Molecular and Biochemical Parasitology, 53 (1992) 17-32

MOLBIO 01744

Alternative splicing of the *Schistosoma mansoni* gene encoding a homologue of epidermal growth factor receptor

Charles B. Shoemaker[a], Hema Ramachandran[a], Abraham Landa[b], Mitermayer G. dos Reis[c] and Lincoln D. Stein[d]

[a]*Department of Tropical Public Health, Harvard School of Public Health, Boston, MA, USA;* [b]*Department of Immunology, Inst. De Investigaciones Biomédicas, UNAM, México, D.F., Mexico;* [c]*Centro de Pesquisas Goncalo Moniz, FIOCRUZ/UFBa, Bahia, Brazil; and* [d]*Deptartment of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*

(Received 20 November 1991; accepted 30 January 1992)

The complete coding DNA for a *Schistosoma mansoni* homologue of the epidermal growth factor receptor (SER) was characterized from cDNA clones obtained by homology to the tyrosine kinase domain of *erbB*. The DNA sequence predicts a 200-kDa translation product that contains a secretory leader, a cysteine-rich extracellular domain, a hydrophobic transmembrane sequence, and an intracellular tyrosine kinase domain. The SER transcript is present in cercariae and adult schistosomes.

In addition to SER transcripts, schistosomes produce at least 3 variant transcripts encoding truncated SER products that include the secretory leader and a small portion of the extracellular domain followed by short sequences of unrelated, C-terminal amino acids. Based on these sequences, 2 of the variant mRNAs (class 2 and 5) appear to encode soluble, secreted proteins while one (class 4) encodes an SER variant protein with a hydrophobic C-terminus that may serve as a membrane anchor. Class 2 SER variant transcripts are present at levels comparable to SER transcripts in adult worms but are not detected in cercariae. Class 4 and 5 SER variant transcripts are also found within adult worms but at lower levels. Genomic cloning and characterization demonstrate that the variant SER transcripts arise through alternative splicing of the SER gene.

Key words: Schistosome; *Schistosoma mansoni*; EGF; EGF receptor; Alternative splicing; Membrane protein

Introduction

Mammalian EGFRs are among the most well-characterized of the mitogenic receptors [1,2]. They are integral surface membrane proteins and have cysteine-rich extracellular domains that recognize both epidermal growth factor (EGF) and transforming growth factor-α. Upon ligand binding, the intracellular tyrosine kinase of EGFR becomes activated, initiating an incompletely understood signaling pathway that results in cell proliferation. Some oncogenes, such as erbB and neu, are modified members of the EGFR gene family [3,4] and over-expression of their normal cellular gene, EGFR or c-*neu*, can result in cell transformation [5–7].

Homologues of EGFR have been identified within eukaryotic organisms spanning a broad evolutionary range. A receptor tyrosine kinase closely related to EGFR has been demonstrated to be a malignant melanoma-inducing oncogene from *Xiphophorus* fish [8]. The *Drosophila* EGFR homologue has been isolated based on homology with erbB [9] and

Correspondence address: C.B. Shoemaker, Dept. of Tropical Public Health, Harvard School of Public Health, 655 Huntington Ave., Boston, MA 02115, USA.

Note: Nucleotide sequence data reported in this paper have been submitted to the GenBank™ database with the accession numbers M86396, M86397, M86398 and M86399.

Abbreviations: EGFR, epidermal growth factor receptor; SER, *Schistosoma mansoni* EGFR homologue; DER, *Drosophila melanogaster* EGFR homologue; GTC, guanidine isothiocyanate; PCR, polymerase chain reaction; TPI, triose phosphate isomerase; ORF, open reading frame; HGPRT, hypoxanthine-guanine phosphoribosyl transferase.

found to be allelic to *faint little ball*, a gene essential in embryonic development [10,11]. The *let*-23 gene, which is necessary for inducing the formation of the *Caenorhabditis elegans* vulva, is also an EGFR homologue [12]. Here we report the characterization of cDNA encoding a *Schistosoma mansoni* homologue of mammalian EGF receptor (SER). We also show that alternative RNA splicing generates transcripts encoding several truncated forms of SER.

Materials and Methods

Parasites. A Puerto Rican strain of *S. mansoni* was maintained by passage through *Biomphalaria glabrata* snails and CBA/J mice. Parasites from the cercarial and adult stages were obtained as previously described [13,14].

Preparation of S. mansoni mRNA. Poly (A)$^+$ RNA was isolated directly from parasites using a modification of the method of Cox et al. [15]. One volume of organisms as a packed frozen pellet was resuspended in 2 vols. of lysis buffer (6 M guanidine isothiocyanate (GTC)/10 mM Tris-HCl/1 mM β-mercaptoethanol, pH 7.5) and immediately sonicated using a microtip at 200–300 W until the solution lost viscosity. The supernatant following centrifugation at 2000 × *g* was mixed with 0.2 vols. of pre-swollen, pre-rinsed (in wash buffer: 4 M GTC/10 mM Tris-HCl/1 mM β-mercaptoethanol, pH 7.5) poly(U) Sephadex beads (Gibco, Bethesda, MD) and incubated for at least 1 h at 4°C on a rocking platform. The beads were then washed twice in 5 volumes of wash buffer and loaded into a small RNAse-free column. The beads were washed 4 times with 5 bed volumes of wash buffer and 4 times with 3 bed volumes of salt wash (0.2 M NaCl/10 mM Tris-HCl/1 mM EDTA, pH 7.5). Bound RNA was eluted by several additions of 0.5 bed volumes of elution buffer (90% formamide/1 mM EDTA/15 mM Tris-HCl, pH 7.5). The fractions were ethanol-precipitated and resuspended in RNAse-free water. 1 μl of each fraction was mixed with 10 μl of 1 μg ml^{-1} ethidium bromide and viewed with a UV light-box to ascertain the fractions containing RNA. Poly(A)$^+$ RNA concentration was estimated using the same procedure in comparison to RNA standards.

Preparation of cDNA libraries. Double-stranded cDNA from both adult and cercarial stages was prepared by the method of Gubler and Hoffman [16]. The cDNA was cloned into the λgt11 vector as described by Young and Davis [17]. The cDNA used to prepare the adult library was size fractionated prior to vector ligation by agarose gel electrophoresis and purification of 2–5 kb cDNA as described by Vogelstein and Gillespie [18].

SER cDNA cloning. The 562-bp *Bam*HI fragment containing most of the tyrosine kinase domain of avian erbB cDNA [19] was radiolabeled (Multiprime kit, Amersham Corp., Arlington Heights, IL) as recommended by the supplier. This EGFR homologue probe was hybridized at 50°C in 6 × SSC for 20 h with duplicate replica filters representing a total of about 60 000 phage plaques of the 2–5-kb *S. mansoni* adult male cDNA library [20]. Filters were extensively washed at 50°C in 2 × SSC and subjected to autoradiography for 3 days. Two weak positive signals were obtained, the recombinant phage plaque-purified (ERB3 and ERB4, Fig. 1), and their cDNA inserts were sequenced. The two clones overlap by about 1300 bp in the region of probe homology and have 57% identity with the erbB probe. The overlapping regions have the same sequence except that clone ERB3 has a 140-bp insert (ins1; Fig. 1) within this region relative to ERB4 (and other clones isolated later) that is apparently an unspliced intron (data not shown).

The adult cDNA library was re-screened with a 700-bp cDNA fragment probe obtained from the amino-terminal coding end of the ERB4 clone (SER 5′ probe, Fig. 1). About 25 strongly positive signals were obtained from 300 000 recombinant phage. Ten clones were characterized; 3 contained cDNA inserts that were similar to ERB4 except that they

extended further into the amino terminal coding end and each had a small internal deletion relative to ERB4. One of these, ERB28, was sequenced and lacks a 196-bp sequence found in ERB4 (ins2, Fig. 1) that contains in-frame stop codons. The other 7 characterized cDNA clones identified by the 5′ SER probe appeared nearly identical to one another by restriction mapping, but differed significantly from ERB28. Characterization of these 'class 2' variant cDNA clones, represented by ERB20 in Fig. 1, is described in Results.

The *S. mansoni* cercariae cDNA library was also screened with the 5′ SER probe. Four positive phage plaques were identified from about 100 000 recombinants. Two were apparently identical to one another, based on restriction mapping, but different from SER cDNA clones previously isolated. Characterization of these 'class 4' variant cDNA clones, represented by CERB22 in Fig. 1, is also described in Results.

DNA sequencing. DNA from the plaque-purified recombinant phage was purified [20] and the cDNA inserts subcloned into plasmid vectors. Both strands of the cloned DNA were sequenced for all coding DNA. Sequencing was performed by the method of Sanger et al. [21] following further subcloning of various restriction fragments into M13 vectors [22]. Either the M13 universal primer or various SER-specific oligonucleotide primers were used.

Polymerase chain reaction studies. S. mansoni mRNA was converted to cDNA by reverse transcription [20]. The polymerase chain reaction was performed on 1 ng of cDNA using the following conditions: 20 mM $(NH_4)_2SO_4$/70 mM Tris-HCl/2 mM $MgCl_2$/100 mg ml^{-1} bovine serum albumin/0.1% Triton X-100/1 μM oligonucleotides, pH 8.8. The temperature cycling program employed 30 cycles, each having an annealing time of 1 min at 50°C, an extension time of 3 min at 72°C and denaturation for 1 min at 94°C. Oligonucleotides are 15- or 16-mers with a G:C content near 50%. The PCR reaction products were characterized by TBE-agarose gel electrophoresis [20].

Preparation of S. mansoni genomic DNA. Genomic DNA was isolated essentially as described by McCutchan et al. [23] with the following modifications. Frozen parasites were ground in a mortar and pestle sitting in dry ice. Following ethanol precipitation, the DNA was recovered by spooling onto a glass rod, washed once in ice cold 70% ethanol, and dried. The DNA was solubilized in TE (10 mM Tris-HCl/1 mM EDTA, pH 8.0) and quantified by ultraviolet light absorbance at 260 nm.

Preparation of genomic library. S. mansoni genomic DNA (5 μg) was digested with *Sau*3A (0.4 units) and one-third of the reaction was quenched by addition of phenol after 5, 10 and 15 min. These conditions had been precalibrated so that the average size after pooling of the 3 time points was about 20 000 bp. After phenol extraction, the DNA was ethanol precipitated and solubilized in TE. Library construction was performed by the partial end-filling method of Zabarovsky and Allikmets [24]. This method prevents self-ligation while permitting ligation of the genomic *Sau*3a DNA fragments into a partially end-filled *Sal*I restriction cleavage site of the vector; in our case the Syrinx 2A λ phage vector constructed by Lutz et al. [25]. After ligation and packaging, the recombinant phage (about 400 000) were amplified as described by Sambrook et al. [20].

Northern and Southern blotting. Poly(A)$^+$ RNA was resolved by formaldehyde-agarose gel electrophoresis while genomic DNA restriction digests were resolved by TAE-agarose gel electrophoresis [20]. In each case, nucleic acids were transferred to a Nytran filter (Schleicher and Schuell) by capillary transfer in 10 × SSC.

Probes were prepared by a modification of the random primer procedure of Feinberg and Vogelstein [26]. PCR was used to generate amplified DNA fragments representing: (1) the region of SER (codons 27–265) common to all

SER transcript classes; (2) class 2-specific cDNA (750 bp from the 3′ untranslated region of ERB20); or (3) *S. mansoni* triose phosphate isomerase cDNA (550 bp from a TPI cDNA clone; ref. 27). These fragments were labeled using the Multiprime kit as above, except that the random primers were replaced by 50 ng of each of the oligonucleotide primers that had been used to generate the PCR fragment. The radiolabeled DNA fragment was then purified on a 1 × 10-cm CL4B column (Pharmacia LKB, Uppsala, Sweden) prior to hybridization [20].

Results and Discussion

Cloning cDNA encoding an S. mansoni epidermal growth factor receptor homologue. An adult *S. mansoni* cDNA library was screened at low stringency with a probe prepared from the tyrosine kinase domain of the EGFR family oncogene, *erbB*. Two positive clones, ERB3 and ERB4 (Fig. 1), overlap at the region of tyrosine kinase homology and, together, contain about 6000 bp of unique DNA. Within the overlapping region, the DNA sequences are identical except for a 139-bp insert within ERB3 (ins1, Fig. 1). The insert is found after bp 3378 (GenBank M86396) and contains typical eukaryotic splice donor and acceptor sites that are strongly suggestive of an unspliced intron. From the DNA sequence and the homology with other EGFR homologues, we conclude that the 2 cDNA clones together contain nearly the complete coding DNA for SER but lack a portion of the amino-terminal coding end (see Materials and Methods). A small fragment of ERB4 was used as a probe to re-screen the cDNA library in an effort to obtain the coding DNA for the SER amino terminus. From this screen, a new clone, ERB28 (Fig. 1), was obtained that contains about 300 bp of additional sequence at the amino terminal coding end. Near the beginning of ERB28 is an open reading frame (ORF) that encodes a methionine followed by 18 hydrophobic amino acids suggestive of a secretory leader. This ORF continues through the entire 4-kb insert and combines with the ERB3 cDNA clone to encode the complete SER protein (GenBank M86396).

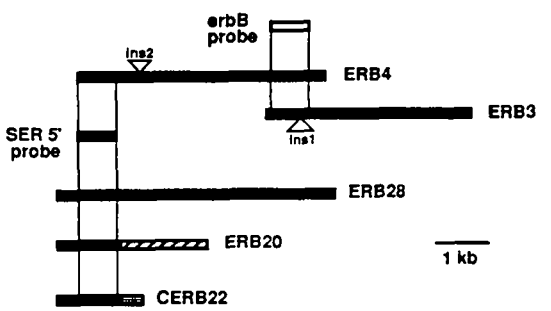

Fig. 1. Strategy for cloning of SER cDNA. The open box represents the original *erbB* probe used to identify the first 2 SER cDNA clones, ERB3 and ERB4 (black boxes). The regions homologous to the probe are oriented below (within vertical lines). Two sequence inserts, relative to the other clones, are shown as triangles. Insert ins1 from ERB3 and ins2 from ERB4 are discussed in the Results and Discussion. A second probe (SER 5′ probe) was prepared from the 5′ end of the ERB4 clone and used to identify additional cDNA clones, including the three represented below. ERB clones derive from the *S. mansoni* adult, size-selected cDNA library while the CERB clone derives from the cercarial cDNA library (see Materials and Methods). The positions where ERB20 and CERB22 differ from the SER sequences are shown as pattern-filled boxes.

Several cDNA clones were also obtained, based on homology to the SER 5′ probe, that differ from SER in their carboxy-terminal coding DNA and 3′ untranslated regions (see Materials and Methods). They are represented in Fig. 1 by clones ERB20 and CERB22. In addition, ERB4 was found to have a 195-bp variant insert (ins2, Fig. 1). These variant SER cDNAs are discussed later.

Amino acid sequence of SER. Combining sequence from cDNA clones ERB28 and ERB3, an ORF exists that predicts an SER protein of 1717 amino acids (Fig. 2). Substantial amino acid sequence homology between SER and the EGF receptor family of proteins can be found throughout. This alignment together with previously proposed structure models for the EGFR family [1,9] were used to define the SER domain assignments as shown in Fig. 2.

The tyrosine kinase domain is generally the most conserved domain within the EGFR

family of proteins and, based on computer comparisons, this was also the case for SER. Indeed the tyrosine kinase domain of SER is more closely related to other members of the EGFR family than to tyrosine kinases of the other types represented in the Swiss-Prot data base release 17 and thus clearly belongs within this subfamily. The alignment between the tyrosine kinase domains of SER, human EGFR (HER) and the EGFR homologues from *Drosophila* (DER) and *Caenohrabditis* (*let*-23) is shown in Fig. 3. All of the residues that were identified as virtually invariant within the protein kinase family by Hanks et al. [28] are conserved within each of the EGFR homologues, including SER. In addition,

Leader	MNFLLPTLVIIFALNCWSL	19
NH2 terminus	KFSISEYPWIKACRVWERDCSKPNPKHIQLTYIKFLYGGCTHIIGNLVICG	70
	LEKLENGSDPDLSFLEKIEDVSGYVYIGQNSVKTISLPSLKVIRGEPGYR	120
	IMNTSAALVISRNSLEILDLRSLTAIQRNDIVALNNQFLCNFGFTIDWEQ	170
	IFEDNRKQMFIPDRKEKTVSHAGCDIALRKYTDD	204
Cys domain 1	RTKHSCHGSCPVVNGRGYCWGPKPEMCQKMLKCANNPDNYCLGGRATTQPCLE	257
	ECLGGCETRPGNCRACKHAMNDGKCVSQCPPPLIVSREESRTVANPEFKYNFH	310
	DICVKNCPAPFLKSDSYCVIECDLNTQIPVNGTCKDCPKSGCPEHC	356
Intradomain 1 (LBD)	KEETIFVNGSLNILQSSSLRKFKSCVYYTGGLYISKESFQKSSLFPDPIQNVN	409
	ELYNLLHLKSIVGYIYFDLREAPEELKNLTFLENLESVVLEVKSQSPGAVITI	462
	MNGENIESFGFKSLTNIGGYVYLKNMPKLCYISALTKMLPVRMIDVQDEELCA	515
Cys domain 2	KRGHVCHSECLPELGCWGADANMCAHCCGLKAGEYCVSRCTDHPGFYELPTP	567
	FNHSILGKTTNNPVCRTLPLTKSDMAEMDEQAIIASVIPSKTCAICHPECAQ	619
	TCYGPNANQCVGECKHYQHGDTCLPECPRNTYIDPQTRHCLPCNESCSHIL	670
	TTGQNQLCSGPGNFLGLGGCETCWTVIQDKITNKYQCLPDDCPP	714
Intradomain 2	KHYTESYQTQDFINKEKIIVSSHSQVKKGELGGMI	749
Cys domain 3	RVCKPCHPFCDLCTANGTHASICHSCTHWWFKSECVEICPPAETYSLAGSD	800
	KELDNQEMFENDLITLSNNTQFSSNQLKEFTFLSNVSTTTNASIDDNQQE	850
	YKTSSLAAPVFLIKLKRTQRRCLLCHEQCIQGCSGPGPEDCVKCRNYQII	900
	LDEETNKFVCNSSCPEDRNHIFHGMCLTAEQNARLSGQTARELRNR	946
TM domain	ILIGVSVSVFIIIALVTIILVVCL	970
Kinase domain	KRKAEAEKIREQLRSAYTNLLEPDMKTQSVSREPNMGRLEMINQDDLFCD	1020
	FNSAPLGTGSFGAVYKGVWKVPKHALLRYNWHRGAQLDVAIKVILNDSPE	1070
	CSVTANPSSPFEAGNSSYSEEEAKRASVRANIEELLQEAKIMASVMHRHC	1120
	LPLIGICLSSERHCLVSIFVELGALDRYVKQHADELNSLTLLSWAEQIAD	1170
	GMSYLEMRGIIHRDLAARNVLVQTREHVQITDFGLAKMLERRDEDSVIVK	1220
	AGRVPIRWLAIETLQYGIYSHKTDVWSYGVTLWEIFTFGKRPYEDVDTVD	1270
	IKDHVIKGGRLTQPDICTLDVYMVLVKCWMEDYESRPTFIELMRTFNTFC	1320
	KTPGRYLYIEGD	1332
C-terminus	QYAINYFHNTNSGSGNFSSESHELQPMLSVRGIPDGGTTPHRNNSLHR	1380
	HHTMLTEPSMTRPYSSGKLLRALSDQPSERSDLFSVGQTGEHTETQLLLP	1430
	IRSNNSNSGNHSTWQSRQHGVGTGASSNTSDTNFSGLGRIWKRSNFHDKV	1480
	NTDNTSLLKQRKAPLASREDSWLNDIPRQSDHSESRSSVALSDPATASTT	1530
	AKTSEWSGISPYNQSRNHSNFGGKNLEEMNTKFSPHTHGNSSNSFGKNTV	1580
	SDYLLDPPPPPPVPRGLPDEYLQPKTKNPTNTSHAYSSALSKSMNYTELG	1630
	PPMVTNPSTTRDEYLSPNMQPPEEYLSPISGGFSVTNPEYLMETYGHPQQ	1680
	YPEPNTLKQSTNPASDSNSPLKFQDKTNQNESSSLEL	1717

Fig. 2. Complete amino acid sequence of SER predicted from the cDNA sequence. The domain separation is based on homologies to other characterized EGF receptors and homologues (see text). Cysteine residues are shadowed. Amino acid numbers are shown to the right. LBD, putative ligand binding domain; Cys, cysteine; TM, transmembrane.

residues that are uniquely conserved within the tyrosine kinase family of protein kinases are also conserved in SER. Together, these observations strongly suggest that SER contains a functional tyrosine kinase. Overall, SER tyrosine kinase domain homology (identity) to HER, DER and *let*-23 is 46%, 42% and 41% respectively. When compared with the other EGFR homologues, SER contains 2 sizeable inserts near the amino-terminus of the tyrosine kinase domain. The first is between subdomains I and II of the Hanks consensus, a region known to be variable in protein tyrosine kinases. The second large insert separates subdomains II and III and adds nearly 30 amino acids to a region varying by a maximum of two amino acids among the other characterized tyrosine kinases.

In general, the extracellular, ligand-binding domains of EGFR homologues from distantly related animals are poorly conserved. Yet several cysteine-rich regions, sharing distinct similarities with respect to their cysteine motifs, have been consistently noted within this domain [9,12]. Proteins of the EGFR family from mammals, birds and fish [1,8,29] have two roughly equal sized cysteine-rich regions. In EGFR homologues from arthropods and nematodes, the second cysteine-rich region is much larger than the first [9,12] and can be viewed, instead, as consisting of 2 distinct regions [30]. Extending these observations to SER, both of the extracellular, cysteine-rich regions are apparent (Fig. 2) and, as with the other lower eukaryotes, the second cysteine-rich region is larger than the first. Overall the extracellular domain of SER displays 26% homology with HER and 22% with DER.

The EGFR homologue cysteine-rich regions

```
HER-666    EPLTPSGEAPNQALLRILKETEFKKIKVLGSGAFGTVYKGLWIPEGEK-VKIPVAIKELRE
DER-867    EPLRPSNIGANLCKLRIVKDAELRKGGVLGMGAFGRVYKGVWVPEGEN-VKIPVAIKELLK
LET-864    PIDASVRPNMSRICLIPSSELQTKLDKKLGAGAFGTVFAGIYYPKRAKNVKIPVAIKVFQT
SER-1001   EPNMGRLEMINQDDLFC-----DFNSAPLGTGSFGAVYKGVWKVPK>12<QLDVAIKVILN
PROTKIN    ..............................G.G....V.................A.K....

HER-726    ATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPFGCLLD-YVREHKDNI
DER-927    STGAESSEEFLREAYIMASEEHVNLLKLLAVCMSSQMMLITQLMPLGCLLD-YVRNNRDKI
LET-924    DQSQT--DEMLEEATNMFRLRHDNLLKIIGFCMHDDGLKIVTIYRPLGNLQNFLKLHKENL
SER-1067   DSP>33<EELLQEAKIMASVMHRHCLPLIGICLSSERHCLVSIFVQLGALDRYVKQHADEL
PROTKIN    ............E.................................................

HER-786    GSQYLLNWCVQIAKGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGA-EEKEY
DER-987    GSKALLNWSTQIAKGMSYLEEKRLVHRDLAARNVLVRLL-AGEDHDFGLAKLLSS-DSNEY
LET-983    GAREQVLYCYQIASGMQYLEKQRVVHRDLATRNVLVKKFNHVEITDFGLSKILKH-DADSI
SER-1158   NSLTLLSWAEQIADGMSYLEMRGIIHRDLAARNVLVQTREHVQITDFGLAKMLERRDEDSV
PROTKIN    ..........................D.aarN............DFG.............

HER-846    HAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEK
DER-1046   KAAGGKMPIKWLALECIRNRVFTSKSDVWAFGVTIWELLTFGQRPHENIPAKDIPDLIEV
LET-1043   TIKSGKVAIKWLAIEIFSKHCYTHASDVWAFGVTCWEIITFGQSPYQGMSTDSIHNFLKD
SER-1218   IVKAGRVPIRWLAIETLQYGIYSHKTDVWSYGVTLWEIFTFGKRPYEDVDTVDIKDHVIK
PROTKIN    .......pikw...E...........D....G..............................

HER-906    GERLPQPPICTIDVYMIMVKCWMIDADSRPKFRELIIEFSKMARDPQRYLVIQGD
DER-1106   GLKLEQPEICSLDIYCTLLSCWHLDAAMRPTFKQLTTVFAEFARDPGRYLAILGD
LET-1103   GNRLSQPPNCSQDLYQELLRCWMADPKSRPGFEILYERFKEFCKVPQLFLENSNK
SER-1278   GGRLTQPDICTLDVYMVLVKCWMEDYESRPTFIELMRTFNTFCKTPGRYLYIEGD
PROTKIN    ..........................R.........................
```

Fig. 3. Tyrosine kinase domain homologies with other EGFR homologues. The sequence source and amino acid number are shown to the left. HER, human EGFR [4]; DER, *Drosophila* EGFR homologue [9]; LET, let-23 EGFR homologue [12]. Conserved residues are shaded. Gaps introduced to maximize homology are shown as dashes. Inserts are shown as <#> where # is the number of amino acids inserted. PROTKIN is the protein kinase family consensus of Hanks et al. [28] with invariant residues shown as capital letters and residues highly conserved only in tyrosine kinases in lower case.

retain distinct homologies to one another and have been postulated to result from a primordial duplication event [4]. Furthermore, within each of the cysteine-rich regions, a sub-repeat of about 40 amino acids has been previously noted [31]. With the availability of EGFR homologue amino acid sequence data from diverse taxa, the significance of this sub-repeat was re-explored. As shown in Fig. 4, a repeating cysteine motif is apparent within the cysteine-rich regions from each of the EGFR homologues and a consensus can be built that includes other conserved residues within the motif. Amino acid sequence conservation is more apparent between sub-repeats at similar locations in homologues

```
CONSENSUS  ..C..CHP.C...GCTGPG...C..C........CV..CP.

HER-161    GSCQKCDPSCPNGSCWGAGEENCQKLTKIICAQQCSGRCRG
XMR-165    RQCQKCDHGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRCRG
DER-207    RECPKCHESCTH-GCWGEGPKNCQKFSKLTCSPQCAGGRCY
LET-215    KSMAKCHESCND-KCWGSGDNDCQRVYRSVCPKSC-SQCFY
SER-205    RTKHSCHGSC>7<YCWCPKPEMCQKMLKCAN>5<CLGGRAT

HER-203    SPSDCCHNQCAA-GCTGPRESDCLVCRKFRDEATCKDTCPP
XMR-207    KPIDCCNEHCAG-GCTGPRATDCLACRDFNDDGTCKDTCPP
DER-249    KPRECCHLFCAG-GCTGPTQKDCIACKNFFDEAVSKEECPP
LET-258    SSYECCDSACLG-GCTGHGPKNCIACSKYELDGICIETCPS
SER-250    ATTQPCLEECLG-GCETRPG-NCRACQHAMNDGKCVSQCPP

HER-477    ATGQVCHALCSPEGCWGPEPRDCVSCRNVSRGRECVDKCNL
XMR-474    TENQTCNNECSEDGCW-PGPTMCVSCLHVDRGGRCVASCNL
DER-519    KNGTICSDQCNEDGCWGAGTDQCLTCKNFNFNGTCIADCGY
LET-515    TEQRVCDKNCNKRGCWGKEPEDCLECKTWKSVGTCVEKCDT
SER-516    KRGHVCHSEC>4<GCWGADANMCAHCCGLKAGEYCVSRCTD

HER-529    SECIQCHPEC>7<TCTGRGPDNCIQCAHYIDGPHCVKTCPA
XMR-525    GRCVQCHQEC>7<TCYGPGPANCSKSAHFQDGPQCIPRCPH
DER-569    RTCKICHPECR--TCNGAGADHCQECVHVRDGQHCVSECPK
LET-565    MKCERCSPECE--TCNGLGELDCLTCRHKT>11<CVHDCPV
SER-608    KTCAICHPELAQ-TCYGPNANQCVGECKHYQ>4<CLPECPR

HER-591    HVCHLCHPNCTY-GCTGPGLEGCP
XMR-587    GQCQPCHQNCTQ-GCSGPGLSGCR
DER-614    GVCRECHATCD--GCTGPK>7<CTTCNLAI>12<KDDKCPD
LET-618    NVCEKCHPTCYDNGCTGPD>7<CKQCKYAVKYENDTIFCLQ
SER-657    RHCLPCNESC>11<CSGPG>7<CETCWTVI>9<CLPDDCPP

DER-686    AVCRKCHPLCE--LCTNYG>5<CSKCTHYKRREQCETECPA
LET-691    TMCEKCSISCK--TCSSAG>6<KCVCKHVEY>8<CHDQCPV
SER-750    RVCKPCHPFCD--LCTANG>5<CHSCTHWWFKSECVEICPP

DER-735    RECFQRHPECN--GCTGPGADDCKSCRNFK>17<CTSKCPL
LET-748    TVCKKCHHEC>4<HCANGQSTGCQKCKNFTV>7<CVSECPK
SER-870    RRCLLCHEQCIQ-GCSGPGPEDCVKCRNYQ>10<CNSSCPE

CONSENSUS  ..C..CHP.C...GCTGPG...C..C........CV..CP.
```

Fig. 4. Conserved cysteine-rich domain sub-repeat of EGFR homologues. The sequence source and amino acid number are shown to the left. HER, human EGFR [4]; XMR, *Xiphophorus Tu* locus [8]; DER, *Drosophila* EGFR homologue [9]; LET, let-23 EGFR homologue [12]. Conserved residues are shaded. Gaps introduced to maximize homology are shown as dashes. Inserts are shown as < # > where # is the number of amino acids inserted. The consensus shows residues that are conserved in 10 or more of the 31 sub-repeats.

from different species than between sub-repeats within the same proteins, suggesting that evolutionary duplication of the element occurred within a primordial organism prior to divergence of the *Platyhelminths*.

The carboxy-terminal domain of SER has almost no significant homology to previously characterized EGFR homologues. It is distinguished by the disproportionate number of prolines, asparagines and hydroxylated amino acids. Probably as a fortuitous result, this intracellular domain, which should not contain carbohydrate, has numerous potential *N*-linked glycosylation sites. There are several tyrosines near the C-terminus that may be phosphorylation sites but the similarity to other EGFR homologues is too poor to permit specific predictions.

In higher eukaryotes, EGF receptors transfer a mitogenic signal from extracellular EGF or transforming growth factor-α through the membrane and into the cell [32]. Because of their oncogenic potential, it is likely that EGFR homologues such as mammalian *neu* [3] and *Xiphophorus Tu* [8] also have a role in mitogenic signal transfer. The genes for the EGFR homologues DER and *let*-23 both serve roles in determining cell fate during development [10–12], presumably through passage of an extracellular signal into the target cell expressing the receptor. Lacking the genetic tools available in other biological systems, approaches to the study of SER function in schistosomes are limited. Studies that determine the anatomical expression sites of SER might provide insight into function. It should also be possible to express functional SER within recombinant host cells to test SER for tyrosine kinase activity. Such a functional expression system would make possible an assay to detect ligands that stimulate the autophosphorylation activity of SER.

If recombinant SER has tyrosine kinase activity, it might aid attempts to establish a schistosome cell line, something that has eluded investigators up to now. Certain modifications to, or over-expression of, EGFR and *neu* can result in transformation within responsive cells [3,5,6,33]. Analogous modifications to the *Drosophila* EGFR homologue, DER, enhanced in vivo tyrosine kinase activity [34]. Perhaps introduction of a vector capable of over-expressing a similarly modified SER might help to immortalize schistosome cells. Chimeras of the EGFR ligand binding domain and neu tyrosine kinase domain can produce EGF-dependent transformation of cells [35] and chimeras of the *neu* ligand binding domain and DER tyrosine kinase domain functioned in insect cells [34]. A vector over-expressing a chimera between SER and the extracellular domain of EGFR might facilitate development of an EGF-dependent schistosome cell line.

It will be interesting to determine whether schistosome SER recognizes EGF within the mammalian host. A protein that binds EGF and that is recognized by antiserum against mammalian EGF receptor has been identified in trypanosomes [36]. But the poor homology between the extracellular domains of SER and HER, and especially poor homology (18%) within the putative ligand binding domain between the two cysteine-rich regions [37], would seem to suggest that SER recognizes a different ligand than the EGF ligand of HER. This issue can be addressed if it is possible to obtain functional, recombinant expression of SER within eukaryotic host cells.

SER variant cDNA clones. Three different variant SER cDNA classes were identified during the isolation of the complete SER coding DNA. One of the variant classes, class 2, was the predominant type of clone obtained from the adult male cDNA library when it was screened with a 5′ SER probe (see Materials and Methods). These clones, represented by ERB20 (Fig. 1), have about 1000 bp of identity with the *S. mansoni* EGFR homologue, SER (class 1), at the amino terminal coding end. At that point they diverge and the subsequent 1400 bases, ending in poly(A), bear no resemblance to class 1 cDNA (GenBank M86397). If class 2 transcripts are translated, the product would contain the amino terminus and most of the first cysteine-rich region of class 1 SER but have 15 unique amino acids at

the carboxy-terminus (Fig. 5). Since the polypeptide would contain the SER secretory leader but lack the transmembrane domain of SER, the class 2 product should be a secreted protein of about 35 kDa.

Another variant class, class 4, was identified

A.

```
Leader          MNFLLPTLVIIFALNCWSL                                        19

NH2 terminus    KFSISEYPWIKACRVWERDCSKPNPKHIQLTYIKFLYGGCTHIIGNLVICG        70
                LEKLENGSDPDLSFLEKIEDVSGYVYIGQNSVKTISLPSLKVIRGEPGYR        120
                IMNTSAALVISRNSLEILDLRSLTAIQRNDIVALNNQFLCNFGFTIDWEQ        170
                IFEDNRKQMFIPDRKEKTVSHAGCDIALRKYTDD                        204

Cys domain 1    RTKHSCHGSCPVVNGRGYCWGPKPEMCQKMLKCANNPDNYCLGGRATTQPCLE     257
                ECLGGCETRPGNCRACKHAMNDGKCVSQCPPPLIVSREESRTVANPEFKYNFH     310
                DICVKNCPAPFLKSDSYCVIECDLNTQIPVNGTCKDCPKSGCPEHC... CLASS 1 356
                       |
                       ATKTELPQFLFHDDP- CLASS 2
                       |
                       GVKRLRKQVLIQFSILVFYSLFLH- CLASS 4
                                                  |
                                                  RNMSTCRVDNVI- CLASS 5
```

B.

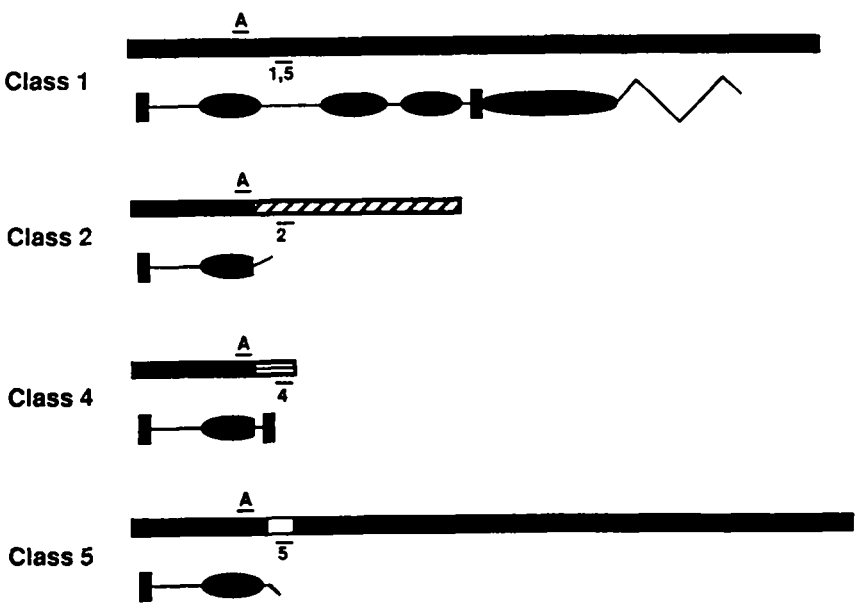

Fig. 5. Amino acid sequences and structures of the SER variants predicted from their cDNA sequence. (A) Amino acid sequence of the amino terminus of SER from Fig. 2 showing also the sequences of class 2, 4 and 5 and the position at which they diverge from SER. (B) Representation of the SER class 1, 2, 4 and 5 cDNA and translation products. Black boxes represent SER class 1 cDNA sequences while pattern-filled boxes represent SER variant cDNA sequences. Below the cDNA diagrams are representations of their translation products. Thickened vertical lines show the positions of secretory leader sequences or transmembrane sequences. Small ovals symbolize cysteine-rich regions and the large oval signifies the tyrosine kinase domain. Opposing PCR oligonucleotide positions (see text) are shown as short lines over or under the cDNA boxes.

within the *S. mansoni* cercariae cDNA library. This library, which was not size selected, was also screened with the 5′ SER probe. Two essentially identical 1400-bp clones were identified that differed from SER cDNA clones previously isolated from the *S. mansoni* adult library. One of these, CERB22 (Fig. 1), was sequenced. Like class 2, this cDNA is identical to class 1 for the first 1000 bp and diverges at precisely the same position, but the subsequent 3′ sequence is unique and ends with a poly(A) sequence (GenBank M86398). Class 4 transcripts (Fig. 5), if translated, should produce a translation product with the same SER amino acids as class 2, followed by 24 class 4-specific amino acids and resulting in a 37-kDa protein. Interestingly, the C-terminal 17 amino acids of the class 4 product are hydrophobic and suggestive of a membrane anchor.

The third SER variant cDNA, class 5, is represented by the original ERB4 cDNA. This cDNA contains a 196-bp sequence (ins2, Fig. 1) that is inserted 108 bp downstream from the point of class 2 and 4 divergence (GenBank M86399) and contains stop codons in all reading frames. The predicted translation product of class 5 SER transcripts is, thereby, slightly larger than the class 2 and 4 products (40 kDa) and consists of the amino terminus and virtually all of the first cysteine-rich region of class 1 SER, terminating with 12 unique amino acids (Fig. 5).

All 3 variant cDNAs, classes 2, 4 and 5, represent bona fide variant mRNAs, as demonstrated by the transcription and genomic analysis presented and discussed below.

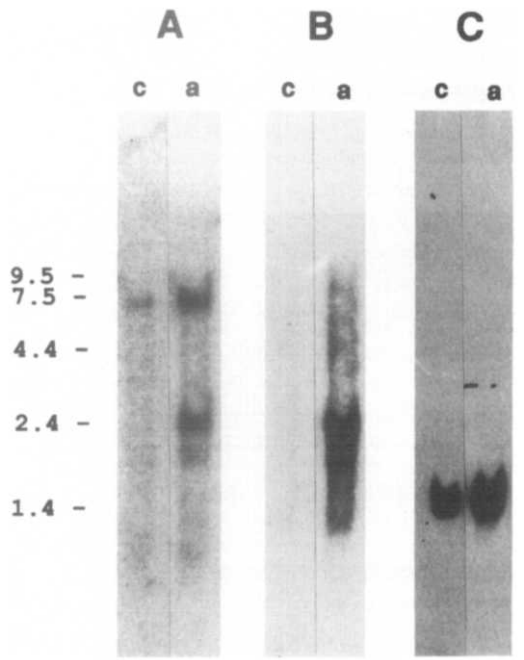

Fig. 6. Northern blots of schistosome mRNA. 1 μg of poly(A)$^+$ RNA from cercariae (c) or adult (a) mRNA was subjected to electrophoresis in a formaldehyde-agarose gel and transferred to a Nytran filter. The same filter was hybridized to the following probes: (A) an SER probe common to all transcript classes; (B) an SER class 2-specific probe; (C) a TPI probe.

Transcription of SER and SER variant mRNAs. Poly(A)$^+$ RNA was prepared from cercariae and adult schistosomes and resolved on formaldehyde agarose gels. After filter transfer, the Northern blot was hybridized to a series of different probes. The first probe, derived from SER sequences common to all 4 of the SER classes (Fig. 6A), recognized a 6.5-kb mRNA in both mRNA preparations. 6.5 kb is the size expected for class 1 transcripts based on the cDNA clones, although these transcripts may not resolve from class 5 transcripts on this Northern. A 2.5-kb, putative class 2 transcript was as abundant as the 6.5-kb transcript in the adult mRNA, but was not detected in cercarial mRNA. Neither of the schistosome mRNA preparations reveal the 1.5-kb transcript expected for class 4, indicating that it is, at most, a minor SER transcript.

Since class 2 cDNA was well represented in the adult male cDNA library, it seemed likely that the 2.5-kb transcript was class 2. This was confirmed when the Northern blot was rehybridized with a probe that recognizes 3′ untranslated sequences unique to class 2 (Fig. 6B). The class 2-specific probe again recognizes the 2.5-kb mRNA species, but even after long exposure times it is not detected in cercariae. We conclude that, while the class 1 SER transcript is present within both cercariae and adult, expression of the class 2 variant begins sometime after cercarial transformation into

schistosomula. A third probe, specific for *S. mansoni* triose phosphate isomerase (TPI), which, based on protein data, should be expressed at similar levels in cercariae and adult schistosomes (D. Harn, personal communication) is shown in Fig. 6C.

The polymerase chain reaction (PCR) was used to test for the presence of class 4 and 5 transcripts in cercariae and adult cDNA. Within each PCR reaction, one oligonucleotide was complementary to the sequences common to all SER cDNAs and the second, opposing PCR oligonucleotide was complementary to a region unique to one of the SER cDNA classes (see Fig. 5B). Amplified products representing all 4 classes were detected using *S. mansoni* adult cDNA but, with cercarial cDNA, only class 1 and class 5 products were visible on ethidium bromide stained agarose gels. The class 2 and 4 products could be detected when the products were analyzed by Southern hybridization using a radiolabeled, SER-specific probe (data not shown).

The oligonucleotides (A and 1,5 from Fig. 5B) used to detect class 1 transcripts by PCR also hybridize to class 5 transcripts on each side of the small DNA insert (ins2, Fig. 1) that distinguishes these 2 transcripts. This one pair of oligonucleotides in a single PCR reaction will amplify a different and distinct product for both transcript classes and thus the molar ratio of these products will meaningfully reflect the class 1 and class 5 transcript ratio within the template cDNA. In all PCR reactions, with both adult and cercarial cDNA, the class 1 product was present in at least a 10-fold excess over the class 5. We conclude therefore that class 1 is the predominant transcript and represents the major portion of the 6.5-kb species within the SER Northern blots.

SER variant forms are generated by alternative splicing. An *S. mansoni* genomic library was screened with SER cDNA probes homologous to the region encoding the first 830 amino acids of SER. Several positive clones were obtained and five were characterized by restriction mapping and Southern blotting. Four of the genomic clones overlapped, permitting deduction of the SER gene structure encompassing the entire region in which SER transcript diversity was observed (Fig. 7A).

All of the coding exons within clones GERB1, GERB18, GERB28 and GERB13 were completely sequenced to ascertain the intron locations and to determine the sequences at the intron/exon boundary. One genomic clone, GERB1, contains 4 exons that together encode SER translation initiation through codon 143. The 4 overlapping genomic clones encompass the 1670 bp of SER coding DNA encoding residues 234 to 790. Exons containing class 2- and class 4-specific coding DNA, and their 3′ untranslated sequences, are located between the SER exons labeled 'f' and 'g' in Fig. 7A. Sequences homologous to the 196 bp insert, which characterize class 5 transcripts, are present between exons 'g' and 'h'.

Characterization of the amino-terminal coding region of the SER gene resulted in several observations. First, the extreme 5′ end of SER mRNA was found to be encoded by 4 exons that are separated by three very small introns. These introns interrupt the codons for SER amino acid residues 12, 31 and 75 (Fig. 2) and have sizes of 34, 31 and 36 bp, respectively. Craig et al. [38] reported that the first 4 introns of the *S. mansoni* hypoxanthine-guanine phosphoribosyl transferase (HGPRT) gene are each between 31 and 42 bp, making them among the smallest introns to be characterized. The first intron of the *S. mansoni* TPI gene is only 42 bp (C. Shoemaker, unpublished observation) and, together with the SER and HGPRT results, suggest that small introns at the beginning of *S. mansoni* genes may be a common characteristic of this organism.

In contrast to the first few introns, all of the subsequent SER introns characterized to date are quite large. In fact, more than 50 kb of the SER gene has been characterized and this region encodes less than 40% of the SER mature transcript. The large size of the SER gene is consistent with observations made for the HER [39] and DER [30] genes but is significantly larger than observed for any of

the several other *S. mansoni* genes that have been characterized.

Analysis of the intron positions within the 5′ region of the SER gene permits additional notable observations. For example, neither the corresponding human (HER) nor *Drosophila* (DER) genes have an intron analogous to the first intron of SER within the leader peptide coding sequence. The second intron in the SER gene, however, is located in a precisely comparable position to the first intron of the human gene [39] and to the *Drosophila* intron that is involved in alternative splicing [30]. We have seen no evidence for alternative splicing at this intron in SER. Introns following exons d and e in Fig. 7A are present at the codons for SER amino acid residues 234 and 272. These residues are at nearly identical positions within the first and second cysteine 'sub-repeats' (Fig. 4) and this observation further suggests that this sub-repeat motif derives from a primordial sequence that has become replicated.

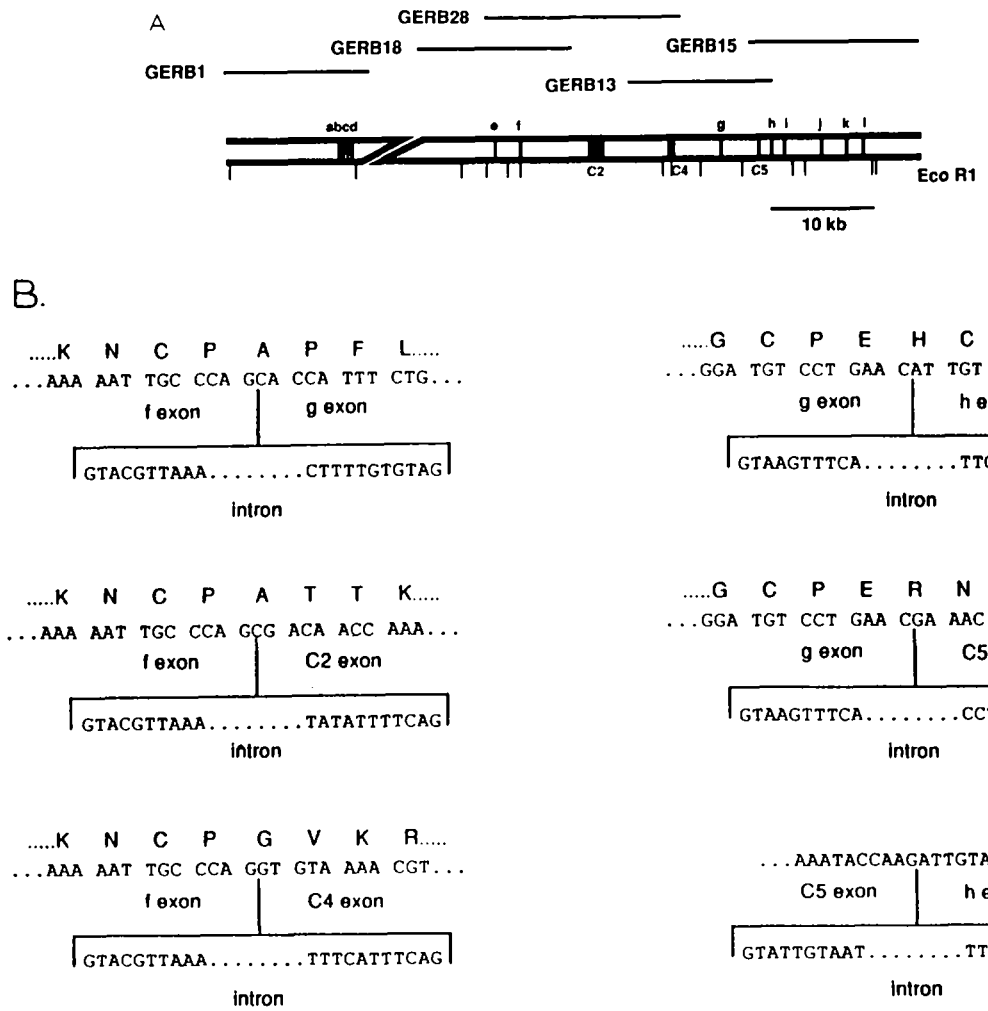

Fig. 7. SER gene organization and sequence of exon/intron boundaries for alternatively spliced exons. (A) The SER gene structure is shown below the relative positions of the 5 genomic clones used for gene characterization. Exons are represented as thickened vertical lines and are labeled above or below. *Eco*RI restriction cleavage sites are shown below as thin vertical lines. (B) Sequences of alternatively spliced exon/intron boundaries from the SER gene. The cDNA sequences and their translation (if coding) are shown above the intron sequences with the intron position indicated. Dots indicate that additional sequences are present that are not shown. The 3 splicing options found for the f exon are shown to the left. The 2 splicing options for both the g and h exons are shown on the right.

DNA sequence of the intron/exon boundaries for the SER exons f, g and h, as well as the class 2, 4 and 5-specific exons (C2, C4 and C5 in Fig. 7A) are shown in Fig. 7B. The intron between exons f and g interrupts the SER coding DNA at precisely the position at which the class 2 and 4 cDNAs diverge from class 1 cDNA. The C2 and C4 exons within this intron are completely contiguous with the corresponding class 2 and class 4 cDNAs up to the polyadenylation site, demonstrating that no additional introns are present within these exons. SER exon f ends with a canonical splice donor site and exon g, as well as C2 and C4, each have typical splice acceptor sites. These results strongly suggest that class 2 and class 4 transcripts arise by the alternative splicing of SER exon f to C2 or C4 and stop as a consequence of the class-specific polyadenylation sites that terminate the exons.

The 196-bp class 5 'insert' also appears to result from alternative splicing. These sequences are present within the intron following exon g of the SER gene (Fig. 7A) and are surrounded by canonical splice acceptor and donor sites (Fig. 7B). When the exon g splice donor joins this exon C5 instead of exon h, a new splice donor site is utilized to join the resulting 196-bp exon to the splice acceptor for SER exon h. This gives rise to the class 5 transcript.

Although the existence of the class 2, 4 and 5 exons within the SER gene strongly suggests that the observed transcript diversity arises through alternative splicing of the SER gene, it does not exclude the possibility that the variant transcripts might originate from unique genes. Several additional observations though confirm the alternative splicing mechanism. The complete sequences of the class 2, 4 and 5 exons were obtained and found to be identical to the sequences of the corresponding cDNAs. It is difficult to conceive that, were distinct genes present, they would have 100% identity within these 3′ untranslated sequences. Secondly, all the SER genomic clones isolated and characterized are consistent with the conclusion of a unique SER gene. Finally, a Southern blot of restriction enzyme digests of *S. mansoni* genomic DNA, probed with an SER class 1 cDNA fragment from the region of transcript diversity, is also consistent with the existence of only the one SER gene presented in Fig. 7A (data not shown).

During the last several years, alternative splicing has been revealed as a common mechanism employed by higher eukaryotes to generate transcript diversity. The mechanism has been demonstrated for a number of receptor genes, mostly in mammals, in which both secreted and membrane-bound forms of a receptor are expressed from the same gene by generating different mRNAs through an alternative splicing pathway [40–44]. The secreted form of the receptor generally consists of a nearly complete extracellular ligand-binding domain that lacks the transmembrane domain and C-terminus of the membrane-bound form of the receptor. It has been postulated that these soluble forms are still capable of binding ligand and, thereby, act in some way to regulate the response to the ligand [40,41,44].

Alternative splicing has also been observed within EGF receptor genes and their homologues. Several different EGFR transcript sizes are expressed in rats and one of the transcripts was shown to result from alternative splicing and encode a secreted form of the receptor [44]. The other rat transcripts and the differently sized EGFR transcripts found in humans [4] probably result from a similar mechanism. The *Drosophila* EGF receptor homologue (DER) gene undergoes alternative splicing of its 5′ exons to produce variant mRNAs encoding DER proteins that differ only at their amino terminus and have similar tissue distribution during development [30]. Multiple transcript sizes have also been observed for the EGFR homologues from *Xiphophorus* [8] and *Caenohrabditis* [12] that may be produced by an alternative splicing pathway.

In *S. mansoni*, the predicted products of the 3 variant SER transcripts would contain only the secretory leader, the amino terminal domain and most, or all, of the first cysteine-rich region. The most highly expressed SER variant mRNA in adult worms, class 2, as well

as the class 5 variant, apparently encodes soluble, secreted products. Another variant, class 4, encodes a truncated SER with a hydrophobic carboxy-terminus that may serve as a membrane anchor. Unlike most truncated, variant receptors that have been studied in higher eukaryotes, the SER variants would lack the putative ligand binding domain predicted by homology with the mammalian EGFR [29].

What role(s) might truncated SER products play? Without the capacity to bind ligand, the products, if expressed, should have no ability to regulate the cellular response to ligand through interference with normal receptor binding as proposed for truncated receptors in other systems. It is possible that the variant splicing is simply an accident that is tolerated as an insignificant metabolic burden for a parasite living in a nutrient-rich environment. The stage specificity of SER variant mRNA transcription and the limited diversity of the alternatively spliced transcripts argue against an accidental occurrence. One possibility is that the products of the truncated SER transcripts serve to interfere with a potentially damaging immune response against the receptor form of SER. Secreted, or membrane-bound, truncated forms of SER might operate, for example, by inducing antibodies against ineffective epitopes which then act as 'blocking antibodies' that somehow 'hide' epitopes which might otherwise induce a damaging immune response. Such a schistosome immune avoidance mechanism has been postulated by Grzych et al. [45] and Butterworth [46]. Similarly, the variant SER forms might serve to absorb otherwise damaging antibodies directed against SER. Another possible role for the truncated products are that they serve as ligands for a different receptor in a manner similar to EGF, which itself derives from a receptor-like precursor [31].

Acknowledgements

The authors wish to thank Dan Gannett for providing the *erbB* DNA used in preparing probes. We thank Tomas Sanchez and Gail Benson for technical assistance with DNA sequencing and PCR studies. The help of Jim Tobin and the current members of the lab in critical reading of the manuscript is also greatly appreciated.

This investigation received financial support from the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases and the John and Catherine McArthur Foundation.

References

1 Schlessinger, J. (1988) The epidermal growth factor receptor as a multifunctional allosteric protein. Biochemistry 27, 3119–3123.
2 Stoscheck, C.M. and King, L.E. (1986) Functional and structural characteristics of EGF and its receptor and their relationship to transforming proteins. J. Cell Biochem. 31, 135–152.
3 Bargmann, C.I., Hung, M. and Weinberg, R.A. (1986) The *neu* oncogene encodes an epidermal growth factor receptor-related protein. Nature 319, 226–230.
4 Ullrich, A., Coussens, J.S., Hayflick, T.J., Dull, A., Gray, A.W., Tam, J., Lee, J., Yarden, Y., Libermann, T.A., Schlessinger, J., Downward, J., Mayes, E.L.V., Whittle, N., Waterfield, M.D. and Seeburg, P.H. (1984) Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature 309, 418–425.
5 DiFiore, P.P., Pierce, J.H., Fleming, T.P., Hazan, R., Ullrich, A., Richter King, C., Schlessinger, J. and Aaronson, S.A. (1987a) Overexpression of the human EGF receptor confers an EGF-dependent transformed phenotype to NIH 3T3 cells. Cell 51, 1063–1070.
6 DiFiore, P.P., Pierce, J.H., Kraus, M.H., Segatto, O., King, C.R. and Aaronson, S.A. (1987b) erbB-2 Is a potent oncogene when overexpressed in NIH/3T3 cells. Science 237, 178–182.
7 Velu, T.J., Beguinot, L., Vass, W.C., Willingham, W.C., Merlino, G.T., Pastan, I. and Lowy, D.R. (1987) Epidermal growth factor-dependent transformation by a human EGF receptor proto-oncogene. Science 238, 1408–1410.
8 Wittbrodt, J., Adam, D., Malitschek, B., Mäueler, W., Raulf, F., Telling, A., Robertson, S.M. and Schartl, M. (1989) Novel putative receptor tyrosine kinase encoded by the melanoma-inducing *Tu* locus in *Xiphophorus*. Nature 341, 415–421.
9 Livneh, E., Glazer, L., Segal, D., Schlessinger, J. and Shilo, B. (1985) The *Drosophila* EGF receptor gene homolog: conservation of both hormone binding and kinase domains. Cell 40, 599–607.
10 Price, J.V., Clifford, R.J. and Schüpbach, T. (1989) The maternal ventralizing locus *torpedo* is allelic to *faint little ball*, an embryonic lethal, and encodes the *Drosophila* EGF receptor homologue. Cell 56, 1085–1092.

11 Schejter, E.D. and Shilo, B. (1989) The Drosophila EGF receptor homolog (DER) gene is allelic to *faint little ball*, a locus essential for embryonic development. Cell 56, 1093-1104.

12 Aroian, R.V., Koga, M., Mendel, J.E., Ohshima, Y. and Sternberg, P.W. (1990) The let-23 gene necessary for *Caenorhabditis elegans* vulval induction encodes a tyrosine kinase of the EGF receptor subfamily. Nature 348, 693-699.

13 Ramalho-Pinto, F.G., Gazzinelli, G., Howells, R.E., Mota-Santos, T.A., Figuierido, E.A. and Pellegrino, J. (1974) *Schistosoma mansoni*: defined system for stepwise transformation of cercariae to schistosomula in vitro. Exp. Parasitol. 36, 360-372.

14 Warren, K.S. and Peters, P.A. (1967) Comparison of penetration and maturation of *Schistosoma mansoni* in the hamster, mouse, guinea pig, rabbit and rat. Am. J. Trop. Med. Hyg. 16, 718-722.

15 Cox, R.A. and Smulian, N. (1983) A single-step procedure for the isolation of individual mRNA species from crude lysates of *Physarum polycephalum*. FEBS Lett. 155, 73-80.

16 Gubler, J. and Hoffman, G.R. (1983) A simple and very efficient method for generating cDNA libraries. Gene 25, 263-269.

17 Young, R.A. and Davis, R.W. (1983) Yeast RNA polymerase II genes, isolation with antibody probes. Science 222, 778-782.

18 Vogelstein, B. and Gillespie, D. (1979) Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA 76, 615-619.

19 Nilsen, T.W., Maroney, P.A., Goodwin, R.G., Rottman, F.M., Crittenden, L.B., Raines, M.A. and Kung, H. (1985) c-*erbB* activation in ALV-induced erythroblastosis: novel RNA processing and promoter insertion result in expression of an amino-truncated EGF receptor. Cell 41, 719-726.

20 Sambrook, J., Fritsch, E.M. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

21 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

22 Norrander, J., Kempe, T. and Messing, J. (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene 26, 101-106.

23 McCutchan, T.F., Simpson, A.J.G., Mullins, J.A., Sher, A., Nash, T.E., Lewis, F. and Richards, C. (1984) Differentiation of schistosomes by species, strain, and sex by using cloned DNA markers. Proc. Natl. Acad. Sci. USA 81, 889-893.

24 Zabarovsky, E.R. and Allikmets, R.L. (1986) An improved technique for the efficient construction of gene libraries by partial filling-in of cohesive ends. Gene 42, 119-123.

25 Lutz, C.T., Hollifield, W.C., Seed, B., Davie, J.M. and Huang, H.V. (1987) Syrinx 2A: an improved lambda phage vector designed for screening DNA libraries by recombination in vivo. Proc. Natl. Acad. Sci. USA 84, 4379-4383.

26 Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6-13.

27 Shoemaker, C., Gross, B., Gebremichael, A. and Harn, D. (1992) cDNA cloning and functional expression of the protective *Schistosoma mansoni* antigen, triose phosphate isomerase. Proc. Natl. Acad. Sci. USA 89, 1842.

28 Hanks, S.K., Quinn, A.M. and Hunter, T. (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241, 42-52.

29 Lax, I., Johnson, A., Howk, R., Sap, J., Bellot, F., Winkler, M., Ullrich, A., Vennstrom, B., Schlessinger, J. and Givol, D. (1988a) Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mouse cells, and differential binding of EGF and transforming growth factor alpha. Mol. Cell. Biol. 8, 1970-1978.

30 Schejter, E.D., Segal, D., Glazer, L. and Shilo, B. (1986) Alternative 5' exons and tissue-specific expression of the *Drosophila* EGF receptor homolog transcripts. Cell 46, 1091-1101.

31 Pfeffer, S. and Ullrich, A. (1985) Epidermal growth factor: is the precursor a receptor? Nature 313, 184.

32 Yarden Y. and Ullrich, A. (1988) Molecular analysis of signal transduction by growth factors. Biochemistry 27, 3113-3118.

33 Hudziak, R.M., Schlessinger, J. and Ullrich, A. (1987) Increased expression of the putative growth factor receptor $p185^{HER2}$ causes transformation and tumorigenesis of NIH3T3 cells. Proc. Natl. Acad. Sci. USA 84, 7159-7163.

34 Wides, R.J., Zak, N.B. and Shilo, B. (1990) Enhancement of tyrosine kinase activity of the *Drosophila* epidermal growth factor receptor homolog by alterations of the transmembrane domain. Eur. J. Biochem 189, 637-645.

35 Lehväslaiho, H., Lehtola, L., Sistonen, L. and Alitalo, K. (1989) A chimeric EGF-R-neu proto-oncogene allows EGF to regulate neu tyrosine kinase and cell transformation. EMBO J. 8, 159-166.

36 Hide, G., Gray, A., Harrison, C.M. and Tait, A. (1989) Identification of an epidermal growth factor receptor homologue in trypanosomes. Mol. Biochem. Parasitol. 36, 51-60.

37 Lax, I., Burgess, W.H., Bellot, F., Ullrich, A., Schlessinger, J. and Givol, D. (1988) Localization of a major, receptor-binding domain for epidermal growth factor by affinity labeling. Mol. Cell. Biol. 8, 1831-1834.

38 Craig III, S.P., Muralidhar, M.G., McKerrow, J.H. and Wang, C.C. (1989) Evidence for a class of very small introns in the gene for hypoxanthine-guanine phosphoribosyl-transferase in *Schistosoma mansoni*. Nucleic Acids Res. 17, 1635-1647.

39 Haley, J., Whittle, N., Bennett, P., Kinchington, D., Ullrich, A. and Waterfield, M. (1987) The human EGF receptor gene: structure of the 110-kb locus and identification of sequences regulating its transcription. Oncogene Res. 1, 375-396.

40 DiStefano, P.S. and Johnson, E.M. (1988) Identification of a truncated form of the nerve growth factor receptor. Proc. Natl. Acad. Sci. USA 85, 270-274.

41 Koike, S., Horie, H., Ise, I., Okitsu, A., Yoshida, M., Iisuka, N., Takeuchi, K., Takegami, T. and Nomoto, A. (1990) The poliovirus receptor protein is produced both as membrane-bound and secreted forms. EMBO J. 9, 3217-3224.

42 Leung, D.W., Spencer, S.A., Cachianes, G., Hammonds, R.G., Collins, C., Henzel, W.J., Barnard, R.,

Waters, M.J. and Wood, W.I. (1987) Growth hormone receptor and serum binding protein: purification, cloning and expression. Nature 330, 537 543.
43 Loosfelt, H., Misrahi, M., Atger, M., Salesse, R., Thi, M.T.V., Joliver, A., Guiochon-Mantel, A., Sar, S., Jallal, V., Garnier, J. and Milgrom, E. (1989) Cloning and sequencing of porcine LH-hCG receptor cDNA: variants lacking transmembrane domain. Science 245, 525-528.
44 Petch, L.A., Harris, J., Raymond, V.W., Blasband, A., Lee, D.C. and Earp, H.S. (1990) A truncated, secreted form of the epidermal growth factor receptor is encoded by an alternatively spliced transcript in normal rat tissue. Mol. Cell. Biol. 10, 2973-2982.
45 Grzych, J., Capron, M., Dissous, C. and Capron, A. (1984) Blocking activity of rat monoclonal antibodies in experimental schistosomiasis. J. Immunol. 133, 998 1004.
46 Butterworth, A.E. (1987) Immunity in human schistosomiasis. Acta Trop. 44 (Suppl. 12), 31-40.