Diffential Gene Expression in Vertebrate Embryos

Igor B. Dawid

Published, JBC Papers in Press, January 21, 2009, DOI 10.1074/jbc.X800018200

The invitation to contribute to the Reflections series in the Journal of Biological Chemistry is an honor that I greatly appreciate. It is humbling to find oneself in such distinguished company and a challenge to the literary limitations reflected all too obviously in the scientific writing I usually engage in. What I would like to do is to describe my own professional path in the context of the evolution of the field. In this I would like to focus on the story of how embryology, after joining forces with biochemistry and molecular biology, became to be usually called developmental biology and how the focus of my work and that of much of the field revolved around differential gene expression in different cell types, tissues, and developmental stages. As always in biology, it is a tale of new questions and concepts and of advancing technology that allowed old questions to be answered and new ones to be posed in a rational and approachable manner.

Becoming a Developmental Biologist: The Carnegie Years

I studied chemistry in Vienna, proceeding to a thesis in biochemistry that led to a Ph.D. in 1960. The way in which I reached the Department of Embryology of the Carnegie Institution after a postdoctoral stay at the Massachusetts Institute of Technology was both fortuitous and fortunate, as I have related elsewhere (1). The Carnegie Department in Baltimore, then headed by Jim Ebert, was a great place to be for a biochemist who wanted to enter developmental biology and establish his research in this field. Many factors made this an outstanding environment, first and foremost the wonderful colleagues at the Department, especially Don Brown, who had arrived a few years earlier and provided me with much guidance and stimulation. The problem that fascinated us was how does a fertilized egg develop into a complex organism with its varied cell types and complex shape? Clearly, genes had something to do with it, but how did they act to generate developmental variety? One point of interest was if and to what extent the genome itself might change during differentiation. Although the constancy of the genome through development was generally accepted at the time, exceptions did come up. My earliest work in the field dealt with “extra” DNA that was suspected to occur in the egg cytoplasm of our favorite model system, the frog Xenopus laevis. I could show that it is mitochondrial DNA, resolving the riddle of egg DNA and contributing to the work in several laboratories that at the time established the existence of DNA in this organelle (2, 3). A few years later, Don and I (4) and Joe Gall (5) independently showed that genes encoding ribosomal RNA were amplified in oocytes, presumably to be able to support the prodigious rate of rRNA synthesis in these cells. This was a developmental result, showing that, under some circumstances, gene amplification did occur to support developmental needs. Although amplification was later demonstrated for certain protein-coding genes in Drosophila oocytes as well (6), gene amplification is not a widespread mechanism for differential gene expression in development.

The generalization that different cells and tissues carry a constant genome but a variable population of proteins implies differential gene expression. Looking directly at individual protein-
coding genes, notably in animal cells, seemed impossible during the first decade of my time at the Carnegie Institution. Still, progress was made in identifying and characterizing individual abundant mRNAs from animal sources by translating globin mRNA in cell-free extracts or by injection into frog oocytes (7, 8) or by taking advantage of the beginnings of sequencing technology in the case of silk protein mRNA (9). The grand breakthrough came, of course, with cloning technology. This is not the place to attempt a history of this development but to relate how it affected me. I recall that I first learned about the successful cloning of vertebrate DNA in bacteria because the *Xenopus* ribosomal DNA used in this experiment had been donated by Don Brown to the group at Stanford who carried out this work (10). In the context of subsequent work on differential gene expression, the next major advance was the introduction of cDNA cloning (11). This technique made it possible to isolate individual mRNAs but also to generate a library representing the entire mRNA population in a cell or tissue, in other words, to look at the expressed sequences in the genome in a global way. Cloning and sequencing (12, 13) changed biology. My work, as that of so many others, was changed and enhanced by these developments in the most profound way. These techniques allowed me to transfer my interest from mitochondrial and ribosomal genes to protein-coding genes.

**Moving to the National Institutes of Health: Beginning to Work with cDNA Libraries**

By that time, in the late 1970s, I moved from the Carnegie Department of Embryology in Baltimore to the National Institutes of Health (NIH). Thanks to the efforts of Bob Goldberger, I joined the Laboratory of Biochemistry at the National Cancer Institute and a few years later switched to the National Institute of Child Health and Human Development (NICHD), which has remained my home ever since. The greatly appreciated support from the Institute has made it possible for me and my colleagues to pursue our interests for all this time. Right around the time of our move to NIH, we started to study differential gene expression in development in a general way by applying cDNA cloning technology to look at changing mRNA populations in the frog embryo. As the methods for preparing cDNA libraries improved, the idea to interrogate various libraries with probes derived from RNA populations of different developmental stages, multiple tissues, or cells exposed to various conditions was bound to arise. It took a few years to turn this idea into praxis. Although cDNA cloning was announced in 1976 (11), comparisons of RNA populations continued to be made by using hybridization kinetics throughout much of the decade. Starting in 1979 and 1980, several publications appeared that made use of cDNA libraries to investigate developmental or physiological changes in RNA populations. A variety of biological systems were subjected to this approach: developmental changes in *Dictyostelium* (14, 15) and *Aspergillus* (16), effects of starvation in yeast (17), and changes induced by auxin in plants (18) and in embryonic development in sea urchins (19) and in *Xenopus* (20).

Our work with cDNA cloning began with studies on vitellogenin mRNA in *Xenopus* by Walter Wahli (21) and, directly relevant to the theme of this article, with the work initiated by Mark Dworkin after he joined the lab as a postdoctoral fellow in 1978 (20). Later, Mark continued his research at Columbia University and in Vienna, Austria, but died of cancer at a tragically young age. At the time Mark came to the lab, it was known that eggs and embryos of frogs and other animals contained a large population of mRNAs whose general properties had been measured by renaturation kinetics, as summarized elegantly in Eric Davidson’s book *Gene Activity in Early Development* (22). These studies revealed the total complexity of the embryonic mRNA population and the overall distribution and developmental changes of classes of RNAs of certain ranges of abundance; yet for all the power of these global analyses, it seemed likely that they masked variations in the behavior of individual RNAs that could be revealed only by looking at them one at a time. To do this, we generated cDNA libraries from different embryonic stages and then randomly selected about 800 clones for further study. These clones were grown and transferred to filters (23) to obtain what we would now call in hindsight a macroarray. To probe these arrays, we prepared RNA from five embryonic stages and two tissues, in some cases using different cell fractions for RNA isolation, and synthesized radioactively labeled cDNA from each RNA sample. Hybridization of the arrays with these probes allowed us to compare the RNA populations in the different embryonic stages in a semiquantitative way. Given the technical limitations of these procedures, only moderately to highly abundant RNAs/cDNAs yielded a detectable signal, and as a result, only ~30% of the cDNA clones could be visualized with any of the probes that we used. Nevertheless, the experiment led to some interesting conclusions, as illustrated in three panels taken from the original publication (Fig. 1). First, it was clear that many of the clones that could be visualized were present at all stages tested, presumably representing widely expressed or so-called housekeeping genes. Second, some clones were differentially expressed; although this was expected, it demonstrated the feasibility of identifying many differentially...
Subtractive Cloning: An Effective Approach to Studying Differential Gene Expression

The approach of hybridizing cDNA libraries with labeled cDNA preparations was capable of revealing differences between RNA populations and could lead to the isolation of differentially expressed genes, but a major limitation was the fact that only moderately to highly abundant RNAs could be visualized in this method. Thus, the patterns observed were dominated by abundant RNAs, often by widely expressed species corresponding to housekeeping genes that are represented in multiple copies in standard cDNA libraries. A major advance that removed this limitation came through a combination of cDNA cloning with methods based on the kinetics of hybridization of nucleic acids in solution that had been worked out mostly by Roy Britten, Eric Davidson, and their colleagues (22, 24, 25). When complementary nucleic acid strands are annealed, the rate of duplex formation is proportional to the concentration of the two strands for each particular sequence. Solution hybridization in combination with cDNA synthesis and cloning is the basis of subtractive cloning for the isolation of differentially expressed genes. The method was developed independently in our lab by Tom Sargent, who continues to be my colleague as a Senior Investigator at NICHD, and by Mark Davis, at that time also at NIH, and colleagues (26–28). In our case, it involved the synthesis of cDNA from mRNA isolated from gastrula stage Xenopus embryos, followed by hybridization of the cDNA in solution with an excess of RNA from eggs. cDNA that failed to hybridize, i.e. that is not or is very rarely represented in egg RNA, was isolated, rendered double-stranded, and cloned in a plasmid vector. The great advantage of this approach compared with direct interrogation of conventional cDNA libraries is that copies of genes expressed in both stages are eliminated or greatly reduced, more effectively the more highly expressed they are. This is important as abundant RNAs dominate standard libraries, confounding efforts to identify low abundance differentially expressed genes. Thus, the subtractive method not only gives direct access to differentially expressed genes but at the same time expands the range of genes that can be effectively screened to the class of low abundance transcripts.

By now, subtractive cloning appears to be an obvious and simple method, but it presented substantial technical challenges at the time. The successful resolution of these challenges gives an illustration of two major uses for this technique. One aim is to clone a particular gene known to be differentially expressed; this was the goal that Mark Davis and colleagues achieved in isolating the T cell receptor (27, 28). In Tom Sargent’s studies in our lab (26, 29), the aim was to achieve a broad comparison of gene expression in different stages of frog embryogenesis, trying to characterize patterns of gene activity and to identify genes that have a role in development. Our initial studies showed that frog gastrulas contain many gene transcripts not represented in the maternal RNA pool (Fig. 2); we named them DG genes, for differentially expressed in gastrula. Although some of the gastrula RNAs continue to be expressed in later development, others persist for a short only, suggesting that they have a time-limited role in the embryo. Among the RNAs that attracted our attention in the initial work, some were studied extensively in subsequent years. Of these, DG42 seemed of developmental interest because of its regionally and temporally specific expression (30), but its identity remained unclear for a long time. Eventually, through the efforts of others laboratories, it proved to represent the gene encoding vertebrate hyaluronan synthase, an important gene that had
long resisted cloning efforts (31–33). Our own work was focused for several years on a group of genes with differential expression patterns that proved to encode cytokeratins. Some cytokeratins in this group are specifically expressed during embryonic and tadpole development in the frog, and they were useful as epidermal markers, which we explored in work involving Milan Jamrich, Jeff Winkles, and others (34, 35). A particularly interesting application of the DG library was its use in isolating Mix.1, the first gene known to be induced by a signaling factor in the embryo (36). In this experiment, the subtractive approach was used twice. Explants from Xenopus embryos were treated with activin, a factor that is able to induce endoderm and mesoderm in the naïve explant, and the genes induced by activin were enriched by subtraction of the cDNA generated from induced explants with ovarian RNA. The enriched cDNA was used to probe the DG library, allowing the isolation of several activin-induced genes, among which the homeobox gene Mix.1 was the primary focus of attention. A similar approach for the identification of inducible genes has generated useful results in subsequent years.

In Situ Hybridization as a Means to Find Differentially Expressed Genes

Many ideas and many techniques have contributed to the rise of biology over the past half-century. One of the most widely useful techniques is in situ hybridization, first applied to localize specific DNA sequences by Gall and Pardue (41, 42) and subsequently used to localize individual RNAs in cytological sections and in three-dimensional objects such as embryos by whole mount techniques (43–45). Fine cytological detail is achievable in this technique, making it possible to study gene expression in an embryo at a cellular or even subcellular level. The approach usually taken is to isolate a gene of interest in one of a number of ways and then study its expression pattern by the in situ hybridization technique. Using in situ hybridization as a method to identify genes of developmental interest and to survey the broad range of expression patterns that occur was first done, to my knowledge, by Niehrs and colleagues using Xenopus embryos (46). Our laboratory, in particular Tetsu Kudoh, Michael Tsang, Neil Hukriede, and others (47), and the laboratory of Thisse and Thisse (ZFIN Zebrafish Model Organism Database ID ZDB-PUB-010810-1) applied this approach to the zebrafish embryo. Given the considerable spatial resolution that can be obtained (for illustration, see Fig. 3), the method reveals a wealth of patterning data that are useful at different levels of analysis. First, a catalog of expression patterns is
markers now available, be they different cell types and tissues. The wealth of molecular available. Second, the approach generates new markers for ing on a new gene, its expression pattern may already be obtained that accelerates research: when you start work-

FIGURE 3. Montage of expression patterns as visualized by in situ hybridization for a selection of cDNA clones from the zebrafish. Each image represents an embryo hybridized with a different cDNA clone. Different stages from the gastrula to 3-day-old larvae are repre-

sented in this montage, which was generated to illustrate the wide variety of patterns that are obtained. For example, genes expressed in the head or tail, the somites, or the nervous system and in multiple distinct patterns within the nervous system have been visualized. The figure was assembled by Michael Tsang based on data obtained in our screen (the entire data set is available at the ZFIN Zebrafish Model Organism Database) (47).

obtained that accelerates research: when you start work-

ing on a new gene, its expression pattern may already be available. Second, the approach generates new markers for different cell types and tissues. The wealth of molecular markers now available, be they in situ probes or antibod-

ies, has ushered in a higher level of anatomical definition than had been possible previously. Third, this is an approach to functional genomics, i.e. to infer functions of novel genes. Although pattern alone cannot provide reli-

able functional information, it can provide guidance to further experiments. Fourth, access to functional information is facilitated in the particular situation in which a novel gene is recognized as a member of a synexpression group (46). This term refers to a group of genes that have similar complex expression patterns in development, and in many instances, members of a synexpression group prove to be components of a common functional pathway. This allows identification of novel components in various pathways. For example, a feedback inhibitor of Fgf signaling named Sef was discovered in this way (48, 49).

The results of in situ hybridization screens can be help-

ful in drawing attention to previously unexplored relations-

ships, even when the genes under study have been de-

scribed in earlier work. Such an example is provided by Tetsu Kudoh’s work in our laboratory that took its cue from observing the expression of the cyp26 gene in the anterior neural ectoderm of the zebrafish embryo (50). This gene encodes all-trans-retinoic acid 4-hydroxylase, an enzyme that degrades retinoic acid. As retinoic acid is known to affect brain development in guiding differentiation along the anterior-posterior axis, the expression pattern of cyp26 suggested to us that this gene had a role in defining the retinoic acid gradient along this axis. This idea led to a study in which we explored the role of retinoic acid, Wnt, and Fgf in the organization of the early nervous system in the zebrafish, with a particular focus on the inte-

gration of these different signaling pathways in the process (50).

Application of DNA Microarray Technology to the Study of Embryogenesis

Hybridization of a nucleic acid probe to nucleic acid fixed on a support has a long history, from the early efforts of Bolton and McCarthy (51) to the “paradigm changing” introduction of nitrocellulose as support by Nygaard and Hall (52) and to the introduction of gel-to-membrane transfer by Southern (53). The idea of applying target DNAs in the form of dots to a solid support likewise was introduced some time ago (54, 55), but the approach changed dramatically when, based on the exploitation of genome and EST sequencing efforts, the modern microar-

ray procedure was introduced (56, 57). Given the high density of probes on the slide or chip, it becomes possible to compare gene expression patterns in different cells for many (ultimately all) of an organism’s genes at the same time. Multiple wide reaching applications have been developed for this technology.

Our use of the microarray technology got started in ear-

nest when Affymetrix® generated chips for X. laevis and the zebrafish. Although the genome sequences of the zebrafish and Xenopus tropicalis, a species closely related to X. laevis, were at the time incomplete (and continue to be so) and the arrays therefore represent only about half of the genes in the two species, a great deal of information can be generated in such experiments (58 – 61). We car-

ried out several types of analyses on embryonic material from frogs and fish. In particular, experiments that com-

pared different regions of Xenopus embryos proved useful. Hui Zhao and Kosuke Tanegashima separated gastrula and early neurula embryos into several parts by microdis-

section, compared the different RNA populations by microarray analysis, and subjected individual genes with differential expression to further analysis. This approach yielded information on a guanine nucleotide exchange fac-

tor with a role in the Wnt-planar cell signaling pathway that controls cell movements in gastrulation (62) and revealed the role of the transmembrane protein Lrig3 in modulating signaling pathways to control neural crest for-

mation in the Xenopus embryo (63).
Using the zebrafish, we have carried out several microarray experiments, including an as yet unpublished study describing the transcriptome of the pineal gland in a global way (R. Toyama, X. Chen, N. Jhawar, J. A. Epstein, Y. Gothilf, D. Klein, and I. B. Dawid, unpublished data). In another project, we focused on targets of the Fgf signaling pathway in early development. Sung-Kook Hong injected embryos with mRNA encoding Fgf or a dominant-negative form of the Fgf receptor, and the RNA populations isolated from both classes of embryos were compared with controls by microarray analysis. As a result of these experiments, we identified two factors, Ier2 and Fibp1, whose possible functions in embryogenesis had not been characterized. Ier2 and Fibp1 proved to be involved downstream of Fgf signaling in the establishment of left/right asymmetry in the zebrafish (S.-K. Hong and I. B. Dawid, unpublished data). This example again illustrates how differential expression analysis can assist in functional genomics to help identify novel factors and associate novel functions with previously described proteins.

The Future of Differential Expression Analysis

The pace of progress in biology over the past half-century has been so rapid and the advances so profound that predictions are bound to be wrong, especially coming from someone whose general inclination to caution has invariably produced overly conservative estimates of future developments. Still, there is little doubt that differential expression analysis, increasing in range, resolution, and precision, will continue to be an important approach in many fields of biological research, notably in developmental biology but also in other applications such as diagnosis of diseases and searches for targets for pharmacological intervention. Among currently used techniques, subtractive cloning remains a useful tool 25 years after its initiation. DNA microarray methods have generated a broad range of applications that likely will grow further for the foreseeable future, especially as uses in diagnosis, haplotype mapping, and other areas expand. When applied to cDNA copies of RNA populations, new ultrahigh-throughput sequencing methods have the potential to supply the most complete information about the genes that are expressed in the material of interest (64). Effective comparisons can then be carried out between different cells and tissues or embryonic stages and between different metabolic or differentiation states of cells and embryos. It seems that with this methodology we are poised to achieve the ultimate aim of being able to describe the differential expression of all genes in a developing organism. The continuous refinement of the description of differential gene expression will enrich efforts to elucidate the transcriptional control mechanisms and intercellular signaling pathways that combine to promote the differentiation and patterning of the embryo.

A Personal Conclusion

On rereading what I have written, it appears less clearly a memoir than some of the articles in the Reflections series, having more the character of a thematic piece, yet I am trying to focus on the work we have done in the area of differential gene expression while placing our work into a broader context. To achieve some degree of thematic unity, I mentioned briefly or not at all other areas that occupied my interest over the years. This also accounts for the fact that I did not mention so far several colleagues who featured in the development of my laboratory’s research, such as Peter Wellauer, whose elegant electron microscopy explored ribosomal RNA and DNA structure (65, 66); Susan Haynes, Brian Mozer, and Alex Mazo, who first isolated the Drosophila trl and fsh genes (67–69); Paolo Di Nocera and Brian Kay, who studied repetitive DNA elements (70, 71); Masanori Taira, who developed our work on Lim homeodomain transcription factors (72, 73); and Jean-Pierre Saint-Jeannet and more recently Raymond Habas, who studied Wnt signaling (74–76). These and many more former and present colleagues have greatly enriched my life in science and beyond.

I feel that I have been most fortunate in having the opportunity to engage in a career of research in biology. Looking at it from the current vantage point of single-digit pay lines, one might think that scientists of my generation had an easy time of it. In some ways that may well be true, even though at the time the field seemed highly competitive, and the mountain to climb toward an independent research career appeared steep. It certainly has been a propitious time to work in biology, developmental biology in particular; the field was in a sustained growth phase, with exciting advances following each other in rapid succession, and broad support was available for the work in this country and elsewhere. Now, biological research has more the appearance of a “mature” enterprise, with obvious limitations on growth; yet the excitement has not left the field, progress is as rapid as ever, and many opportunities for fundamental insights and for practical applications arise continuously. It remains, I believe, a great field for a young scientist to pursue a career as a principle investigator, albeit not for every person who enters the field; a series of talents is needed to succeed in this age, some of them different from those demanded of us. The basic requirements for a research career, an interest in science, an understanding of what the important questions are, and
an ability to do the “right” experiment, have not changed. For those who have these talents, a career in science remains a great choice. It certainly has been for me.

Acknowledgments—This work was supported by the National Institutes of Health Intramural Research Program of the NICHD. I thank Michael Tsang for the assembly of Fig. 3.

Address correspondence to: idawid@nih.gov.

REFERENCES
