A new approach to the study of haematopoietic development in the yolk sac and embryoid bodies

M. Jorge Guimarães1,*, J. Fernando Bazan1, Albert Zlotnik1, Michael V. Wiles2, J. Christopher Grimaldi1, Frank Lee1,† and Terrill McClanahan1

1DNAX Research Institute of Molecular and Cellular Biology, 901 California Avenue, Palo Alto, CA 94304, USA
2Basel Institute for Immunology, Grenzacherstrasse 487, CH4005 Basel, Switzerland

*Author for correspondence
†Present address: Millenium Pharmaceuticals, 640 Memorial Drive, Cambridge, MA 02139, USA

SUMMARY

To understand the mechanisms that control the differentiation of uncommitted mesoderm precursors into haematopoietic stem cells (HSCs) and the activation of haematopoiesis, we conducted a study to identify genes expressed at the earliest stages of both in vivo and in vitro haematopoietic development. Our strategy was to utilize Differential Display by means of the Polymerase Chain Reaction (DD-PCR) to compare patterns of gene expression between mRNA populations representing different levels of haematopoietic activity obtained from the mouse embryo, embryoid bodies (EBs) and mouse cell lines. We report the molecular cloning of two groups of genes expressed in the yolk sac: a group of genes expressed in the day-8.5 yolk sac at higher levels than in the day-8.5 embryo proper and up-regulated during EB development, and another group of day-8.5 yolk sac genes not expressed in the day-8.5 embryo proper or in EBs. Specifically, we describe the molecular cloning of the first nucleobase permease gene to be found in vertebrates, yolk sac permease-like molecule 1 (Yspl1). The Yspl1 gene has the unique property of encoding both intracellular, transmembrane and extracellular protein forms, revealing novel aspects of nucleotide metabolism that may be relevant during mammalian development.

Key words: cell differentiation, gene expression, haematopoiesis, haematopoietic stem cells, yolk sac, permeases, polymerase chain reaction

INTRODUCTION

Haematopoiesis is the process by which all blood cells are formed from multipotential, undifferentiated haematopoietic stem cells (HSCs). The first active site of haematopoiesis occurs in the yolk sac, both in avians (Dieterlen-Lièvre et al., 1979) and in mammals (Moore and Metcalf, 1970), at approximately day 7.5 of gestation. In mammals, it has been suggested that all haematopoietic activity results from the colonization of the embryo with cells that migrate from the yolk sac to the fetal liver after the activation of circulation by day 8.5. These cells would later colonize the bone marrow and be responsible for the formation of blood cells for the entire life of the organism (Moore and Metcalf, 1970). HSC activity was defined in these studies as the formation of day-8 spleen colony-forming units (CFU-S) - macroscopic colonies on the spleen of recipient mice at 8 days posttransplantation (McCulloch and Till, 1964), an assay later shown to detect only committed haematopoietic progenitors (Jones et al., 1990). Recent studies in the mouse showed that the splancnopleura (Godin et al., 1993), para-aortic (Cumano et al., 1993) or AGM region (delimited by the aorta, gonada and mesonephros, Medvinsky et al., 1993), contains haematopoietic activity earlier than the fetal liver. Moreover, complete multilineage, long-term repopulation of irradiated mice with cells derived from the AGM region was reported (Muller et al., 1994).

ES cells are derived from the inner cell mass of blastocysts and appear to resemble the primitive ectoderm of the postimplantation embryo (Evans and Kaufman, 1981; Martin, G. R., 1981; Robertson et al., 1986). Culture systems for ES cells that allow their differentiation in vitro into embryoid bodies (EBs) containing haematopoietic activity have been described (Doetschman et al., 1985; Lindenbaum and Grosveld, 1990; Wiles and Keller, 1991; McClanahan et al., 1993; Nakano et al., 1994; Johansson and Wiles, 1995).

Our goal was to identify differentially expressed genes at the early stages of haematopoietic development, both in vivo and in vitro, to find candidate genes involved in haematopoietic commitment. The application of the technique of Differential Display by PCR (DD-PCR) allowed the comparison of gene expression patterns between multiple mRNA populations (Liang and Pardee, 1992; Guimarães et al., 1995). Here we report the molecular cloning of two groups of genes preferentially expressed in the yolk sac: Group 1 genes are up-regulated during EB development, whereas Group 2 yolk sac genes are not expressed in EBs or the day-8.5 embryo proper. Specifically, we describe a Group 2 gene which constitutes the first vertebrate member of...
a family of nucleobase permeases described in yeast and bacteria. Moreover, we demonstrate the existence in the yolk sac of cDNAs that encode intracellular, transmembrane and extracellular protein forms of this novel yolk sac gene designated Yspl1, yolk-sac permease-like molecule 1. We describe an efficient strategy for the identification of differentially expressed genes during early development that can be extended to inspire similar models for the study of other developmental systems.

**MATERIALS AND METHODS**

**Animals**

Timed pregnant ICR female mice were used (Harlan Sprague Dawley, Indianapolis, Indiana). The morning of the day when the vaginal plug was found was designated as day 0. Somite pairs were carefully counted and 3- to 6-somite stage embryos were designated as 8.0-day embryos (Hogan et al., 1986). Mice were killed by CO2 narcosis. Embryos were dissected under a dissecting microscope and samples were collected in phosphate-buffered saline (PBS) + 5% (vol/vol) fetal calf serum (FCS; Sigma, St. Louis, MO) and kept on ice during dissection, centrifuged at 1,000 rpm for 5 minutes, supernatants were aspirated and pellets were frozen in dry ice before storage at −80°C.

Three week old Balb/c male mice were killed, as above, for the extraction of multiple organs and peri-visceral abdominal fat from the splenic region. Femurs and tibiae were flushed with PBS + 5% FCS and bone marrow cells filtered through a 70 μm nylon Cell Strainer (Becton Dickinson, Franklin Lakes, N. Jersey), centrifuged, and processed as above.

**Cell culture conditions**

ES cells were used and EBs were prepared as previously described (McClanahan et al., 1993). STO cells (American Type Cell Culture - ATCC CRL 1503) were maintained in Dulbecco’s Modified Eagle’s medium + 5% FCS. N2a cells (American Type Cell Culture - ATCC CCl 131) were kept in Dulbecco’s Modified Eagle’s medium + 10% horse serum (GIBCO BRL, Grand Island, NY) containing 100 U/ml penicillin, 100 μg/ml streptomycin, FDCPmixA4 cells (American Type Cell Culture; Ford et al., 1971) were maintained in Iscove Modified Eagle medium + 20% horse serum containing 50 mM 2-mercaptoethanol, 100 U/ml penicillin, 100 μg/ml streptomycin and 100 U/ml mouse IL-3 (DNAX Research Institute, Palo Alto, CA). All cell lines were plated in tissue culture dishes at a concentration of 1×10⁵ cells/ml and kept at 37°C in a humidified atmosphere with 5% CO2.

**RNA extraction**

RNA was isolated from cells, embryonic tissues, and adult organs using RNAzol solution (Tel-test, Inc, Friendswood, TX) according to manufacturer’s instructions. Heart and testis total RNAs were purchased from Clontech, Palo Alto, CA.

**Differential display by PCR (DD-PCR)**

Reverse transcription and PCR conditions were as suggested by GenHunter Corporation (Brookline, MA). All reagents were from the RNAmap kit (Genhunter), except AmpliTaq DNA polymerase (Perkin Elmer-Cetus, Norwalk, CT) and [32P]dATP (Amersham, Arlington Heights, IL). A duplicate reverse transcription reaction was performed for each sample; PCR reactions were performed with these duplicate cDNAs and run side by side in the same polyacrylamide gel in order to evaluate the reproducibility of the results.

A DNA ladder was prepared by labeling the 5’ end of DNA Marker V (Boehringer Mannheim) for sizing purposes. The PCR products were run on denaturing 6% polyacrylamide gels. Gels were dried and exposed to Kodak O-mat film (Eastman Kodak, Rochester, NY) for 2 to 5 days.

Specific bands were cut from the gels after long runs for better resolution of the bands, reamplified PCR products were gel extracted, cloned and mini prep DNA was obtained as described by Guimarães et al. (1995). Three independent clones derived from each polyacrylamide gel slice were sequenced in order to exclude the possibility of more than one PCR product being represented as a single band in the polyacrylamide gel (Guimarães et al., 1995).

**Northern blot analysis**

Large preparations of plasmid DNA containing the DD-PCR products were done using the QIAGEN Plasmid Maxi Kit (QIAGEN) following the manufacturer’s instructions. Plasmid DNA was cut with EcoRI (Boehringer Mannheim) or BssXI (Biolabs, New England), gel extracted with the QIAEX gel extraction kit (QIAGEN) and random primed with [32P]dCTP (Amersham) using the Prime-it II kit (Stratagene, La Jolla, CA) all in accordance with manufacturer’s instructions.

10-20 μg of total RNA were run in formaldehyde gels (Sambrook et al., 1989) and transferred to Nytran membranes (Schleicher & Schuell, Keene, NH) by standard methods (Sambrook et al., 1995), and blots were hybridized and washed at 65°C as described (McClanahan et al., 1993).

**Full cDNA cloning, sequence and structural analysis**

Poly(A)+ RNA was obtained and cDNA libraries were made and screened as described (Guimarães et al., 1995). Group II genes (see below) were subcloned into the pMET7 plasmid (DNAX) after a small scale λ DNA prep (Grimaldi and Grimaldi, 1989). The DNA was fully sequenced as described (Guimarães et al., 1995). The complete sequence of Mouse Yspl1 Form 1 cDNA (longer transcript) has been deposited in GenBank (Benson et al., 1994) with accession number U25739.

The FASTA (Pearson and Lipman, 1988) and BLAST (Altschul et al., 1990) programs were used to comb nonredundant protein and nucleotide databases (Benson et al., 1994; Bairoch and Boeckman, 1994) with the resultant cDNA and encoded protein sequences. The sensitive search strategies of Altschul et al. (1994) and Koonin et al. (1994) served as examples of how to locate distant structural homologues of protein chains. Multiple alignments of collected homologues were carried out with ClustalW (Thompson et al., 1994) and MACAW (Schuler et al., 1991).

The membrane topologies of Yspl1 and a cohort of putative homologues were analyzed by a variety of methods that sought to determine the consensus number of hydrophobic membrane-spanning helices and the likely cytoplasmic or extracellular exposure of the hydrophilic connecting loops. For single sequence analysis, the ALOM and MTP (Klein et al., 1985; Hartmann et al., 1989) programs were accessed from the PSORT World-Wide Web site (Nakai and Kanehisa, 1991, 1992); in turn, the TopPredII program (Claros and von Heijne, 1994; MacIntosh PPC version) was used to parse chains into probable hydrophobic transmembrane and loop regions, and further predict the localization of these latter regions by prevalence of charged residue types (von Heijne, 1992; Sippos and von Heijne, 1993). MEMSAT (Jones et al., 1994; MS-DOS PC version) was likewise used to fit individual sequences into statistically based topology models that render judgement on membrane spanning and loop chain segments. Two Web-accessible programs that are able to make use of evolutionary data by analyzing multiply aligned sequences are PHD (Rost et al., 1994, 1995) and TMAP (Persson and Argos, 1994); the former utilizes a neural network system to accurately predict the shared location of helical transmembrane segments in a protein family.

**PCR analysis**

Total RNA (5 μg) was reverse transcribed and conditions for PCR
reactions were as described by McClanahan et al. (1993). Primers were chosen to flank intron sequences. The primers for AIC2B PCR analysis were as described by McClanahan et al. (1993), primers for Brachyury as described by Keller et al. (1993), and βH1-globin primers were as described by Johansson and Wiles (1995). For Hypoxanthine-phosphoribosyl-transferase (Hprt), s.p. - 5’ GTAATGACTAGTCAACGGGGGAC 3’; a.p. - 5’ CCAGCAAGGCTTTGCAACTTAACCA 3’; i.p. - 5’ TCCCTGGGATGTCGCCAGTC 3’; expected size of the PCR product (e.s.p.p.): 214 base pairs (bp). For α-fetoprotein (Gorin et al., 1981): s.p. - 5’ GAAGGAACAGCCGCCGATGAAG 3’, a.p. - 5’ GACCAGAAAGAACAGGGGAC 3’, i.p. - 5’ GTACGGAAGAAGATGTG 3’; e.s.p.p.: 503 bp. For Clone 240 (Yspl1): s.p. - 5’ AGAGCAAGATCCAGACCCAGCAC 3’, a.p. - 5’ GATCACCTCTCCATCCATTCCATC 3’, i.p. - 5’ GTGCAAGACACTGCTAC 3’; e.s.p.p.: 191 bp. EBs were developed in the absence of serum as described by Johansson and Wiles (1995) for the PCR analysis of Yspl1 expression. Identity of the PCR product was confirmed by size on agarose gels and hybridization with an i.p. after transfer to a nylon membrane (Southern, 1975). Probes were labeled at the 5’ end with [32P]ATP using polynucleotide kinase (Boeheringer Mannheim). Blots were hybridized in 0.5 M NaHPO4, pH 7.2, 7% sodium dodecyl sulfate (SDS) and 0.5 M EDTA, pH 8.0 at 42˚C for 24 hours, then washed in 6x SSC, 0.2% SDS at 42˚C for 1 hour with one change of buffer, and exposed to Kodak O-mat film (Eastman Kodak) at ~80˚C.

**Protein expression of Yspl1**

Constructs for the expression of Forms 3 and 4 of Yspl1 were made in which a tag (FLAG) sequence (Hopp et al., 1988) was introduced in the protein. The open reading frame of cDNAs 240-4 and 240-210B, corresponding to Forms 3 and 4 of Yspl1 (see below) was amplified by PCR to introduce the FLAG peptide sequence (IBI, New Haven, CT) at the C terminus of both protein Forms 3 and 4 of Yspl1. Primers used for Form 3 Yspl1 - s.p.: 5’ ACTTCTCGAGGCACCATGGTCTCTGCTGTCA 3’; a.p.: 5’ CTAGTTCTAGATTACCTGTCTGCTCTTGTAGCTTGGACCTAACTCCCATCTCTGCTAGGTCTA-3’. Primers used for Form 4 Yspl1 - s.p.: 5’ ACTTCTCGAGGCACCATGGTCTCTGCTGTCA 3’; a.p. as described above for Form 3. PFU enzyme (Stratagene) was used with 12 cycles PCR: 94˚C 30 seconds; 55˚C 1 minute; 72˚C 4 minutes. These constructs were cloned into the PME18X vector (DNAX) using XhoI and XbaI sites incorporated into the 5’ and 3’ primers, respectively.

**Results**

**Identification of genes differentially expressed in the mouse embryo and in EBs**

Our strategic approach was the direct comparison of gene expression between the head primordium of 3- to 6-somite stage (day-8.0) mouse embryos, which is deprived of circulation at this stage (Godin et al., 1993), and regions of the developing embryo containing haematopoietic activity, circulating blood or suspected to have the potential to develop haematopoietic activity. For this purpose we used, respectively, the day-8.5 yolk sac, the day-8.5 embryo proper and the posterior region of day-8.0 embryos (from which the head primordium and the yolk sac were removed) (Fig. 1A).

Additionally, we compared the gene expression of ES cells, day-3, day-6 and day-9 EBs from a culture system in which

![Fig. 1. Strategy followed in this study. (A, top) Schematic representation of a mouse embryo. YS - Yolk sac, H - Head primordium, T, posterior region; FL, fetal liver; AGM, AGM region (see text). (Middle) Schematic representation of the in vitro culture system of embryoid bodies (EBs) derived from embryonic stem (ES) cells. d3, d6 and d9, day 3, day 6 and day 9 EB development. (Bottom) Mouse cell lines: fibroblastic cell line STO, neuronal cell line N2a and haematopoietic cell line FDCPmixA4 (as described in the text). Grey areas represent cells or tissues for which haematopoietic activity has been demonstrated. (B) PCR analysis of the expression of the common β chain of the Interleukin 3 (IL-3) receptor (AIC2B); W, early day-8.5 embryo proper; R, late day-8.5 embryo proper; Y.S., day-8.5 yolk sac; H, head primordium of day-8.0 embryos; T, posterior region of day-8.0 embryos; E.S., d3, d6, d9, STO, N2a and FDCPmixA4 as above; RT(W), RT(d3) and RT(d9), reverse transcriptase controls (samples for which no reverse transcriptase enzyme was added in the cDNA synthesis reaction); H2O, no cDNA added.**

---

**Image References:**
- Fig. 1A: Schematic representation of a mouse embryo. YS - Yolk sac, H - Head primordium, T, posterior region; FL, fetal liver; AGM, AGM region (see text).
- Fig. 1B: PCR analysis of the expression of the common β chain of the Interleukin 3 (IL-3) receptor (AIC2B); W, early day-8.5 embryo proper; R, late day-8.5 embryo proper; Y.S., day-8.5 yolk sac; H, head primordium of day-8.0 embryos; T, posterior region of day-8.0 embryos; E.S., d3, d6, d9, STO, N2a and FDCPmixA4 as above; RT(W), RT(d3) and RT(d9), reverse transcriptase controls (samples for which no reverse transcriptase enzyme was added in the cDNA synthesis reaction); H2O, no cDNA added.
we previously described the activation of genes involved in haematopoiesis (McClanahan et al., 1993). We expected that ES cells would not express genes involved in haematopoietic commitment and predicted that genes that would be important in the mesoderm-haematopoietic transition could be activated as early as day-3 EBs. Genes involved in early haematopoietic development were expected to be expressed by day 6 and day 9 of EB development (Fig. 1B).

Finally, we used three different cell lines: the fibroblastic line STO, the neuronal line Neuro-2a, and the multipotential precursor haematopoietic cell line FDCPmixA4. Genes potentially involved in the mesoderm-haematopoietic transition, but not haematopoietic-specific, could potentially be expressed in STO, since it is of mesodermic origin. ES cells, the head primordium of day-8.0 embryos, and the N2a cell line, derived from a spontaneous mouse neuroblastoma, were used as negative populations. FDCPmixA4 is an IL-3-dependent non-transformed cell line that can undergo multilineage myeloid, lymphoid or osteoclastic differentiation (Ford et al., 1971). The multipotentiality of this cell line predicts that genes involved in the earliest events of haematopoietic determination may be expressed in these cells (Fig. 1C).

Haematopoietic activity was assessed in all samples by examining the expression of the common β chain of the IL-3 receptor (AIC2B) (Kitamura et al., 1991) by PCR. As shown in Fig. 1D, AIC2B mRNA is expressed in the body of early and late 8.5-day embryos, 8.5-day yolk sac and the posterior region of day-8.0 embryos. As shown previously (McClanahan et al., 1993), expression of AIC2B mRNA is first observed by day 6 of EB development. Only FDCPmixA4 showed significant expression levels of the receptor among the cell lines used. For the qualitative evaluation of all these cDNAs, the expression of HPRT was used (data not shown).

The analysis of the DD-PCR results obtained after 20 PCR primer combinations (C/AP1 through AP5, G/AP1 through AP5, A/AP1 through AP5, T/AP1 through AP5) lead to the identification of a group of 5 bands (Clones 165, 260, 305/310, 560 and 1000 shown in Fig. 2)

Fig. 2. Definition of Group 1 developmentally regulated genes: genes preferentially expressed in the yolk sac and in the haematopoietic cell line FDCPmixA4, which are upregulated during embryoid body development. On the left are shown the DD-PCR results and on the right the northern blot analysis for each DD-PCR product. W, early day-8.5 embryo proper; R, late day-8.5 embryo proper; YS, day-8.5 yolk sac; H, head primordium of day-8.0 embryos; T, posterior region of day-8.0 embryos; ES, ES cells; d3, d6, d9 – day-3, day-6 and day-9 EBs; STO, N2a and FDCPmixA4 as described in the text.
with an expression pattern similar to the one we had predicted for genes involved in haematopoietic development: bands expressed in day-8.5 yolk sac at higher levels than the day-8.5 embryo proper or the head primordium of day-8.0 embryos, of lower intensity in ES cells than in day-3 and/or day-6 EBs, and, in general, with higher expression in FDCPmixA4 cells than in STO and N2a cells (Group 1).

Additionally, a different group of yolk sac bands (Clones 240, 320 and 460) was identified which were not expressed in EBs (Group 2) (Fig. 3).

**Similar patterns of gene expression were obtained by DD-PCR and by northern blot analysis**

In order to confirm the differential expression results obtained by DD-PCR, we conducted northern blot analysis on the same RNA populations that were used to generate cDNAs for DD-PCR. The head primordium and the posterior region of day-8.0 embryos were omitted because insufficient RNA was available for northern analysis. The cloned PCR products were used as probes. As shown in Figs 2, 3, the expression patterns of the two groups of genes defined by DD-PCR were reproduced by northern analysis.

**Differential expression of Group 1 and Group 2 genes in intraembryonic sites of haematopoiesis**

In order to characterize further their tissue distribution, particularly in intraembryonic sites of haematopoiesis, we performed northern blot analysis of both groups of genes.
All Group 1 genes were expressed in a variety of fetal and adult tissues including the AGM region, fetal liver and yolk sac (Clones 260 and 305/310 are shown in Fig. 4). Clone 305/310 had its highest level of expression in haematopoietic tissues: day-11.5 AGM, day-11.5 and day-15 yolk sac and adult bone marrow. The results of the expression of Clones 165 and 1000 will be published elsewhere (Guimarães et al., unpublished data).

However, the expression of Group 2 genes was undetectable in intraembryonic sites of haematopoiesis (day-11.5 AGM region and day-15.5 fetal liver) (see Fig. 5). A time course of expression in the yolk sac from day-8.5 to day-15 revealed that these genes were differentially expressed over time in that tissue. Two smaller transcripts of Clone 240 were found at days 11.5 and 15, Clone 460 was only expressed in day-8.5 yolk sac, and Clone 320 was highly expressed in day-8.5 and 11.5 yolk sac.

**Molecular cloning of a sialic acid esterase, selenophosphate synthetase and a permease-like molecule**

The DD-PCR products from both groups of genes were submitted to careful sequence analysis aimed at recognizing identical, closely similar or distant homologues from nonredundant DNA and protein databases (Benson et al., 1994; Bairoch and Boeckmann, 1994) in order to determine their potential biological structure and function (for review, see Altschul et al., 1994). In this manner, Clone 560 in Group 1 was identified as encoding murine coproporphyrinogen oxidase, an enzyme required for the synthesis of the porphyrin ring and therefore essential for the formation of haeme (Martin et al., 1985). Clone 320 in Group 2 appears to encode the mouse homologue of calbindin-D9K, a cytoplasmic calcium-binding protein (Jeung et al., 1992); the expression of calbindin-D9K in the mouse yolk sac has been previously described (Bruns et al., 1985). The remaining novel sequence fragments from the DD-PCR screen did not elicit any close relatives from the available databases.

In order to characterize these remaining clones further, full length cDNAs corresponding to the cloned DD-PCR fragments were isolated. The protein encoded by the full sequence of Clone 165 contains regions that closely match two short peptide fragments previously isolated from a rat liver sialic acid-specific esterase (Butor et al., 1993; Guimarães et al., unpublished data). Clone 1000, also from Group 1, is predicted to be the mouse homologue of *E. coli* selenophosphate synthetase, an enzyme that is capable of activating selenium metabolism by synthesizing monoselenophosphate from selenide and ATP (Veres et al., 1994; Guimarães et al., unpublished data). However, the full length sequences of Clones 260, 305/310 and 460 have not yet been found to have any relationship to extant protein families.

**Fig. 4.** Northern blot analysis of expression of Clones 260 and 305/310 in fetal and adult tissues. Both genes are expressed in the day-11.5 AGM region at higher levels than in the day-11.5 yolk sac. Their expression in day-15 YS is upregulated compared to day 11.5. Clone 305/310 is preferentially expressed in the adult bone marrow. YS, yolk sac; AGM, AGM region; FL, fetal liver; H, head of the embryo; Plac, placenta; BM, bone marrow; S, Muscle, skeletal muscle; A, Fat, perivisceral abdominal fat, FDCP, FDCPmixA4 cell line, STO and N2a as described; 28S, ribosomal RNA.

**Fig. 5.** Northern blot analysis of the expression of Group 2 genes in the yolk sac and early intraembryonic sites of haematopoiesis. These genes are differentially expressed during yolk sac development. None is expressed in intraembryonic sites of early haematopoiesis. Four different messages were identified for Clone 240 (sizes shown on the right).
Fig. 6. Nucleotide and deduced protein sequence of Clone 240 (Yspl1). The four protein forms derived from the Yspl1 gene are indicated by arrows. The grey boxes, labeled TM1-12, represent presumptive membrane spanning hydrophobic segments. The hydrophobic sequence labeled as box X represents an inserted domain in the canonical nucleobase permissive fold of Yspl1 - it may also be membrane associated but should not alter the consensus topological features of Yspl1.
The screening of a 8.5 day-yolk sac cDNA library by hybridization with clone 240 lead to the isolation of four different cDNAs: 240-7 (2.1 kb), 240-205 (1.8 kb), 240-4 (0.7 kb) and 240-10B (0.6 kb). The size of these cDNAs is in agreement with the northern blot analysis results shown in Fig. 5 and correspond, respectively, to the protein Forms 1, 2, 3 and 4 of Clone 240 shown in Fig. 6. The frequency of Clone 240 positive forms, all 4 cDNA forms together, was approximately 1/63,000. Forms 3 and 4 cDNAs of Clone 240, undetectable in day-8.5 yolk sac by northern analysis, were represented in the day-8.5 yolk sac-derived cDNA library with a frequency three times lower than that of Forms 1 and 2 cDNAs. The deduced protein sequence of the longer cDNA sequence of Clone 240-7 immediately revealed in the jagged contours of its hydrophilic profile (Kyte and Doolittle, 1982) that it was probably an integral membrane protein with 10-13 transmembrane (TM) helices. Accordingly, databank searching disclosed a faint but significant chain similarity to a family of nucleobase permeases described in bacteria and yeast (Diallinas et al., 1995) (Fig. 7A). Due to its preferential expression in the yolk sac, we designated this novel gene product as Yolk permease-like molecule 1 (Yspl1).

### Fig. 7. Identification of Yspl1 as a nucleobase permease.

(A) A nucleobase permease family. Portions of a ClustalW-derived (Thompson et al., 1994b) sequence alignment between Yspl1 and permease homologues are presented in regions that display family-specific sequence patterns. TM helices are labeled as in Fig. 6. Shading indicates identities (dark) or conserved (light) residues in a column. The sequences include an E. coli uracil permease (EcUrA); GenBank Acc. no. X73586; Andersen et al., unpublished data) and a permease-like gene (EcORF; GenBank Acc. no. L10328; Burland et al., 1993), pyrimidine permeases from B. subtilis (BsPyrP; GenBank Acc. no. M59757; Diallinas et al., 1995) and B. caldolyticus (BcPyrP; GenBank Acc. no. X76083; Ghim and Neuhard, 1994), xanthine permease from B. subtilis (BsXpt; GenBank Acc. no. X71807; Saxild et al., 1995) and a similar hypothetical protein (BsORF; GenBank Acc. no. X73124; Schneider et al., 1993), and two uric acid-xanthine permeases from A. nidulans (AnUAPA; GenBank Acc. no. X79796; Diallinas et al., 1995). Mouse Yspl1 is abbreviated MoYSPL1. (B) Membrane topology of Yspl1. (Top) A TopPredII (Claros and von Heijne, 1994) profile of the Yspl1 sequence showing peaks that reach beyond ‘Putative’ or ‘Certain’ baselines. Peaks representing the consensus twelve TM segments are labeled above, as is the hydrophobic X region outlined in Fig. 6. (Bottom) Schematic arrangement of the TM helices in the membrane. The Yspl1 chain weaves through the membrane in an in-out fashion determined by TM regions and charged residue bias (von Heijne, 1994). There are no N-glycosylation sites in the exposed, extracellular face of the molecule; however, cysteine residues capable of participating in disulfide links are marked by dark points. Start positions for Forms 1-4 are again indicated by arrows. Notably, Form 3 protein commences with the hydrophobic sequence of TM12 which could serve as a cleavable signal peptide, while a Form 4 molecule would be purely cytoplasmic.

### Protein structure and expression of Yspl1

The overriding structural feature of the 611 amino acid Yspl1 chain is its patchwork of 20-30 residue hydrophobic segments separated by hydrophilic sequences of varying length (Kyte and Doolittle, 1982). The weave of the chain through the membrane is dictated by these presumed membrane-spanning, hydrophobic helices (for review, see von Heijne, 1994). An accurate prediction of the membrane topology of Yspl1...
Yspl1 then depends on the correct number and sequence location of TM segments as well as their orientation in the membrane - protein termini and TM-linking regions may either be surface exposed or cytoplasmic (Hartmann et al., 1989; Sippos and von Heijne, 1993; Jones et al., 1994; von Heijne, 1994).

The introduction of evolutionary information in the form of sequence homologues simplifies the structural analysis considerably for related molecules which share a common structural framework in spite of considerable sequence divergence (Chothia and Lesk, 1986). This concept can be effectively extended to the strong prediction of TM regions across an aligned protein family, whereas any single sequence may provide an uncertain topology (Persson and Argos, 1994; Rost et al., 1995). In the case of Yspl1, a number of sequence homologues were first assembled by comparative matching to protein and translated nucleotide databases (Altschul et al., 1994; Koonin et al., 1994). These distant relatives of Yspl1 prominently include bacterial uracil permeases (Andersen et al., unpublished data; Burland et al., 1993; Ghim et al., 1994; Quinn et al., 1991; Saxild et al., 1995; Schneider et al., 1993) and fungal uric acid-xanthine permeases (Gorfinkel et al., 1993; Diallinas et al., 1995) (Fig. 7A). These varied purine and pyrimidine – generically, nucleobase – permease sequences were subjected to parallel analyses by a suite of computer programs that have greatly improved on the initial Kyte and Doolittle (1982) hydrophatic profile as a means of predicting the topology of integral membrane proteins. Four algorithms (ALOM, MTOP, MEMSAT and TopPredII) (Klein et al., 1985; Hartmann et al., 1989; Jones et al., 1994; Claros and von Heijne, 1994) were used to individually predict TM extensions and orientations; these predictions were pooled and mapped onto the multiple sequence alignment produced by ClustalW and MACAW (Thompson et al., 1994b; Schuler et al., 1991). Furthermore, these multiple aligned sequence files were used as input to PHD and TMAP (Rost et al., 1995; Persson and Argos, 1994) for a familial prediction of shared TM regions. Structural features that persisted in this two-step analysis were most likely to be shared topological traits present in all members of this permease family from bacteria to vertebrates.

The TM analysis for Yspl1 and homologues suggested the presence of twelve consensus helical TM segments (Fig. 7B). This result likely places this nucleobase permease family into a larger superfamily of 12-TM helix transporters (Nikaido and Saier, 1992; Uhl and Hartig, 1992; Griffith et al., 1992; Marger and Saier, 1993; Henderson, 1993; Maloney, 1994) that act as discriminating protein channels for a wide variety of solutes (Kramer, 1994; Hediger, 1994). There are few amino acid motifs that survive between branches of this vast superfamily of 12-TM molecules (Griffith et al., 1992; Henderson, 1993; Marger and Saier, 1993; Goswitz and Brooker, 1995). The prominent sequence patterns that describe the Yspl1-like family (Fig. 7A) cover TM4, TM9 and TM10; these motifs are distinct from other topologically analogous 12-TM patterns (Griffith et al., 1992). Permease specificity within a universal TM framework may be modulated by these different structural motifs (see for example Bloch et al., 1992). Nevertheless, profile methods (Gribkov et al., 1987; Thompson et al., 1994a) do reveal a faint but significant global similarity (not shown) to a functionally similar family of 12-TM molecules formed by the allantoin and uracil permeases of yeast (Jund et al., 1988; Rai et al., 1988; Yoo et al., 1992) and the cytosine permeases of E. coli and yeast (Danielsen et al., 1992; Weber et al., 1990). Yspl1 remains the sole vertebrate member of a now expanded nucleobase permease family.

The different forms of Yspl1 represent molecules with staggered N termini and common C termini (Fig. 6) of which the longest (611 residue) chain, as discussed previously, forms a prototypical 12-TM transporter. Form 2 which lacks TM1 and the hydrophobic 'X' stretch (Figs 6, 7B) would still encode a truncated permease with eleven TM helices. Form 3 cDNA encodes a protein of 130 residues (Mr 14×10^3) with one hydrophobic TM12 stretch that could then resemble an N-terminal secretion peptide for a domain formed by the C-terminal, predicted cytoplasmic segment of Yspl1. Finally, an 82 amino acid Form 4 (Mr 9×10^3) would be formed by a shortened version of this cytoplasmic domain, lacking any hydrophobic segments. In order to determine if Form 3 Yspl1 would be secreted or intracellular, and to confirm the cytoplasmic localization of Form 4 protein, we expressed constructs of Yspl1 in which a FLAG epitope tag was introduced at the C termini of both Form 3 and 4 proteins (see Materials and Methods). Western blot analysis using the M2 antibody directed to this tag sequence shows that Form 3 Yspl1 can be detected both inside and outside cells whereas Form 4 is purely intracellular (Fig. 8).

**Yspl1 is a unique marker of yolk sac development**

To determine if Yspl1 is expressed in EBs at levels undetectable by northern hybridization, we conducted 30 cycle PCR analysis comparing the expression of Yspl1 with that of a mesoderm marker (*Brachyury*), a haematopoietic marker (*βH1-globin*) and a commonly used marker of yolk sac differentiation (α-fetoprotein). In contrast to the *Brachyury* (detectable at day 3 and day 6), *βH1-globin* (detectable from day 6) and α-fetoprotein (detectable from day 9 of EB development), Yspl1 was not found to be expressed in EBs developed up to day 9 under standard conditions (see Materials and methods) or day-5.0 EBs developed in CDM (Chemically Defined Medium) alone (Johansson and Wiles, 1995), or in the presence of Activin-A or BMP-4 (mediators of the expression of mesoderm and haematopoietic markers, respectively; Johansson and Wiles, 1995; and data not shown).

![Fig. 8. Protein expression of Form 3 and Form 4 cDNAs of Yspl1.](image-url) Western blot obtained using the M2 antibody directed to the FLAG sequence introduced at the C terminus of the Yspl1 protein constructs. Protein Form 3 of Yspl1 can be observed both intracellularly (lysates) and extracellularly (media). Protein Form 4 of Yspl1 is intracellular. PME, PME18X plasmid without insert was transfected into COS-7 cells and used as a negative control for the expression of Yspl1.
DISCUSSION

We report the molecular cloning of two groups of genes preferentially expressed in the yolk sac. Specifically, we identified a set of genes expressed in the day-8.5 yolk sac at higher levels than in the day-8.5 embryo proper, which are expressed in EBs (Alc2B, coproporphyrinogen oxidase, lumenal sialic acid esterase, selenophosphate synthetase, Clones 260 and 305/310, βH1-globin and α-fetoprotein), and another set of genes expressed in the day-8.5 yolk sac, but not in the day-8.5 embryo proper, which are undetectable in EBs, as analyzed here, up to day 9 of EB development, and in intraembryonic sites of haematopoiesis (Ysp1, Calbindin-D9k and Clone 460). This discrepancy could result from a complete failure or a delay in reproducing certain yolk sac developmental events in EBs and suggests that the haematopoietic potential derived in vitro from ES cells may be independent of the complete reproduction of yolk sac developmental programs. However, EBs derived from ES cells can produce embryonic globin and nucleated erythroblasts (Doetschman et al., 1985), yolk sac-specific events during in vivo development (Brotherton et al., 1979). The relevance of these findings can be related to the fact that intraembryonic haematopoiesis is only detected shortly after the activation of circulation, when haematopoietic events are already occurring in the yolk sac. It is therefore particularly difficult to determine if such activity results from the colonization of the embryo by progenitor cells that originated in the yolk sac or whether haematopoietic activity arose independently within the embryo. W/W mutant mice (Bennett, 1956) and the c-myb (Mucenski et al., 1991) and Pu.1 (Scott et al., 1994) knock-out mice provide evidence for a distinction between yolk sac and embryo-derived haematopoiesis. Yolk sac erythropoiesis is unaffected in any of these mutant mice. However, in each of these models, by day 15 the fetus becomes detectably anemic and severe anemia rapidly develops, resulting in prenatal lethality. The demonstration that certain yolk sac-specific events cannot be reproduced in EBs, or are clearly delayed in relation to the generation of haematopoietic activity, may provide clues to the resolution of this question.

We have reported the molecular cloning of Ysp1, a novel yolk sac gene that EB developmental programs do not recapitulate. Ysp1 represents the first member, to be found in vertebrates, of a family of nucleobase permeases described in bacteria and yeast. Interestingly, the intracellular, transmembrane and extracellular forms of the protein are represented by different cDNAs expressed in the yolk sac. This is the first report of intracellular and extracellular protein forms of a nucleobase or nucleoside permease. The characterization of the function of the Ysp1 gene products is likely to reveal novel aspects of nucleotide metabolism which may play a role in mammalian development.

The number of DD-PCR primer combinations under the conditions that we have used in this study represent about 1/15 of the number required for the observation of the complete repertoire of mRNA species (Guimarães et al., 1995). Based on this number, we can predict the existence of a total number of about 45 genes specifically expressed in the day-8.5 yolk sac. This estimate substantiates efforts to isolate additional novel yolk sac genes.

We have established a strategy to understand the relationship between haematopoietic events observed in the in vitro culture system of embryoid bodies with those of in vivo mouse development. Similar strategies could be devised for the dissection of tissue-specific events involved in cell lineage commitment and understanding the development of systems other than the haematopoietic.

We gratefully acknowledge Professor J. M. Pina Cabral for guidance and Dr P. Vieira, Dr A. O’Garra, Dr G. Hardiman, Dr A. Vicari, Dr E. Ching and Dr J. Chiller for encouragement, Dr S. Dalrymple for help in the dissection of mouse embryos, D. Liggett and B. Johansson for excellent technical assistance, and Professor M. Teixeira da Silva, Professor M. de Sousa, Dr A. Cumano and Professor A. Coutinho for advice. M. Jorge Guimarães is supported by a grant from JNICT, Portugal (CIÊNCIA/BD/2685/93). The Basel Institute for Immunology was founded and is supported by Hoffmann-La Roche Ltd, CH-4005 Basel, Switzerland. DNAX Research Institute of Molecular and Cellular Biology, Incorporated, is supported by the Schering-Plough Corporation.

REFERENCES

Genbank Acc. no. X63656.
Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. and Kemler, R.
that the protein encoded by the AIC2B cDNA is a subunit of the murine GM-CSF receptor. Proc. Natl. Acad. Sci. USA 88, 5082.


(Accepted 4 July 1995)