Expression of a large number of novel testis-specific genes during spermatogenesis coincides with the functional reorganization of the male germ cell

CHRISTER HÖÖG*

Department of Cell and Molecular Biology, Medical Nobel Institute, Karolinska Institute, Stockholm, Sweden

ABSTRACT Structural and functional changes, essential for the formation of mature male germ cells, are known to take place at specific stages of the mammalian spermatogenic process. To identify novel genes that are involved in this developmental process, we have initiated a large-scale cDNA sequencing project (Höög, Nucleic Acids Res. 19: 93-98, 1991; Starborg et al., Mol. Reprod. Dev. 33: 243-251, 1992; Yuan et al., Biol. Reprod., 1995). Five-hundred and forty cDNAs have been isolated from testicular cDNA libraries and partially sequenced, 355 of which were found to represent genes previously not described in the literature. In addition, a number of cDNAs was found to be related to genes previously identified only in lower eukaryotes, suggesting that these murine genes encode functions that are evolutionary conserved. One of these murine cDNAs was related to the Aspergillus nidulans BimE gene, a putative cell cycle checkpoint regulator (Starborg et al., J. Biol. Chem., 1994). Southern blot analysis revealed that the murine BimE-related gene is strongly conserved in mammals. RNA blotting experiments of 361 novel murine cDNAs have identified 52 cDNAs that are expressed only during spermatogenesis, 36 of which are expressed only in spermatids, and 16 cDNAs that are expressed in both spermatocytes and spermatids. A survey of the literature revealed 40 mammalian genes that have previously been shown to be expressed mainly during spermatogenesis, and together with our results, they define three dominating temporal patterns of gene expression during spermatogenesis, each pattern coinciding with known functional or structural changes occurring during this differentiation process.

KEY WORDS: cDNA sequencing, gene expression, spermatogenesis, testis

Introduction

Mammalian spermatogenesis provides an excellent model system for studying cellular differentiation (reviewed in Bellvé, 1979; Hecht, 1986; Handel, 1987; Willison and Ashworth, 1987; Russel et al., 1990). During spermatogenesis, mitotically dividing spermatogonia are first transformed into spermatocytes undergoing meiosis and then into spermatozoa carrying a haploid nuclear content. Spermatogonia are first formed during fetal life, but do not begin to differentiate until the first week after birth (Nebel et al., 1981). Type A spermatogonia are mitotically active throughout adult life and a fraction of them continuously becomes committed to meiosis. Meiotic prophase lasts for approximately 10 days and can be divided into four stages: leptotene, zygotene, pachytene and diplotene, and is completed by two reductive divisions. During spermiogenesis, which takes approximately two weeks, the round haploid cell differentiates into a motile spermatozoon, culminating with the release of the mature germ cell into the lumen of the seminiferous tubule. The maturation of male germ cells in the seminiferous epithelium of testis has been shown to occur in close contact with a somatic cell type, the Sertoli cell (reviewed in Russel and Griswold, 1993). Male germ cells undergo many novel stage-specific ultrastructural alterations during spermatogenesis. For example, during meiosis two unique macromolecular structures, the synaptonemal complex and the recombination nodule, become associated with the chromosomes. These two nuclear structures are assumed to be essential for the reduction of the ploidy level and for the generation of new combinations of genes, two essential tasks achieved during meiosis (reviewed in von Wettstein et al., 1984). However, the most dramatic morphological changes take place during spermiogenesis and involve the formation of the head and the tail of the sperm, the condensation of the nucleus and the elimination of excess cytoplasm as part of the residual body (reviewed in Bellvé and O’Brien, 1993). The sperm head contains several novel structures, e.g. the acrosome, the calyx and components required for sperm-egg interactions, whereas

Abbreviations used in this paper: MTEST, mouse testicular expressed sequence tags; EST, expressed sequence tags.
the flagellum and the manchette represent structures unique to the sperm tail.

It is likely, due to the novelty of the spermatogenic process, that many germ cell-specific gene products are involved in this differentiation process. Structural and functional changes that take place in male germ cells during spermatogenesis occur in parallel with changes in the expression of total RNA and stagespecific proteins (Monegi, 1965; Boltani et al., 1980; Kramer and Erickson, 1982; Gold et al., 1983b; Stern et al., 1983). Furthermore, approximately 20 murine genes have been shown to be expressed almost exclusively during spermatogenesis, including genes encoding structural and regulatory proteins, as well as enzymes (reviewed in Willison and Ashworth, 1987; Propst et al., 1988; Thomas et al., 1989; Erickson, 1990; Wolgemuth and Watrin, 1991). In addition, a number of genes undergo stage-specific splicing events during spermatogenesis, resulting in the production of novel transcripts (Willison and Ashworth, 1987), e.g. two unique forms of lamins have recently been shown to be produced during meiosis (Furukawa and Hotta, 1993; Furukawa et al., 1994). As it is likely that the identification of additional germ cell-specific cDNAs gives further insight into the molecular details of this differentiation process, several different methods have been used to isolate genes that are active predominantly in germ cells. One such approach, differential cDNA screening, has led to the identification of a number cDNA clones that are expressed only in germ cells (Kleene et al., 1983; Dudley et al., 1984; Thomas et al., 1989). In these studies, cDNA libraries were screened with labeled cDNA prepared from male germ cells, selecting primarily for highly abundant cDNA molecules. An alternative approach, which also facilitates the isolation of rare cDNA molecules, was recently developed (Adams et al., 1991; Höög, 1991). In this strategy, called the expressed sequence tag (EST) approach (Adams et al., 1991), a very large number of cDNA clones are isolated at random from cDNA libraries and partially sequenced. More than 10000 different genes have been identified from different organisms using the EST approach (reviewed in Southern, 1992; Sikela and Auffray, 1993). However, to investigate the relative expression level of individual transcripts in different tissues, or at different stages of a differentiation process, a very large number of cDNAs would have to be isolated using this random cDNA isolation approach.

We have initiated a project combining the EST approach with Northern blot analysis, with the primary goal to identify novel germ cell-specific cDNA molecules (Höög, 1991; Starborg et al., 1992; Yuan et al., 1995). We have partially sequenced 540 cDNAs isolated from testicular cDNA libraries and identified 355 cDNA clones, representing genes previously not described in the literature. Several cDNA clones were also isolated that were found to be related to genes that have previously been identified only in lower eukaryotes, suggesting that further analysis of this group of cDNAs could give valuable information about functions conserved throughout evolution. One of these murine cDNA clones has been shown to be related to a cell cycle checkpoint regulator, BimE (Starborg et al., 1994). It is shown here that the murine BimE-related gene is strongly conserved in mammals, suggesting that it performs an important function in higher eukaryotes. A survey of the literature identified 40 mammalian genes that have been shown to be expressed mainly during spermatogenesis. We have analyzed the expression of 361 novel murine cDNAs by RNA blotting experiments and identified 52 cDNAs that appear to be expressed only in germ cells, 16 of which were found to be expressed in both meiotic and haploid cells, and 36 cDNAs that were found to be expressed only in haploid cells. These results strongly suggest that the majority of them participate in the formation of the haploid male germ cell.

### Table 1

<table>
<thead>
<tr>
<th>Match category</th>
<th>(No.)</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No database match</td>
<td>355</td>
<td>68</td>
</tr>
<tr>
<td>Mouse match</td>
<td>67</td>
<td>12</td>
</tr>
<tr>
<td>Non-mouse match</td>
<td>66</td>
<td>12</td>
</tr>
<tr>
<td>EST match</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>Repeat sequences</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>Mitochondrial rRNA</td>
<td>3</td>
<td>&lt;1</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

Partial cDNA sequences taken from Höög (1991) and Li et al. (1994) were compared to the EMBL data library using the program FASTA. MTESTs for which no identical mouse sequences were identified have been submitted to the EMBL data library (accession numbers Z31021-Z31347; X61801-X61899; X58485; X80169).

---

**Fig. 1. Southern blot analysis of genomic DNA from different organisms.**

(A) A filter containing BALB/c mouse kidney genomic DNA, digested with three restriction enzymes, BamH1, HindIII and EcoR1, was hybridized with a 992 bp DNA fragment from the TSG24 cDNA sequence (Starborg et al. in press). After hybridization, the filter was washed as described in Materials and Methods, and exposed to X-ray film for 3 days. Size markers in kilobases are indicated to the left. (B) A filter containing EcoR1 digested genomic DNA from human, Rhesus monkey, Sprague Dawley rat, BALB/c mouse, cow, dog, rabbit and chicken, was hybridized, washed and exposed as described in (A). Size markers in kilobases are indicated to the right.
Fig. 2. Expression profile of functional categories of ESTs and MTESTs expressed in testis and brain. 106 MTESTs from mouse testis (Yuan et al., 1995) and 879 ESTs from human brain (Adams et al., 1993), were divided into functional categories as described previously (Okubo et al., 1992; Adams et al., 1993a). The relative number of genes in each functional category has been calculated and is displayed as a percentage. The number of genes in each category is indicated above each bar. Left bar in each pair represents brain data.

Results and Discussion

Sequencing of cDNA clones and comparison to databases

We have isolated 540 cDNAs from juvenile and adult testicular cDNA libraries, and partially sequenced them from their 5' end by using a fluorescently labeled vector-specific sequencing primer integrated into a cycle sequencing protocol (Höög, 1991; Yuan et al., 1995). These partial cDNA sequences were called mouse testicular expressed sequence tags (MTESTs).

The 540 MTESTs were analyzed for similarities with sequences deposited in the EMBL sequence data library and the results have been summarized in Table 1. Since only a small number of the estimated 50000-100000 murine genes have been characterized yet, it could be expected that the majority of the partially sequenced cDNAs should represent genes previously not characterized. Indeed, 68% of the MTESTs were found to represent novel genes previously not described in the literature (Table 1). Another 12% was found to be identical to mouse sequences already characterized, whereas 12% was found to be similar to sequences previously described in other organisms. Other EST sequencing projects, analyzing gene activity in different mammalian tissues, have given very similar results, i.e. approximately 60% of all randomly isolated cDNA clones appears to represent novel genes (Adams et al., 1991, 1992, 1993a,b; Khan et al., 1992; Okubo et al., 1992).

The most exciting outcome of the EST approach has been the discovery of a large number of mammalian cDNA clones that display similarities to genes previously identified only in lower eukaryotes, suggesting that the functions of the encoded proteins are conserved. We have isolated several murine cDNA clones that have previously been identified only in sea urchins, Drosophila, chicken and in yeast, e.g. genes encoding an actin filament-associated protein, a cytoplasmic dynein, a microtubule-associated protein, kinesin-2, a ubiquitin-conjugating protein and one gene affecting the production or stability of mRNAs, suppressor of forked (Yuan et al., 1995). One of the analyzed cDNA clones was found to be related to the Aspergillus nidulans BimE gene, a gene that has been shown to be essential for cell cycle progression (Osmani et al., 1988; Engle et al., 1990). The murine cDNA clone related to BimE, TSG24, has been completely sequenced and shown to encode a 216 kD protein that is uniformly expressed throughout the cell cycle (Starborg et al., 1994). To investigate the evolutionary conservation of this gene in other organisms, a set of Southern blot experiments was performed. Mouse genomic DNA was digested with several different restriction enzymes and hybridized with a labeled DNA fragment isolated from the TSG24 cDNA clone (Fig. 1A). The simple band pattern obtained using reduced stringency conditions, strongly suggests that the TSG24 gene occurs as a single copy gene in the mouse haploid genome. Low-stringency hybridization with the TSG24 DNA fragment was also performed on EcoR1 digested DNA isolated from other organisms, revealing that the TSG24 gene was conserved in humans, monkeys, rats, dogs, cows and rabbits, i.e. in all mammals investigated (Fig. 1B). No cross-hybridizing band was observed in DNA isolated from chickens. The sequence conservation of the TSG24 gene strongly suggest that the encoded protein performs an important cellular function.

Comparative expression analysis

Of the 540 MTESTs identified here (Höög, 1991; Starborg et al., 1992; Yuan et al., 1995), only a small number of them has previously been shown to be expressed in testis, reinforcing the fact that the EST approach is a powerful method to rapidly describe the transcriptional and functional activities of a tissue (Okubo et al., 1992; Adams et al., 1993a). The expression profile of mouse testis (Yuan et al., in press) was compared with EST data described for human brain (Adams et al., 1993a), to find out whether this type of data could be used to identify differences in gene expression patterns (Fig. 2). Many similarities were observed between testis and brain, i.e. the relative frequency of expressed genes encoding cell surface proteins, or proteins involved in transcription, translation as well as in signal transduction, was approximately the same.

Interestingly, some striking differences in the expression patterns between brain and testis could also be observed (Fig. 2). For example, more than three times as many testicular cDNAs were found to encode proteins involved in the organization of the genome (DNA metabolism). This observation is in agreement with the fact that developing germ cells, in contrast to brain cells, are known to undergo dramatic nuclear changes (reviewed in Bellvé, 1979; Hecht, 1986; Handel, 1987; Willison and Ashworth, 1997; Russel et al., 1990). The result of this comparison suggests that it should be possible to use EST sequence data to identify differences in gene expression patterns between different tissues and to link these differences to particular functional or structural processes unique to these tissues.
Identification of male germ cell-specific cDNAs

The goal of this project has been to identify genes that are expressed predominantly in germ cells to better understand functions that are unique to germ cells. The expression of 361 MTESTs in different tissues, in isolated Sertoli cells and in interstitial testicular cell types, has been investigated using Northern blot methodology (Starborg et al., 1992; Yuan et al., 1995). As an internal control of these experiments, genes that have previously been shown to be expressed predominantly in germ cells were included. Fifty-two novel murine MTESTs were found to have expression patterns similar to control genes, suggesting that these MTESTs represent genes that are expressed predominantly in male germ cells (Starborg et al., 1992; Yuan et al., 1995). To be able to investigate the temporal expression pattern of the 52 novel MTESTs during spermatogenesis, RNA was isolated from purified germ cells. Spermatocytes and round haploid spermatids were obtained from mouse testis after centrifugal elutriation and Percoll density gradient centrifugation (Bellvé et al., 1977; Meistrich et al., 1981; Heyting et al., 1985; Bucci et al., 1986). The purity of the two fractions was determined microscopically and both fractions were found to be more than 95% homogenous (Fig. 3). Thirty-six MTESTs were found to be expressed only in haploid cells, as revealed by Northern blot analysis, whereas the remaining 16 MTESTs were found to be expressed in both spermatocytes and spermatids (Starborg et al., 1992; Yuan et al., 1995).

A comparison of our data to previously published data, should give an insight into the global patterns of gene expression during spermatogenesis. A survey of the literature revealed 40 mammalian genes that have been shown to be expressed mainly during spermatogenesis. Eleven of these genes have been shown to be expressed only in haploid cells during spermiogenesis, another 18 genes in both meiotic and haploid cells, 9 genes only during meiosis, whereas 2 genes have been shown to be expressed throughout spermatogenesis. Interestingly, no germ cell-specific genes have yet been shown to be expressed only in spermatogonial cells, suggesting that the differentiation of this cell type mainly involves ubiquitously expressed genes. The temporal expression patterns of the germ cell-specific cDNAs identified in the literature, as well as of those anonymous cDNAs identified in our studies, suggest that the majority of them are involved in the transformation of the meiotic cell into a haploid cell, as well as in the functional reorganization of the haploid cell. The number of germ cell-specific genes activated at specific stages of spermatogenesis therefore appears to be directly correlated to the degree of structural and functional activities that are taking place at these stages. Thomas et al. (1989) have previously defined, using 18 cDNAs isolated from testis, four different phases of gene activity during spermatogenesis. In addition to the expression patterns described here, they have also shown that gene activation during meiosis can be divided into an early and a late phase.

Coordination of gene expression during spermatogenesis requires both stage- and cell-specific regulatory activities and are thought to be achieved at both the transcriptional and the translational level. Translational regulatory mechanisms have been shown to control the temporal appearance of several polypeptides during spermatogenesis (reviewed in Erickson, 1990; Hecht, 1990). Analysis of the translational activity of genes that are transcriptionally active in both spermatocytes and spermatids has revealed two different translational patterns, i.e. mRNAs that are translated in both spermatocytes and spermatids (Ayer-LeLievre et al., 1988), and mRNAs that are translated only in spermatids (Gold et al., 1983a; Kwon and Hecht, 1991). We have...
carefully analyzed the expression of the haploid-specific cDNAs identified by Yuan et al. (1995), during spermiogenesis. The expression of these genes was first detected in early round spermatids and the transcript level remained high also in the transcriptionally inactive elongated spermatids, suggesting that they are subject to translational control (Penttilä et al., 1995). Further analysis of these sequences should be valuable for our understanding of this novel regulatory process.

Today, little is known about the mechanisms of transcriptional regulation of gene expression during spermatogenesis. It is likely that genes expressed coordinately at specific stages of spermatogenesis are regulated by a common set of transcription factors, and that these factors interact with their target genes through common cis elements present in the promoter-regions of these genes, similar to what has been shown to occur in somatic cells (reviewed in Mitchell and Tjian, 1989). A number of transcription factors have been implicated in the regulation of genes expressed during spermatogenesis (Wolgemuth et al., 1987; Cunliffe et al., 1990; Nagamine et al., 1990; Rogers et al., 1991; Chowdury et al., 1992; Noce et al., 1992; Andersen et al., 1993; Boissonneaut and Lau, 1993; Delmas et al., 1993; Ito et al., 1993). Comparison of promoter sequences from genes expressed during spermatogenesis has revealed common sequence elements, and the use of transgenic mice, as well as the development of in vitro transcription systems, have made it possible to correlate promoter structures with DNA-protein interactions occurring only in spermatids (Bunick et al., 1990; Tamura et al., 1992; van der Horn and Tarnasky, 1992; Howard et al., 1993; Zambrowicz et al., 1993, 1994; Zambrowicz and Palmiter, 1994). Less is known about the regulatory mechanisms that control gene expression during meiosis. The SCP1 gene, encoding a synaptonemal complex protein, has been shown to be expressed only in spermatocytes (Mewissen et al., 1992). We have isolated the promoter region of the SCP1 gene and are currently characterizing this promoter in transgenic mice, to define the region(s) that mediates meiosis-specific gene expression (Liu and Höög, unpublished results).

We have identified 52 novel murine MTESTs that appear to be expressed only in germ cells (Höög, 1991; Starborg et al., 1992; Yuan et al., 1995). One of them, MTEST638, was found to be 50% identical at the amino acid level to an intermediate filament-like sea urchin protein, tektin A1 (Yuan et al., 1995). The tektin A1 protein forms an extended polymer that interacts with the flagellar microtubules found in mature sea urchin sperm (Norrander et al., 1992). The expression pattern of the related murine gene suggests that it could perform a similar function in murine spermatozoa. To be able to define the functions for the remaining germ cell-specific MTESTs, several different experimental approaches are in progress. Sequencing of the full-length cDNAs for these MTESTs will reveal whether they contain conserved sequence motifs that would give an insight into the function of these genes. Antibodies are also currently being developed against the open reading frames found within these MTESTs. The antibodies will be used to determine the subcellular localization of the encoded proteins within the developing germ cells, and to determine whether these proteins are part of germ cell-specific structures. Due to the conservation of the spermatogenic process in mice and humans, parallel analysis of genes and gene products from these two model organisms should also give valuable insight into the human spermatogenic process.

### Materials and Methods

#### DNA analysis

Isolation and sequencing of cDNA clones from mouse juvenile and adult testis has been previously described (Höög, 1991; Yuan et al., in press). Genomic filter hybridization was performed using two commercial filters (Clontech) containing different DNA samples. The filters were hybridized according to standard procedures (Sambrook et al., 1989). The probe consisted of a PCR amplified fragment (Saiki et al., 1988) extending from position 5161 to position 6053 of the TSG24 cDNA.
sequence (Starborg et al., 1994). After hybridization, the filters were washed twice with 0.3 M NaCl, 3 mM sodium citrate, pH 7.0 and 0.05% SDS, for 10 min at 50°C, and exposed to X-ray film at -70°C with an intensifying screen for 3 days.

isolation of testicular cell types

Testes taken from 4-5-week-old mice were decapsulated and cell suspensions were obtained by trypsinization of the seminiferous tubules (Bellvé et al., 1977). Pachytene spermatocytes and round haploid cells were isolated by centrifugal elutriation, followed by Percoll density gradients (Bellvé et al., 1977; Meistrich et al., 1981; Heyting et al., 1985; Bucci et al., 1986). A sample from each fraction was fixed in Carnoy’s fixative, stained with May and Grunewald’s stain and with Giemsa stain. The purity of the fractions was determined microscopically and each cell fraction was found to be more than 95% homogeneous.

Acknowledgements

I thank Jian Guo Liu and Li Yuan for providing information prior to publication. I also thank Maria Starborg for help with the genome blot. I thank Jihan Guo Liu and Li Yuan for providing information prior to publication. I also thank Maria Starborg for help with the genome blot.

References


HOWARD, T., BALOGH, R., OVERBEEK, P. and BERNSTEIN, K.E. (1993). Sperm-specific expression of angiotensin-converting enzyme (ACE) is mediat-


