Developmental Biology 416 (2016) 149–161

Contents lists available at ScienceDirect

Developmental Biology

journal homepage: www.elsevier.com/locate/developmentalbiology

# A conserved alternative form of the purple sea urchin HEB/E2-2/E2A transcription factor mediates a switch in E-protein regulatory state in differentiating immune cells

Catherine S. Schrankel [a,b,1], Cynthia M. Solek [a,c,1,2], Katherine M. Buckley [a,b], Michele K. Anderson [a,b], Jonathan P. Rast [a,b,c,*]

[a] Biological Sciences, Sunnybrook Research Institute, Toronto, ON, Canada
[b] Department of Immunology, University of Toronto, Toronto, ON, Canada
[c] Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada

ARTICLE INFO

*Article history:*
Received 2 March 2016
Received in revised form
26 May 2016
Accepted 27 May 2016
Available online 2 June 2016

*Keywords:*
Hematopoiesis
Alternative splicing
Echinoderm
Immunity
Deuterostome
bHLH

ABSTRACT

E-proteins are basic helix-loop-helix (bHLH) transcription factors with essential roles in animal development. In mammals, these are encoded by three loci: *E2-2* (*ITF-2/ME2/SEF2/TCF4*), *E2A* (*TCF3*), and *HEB* (*ME1/REB/TCF12*). The *HEB* and *E2-2* paralogs are expressed as alternative (Alt) isoforms with distinct N-terminal sequences encoded by unique exons under separate regulatory control. Expression of these alternative transcripts is restricted relative to the longer (Can) forms, suggesting distinct regulatory roles, although the functions of the Alt proteins remain poorly understood. Here, we characterize the single sea urchin E-protein ortholog (*SpE-protein*). The organization of the *SpE-protein* gene closely resembles that of the extended *HEB/E2-2* vertebrate loci, including a transcript that initiates at a homologous alternative transcription start site (*SpE-Alt*). The existence of an Alt form in the sea urchin indicates that this feature predates the emergence of the vertebrates. We present additional evidence indicating that this transcript was present in the common bilaterian ancestor. In contrast to the widely expressed canonical form (*SpE-Can*), *SpE-Alt* expression is tightly restricted. *SpE-Alt* is expressed in two phases: first in aboral non-skeletogenic mesenchyme (NSM) cells and then in oral NSM cells preceding their differentiation and ingression into the blastocoel. Derivatives of these cells mediate immune response in the larval stage. Inhibition of SpE-Alt activity interferes with these events. Notably, although the two isoforms are initially co-expressed, as these cells differentiate, *SpE-Can* is excluded from the *SpE-Alt*$^{+}$ cell population. This mutually exclusive expression is dependent on SpE-Alt function, which reveals a previously undescribed negative regulatory linkage between the two E-protein forms. Collectively, these findings reorient our understanding of the evolution of this transcription factor family and highlight fundamental properties of E-protein biology.



## 1. Introduction

E-proteins are basic helix-loop-helix (bHLH) transcription factors that are present throughout Metazoa and regulate fundamental biological processes including myogenesis, neurogenesis, cell-cycle, sex determination and vertebrate hematopoiesis (Massari and Murre, 2000; Kee, 2009; de Pooter and Kee, 2010). There are three mammalian E-protein paralogs: *E2A* (TCF3), *E2-2* (ITF-2/ME2/SEF2/TCF4), and *HEB* (ME1/REB/TCF12) (Murre et al., 1989a; Henthorn et al., 1990; Hu et al., 1992). These regulatory proteins share a conserved, modular structure (Murre et al., 1989a, 1989b) in which the basic region binds DNA at E-box consensus sites (Ephrussi et al., 1985), and transactivation domains (AD1-3) (Aronheim et al., 1993; Chen et al., 2013) recruit co-activators and chromatin modifying enzymes (Kee, 2009). These broadly expressed class I bHLH factors mediate specific activity by binding as homodimers or heterodimers with a complex array of partners, including other E-proteins or class II bHLH factors (*e.g.*, Scl and MyoD) that exhibit more tissue-restricted expression patterns (Murre et al., 1994). In contrast, members of the class IV bHLH factor family (Id, inhibitor of DNA binding), which lack the DNA-binding domain, antagonize E-protein activity by forming

* Correspondence to: Sunnybrook Research Institute, Department of Medical Biophysics and Department of Immunology, University of Toronto, 2075 Bayview Avenue, Rm. S126B, Toronto, ON, Canada M4N 3M5.
*E-mail address:* jrast@sri.utoronto.ca (J.P. Rast).
[1] These authors contributed equally to this work.
[2] Present address: Montreal Neurological Institute, McGill University, Montreal, QC, Canada.

http://dx.doi.org/10.1016/j.ydbio.2016.05.034
0012-1606/

heterodimers that are unable to bind DNA (Sun et al., 1991).

In vertebrates, E-protein activity is further modulated by alternative (Alt) isoforms that are generated from secondary transcription start sites (TSS). These transcripts initiate at alternative (Alt) exons that are located between exons 8 and 9 of canonical *E2-2* and *HEB* loci and are spliced to the remaining exons in each locus (Skerjanc et al., 1996; Wang et al., 2006). The Alt exon is excluded from canonical forms of these transcripts. The truncated HEB-Alt and E2-2-Alt proteins include a unique N-terminal 23 amino acid Alt sequence and the AD2, AD3, and bHLH domains, but lack the upstream sequence that includes the AD1 region. The amino acid sequence encoded by the Alt exon is 80% identical among jawed vertebrate paralogs, suggesting a conserved function in E-protein activity (Wang et al., 2006; Braunstein and Anderson, 2012).

Alternative TSS and exon usage is widely observed among transcription factors, although the functions of most internal promoters remain unresolved (Davuluri et al., 2008; Zhu and Halfon, 2009; Brown et al., 2014). Although E-proteins are widely studied, functional data on Alt isoforms remain limited. Both *HEB-Alt* and *E2-2-Alt* exhibit tightly restricted spatiotemporal expression patterns in comparison to the more widely expressed canonical forms. *E2-2-Alt* (ITF-2A) was originally identified in a muscle cell line; analysis of the cardiac alpha-actin promoter indicates that E2-2-Can and E2-2-Alt have opposing regulatory effects (Skerjanc et al., 1996). HEB-Alt is essential for early T cell specification and commitment downstream of Delta/Notch signaling (Wang et al., 2006, 2010; Braunstein and Anderson, 2010; Braunstein et al., 2010). Overexpression of *HEB-Alt* is also linked to metastasis and poor prognosis in colorectal cancers (Thorsen et al., 2011). Investigations into the regulation and function of Alt forms of vertebrate E-proteins are complicated by compensatory mechanisms among paralogs, combinatorial diversity of cognate bHLH dimerization partners, and the overall morphological complexity of these systems.

As an invertebrate deuterostome, the purple sea urchin (*Strongylocentrotus purpuratus*) shares a regulatory heritage with vertebrates, including many aspects of E-protein biology. The sea urchin genome encodes a single *E-protein* ortholog (*SpE-protein*; Hibino et al., 2006; Howard-Ashby et al., 2006; Sodergren et al., 2006) that is ubiquitously expressed in the developing embryo (Solek et al., 2013). Importantly, the sea urchin has single orthologs of transcription factors that interact with E-proteins in other systems (*e.g., Scl, Gata-1* and *Id*; Wadman et al., 1997; Tripic et al., 2009). These factors are co-expressed in immune cells (including pigment cells and several types of blastocoelar cells) and mesodermal precursors. We have previously investigated the regulatory functions of the sea urchin orthologs of *Gata1/2/3* and *Scl/Lyl/Tal-2* in the context of mesodermal development (Solek et al., 2013). To further explore the regulatory capability of the bHLH factor Scl, we here characterize the role of its potential binding partner, *SpE-protein*.

Pigment cells and blastocoelar cells arise from a ring of non-skeletogenic mesoderm (NSM) specified in part by Delta-Notch signaling (Sherwood and McClay, 1997; Sweet et al., 2002; Materna and Davidson, 2012). Other NSM derivatives include the coelomic pouches and esophageal muscles (Andrikou et al., 2013, 2015; Materna et al., 2013). By 24 hours post-fertilization (hpf), the NSM region is partitioned along the oral/aboral axis (Ruffins and Ettensohn, 1996). Cells from the aboral region differentiate into pigment cells, which have immune functions in the larva (Hibino et al., 2006; Solek et al., 2013; Ho et al., 2016). The NSM precursors of these cells are patterned by the transcription factor glial cells missing (gcm) (Ransick et al., 2002; Ransick and Davidson, 2012) and exhibit transiently increased expression of Id (the ortholog of the vertebrate *Id1/2/3/4* factors; Solek et al., 2013). These cells quickly express differentiation markers, undergo an epithelial to mesenchymal transition (EMT), and migrate through the blastocoel to embed in the outer epithelium by 40 hpf (Gibson and Burke, 1985; Saunders and McClay, 2014).

In contrast to the pigment cell lineage, the oral NSM cells differentiate into a heterogeneous set of cell types (Tamboline and Burke, 1992; Ho et al., 2016). At 24 hpf, the oral field of NSM (10–14 blastocoelar cell precursors) is marked by co-expression of single orthologs of vertebrate *Gata1/2/3* and *Scl/Lyl-1/Tal-1* (Solek et al., 2013). These cells remain undifferentiated at the tip of the developing archenteron prior to EMT and ingression into the blastocoel (~45 hpf). Coincident with ingression, *Gata1/2/3* is downregulated and expression of cell-type specific markers is initiated. These cells differentiate into four morphologically distinct cell types (collectively known as blastocoelar cells) that populate the larva blastocoel and exhibit immune characteristics, including surveillance-like motility, rapid migration to the gut in response to microbial disturbance, and phagocytosis of bacteria that have entered the blastocoel (Hibino et al., 2006; Solek et al., 2013; Ho et al., 2016). This system of mesenchymal cell development within the morphologically simple sea urchin embryo involves minimal cell division and provides a tractable system in which to explore the regulation of E-protein isoforms and alternative promoter usage in the context of reduced paralog diversity.

Here, we characterize the sea urchin *E-protein* locus, and find that the Alt isoform (*SpE-Alt*), which is conserved with vertebrate *HEB-Alt* and *E2-2-Alt,* exists also in invertebrates. The conservation of an Alt isoform in invertebrates demonstrates that this *E-protein* feature is primitive for the vertebrate paralogs and was present in the common bilaterian ancestor. As in the vertebrates, expression of *SpE-Alt* is highly restricted in time and space relative to the widely expressed canonical form (*SpE-Can*). Functional perturbation of SpE-Alt results in developmental defects, the timing of which indicates a role in mesodermal cell differentiation. Notably, we find that these factors are reciprocally expressed such that as *SpE-Alt* is upregulated, *SpE-Can* transcripts are extinguished in these cells. SpE-Alt is required for this exclusion and subsequent shift in E-protein activity, which reveals a regulatory feedback between the two E-protein isoforms.

## 2. Materials and methods

### 2.1. Embryo and larval culture

*S. purpuratus* specimens were obtained from Point Loma Marine Invertebrate Lab (Lakeside, CA, USA) and maintained in aquaria with artificial seawater (ASW; Instant Ocean) at 13 °C. Eggs were spawned by gentle shaking or by injection of 0.5 M KCl. Fertilized embryos were cultured by standard methods in 0.45 μM-filtered ASW in Petri dishes or stirred flasks at 15 °C. Larvae were fed with *Rhodomonas lens* (5000 cells/ml) every other day beginning at 5 days post-fertilization (dpf). All animal protocols were approved by the Sunnybrook Animal Care Committee.

### 2.2. RNA isolation and expression analysis

To identify the *SpE-Alt* transcript sequence, Rapid Amplification of cDNA Ends (RACE, GeneRacer kit, Invitrogen) was performed on cDNA isolated from 24 hpf mesenchyme blastulae (primer sequences are shown in Table S1). RNA was extracted (Trizol; Life Technologies), column-purified (PureLink RNA Mini Kit; Ambion) and DNase-treated (DNA-free kit; Ambion). Total RNA was reverse transcribed from random hexamers using SuperscriptIII (Invitrogen). Quantitative PCR (qPCR) cycling was performed using SYBR Green PCR master mix (Applied Biosystems) and an Applied Biosystems ViiA7 System. RT-qPCR analysis was performed as previously described (Rast et al.,