Xenopus as a model system to study transcriptional regulatory networks

Tetsuya Koide⁎, Tadayoshi Hayata⁎, and Ken W. Y. Cho†

Developmental Biology Center and the Department of Developmental and Cell Biology, University of California, Irvine, CA 92697-2300

Edited by Eric H. Davidson, California Institute of Technology, Pasadena, CA, and approved February 23, 2005 (received for review November 1, 2004)

Development is controlled by a complex series of events requiring sequential gene activation. Understanding the logic of gene networks during development is necessary for a complete understanding of how genes contribute to phenotype. Pioneering work initiated in the sea urchin and Drosophila has demonstrated that reasonable transcriptional regulatory network diagrams representing early development in multicellular animals can be generated through use of appropriate genomic, genetic, and biochemical tools. Establishment of similar regulatory network diagrams for vertebrate development is a necessary step. The amphibian Xenopus has long been used as a model for vertebrate early development and has contributed greatly to the elucidation of gene regulation. Because the best and most extensively studied transcriptional regulatory network in Xenopus is that underlying the formation and function of Spemann’s organizer, we describe the current status of our understanding of this gene regulatory network and its relationship to mesodermal patterning. Seventy-four transcription factors currently known to be expressed in the mesendoderm of Xenopus gastrula were characterized according to their modes of action, DNA binding consensus sequences, and target genes. Among them, nineteen transcription factors were characterized sufficiently in detail, allowing us to generate a gene regulatory network diagram. Additionally, we discuss recent amphibian work using a combined DNA microarray and bioinformatics approach that promises to accelerate regulatory network studies.

Vertebrate embryogenesis proceeds through a series of inductive events leading to changes in gene regulation. Through this process, signaling molecules produced by one cell population influence the developmental fates of neighboring cells. Defining these signaling processes and understanding the transcriptional regulation driving the subsequent changes in cell fate in the affected cells are primary goals of modern developmental biology. In amphibian development, the formation of mesoderm is the primary inductive event, one that delineates the three germ layers. As part of this process, a region of the embryo known as Spemann’s organizer is also induced. Spemann’s organizer activity was discovered in 1924 when Hilde Mangold heterotopically transplanted this region of the embryo into the flank of a host embryo of equivalent stage (1). An ectopic embryonic axis composed of both transplanted and host cells subsequently developed in the host embryo. Fate mapping revealed that host tissues normally fated to ventrolateral cell lineages were induced to form more dorsal tissue types in the secondary axis, demonstrating that the organizer region has inductive properties involved in development of dorsal axial structures.

This small piece of dorsal mesodermal tissue has understandably fascinated scientists for the decades since Spemann and Mangold’s groundbreaking experiments. We have come a long way since the 1920s in understanding the molecular basis of organizer activity. We now know, for example, that organizer formation and function are mediated by a number of transcriptional regulators expressed specifically in this region, including Goosecoid (Gsc), Xlim1, several members of the forkhead family, Siamois and Twin, Orthodenticle 2 (Otx2), and Xnot (2–9). Expression of all these organizer-specific genes falls under the control of signaling networks modulated by the action of various combinations of Activin/Nodals, bone morphogenetic proteins (BMPs), Wnts, and fibroblast growth factors (FGFs). Discovery of these organizer-specific homeobox genes in the frog embryo played a pivotal role in defining the organizer equivalent regions in chick, zebrafish, and mouse at the molecular level (10–12).

Similarities in sequences and expression patterns between the homologs from these species suggest that the underlying gene function and their regulatory mechanisms of axial induction are conserved between these disparate classes of vertebrates, despite seemingly large differences in the topography and morphogenesis of their early embryos. Equally important to organizer formation in development of the vertebrate body plan is the establishment of ventral marginal mesoderm. This less understood process is regulated by the activity of a variety of transcription factors including Xvent1, Xvent2, and Brachyury (Xbra) (13–15). These genes are predominantly expressed in mesodermal cells within the ventrolateral marginal region, where they specify mesodermal fate and define the boundary between dorsal and ventral mesoderm.

Previously, the gene regulatory network (GRN) approach was applied to endoderm and mesoderm specification in sea urchin embryos, where relationships between nearly 50 transcription factors were generated (16–18). By linking the direct input and output information of each interaction (based on the spatiotemporal expression patterns of the genes, DNA binding information, and perturbation analysis), it becomes feasible to predict the potential outcome of a given reaction. Recently, Loose and Patient (19) generated a GRN diagram of Xenopus mesoderm and endoderm formation by incorporating data on the action of various signaling molecules upon the activity of transcription factors downstream of signaling cascades. The current review is substantively different from that of Loose and Patient (19) because we primarily focus on “direct” interactions between transcription factors and their target genes. Such distinction is useful here if our intention is to predict the outcome of cis-regulatory interactions based on the GRN diagram. Additionally, we discuss the advantages of the Xenopus system to effectively accelerate the rate of the progress in establishing gene regulatory relationships in vertebrates, with great relevance to human biology.

Getting “Organized”

The large amount of work performed at the molecular level over the past 17 years is evidence that the formation of the Xenopus
organizer is one of the most extensively investigated experimental phenomena in developmental biology (2, 20–24). This wealth of information begins to allow us to build a transcriptional network. Table 1, which is published as supporting information on the PNAS web site, lists 74 transcription factors currently known to be expressed in the marginal (equatorial) region of late blastula/early gastrula stage Xenopus embryos and describes their modes of action, DNA binding consensus sequences, target genes, and relevant references.

To model a useful GRN, we believe it is essential to rigorously establish relationships between transcription factors and their downstream target genes. In the present review, for a potential target gene to qualify as a direct target of a given transcription factor, we required that the following conditions must be met. First, expression of a direct target should be affected by perturbations in the activity of the regulatory transcription factor in question in a manner consistent with the transcription factor’s proposed action. Changes in target gene expression after over- or underexpression (e.g., antisense morpholino-mediated translational inhibition) of the transcription factor can be verified in embryos or tissue explants by relative quantitation of target gene mRNA levels (e.g., RT-PCR, Northern blotting. RNase protection, etc.), and visualization of transcripts by in situ hybridization. To determine whether a given transcription factor is an activator or repressor, fusion of the factor to either the herpes virus VP16 transcriptional activation or engrailed repressor domains can be tested for their effects on target gene expression and resultant embryonic phenotypes (25). It has been shown previously that, if the VP16 transcriptional activator domain is fused to a heterologous (DNA-binding) transcription factor, the resulting fusion protein can function as a constitutively active form of the original transcription factor (26). On the other hand, fusion to the repressor domain of the engrailed protein can generate a protein that is a transcriptionally repressive version of the native factor (27).

The second criterion, a fairly obvious and critically important one, is that a putative direct target gene must be expressed temporally and spatially in a manner consistent with the expression of the transcription factor proposed to control it. For example, if the transcription factor is an activator, the target gene should be expressed in the same or overlapping region where the transcription factor is expressed. Conversely, if the transcription factor is a repressor, target gene expression should be excluded from the region where the transcription factor is expressed, unless the action of the repressor is controlled by other means. The third criterion we applied requires that a functionally relevant physical interaction between a transcriptional factor and the regulatory region of the proposed target gene must be established. The last piece of information is often lacking in the published literature, making it difficult to state with confidence whether the induction of a proposed target gene is direct or indirect. Experimentally, this last criterion can be met by using DNase footprinting, gel shifts, or appropriately designed reporter gene assays, performed in combination with mutational analyses. Alternatively, a hormone-regulatable form of a given transcription factor can be used in conjunction with the protein synthesis inhibitor cycloheximide (CHX) to delineate “directness” of control. For instance, a transcription factor can be fused to the glucocorticoid-receptor ligand-binding domain (GR) to create a dexamethasone-regulatable version of the protein. This inducible system has been particularly useful in Xenopus for identification of transcription factor target genes (28–33).

In our analyses, only if the above criteria were met for a proposed transcription factor and its downstream gene(s), were the relationships between these genes assigned as direct targets in Table 1. Therefore, targets of some transcription factors (e.g., Xnot, Xblimp1, and Foxa4a) were not included in our list, because currently available data do not yet satisfy these rigorous standards. It should also be noted that many well-known organizer genes (such as those encoding the secreted signaling inhibitors Chordin, Noggin, and Dkk1) were not included because they are not direct transcriptional regulators. Where possible, gene-specific cis-regulatory modules were built (as shown in Table 2, which is published as supporting information on the PNAS web site) summarizing the relationship between a transcription factor and its targets based on the information in Table 1. This information will be built upon in the future as modules that can be linked together to generate a more sophisticated model of this GRN.

GRNs in Xenopus Meseendoderm

Because the regulatory modules of Gsc, Xbra and Xvent2 are most extensively studied in Xenopus (see Table 2), we attempted to link the three modules to provide a timeline of gene interactions at successive stages of development for different cell and tissue types (illustrated in Fig. 1). Fig. 1A indicates the events occurring in the marginal region of the pregastrula (late blastula) stage embryo. During cleavage stages, Wnt signaling is activated in the dorsal side of the embryo after cortical rotation, and the Tcf/Lef members of the HMG family directly induce two closely related homeobox genes, Siamois and Twin (6, 7). These gene products, in turn, directly bind to the proximal element (PE) of the Gsc promoter (Fig. 2) at the pregastrula stage and activate its transcription (7). Although other homeodomain factors are shown to bind to the PE in vitro, most of these proteins are not yet expressed at this stage (Fig. 2). In addition to the Siamois and Twin input, the organizer and the remaining marginal zone begin to receive Nodal signals from the endoderm. Transcription of Nodal-related genes such as Xnr1, Xnr4, Xnr5, and Derriere is activated by VegT, a maternal transcription factor regulating meseendoderm development (34–37). The Nodal signal is then transduced to the nucleus by using two classes of maternal transcriptional factors, FoxH1 and Wbcrs11, which form complexes with both Smads 2 and 4 to regulate Gsc expression by means of the PE and the distal element (DE), respectively (38, 39).

At about the same time that Gsc expression begins, the expression of Xbra starts in the dorsal marginal zone and quickly encompasses the entire marginal zone by the beginning of gastrulation (15). Therefore, at this stage, Gsc and Xbra are briefly coexpressed within a population of the organizer cells (40, 41). Xbra is directly induced by Nodal, and FGF, and later by zygotic Wnt signaling, but it is not clear whether the induction of Xbra by BMP is direct (15, 42, 43). During this period, the expression of Otx2 and Xlim1 also starts in the organizer, and Mix-related homeobox genes, MixeR, Milk and Mix1, are activated in the mesendoderm by ActiVIN/Nodal signaling. On the other hand, in the ventral marginal zone, BMP signaling induces the expression of Xvent2.

The initial induction of Gsc is mediated by Siamois and Twin, but later Gsc expression is maintained by the transcription factors Otx2 and Xlim1 (Fig. 1B). During gastrulation, both Otx2 and Xlim1 become strongly expressed in the organizer, and each protein helps maintain the expression of Gsc in the organizer and subsequently in the prechordal plate meseendoderm by means of binding to the upstream element, DE, and PE of Gsc. At the same time, the overall expression level of Gsc is also maintained by Gsc itself: Gsc protein binds to the DE and the PE and negatively regulates its own expression (44), perhaps competitively binding to the Otx2 binding site within the DE (see Fig. 2). MixeR and Milk also bind to the DE together with Smads to maintain Gsc expression (45, 46). Thus, activation and maintenance of Gsc expression is mediated by a number of homeobox-containing transcription factors, which act on the DE and the PE. How do they all converge on the same sites to regulate Gsc expression? Unfortunately, we know very little about the extent
to which these transcription factors recognize the target sequences. Although there are multiple core homeobox-binding sequences (ATTA) within the DE and PE (Fig. 2), the precise binding sites and binding order of these homeobox proteins in vivo are not understood.

An important function of Gsc is to suppress the expression of ventralateral marginal zone genes in the organizer, thus ensuring proper specification of the organizer and development of dorsal structures. For instance, Xbra is initially expressed uniformly around the entire marginal zone, and its expression in the
organizer overlaps that of Gsc. However, the Gsc and the Xbra expression domains quickly segregate to generate two nonoverlapping expression domains due to Gsc’s ability to bind to the promoter of Xbra to suppress its expression in the organizer (40, 41). Similarly, Gsc directly represses the expression of zygotic Xwnt8 in the organizer to promote head formation by inhibiting the posteriorizing function of zygotic Xwnt8 (47). Gsc expression is also regulated by Xhex during gastrulation. At the beginning of gastrulation, Gsc and Xhex are coexpressed in the anterior mesendoderm and later have distinct expression domains. This change from coexpression to distinct expression domains occurs because Xhex down-regulates Gsc expression in the anterior endoderm (48).

For the specification of ventrolateral marginal zone, FGF signaling maintains Xbra expression by a positive autoregulatory loop as Xbra directly activates eFGF (49). Xbra also directly activates three other mesodermal genes, namely Xwnt11, Mix.4/Box1, and Bix4 (30, 50), presumably by means of an Xbra-binding site(s) (51). Regulation of Gsc by Xvent2 is also important for defining the organizer and ventrolateral mesodermal boundaries. Xvent2 directly binds to the DE of Gsc to ensure that Gsc is not expressed in the ventral marginal region (52). Additionally, Xvent2 also directly induces the expression of Bmp4, Xvent1, and Xvent2 itself. This induction sets up a positive autoregulatory loop to enhance the expression of Xvent-related genes to further promote ventral specification.

Regional specification of dorsolateral mesoderm also occurs during gastrulation. Myf5 is a basic helix–loop–helix (bHLH) transcription factor that is induced by Wnt and Nodal signaling. The expression of Myf5 is restricted to two symmetrical wings in the dorsolateral domain of the marginal zone (53), and Xvent2 is responsible for the inhibition of Myf5 in the ventral region (54). What controls the dorsal expression boundary is not clear at present.

As discussed above, we now know much about which transcription factors affect which genes. However, detailed temporal (and sometimes spatial) expression data and DNA binding data are not always available. Therefore, we are still left with the question of what combination of transcription factors is required to activate or repress target genes, and when these factors bind to the proposed regulatory sites to exert their effects. Our present knowledge allows us to describe the temporal events of gastrulation at only a relatively crude level, referring to these events as being either pregastrula or early-late gastrula stages (late blastula or early gastrula). Obviously, an effective transcriptional regulatory network diagram will require finer grain detail, which can be accomplished by a significant effort in data acquisition, an effort in which high-throughput approaches could be incorporated.

Advantages of the Xenopus System for Transcriptional Study

It is not difficult to argue that our level of understanding of the control of cell fate specification at the transcriptional level is still in its infancy. However, the ready availability of genomic tools such as cDNA microarrays (55, 56) and Affymetrix Xenopus oligo chips makes the Xenopus system ideally suited for studying transcriptional regulation on a wide scale for a number of reasons. First, Xenopus is one of a very few vertebrate model systems that permit easy cultivation of large numbers of embryos for the routine performance of DNA microarray analyses. Second, Xenopus embryos can be experimentally manipulated at the earliest stages of development to allow gain-of-function and loss-of-function analyses to be conducted with relative ease. Thus, microarray analysis can serve as a powerful adjunct to the rapid identification of genes affected by such manipulations, leading to the discovery of multiple potential downstream target genes in a single experiment. Third, reporter gene analyses are readily performed in embryos, either transiently or by using transgenic approaches (38, 57), in combination with biochemical approaches, to provide in vivo validation of experimental results. Lastly, sequencing of the genome of Xenopus tropicalis is nearly complete (http://genome.jgi-psf.org/Xentr3/Xentr3.home.html). The advantages listed above point to the utility of the Xenopus system for the collection of the large number of data sets required to study transcriptional regulation at the genome-wide level.

Genomic Tools for Transcriptional Regulation Studies

Functional characterization of cis-regulatory elements in Xenopus, as in any system, typically involves rather labor-intensive molecular approaches, including sequential promoter deletions, DNase footprinting, and gel shift analyses. However, the technology that can be brought to bear on such characterization is changing rapidly, and newer genomic approaches may become the norm in the future. We will briefly discuss two technologies that will likely be essential for rapid generation of data sets required to build a picture of the transcriptional regulatory networks operating in developing vertebrate embryos.

Microarray-Based Approach. A microarray-based approach should be useful in rapidly garnering a list of direct target genes regulated by a given transcription factor. An example is provided by the study of VegT, a transcriptional regulator fundamentally important to germ layer specification (58, 59). To better understand the mechanism by which VegT initiates this process, DNA microarray analysis was performed to identify VegT target genes (60). In this study, a glucocorticoid-receptor ligand-binding domain was fused to VegT to create a dexamethasone (Dex)-regulatable version of the VegT protein. Target genes induced (both directly and indirectly) shortly after Dex treatment were identified, and their expression patterns were studied. Among them, 14 were found to be direct VegT target genes because they
were induced even in the absence of ongoing protein synthesis. The promoter regions of these direct target genes, most of which had not been previously linked to VegT regulation, were shown to include VegT binding sites by homology to a consensus binding sequence (61). Although further work is necessary to confirm whether the binding sites identified possess in vivo significance, this approach illustrates the potential utility of the *Xenopus* system to rapidly identify biologically relevant, direct transcriptional targets based on DNA microarray results.

**Phylogenetic Comparisons and Genome Scanning.** The case of the *Drosophila* transcription factor Dorsal probably represents the most successful instance of target gene identification by means of a search for a defined cis-regulatory element. A combination of microarray and bioinformatics methods led to the identification of ~50 genes directly regulated by the Dorsal protein (62, 63). Achievement of this level of success in vertebrates has been challenging. The size of a typical vertebrate genome is over 10 times that of *Drosophila*. Therefore, the number of potential cis-regulatory elements increases at least proportionally and a greater complexity in the organization of these regulatory elements could lead to even greater challenges in the analysis. Searches for simple 6- to 8-bp consensus transcription factor binding sites will identify thousands of potential target sites, most that may not have in vivo significance. Despite this problem, a number of approaches have been reported that allow for the accurate prediction of conserved transcriptional regulatory regions through interspecies sequence comparisons (64–66). Unfortunately, most of the homologies identified using these methods are relatively long (e.g., >100 bp), and the identity of interacting transcription factors is unknown, necessitating further analysis to establish links between putative enhancers and their cognate DNA-binding transcription factors.

Recently, two *Xenopus* research groups reported the use of similar interspecies sequence comparison approaches to identify defined, conserved BMP-responsive elements within the promoters of a group of direct targets for BMP signaling (67, 68). BMP signaling is an ancient, highly conserved signaling pathway in animals. To uncover other as yet unidentified BMP targets in vertebrates, whole genome scans can be performed, and the results can be compared with a list of genes identified by microarray experiments or independently generated by PCR to provide in vivo evidence. This approach differs from searching for a consensus binding sequence de novo among various coregulated genes because it starts from a known consensus sequence. One weakness of the above approaches, however, is that it is difficult to extract information regarding temporal aspects of transcriptional regulation, i.e., determining when a protein is or is not bound to a regulatory site. To obtain such information, it may be useful to explore using a ChIP (chromatin immunoprecipitation) assay (69–71). Once a sufficient amount of information is gathered from the above experiments, we should be able to build a much more intricate model of the GRN controlling vertebrate development, one that may allow us to predict the outcome of loss- and gain-of-function alteration of a given transcription factor, which is our ultimate goal of building the GRN model. We hope this review together with that by Loose and Patient (19) will prove to be useful beginnings toward accomplishing this endeavor.

We thank Dr. Ira Blitz and Mr. Daniel Peiffer for critical reading of the manuscript, and the members of the K.W.Y.C. laboratory for useful discussion. This work was supported by National Institutes of Health Grant HD29507 (to K.W.Y.C.).


