanidine hydrochloride/10 mM Tris/10 mM EDTA/5% β-mercaptoethanol, pH 7·5, after which cesium chloride (CsCl) and sodium sarkosyl were added to concentrations of 40% (w/v) + 0·5% (v/v) respectively. The homogenate was layered onto a 20 mm cushion of 5·7 M CsCl in 0·1 M EDTA, pH 7·5, in a 16 mm × 100 mm polyallomer tube and centrifuged at 100 000 *g* for 16 h at 20°C. Further purification of the total RNA and selection of poly(A) + RNA on oligo(dT) columns was according to the methods described by Maniatis *et al.* (1982).

First strand cDNA synthesis, using mixed primers, was performed with a Reverse Transcriptase System kit from New England Nuclear (Boston, MA., U.S.A.). Second strand synthesis, with hairpin loop self priming, was performed using the same kit. Hairpin loop hydrolysis with S1 nuclease and purification of the cDNA was according to Maniatis *et al.* (1982). Eco R1 sites were methylated and linkers added according to the methods described by Grubler and Hoffman (1983). 'Cold' or 'hot' linkers were generated by phosphorylation of the 5′ end of the oligonucleotide with either ATP or [$\gamma-^{32}$P] ATP (Wu *et al.* 1987). Excess linkers were removed from the ligation mixture by passage over a NACS column (B. R. L. Gaithersburg, MD., U.S.A.).

The cDNA was ligated into λ gtll arms using the Promega 'Protoclone GT' system and packaged using the Promega 'Packagene' system. The library was amplified in *E. coli* Y1090.

Probe Construction

Oligonucleotide probes were synthesised using a Model 380A Automatic DNA Synthesiser (Applied Biosystems Inc., Australia) and the protocols recommended by this company. Non-degenerate probes were derived directly from cDNA sequences already determined while the base sequences of degenerate probes were based on the amino acid sequences of peptides derived from a tryptic digest of the purified enzyme (Moore *et al.* 1988). The methods of Lathe (1985) were utilised to minimise the degeneracy of these probes.

Screening the Library

The λ gtll library was plated out in *E. coli* Y1090 on 150 mm diameter petri dishes at a phage concentration slightly lower than that required to produce confluent lysis. Transfer of plaques to nitrocellulose filters, hybridisation to the labelled probe, purification of positive plaques and the preparation of plate lysate stocks were essentially as described by Maniatis *et al.* (1982).

Clone Sequencing

Insert DNA from individual clones was subcloned into the vector M13tg 130/131 by the method of Kieny *et al.* (1983). The di-deoxy chain termination sequencing method of Sanger *et al.* (1977) was used in conjunction with the [α^{35}S] ATP modification suggested by Biggin *et al.* (1983). Nucleotide and amino acid sequences were compared by use of the MBIS system (Reisner and Bucholtz 1986) which utilises the SEQH and SEQHP programs of Kanehisa *et al.* (1984). The statistical significance of an alignment was tested, within these programs, by comparing it with the alignments of 20 random sequences. The SWISS-PROT protein database used was accessed via the Commonwealth Scientific and Industrial Research Organisation's Molecular Biology Information Service (Canberra, Australia).

Sequencing Strategies

Degenerate probes derived from the sequence of tryptic peptides were used to isolate major clones. Once the base sequences of these clones had been established non-degenerate probes complementary to the terminal sequences of these major clones were synthesised. These probes were used to extend the sequences in the 5′ or 3′ direction, or to provide overlap sequences between major clones.

Table 1. Oligonucleotide probe sequences

Peptides were purified from a tryptic digest of bovine OST (Moore *et al.* 1988). For probes 1–3 the probe sequence was derived from the peptide sequence; degeneracy was minimised by utilisation of codon usage tables and the known instability of C–G base sequences when C is the third codon position. Probes 4–6 were synthesised as the non-degenerate complements of cDNA sequences established earlier in the study. All probes were synthesised using an Applied Biosystems 380A automatic DNA synthesiser and the protocols recommended by the company. In probe 1 inosine was inserted at the positions marked 'I'. Degenerate codons are shown with the alternative nucleotide below the third position in probes 1–3. Peptides are shown N-terminus to C-terminus and probes are shown as the 3′ to 5′ complement

Peptide 29	K G D V G D W K
Probe 1	TTTCCICTACAICCICTAACCTTT
	C G G C
Peptide 3	K F I E Q F H N V E E F E A R P D D
Probe 2	GTATTACACCTCCTCAAACTCCGATCCGGACTACTA
	G G T T G T G T T G
	– L V I V TYPK
	–GACCACTAACACTG
	T
Peptide 31	F D M H Y E Q Q M K
Probe 3	AAACTATACGTAATACTTGTTGTTTACTT
	G G C C C
Probe 4	Based on known cDNA sequence
	ATCACTTCAGAGTGGACGGACACTTCGAAGAAGG
Probe 5	Based on known cDNA sequence
	GACATTTCAAAATTACTTACACTGTATATGGTCA
Probe 6	Based on known cDNA sequence
	CAAACCCTGAGAAAGGACTAACCTGAGTCTGACG

Results

The base sequences of the probes used in this study are shown in Table 1. Also shown are the peptide sequences from which they were derived. The base sequences of probes 4–6 were derived from cDNA sequences of clones isolated with probes 2 and 3.

On a Northern blot (Thomas 1980) of the total mRNA a single size species hybridised to probe 1 and this species was in the range 1·5 kb–3·0 kb (Fig. 1). The size for the message was confirmed when total mRNA was subjected to sucrose density centrifugation. A single species in the fraction sedimenting between 16S and 23S bound to probe 4.

The λ gtll library contained 10^5 recombinants. Screening of this library with probe 1 was unsatisfactory as the probe hybridised to all recombinants. Using the SEQH program strong homologies were found between the probe sequence and several stretches of sequence from the lambda genome, e.g. 16 out of the 24 bases of the probe are identical to the sequence from position 42484 to position 42507 of the lambda genome. Therefore, the longer probes (2–6) were used in the isolation of OST cDNA inserts, none of which were larger than 582 bp. The spatial relationship between the cDNA clones sequenced in this study is

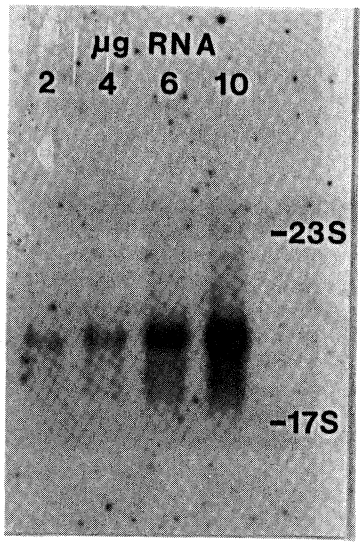

Fig. 1. Northern blot of bovine placental mRNA. Purified mRNA was separated on a 1% agarose gel under denaturing conditions. Following electrophoresis the RNA was transferred to nitrocellulose paper by the Northern blotting technique. The transferred RNA was hybridised to the end-labelled probe 1 (Table 1) and then autoradiographed. The figures down the right-hand side of the gel indicate the position and sedimentation coefficient of prokaryotic ribosomal RNA markers.

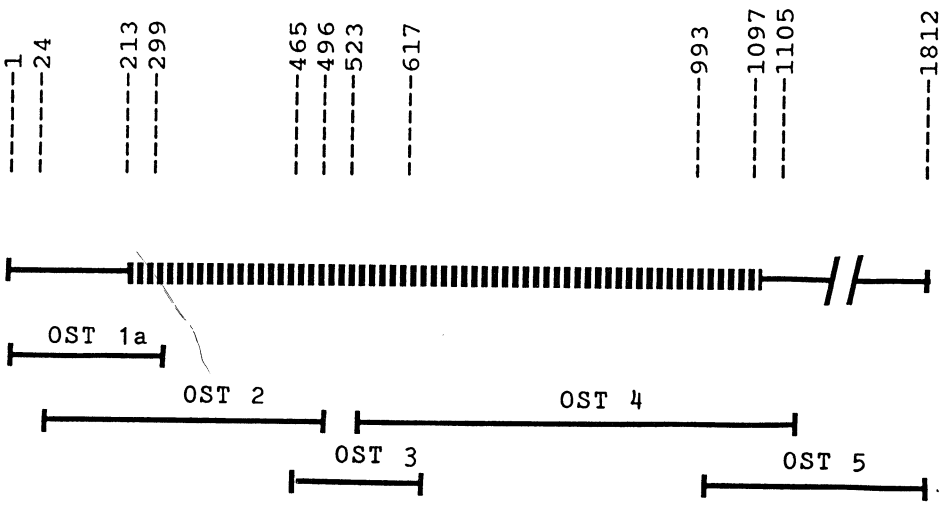

Fig. 2 Map of the cDNA clones representing the complete bovine OST mRNA. cDNA clones OST 2 and OST 4 were isolated by use of degenerate probes derived from the sequence of peptides isolated from a tryptic digest of the enzyme. Clones OST 1, OST 3 and OST 5 were isolated by use of non-degenerate probes derived from the cDNA sequences of clones OST 2 and OST 4. The open reading frame of the message is represented by the hatched box while the nucleotides defining termini of the clones, the open reading frame and the entire message are numbered above the map.

Fig. 3. Full length nucleotide sequence of bovine OST mRNA. The amino-acid sequence of the enzyme deduced from the coding sequence of the mRNA is shown below the nucleotide sequence. Assuming trypsin will not cleave the K-P bond between residues 5 and 6, and ignoring the single basic residue liberated by cleavage after successive basic residues, e.g. residues 22 and 23, there are 34 possible tryptic peptides. The boxed amino-acid sequences represent those tryptic peptides for which sequences were determined during an earlier study (Moore *et al.* 1988) and they are numbered according to their position in the linear array of the 34 peptides. The boxed nucleotide sequences mark the probe hybridisation sites.

```
                                                                  probe 6
  1 CTTCCCCTCCACAGGTGAGGAAGCACTGTTTGGGACTCTTTCCTGATTGGAACTCCAGACTG

 63 GTATTAACAAAGGACTTCTCAGCACCGTTTGAGTCTCGATATTGGTCAGCCCAATTTGGTGA

125 TCTTGAAACCTGAGTGGGGATTGAAGAAAACTGAATTGGATTACCCAGTTGCTTTCACAGGA

187 TCATCTGGACAGTGTACCACTCCACG  ATG AGT TCT TCC AAA CCA TCC TTT TCA
                                Met Ser Ser Ser Lys Pro Ser Phe Ser 9

240 GAT TAC TTT GGC AAA CTT GGT GGA ATA CCA ATG TAT AAA AAA TTT ATC
    Asp Tyr Phe Gly Lys Leu Gly Gly Ile Pro Met Tyr Lys Lys Phe Ile 25
                                Tp2          probe 2
288 GAG CAG TTT CAT AAC GTG GAG GAA TTT GAG GCA AGA CCA GAT GAC CTT
    Glu Gln Phe His Asn Val Glu Glu Phe Glu Ala Arg Pro Asp Asp Leu 41
                                Tp3
336 GTC ATT GTC ACC TAT CCC AAA TCT GGT ACA ACA TGG CTT AGT GAA ATT
    Val Ile Val Thr Tyr Pro Lys Ser Gly Thr Thr Trp Leu Ser Glu Ile 57

384 ATA TGC ATG ATT TAT AAT AAC GGT GAT GTG GAA AAG TGC AAA GAA GAC
    Ile Cys Met Ile Tyr Asn Asn Gly Asp Val Glu Lys Cys Lys Glu Asp 73

432 GTC ATT TTT AAT AGA GTT CCT TAC CTG GAA TGT AGC ACT GAA CAC GTG
    Val Ile Phe Asn Arg Val Pro Tyr Leu Glu Cys Ser Thr Glu His Val 89

480 ATG AAA GGA GTG AAA CAA TTA AAT GAG ATG GCA TCT CCT AGA ATA GTG
    Met Lys Gly Val Lys Gln Leu Asn Glu Met Ala Ser Pro Arg Ile Val 105
                      probe 4
528 AAG TCT CAC CTG CCT GTG AAG CTT CTT CCA GTC TCA TTT TGG GAA AAG
    Lys Ser His Leu Pro Val Lys Leu Leu Pro Val Ser Phe Trp Glu Lys 121
                                          Tp12
576 AAC TGT AAG ATC ATC TAT CTT TCC CGG AAT GCC AAG GAT GTG GTT GTT
    Asn Cys Lys Ile Ile Tyr Leu Ser Arg Asn Ala Lys Asp Val Val Val 137

624 TCT TAT TAT TTT TTA ATT TTA ATG GTG ACT GCT ATT CCA GAT CCT GAC
    Ser Tyr Tyr Phe Leu Ile Leu Met Val Thr Ala Ile Pro Asp Pro Asp 153

672 TCT TTT CAA GAT TTT GTG GAG AAA TTC ATG GAT GGA GAA GTT CCT TAT
    Ser Phe Gln Asp Phe Val Glu Lys Phe Met Asp Gly Glu Val Pro Tyr 169
                                                Tp17
720 GGT TCC TGG TTT GAA CAT ACA AAA TCT TGG TGG GAA AAG AGC AAG AAT
    Gly Ser Trp Phe Glu His Thr Lys Ser Trp Trp Glu Lys Ser Lys Asn 185
                                    Tp18
768 CCG CAA GTG CTA TTT CTT TTC TAT GAA GAC ATG AAG GAG AAT ATC AGA
    Pro Gln Val Leu Phe Leu Phe Tyr Glu Asp Met Lys Glu Asn Ile Arg 201

816 AAA GAG GTG ATG AAA TTG TTA GAA TTT CTG GGA AGG AAG GCA TCA GAT
    Lys Glu Val Met Lys Leu Leu Glu Phe Leu Gly Arg Lys Ala Ser Asp 217

864 GAG CTT GTG GAC AAG ATT ATA AAA CAC ACT TCA TTC CAG GAG ATG AAG
    Glu Leu Val Asp Lys Ile Ile Lys His Thr Ser Phe Gln Glu Met Lys 233

912 AAC AAT CCA TCT ACC AAT TAT ACA ACA TTG CCG GAT GAA GTC ATG AAC
    Asn Asn Pro Ser Thr Asn Tyr Thr Thr Leu Pro Asp Glu Val Met Asn 249
                                Tp27
                                                        probe 1
960 CAA AAA GTA TCT CCC TTC ATG AGA AAA GGA GAT GTA GGA GAC TGG AAG
    Gln Lys Val Ser Pro Phe Met Arg Lys Gly Asp Val Gly Asp Trp Lys 265
                                                    Tp29
                                                                probe 3
1008 AAT CAC TTT ACA GTA GCC CTG AAT GAG AAA TTT GAC ATG CAC TAT GAG
     Asn His Phe Thr Val Ala Leu Asn Glu Lys Phe Asp Met His Tyr Glu 281
                                                                Tp31

1056 CAG CAA ATG AAG GGG TCT ACC CTG AAG TTC CGA ACT AAG ATC TAGGAAG
     Gln Gln Met Lys Gly Ser Thr Leu Lys Phe Arg Thr Lys Ile

1105 GGTTTTCTTAACATCGGAGATTAAATTGTTCTTTCACCCTTCCTCACCTTCTTCATTTTAAG

1167 CTAGTAGATAATGAATTATACAAAAGATCATGATCAGGTTCACATGATTATTGAGATCTTAG

1229 TATGAAAGAGAGATTTTTTTCCTGTGATTATCTCTTTAATATTTTCTTTCCTTTCTTCTTTA

1291 ATTATAAATATTACACAGAGCTATAACCTATGAGAATGATGTAGGTACACACAAATTGTTCA

1353 AGTTGTTAGAGCCTCAGTAAAAATAACCAGACATTCCAAATTATATAACTTTGTGTCTACTT

1415 TTTTCCCATTTATTAGTATATTTGGTATATACATATACCTTTAAATACATATTAATAACTGT

1477 ACATGGAATTTGAAAACCTTGACAAATAGAACTGAAAAGAATAAACATGAGTGTATAACTAG

1539 ATGTATTATTTCTTGGCTGATATTATCTTTCATTAATGCAAAATGACTTAACTTGGAATTTA

1601 CCTTAGTTTAGTGTAAATTTTCAGCATTTTCAGAACTGATTTCCAGATGTCTATAAATAATT

1663 CCAATGAGTACAATTTACATTTTAATATCTTTAGCTAATTTAAAATCGTTATGCTTTAGGGT
                                                                probe 5
1725 TCCTTCTATCCTATACACAATGTAATGAAAGTAAGATCTGTAAAGTTTTAATGAATGTGACA

1787 TATACCAGTAAAATAAAACAAAAGCA
```

shown in Fig. 2. Although clones OST 2 and OST 4 were the first to be sequenced, for ease of explanation the results of clone detection and sequencing are presented in the linear order shown in Fig. 2, not the temporal order in which the clones were isolated.

Probe 6 was a non-degenerate probe derived from the first 34 bases of clone OST 2. After use of this probe, re-analysis of the base sequence of clone OST 2 revealed that the putative single bases at positions 51(A) and 55(C) (Fig. 3), were, in fact, doublets and the probe sequence beyond position 24 (Table 1) was no longer complementary to the cDNA sequence. Despite this, the probe was used successfully to isolate two clones. Clone OST 1a extended from position 1 to position 299. Clone OST 1b spanned positions 1–76 but differed in sequence from OST 1a in nine of the first 18 bases of the molecule as follows:

```
OST 1a  CTTCCCCTCCACAGGTGA----
            : :  :  : :  :  : :  :
OST 1b  AAGCCACACCTCTGGAGG----
```

However, in the absence of additional sequence data from this region, and in view of the results obtained by Richards *et al.* (1979), it seems likely that these differences arose as artefacts of cloning.

Probe 2 was used to isolate eight clones of which five were unique. Clone OST 2 was sequenced and found to be 472 bp long. This clone overlapped with clone OST 1 at position

Table 2. Amino acid composition of oestrogen sulphotransferase deduced from the cDNA nucleotide sequence
The amino acid composition deduced from the cDNA nucleotide sequence given in Figure 3 is shown for comparison with that obtained by direct analysis of the enzyme preparation of highest specific activity from which peptides for probe construction were isolated (Moore *et al.* 1988). N.D., not determined

Amino acid	From cDNA[A]	By analysis
Tryptophan	6	N.D.
Lysine	32	34
Histidine	7	7
Arginine	8	9
Aspartic Acid	17	33
Asparagine	15	
Threonine	13	13
Serine	22	29
Glutamic Acid	24	34
Glutamine	8	
Proline	15	13
Glycine	12	22
Alanine	6	12
Half Cystine	4	N.D.
Valine	24	21
Methionine	14	9
Isoleucine	15	11
Leucine	21	17
Tyrosine	12	10
Phenylalanine	20	16
Total	295	
Mol. wt	34 600	

[A] Including methionine at initiation site.

24 and extended to position 496. All other clones detected with probe 2 were contained within OST 2.

Probe 4 was a non-degenerate probe based on the sequence of the first 34 bases of clone OST 4. It hybridised to a single cDNA species, clone OST 3, of 153 bp, which extended from position 465 to position 617 and provided the overlap between the major clones OST 2 and OST 4.

Probe 3 was used to isolate two clones. Clone OST 4 overlapped with OST 3 at position 523 and extended the sequence to position 1105. The second clone, OST 5, overlapped with OST 4 at position 993 and extended the sequence to position 1812. Although the library was screened with probes from each end of the cDNA (probes 5 and 6) no full length cDNA clone was detected. The library appeared to contain only shorter cDNA fragments. The complete cDNA sequence is shown in Fig. 3.

A terminating codon was present at nucleotide position 156 and there were codons for methionine at nucleotide positions 213 and 270. However, the initiation site was assumed to be position 213 as the methionine coded for at position 270 was contained within a peptide (Tp2) isolated from a tryptic digest of the enzyme (Moore *et al.* 1988). There were termination codons at nucleotide positions 1098 and 1113 and polyadenylation signals at nucleotide positions 1516 and 1798. Potential cAMP-dependent kinase phosphorylation sites were identified at positions 848 and 885 although no phosphate groups were found on the protein.

The amino acid sequence of the enzyme, deduced from the nucleotide sequence of the open reading frame of the message, also is shown in Fig. 3. The alignments, with this sequence, of the tryptic peptides isolated in the earlier study (Moore *et al.* 1988) also are shown in the figure. No peptide sequences found in the tryptic digest are unaccounted for by the cDNA sequence.

The amino acid composition of the enzyme, predicted from the cDNA sequence, is presented in Table 2 for comparison with the composition found by amino-acid analysis of material from the enzyme preparation of Moore *et al.* (1988).

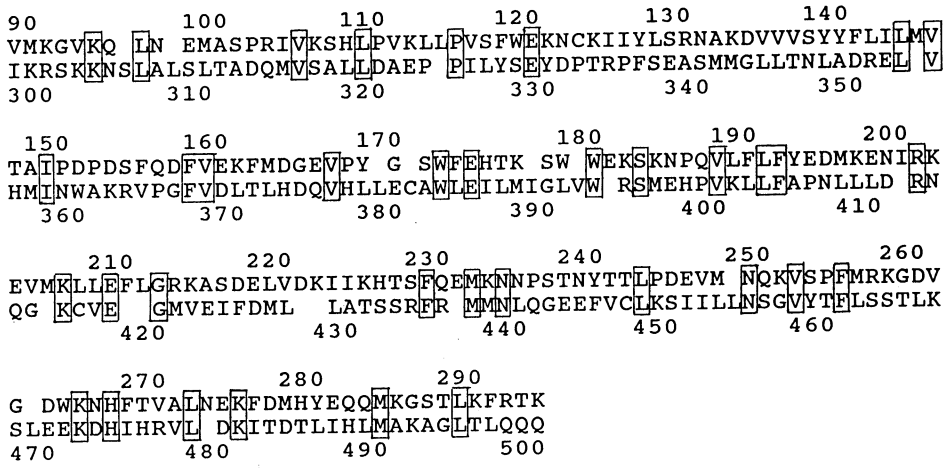

Fig. 4. Alignment of the amino acid sequences of bovine OST and human oestrogen receptor. The amino acid sequence of the ligand binding domain of the human oestrogen receptor mRNA (Petkovitch *et al.* 1987) is shown below portion of the sequence for bovine OST mRNA. Residue numbers are shown above and below the two sequences and homologous residues are boxed.

In Fig. 4 is shown the putative homology between the predicted sequences of the human oestrogen receptor (ER) and the bovine oestrogen sulfotransferase. This alignment was obtained by use of the SEQHP program which introduced spaces into both sequences to maximise the homology. The alignment shown is statistically significant ($P = 0{\cdot}04$).

Discussion

Apart from the very high sulfotransferase activity of the preparation used in this study the data present in Fig. 4 suggest an oestrogen binding role for the protein studied. In the human oestrogen receptor the ligand-binding domain extends from residue 315 to 550 (Petkovich *et al.* 1987). Between the oestrogen receptor and other human steroid receptors there are significant homologies in this region, e.g. 26% (64/243) of residues can be aligned with the progesterone receptor and 28% (68/243) of residues with the glucocorticoid receptor. When phylogenetic distances and different molecular functions are taken into consideration the alignment of 16% (36/220) of residues from the bovine enzyme cDNA sequence with that of the human receptor sequence possibly defines the steroid-binding domain of the oestrogen sulfotransferase. A further similarity between the human receptor (Green *et al.* 1986) and the bovine enzyme is the presence in both of two potential phosphorylation sites; in the enzyme they are located at residues 213 (−RKAS−) and 225 (−KHTS−).

Earlier claims (Adams and Ellyard 1971; Clarke *et al.* 1982) that oestrogen sulfotransferase was structurally similar to serum albumin, and possibly shared a common genetic ancestor, cannot be sustained. Computer assisted comparisons of the primary structures of bovine serum albumin and bovine OST provided no evidence of albumin-like 'domains' in the primary structure of the enzyme and the sequence around none of the 35 half-cystine residues in bovine serum albumin bore any resemblance to the sequences around the four half-cystine residues found in bovine OST.

The cDNA sequence in Fig. 4 is that of a protein with a maximum molecular weight of 34 600. Peptide Tp1 (residues 1–14) was neither sought nor found so the N-terminal residue of the enzyme has not been defined. There is some evidence that it is blocked, possibly at a serine residue; no N-terminal amino acid could be detected when sequencing of the intact protein was attempted. The peptide, Tp2, was found and sequenced (Fig. 3) so the protein can be no shorter than 281 residues. Post-translational loss of part or all of the N-terminal peptide (Tp1) would yield a protein with a molecular weight in the range 33 000–34 600. Earlier estimates of the molecular weight of this enzyme as 72 000 (Adams *et al.* 1974, Clarke *et al.* 1982) must be revised. Similar revision must be made of claims (Clarke *et al.* 1982) that the presence of lower molecular weight species in enzyme preparations could be attributed to proteolytic cleavage of a larger molecule.

Chromatographically, on Sephacryl S300, and electrophoretically, in non-denaturing conditions, the enzyme behaves as a molecule with an apparent molecular weight of around 70 000 (Moore *et al.* 1988). Since only one cDNA species was detected it would seem that the native, active enzyme is a dimer of identical 33 000–34 600 monomers. This is not dissimilar to the putative structure of another sulfotransferase, dehydroepiandrosterone sulfotransferase, which is a dimer of apparently identical monomers of molecular weight 34 500 (Adams and McDonald 1979).

The peptides used in this study were isolated from an enzyme preparation which contained a minor contaminant with a molecular weight of approximately 50 000. The presence of this contaminant is reflected in the disparity of the amino acid compositions shown in Table 2. It seems likely that the earlier workers co-purified, with their enzyme activity, major contaminating proteins with apparent molecular weights of around 70 000. The presence of these contaminants in an enzyme preparation would provide an explanation for the previous reports of OST isoenzymes of low activity and high molecular weight which had superficial similarities to bovine serum albumin.

Acknowledgments

We thank Miss J. Robinson for her assistance in the preparation of this manuscript and the staff of the Department of Medical Illustration, Prince of Wales Hospital, for the preparation of the figures. The contribution made to the study by Dr B. L. McDonald is

gratefully acknowledged. This study was supported in part by the Australian Research Grants Scheme and the N.S.W. Cancer Council.

References

Adams, J. B. (1964). Enzymic synthesis of steroid sulphates II. Presence of steroid sulphokinase in human mammary carcinoma extracts. *J. Clin. Endocrinol. & Metab.* **24**, 988–96.

Adams, J. B., and Ellyard, R. K. (1971). Comparison of the amino acid sequence around the 'active' sulphydryl group of oestrogen sulphotransferase, phosphokinases and serum albumin. *Proc. Aust. Biochem. Soc.* **4**, 93.

Adams, J. B., and Low, J. (1974). Enzymic synthesis of steroid sulphates. X. Isolation of oestrogen sulphotransferase from bovine placenta and comparison of its properties with adrenal oestrogen sulphotransferase. *Biochim. Biophys. Acta* **370**, 189–96.

Adams, J. B., and McDonald, D. (1979). Enzymic synthesis of steroid sulphates. XII. Isolation of dehydroepiandrosterone sulphotransferase from human adrenals by affinity chromatography. *Biochim. Biophys. Acta* **567**, 144–53.

Adams, J. B., Ellyard, R. K., and Low, J. (1974). Enzymic synthesis of steroid sulphates. IX. Physical and chemical properties of purified estrogen sulphotransferase from bovine adrenal glands, the nature of its isoenzymic forms and a proposed model to explain its wave-like kinetics. *Biochim. Biophys. Acta* **370**, 160–88.

Adams, J. B., Pewnim, T., Chandra, D. B., Archibald, L., and Foo, M. S. (1979). A correlation between estrogen sulphotransferase levels and estrogen receptor status in human primary breast carcinoma. *J. Steroid Biochem.* **19**, 1413–17.

Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983). Buffer gradient gel and [^{35}S] label as an aid to rapid DNA sequence determination. *Proc. Nat. Acad. Sci. U.S.A.* **80**, 3963–5.

Bostrom, H., and Wengle, B. (1967). Studies on ester sulphates 23. Distribution of phenol and steroid sulphokinase in adult human tissues. *Acta Endocrinol.* **56**, 691–704.

Clarke, C., Thorburn, P., McDonald, D., and Adams, J. B. (1982). Enzymic synthesis of steroid sulphates. XV. Structural domains of estrogen sulphotransferase. *Biochim. Biophys. Acta* **707**, 28–37.

Green, S., Walter, P., Kumar, V., Krust, A., Bornett, J. M., Argos, P., and Chambon, P. (1986). Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. *Nature (Lond.)* **320**, 134–9.

Grubler, V., and Hoffman, B. J. (1983). A simple and very efficient method for generating cDNA libraries. *Gene (Amst.)* **25**, 263–9.

Hobkirk, R. (1985). Steroid sulfotransferases and steroid sulfate sulfatases: Characteristics and biological role. *Can. J. Biochem. Cell Biol.* **63**, 1127–44.

Kanehisa, M., Klein, P., Greif, P., and DeLisi, C. (1984). Computer analysis and structure prediction of nucleic acids and proteins. *Nucleic Acids Res.* **11**, 417–28.

Kieny, M. P., Lathe, R., and Lecocq, J. P. (1983). New versatile cloning and sequencing vectors based on bacteriophage M13. *Gene (Amst.)* **26**, 91–9.

Lathe, R. (1985). Synthetic oligonucleotide probes deduced from amino acid sequence data. Theoretical and practical considerations. *J. Mol. Biol.* **183**, 1–12.

Leavitt, W. W., Chen, T. J., Do, Y. S., Carlton, B. D., and Allen, T. C. (1978). Biology of progesterone receptors. In 'Receptors & Hormone Action.' (Eds B. W. O'Malley and L. Birnkaumer.) Vol. II, pp. 157–88. (Academic Press: New York.)

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). 'Molecular Cloning. A Laboratory Manual.' (Cold Spring Harbor Laboratory: New York.)

Moore, S. S., Thompson, E. O. P., and Nash, A. R. (1988). Estrogen sulphotransferase: isolation of a high specific activity species from bovine placenta. *Aust. J. Biol. Sci.* **41**, 333–41.

Pack, B. A., Brooks, C. L., Dukelow, W. R., and Brooks, S. C. (1979). The metabolism and nuclear migration of estrogen in porcine uterus throughout the implantation process. *Biol. Reprod.* **20**, 545–51.

Petkovich, M., Brand, M. J., Krust, A., and Chambon, P. (1987). A human retinoic acid receptor which belongs to the family of nuclear receptors. *Nature (Lond.)* **330**, 444–50.

Pewnim, T., Adams, J. B., and Ho, K. D. (1980). A relationship between estrogen sulfurylation and estrogen and progesterone receptor status in human mammary carcinoma. *Cancer Res.* **40**, 1360–2.

Reisner, A. H., and Bucholtz, C. A. (1986). The MTX package of computer programmes for the comparison of sequences of nucleotides and amino acid residues. *Nucleic Acids Res.* **14**, 233–8.
Richards, R. I., Shine, J., Ullrich, A., Wells, J. R. E., and Goodman, H. M. (1979). Molecular cloning and sequence analysis of adult chicken globin cDNA. *Nucleic Acids Res.* **7**, 1137–46.
Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Nat. Acad. Sci. U.S.A.* **74**, 5463–7.
Thomas, P. S. (1980). Hybridisation of denatured RNA and small DNA fragments transferred to nitrocellulose. *Proc. Nat. Acad. Sci. U.S.A.* **77**, 5201–4.
Tseng, L., Lee, L. Y., and Mazella, J. (1985). Estrogen sulphotransferase in human placenta. *J. Steroid Biochem.* **22**, 611–15.
Tseng, L., Mazella, J., Lee, L. Y., and Stone, M. L. (1983). Estrogen sulfatase and estrogen sulfotransferase in human primary mammary carcinoma. *J. Steroid Biochem.* **19**, 1413–17.
Wengle, B. (1966). Distribution of some steroid sulphokinases in foetal human tissues. *Acta Endocrinol.* **52**, 607–18.
Wu, R., Wu, T., and Ray, A. (1987). Adapters, linkers and methylation. In 'Methods in Enzymology'. (Eds S. L. Berger and A. R. Kimmel.) Vol. 153. pp. 343–9. (Acad. Press Inc.: Orlando.)

Manuscript received 23 August 1988, accepted 6 October 1988