Available online at www.sciencedirect.com
ScienceDirect
Developmental Biology 300 (2006) 108–120
DEVELOPMENTAL BIOLOGY
www.elsevier.com/locate/ydbio

The C_2H_2 zinc finger genes of *Strongylocentrotus purpuratus* and their expression in embryonic development

Stefan C. Materna, Meredith Howard-Ashby, Rachel F. Gray, Eric H. Davidson *

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Received for publication 25 May 2006; revised 4 August 2006; accepted 16 August 2006
Available online 22 August 2006

Abstract

The C_2H_2 zinc finger is one of the most abundant protein domains and is thought to have been extensively replicated in diverse animal clades. Some well-studied proteins that contain this domain are transcriptional regulators. As part of an attempt to delineate all transcription factors encoded in the *Strongylocentrotus purpuratus* genome, we identified the C_2H_2 zinc finger genes indicated in the sequence, and examined their involvement in embryonic development. We found 377 zinc finger genes in the sea urchin genome, about half the number found in mice or humans. Their expression was measured by quantitative PCR. Up to the end of gastrulation less than a third of these genes is expressed, and about 75% of the expressed genes are maternal; both parameters distinguish these from all other classes of regulatory genes as measured in other studies. Spatial expression pattern was determined by whole mount *in situ* hybridization for 43 genes transcribed at a sufficient level, and localized expression was observed in diverse embryonic tissues. These genes may execute important regulatory functions in development. However, the functional meaning of the majority of this large gene family remains undefined.

Keywords: Zinc finger; Zf-c2h2; Transcription factor; Genome; Sea urchin; Development

Introduction

Zinc finger motifs are of particular interest in developmental biology because they occur in some prominent transcriptional regulators. Though there are more than seventy classes of zinc-binding motifs listed in the PFAM database, the specific transcriptional regulators fall mainly in the C_2H_2 zinc finger class, in which the zinc atom is complexed by two cysteines and two histidines, and in their structural relatives the C_4 zinc finger class, in which the zinc is complexed with four cysteines. The latter group consists mainly of nuclear hormone receptors and GATA factors (Krishna et al., 2003). In most animal genomes that have been sequenced, C_2H_2 zinc fingers are among the more abundant protein domains. This applies in particular to mammalian genomes, in which C_2H_2 zinc finger genes have been highly multiplied (Lander et al., 2001; Rubin et al., 2000). The C_2H_2 zinc finger genes usually far outnumber the zinc fingers of the C_4 type, of which most genomes contain only a few dozen. A prominent exception is *Caenorhabditis elegans*, where the nuclear hormone receptors have undergone extensive multiplication (Reece-Hoyes et al., 2005) and outnumber C_2H_2 zinc finger genes. Although sequence-specific DNA-binding proteins can be found in other groups of zinc fingers, specific transcription factors are rare, and these proteins are often part of the basal transcription apparatus or DNA repair machinery.

C_2H_2 zinc finger proteins are commonly viewed as transcriptional regulators, but they may be widely used for RNA binding. This is exemplified by the first known zinc finger transcription factor, *Xenopus* TFIIIa, which binds specifically to both DNA and RNA (Lu et al., 2003). Possibly just as typical for genes with higher numbers of zinc fingers is the *Xenopus xfin* gene, which codes for a protein with 37 zinc fingers. It is localized in the cytoplasm and has been shown to bind to RNA (Andreazzoli et al., 1993). Transcriptional regulatory activity has not been demonstrated for this protein. In addition to DNA and RNA binding, zinc finger domains may also be used for protein–protein interactions (Laity et al., 2001). Several examples are known, including well-known transcriptional regulators. For example, two of the five conserved zinc fingers of the

* Corresponding author. Fax: +1 626 793 3047.
E-mail address: Davidson@caltech.edu (E.H. Davidson).

0012-1606/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ydbio.2006.08.032

Gli family transcription factor encoded by the *cubitus interruptus* gene are needed for specific interaction with another factor, converting the protein into its active form (Croker et al., 2006). C_2H_2 zinc finger genes are thought to account for 30%–50% of all transcription factors in metazoan genomes (Adams et al., 2000; Ruvkun and Hobert, 1998). However, except for a minority which is clearly orthologous to known regulatory factors, there are no canonical criteria which suffice to distinguished those C_2H_2 proteins that are dedicated sequence-specific transcription factors from those that bind RNA or perform other functions.

Regulatory genes of the nuclear hormone receptor and GATA classes in the sea urchin genome have been characterized by Howard-Ashby et al. (2006a; this issue). Here, in order to encompass the major remaining class of transcription factors, we identify all C_2H_2 zinc finger genes predicted by the genomic sequence, and determine their activity during development.

Materials and methods

Identification of zinc finger genes

Zinc finger genes were identified in the contig assembly by searching for the C_2H_2 zinc finger motif. We built a calibrated Hidden–Markov model from the PFAM seed alignment (PF00096, www.sanger.ac.uk/Software/Pfam/) and searched the sea urchin genome with *hmmsearch* (http://hmmer.wustl.edu/) accepting only domains with an *E*-value <0.1. After release of the gene predictions, the identified zinc finger genes were mapped onto the GLEAN models (Sea Urchin Sequencing Consortium, submitted) by finding near perfect matches to the calculated QPCR amplicon. On genes that did not perfectly match a GLEAN model, a BLAST search was performed against the remainder of the GLEAN gene predictions. The results were manually inspected and associations validated. Less than perfect matches are due to incorporation of sequences of different haplotypes in the scaffold assembly. Presumed zinc finger genes that did not match any GLEAN model were searched by BLAST against novel predictions from the whole genome tiling array (Samanta et al., submitted), identifying additional gene models for genes that are expressed in early development. Gene models were aligned with the contigs using the *spidey* genomic mapping program (Wheelan et al., 2001), and the match was validated through manual inspection.

Phylogenetic analysis

For identification of orthologous genes, we obtained the set of C_2H_2 zinc finger containing protein sequences of *Homo sapiens*, *Mus musculus*, *Ciona intestinalis*, *C. elegans*, *Drosophila melanogaster*, and *Nematostella vectensis*. Sequences of *Nematostella* were obtained from *Stellabase* (www.stellabase.org), *C. elegans* sequences from *Wormbase* (www.wormbase.org, WSWS156), and all others from *Ensembl* (www.ensembl.org, v.37—February 2006) by motif search for Interpro domain IP:007087. To obtain a nonredundant set of proteins, we kept only the longest protein and discarded shorter isoforms for any given gene. A BLASTP search was performed for each sea urchin protein against this set of C_2H_2 zinc finger proteins. Good hits were confirmed by manual inspection. For such genes the zinc finger region together with surrounding conserved sequence was excised and aligned using the *mafft* alignment program (Katoh et al., 2005) using 1000 iterations. Phylogenetic analysis was conducted using the neighbor-joining method with the MEGA program (Kumar et al., 2004). The calculated distance was Poisson-corrected, gaps were pairwise deleted, and 1000 iterations were used for calculating bootstrap values.

Transcriptional profiling

We performed transcriptional profiling using quantitative PCR (QPCR). QPCR is a comparative method, in which the accumulation of PCR product is monitored for a gene of interest and in the same sample for a given standard, through the use of a double strand-specific fluorescent dye. By choosing a threshold and determining $C_t\Delta$ (the difference in cycle number at which each PCR reaction crosses the threshold) the initial prevalence of a gene can be calculated, since the cycle difference is proportional to the abundance in the original reaction mix (Wong and Medrano, 2005).

For primer design, the sometimes short stretch of sequence containing the zinc finger domains was extended using a BLASTX search against the NCBI database of non-redundant proteins "*nr*" (www.ncbi.nlm.nih.gov). QPCR primers lying within these regions were obtained using the standalone version of *primer*3 (Rozen and Skaletzky, 2000). Primers were chosen to yield an amplicon of between 110 and 140 base pairs. Primers were tested for specificity on genomic DNA by QPCR. It was assumed that all genes dealt with were single copy. In this case, given equal amplification efficiency, all PCR products should accumulate to a given threshold at roughly the same cycle (C_t). Primer pairs that did not produce an acceptable C_t value (one cycle more or less compared to the mean C_t) were not used for transcriptional profiling and were redesigned. A more exact determination of primer efficiency was conducted for a representative set of primer pairs using serial dilutions (Wong and Medrano, 2005), and it confirmed the initial findings. We therefore generally assumed an amplification efficiency of 1.95. The presence of a single specific band was confirmed by gel electrophoresis.

Embryos were grown and harvested at fertilization, and at 6, 12, 18, 24, 36, and 48 h postfertilization. RNA was isolated with the Qiagen RNeasy Mini-Kit. RT reactions were performed with ABI (Foster City, USA) TaqMan cDNA synthesis kit according to manufacturer's instructions. QPCR was conducted on an ABI 7900 HT with ABI SYBR-Green reaction mix, using the following program: 1× (95°C–10 min,) 40× (60°C–30 s, 95°C–1 min). At the end of each program a dissociation curve was collected to confirm that only one product accumulated during the reaction. A no-template control also assured that no primer–dimers had formed. The RNA copy number was determined by calculating the $C_t\Delta$ for a given zinc finger gene with respect to the poly-ubiquitin gene, which was assumed to be represented by 88,000 transcripts per embryo during the developmental stages examined (Nemer et al., 1991). On each plate, each primer–cDNA combination was run in triplicate. The experiment was repeated once with the same cDNA and twice with cDNA from a second animal. For data analysis, the C_ts of wells that obviously did not amplify were omitted. $C_t\Delta$ was calculated for the triplicates of the four independent runs. The mean of the averages from the four runs was used to calculate the number of transcripts per embryo. Error bars were calculated from the standard deviation on the mean.

Whole mount in situ *hybridization*

In order to identify the spatial domain of expression for the higher expressed genes, we conducted *in situ* hybridization with digoxigenin-labeled antisense probes. From the general assumption that about 10 RNA molecules per cell are needed for sufficient staining, it follows that a minimum of several hundred molecules per embryo is needed to obtain a clear stain since spatially restricted regions like the endoderm contain no fewer than 60 cells prior to gastrulation.

For successful *in situ* hybridization using these methods, the probe should be a minimum of 600 base pairs in length. We attempted to obtain primer pairs by using either conserved sequence that is recognizably located within one exon, or, if no sequence of sufficient length could be obtained, by assuming the GLEAN gene models. Templates for *in situ* probes were amplified from cDNA using primers tailed with *Sp6* and *T7* promotors or subcloned into the pGEM-T Easy vector which contains *Sp6* and *T7* promotor sites adjacent to the multiple cloning site. Sequencing confirmed the identity of the gene. After digoxigenin labeling through *in vitro* transcription with Roche *Sp6* or *T7* polymerase, probes were run on a denaturing gel, confirming the size of the transcript. Whole mount *in situ* hybridizations were carried out according to Minokawa et al. (2004).

Results

Identification of C_2H_2 zinc finger-containing genes

We initially set out to identify all transcription factor genes using a BLAST-based approach, searching the trace archive of