© 2024. Published by The Company of Biologists Ltd | Development (2024) 151, dev202512. doi:10.1242/dev.202512

PERSPECTIVE

A blueprint most wonderful, the homeobox discovery

William McGinnis[1,*] and Michael Levine[2]

ABSTRACT

This is a personal, non-linear summary of the discovery of the homeobox, a short DNA sequence encoding a DNA-binding domain conserved in developmental control genes. It is based on our recollections, a few decaying lab notebooks and letters, the early research papers we published, and conversations with a few colleagues who were in Basel at the time. It presents a simple story, when the research we did was anything but, with failed experiments, blind alleys and dumb ideas. Homeobox DNA sequences were independently discovered by Matt Scott and Amy Weiner in Thomas Kaufmann's lab at Indiana University (Scott and Weiner, 1984). The accompanying Perspective from Scott (2024), provides their fascinating story.

A variety of animals have homeoboxes!

We were both postdoctoral fellows in the laboratory of Walter Gehring in Basel, Switzerland. We had complementary expertise, temperaments and personal narratives (e.g. Mike, born in Los Angeles and Bill, Warrensburg, Missouri). As a graduate student at the University of California (UC) Berkeley, Bill had learned how to detect partial sequence homology matches to DNA using low stringency hybridization on genomic DNA Southern blots. As a graduate student at Yale, Mike had learned how to label DNA probes either with biotin or with high levels of ^{32}P radioactivity; this occasionally led to some highly radioactive lab benches in the Gehring lab!

In early June 1983, shortly after the initial discovery of the homeobox sequence in *Drosophila*, we made our first attempt to test whether such sequences also existed in other animal groups. We used the technique of low-stringency nucleic acid hybridization to DNAs contained on a 'Zoo blot' that displayed genomic DNAs from a variety of animal species, including worms, grasshoppers and mammals. Some of the specimens were obtained from local bait and tackle shops and initially misidentified, as Bill had a sketchy knowledge of invertebrates, but we eventually figured out what species belonged to which DNA samples. Bill developed an autoradiogram of the Zoo blot after hybridization with a red-hot radiolabeled *Drosophila* DNA fragment from a homeotic gene (*Antennapedia*; *Antp*) cDNA. Even in the dim red light of the darkroom, he could see a faint pattern of signals on the film, showing where a *Drosophila* homeotic DNA sequence (the homeobox) hybridized within the genomic DNAs represented on the Zoo blot. The modest number of bands (which were validated soon after as they hybridized with other *Drosophila* homeobox probes) suggested that these widely divergent animals, which had a common ancestor about 600 million years ago, might share homeotic gene sequences and functions with *Drosophila*. Bill grabbed the film, still dripping wet with developer, and ran out of the darkroom to show Mike.

We then showed the Zoo blot autoradiogram to everyone in the lab, bragging that it meant that flies and mammals employed similar developmental mechanisms despite their seemingly divergent modes of embryogenesis. At the time we believed, and still do, that we were the first to do experiments showing a common blueprint for 'endless [animal] forms most beautiful and most wonderful' (Darwin, 1859). These findings also provided molecular evidence that fruit flies and mammals had a common animal ancestor, no matter how different their morphology.

Although we didn't realize it at the time, a common ancestry for arthropods and mammal body plans had been proposed 160 years earlier by Etienne Geoffroy St. Hilaire (Geoffroy Saint-Hilaire, 1822), as Stephen J. Gould pointed out to us later (Gould, 1985). Alas, almost no one believed St. Hilaire's proposal in his time, or later. That said, about 25 years later, when Bill was visiting Caltech, Ed Lewis showed Bill a grant proposal from 1980 where he proposed that homeotic genes in a cluster he named the Bithorax complex (BX-C) would be found in humans. In Ed's typically modest fashion, he said he had added that speculation only to increase the chances the grant would be funded. Whatever the reason, he was right.

In any case, our manic pronouncements about the conservation of homeotic genes were quickly moderated by the sober Swiss culture permeating the lab. As we showed the X-ray film around, our lab mates pointed out various ways we could have screwed up the experiment or its interpretation, and doubts crept in. Was it possible that the fly or mammal DNA might be cross-contaminated? Was the right DNA in the right lanes of the gel? Were we reading the banding pattern on the negative correctly? What if the homeotic sequences we used were found in many other *Drosophila* genes that had nothing to do with development? What if the homeobox DNA sequence encoded a boring sequence of RNA or protein that had nothing to do with their developmental functions? And so on, as often happens when you get an astonishing result. We alternated from elated to deflated as we discussed each possibility, and argued about why each was unlikely to be true. However, with many repeats of the experiment, we gained confidence in the results.

The Zoo blot experiment was the culmination of a series of fast-paced studies that identified the homeobox sequence, first in the homeotic gene *Antp*, and then in each of the other seven protein-coding homeotic genes in the Antennapedia and Bithorax gene complexes (now called Hox clusters in flies and other animals). We both had come to the Gehring lab in large part because he was one of the few *Drosophila* biologists who was applying molecular techniques to developmental biology. Walter's lab management style was ideal for us, as he was largely hands-off; if results were coming in, he was happy! Unfortunately, Walter missed the excitement of the early homeobox experiments, as he was enjoying a short sabbatical in France in the Spring of 1983.

[1]Department of Cell & Developental Biology, School of Biological Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0349, USA. [2]Department of Molecular Biology, Princeton University, 119 Lewis Thomas Laboratory, Washington Road, Princeton, NJ 08544-1014, USA.

*Author for correspondence (wmcginnis@ucsd.edu)

W.M., 0000-0001-8474-0388

Development (2024) 151, dev202512. doi:10.1242/dev.202512

The initial discovery

How did we find the first *Drosophila* homeobox sequences that were used for the Zoo blot? There was a scramble to clone animal developmental control genes in the early 1980s. At the time, no one knew what these genes might be doing inside the cells of developing embryos. For us and many others, the most interesting class of developmental control genes were the fly homeotic genes. Mutations in these genes exhibited fantastic phenotypes, whereby entire body parts were transformed into structures normally located elsewhere. Ed Lewis had long studied the BX-C, as mutations in these genes, such as *Ultrabithorax* (*Ubx*), transformed fly abdominal segments into thoracic segments (learn more in Scott, 2024). Interestingly, the genes were arranged on the chromosome in an order that mimicked the order of the body regions they controlled (Lewis, 1978).

Thomas Kaufman and his coworkers had studied a different cluster of genes, the Antennapedia complex (ANT-C), some of which had homeotic functions in the posterior head and anterior thorax (Wakimoto and Kaufman, 1981; Scott et al., 1983). For example, dominant mutations in a fly homeotic gene called *Antp* partially duplicated thoracic structures, such as legs in place of the antennae. From the abnormalities seen in homeotic mutant animals, it appeared that this class of genes had enormous power in controlling morphological traits. The homeotic genes had an air of mystery in the early 1980s, which was conferred by the complicated genetics, the lack of knowledge of their molecular function, and their abstract functions, controlling developmental fates in zones of the body, rather than specific cell types, tissues or organs.

In 1982, Rick Garber and Atsushi Kuroiwa, postdocs in the Gehring lab, isolated cDNAs from the *Antp* gene (Garber et al., 1983) and, soon thereafter, Mike and Ernst Hafen, a senior graduate student in the lab, used an *Antp* cDNA probe to detect RNA in *Drosophila* embryos. Establishing the RNA detection method took months of mind-numbing repetitions and controls. Fortunately, Ernst was a local Basler who knew all the best pubs with the local brew, which helped to propel them through the monotony of working out the method. Frozen sections of *Drosophila* embryos and larvae were hybridized with radioactive (tritiated) DNA probes. The sections were then dipped in a photographic emulsion to detect *Antp* transcripts in the tissue sections. Ernst and Mike were excited to see that the *Antp* cDNA probe detected gene transcripts localized in the future second thoracic segment (Levine et al., 1983), a known site of *Antp* gene activity based on Kaufman's genetic studies. Their efforts to work out the method benefited from exchanges with Michael Akam, who found the same correlation was true for *Ubx*, transcripts of which were localized in posterior thorax and abdominal segments (Akam, 1983).

Mike and Ernst discussed their results with Mike's roommate, Erich Frei, a graduate student in Markus Noll's lab. This led to a key experiment: the hybridization of the *Antp* cDNA probe to embryos with deletions for all the BX-C genes. The *Antp* radioactive signals were no longer restricted to the second thoracic segment but extended into more posterior thoracic and abdominal segments. These results led to the 'posterior prevalence' model, whereby posterior BX-C genes repress the transcription and function of homeotic genes expressed in more anterior regions (Hafen et al., 1984).

Soon after, Bill arrived in the Gehring lab and realized the methods he had proposed in his postdoc fellowship application to clone new homeotic genes using transposable element tagging were hopeless in Basel, so he looked around for different projects. He began a short collaboration with Bernard Mechler, who was cloning a *Drosophila* cancer suppression gene called *lethal(2) giant larvae* [*l(2)gl*]. Bernard and Bill tried to find *l(2)gl* coding sequences by hybridizing more than a dozen cloned sequences from the chromosomal region of *D. melanogaster l(2)gl* to Southern blots with genomic DNA of other *Drosophila* species, reasoning that only coding sequences would be conserved in the other species. Bill made Southern blots with genomic DNA from *D. melanogaster* and two other fly species and concluded this approach wouldn't work for *l(2)gl*.

Later, we went to lunch with Ernst, and Mike suggested in no uncertain terms that it might be a lot more interesting to try a similar experiment with the *Antp* cDNA. At the time, we had no idea whether the *Antp* gene even made a protein and no clue about the structure of its gene product, save that *Antp* encoded an RNA. But we did know, based on research from the Lewis and Kaufman labs, that there seemed to be about a dozen homeotic genes in the Antennapedia and Bithorax gene clusters (we now know there are eight). We also knew that animal genes in clusters sometimes had similar protein-coding sequences (e.g. the globin genes; Lawn et al., 1978), so perhaps the homeotic genes in *Drosophila* were molecular relatives as proposed by Ed Lewis (Lewis, 1978).

In mid-March of 1983, Mike labeled a *D. melanogaster Antp* cDNA, and we did a low stringency Southern blot with DNA from *melanogaster* and two other *Drosophila* species. The results were not promising. There were hundreds of signals on the Southern blot, both in the lanes with *D. melanogaster* DNA as well in the lanes of other fly species, far too many signals to believe we were working with a sequence that was restricted to homeotic genes.

We discussed our results with Rick Garber, who told us that Michael Akam's lab in Cambridge had tried a similar experiment with DNA sequences from the 5′ exon of the *Ubx* gene, which had been isolated in David Hogness's lab at Stanford (Beachy et al., 1985). That piece of *Ubx* DNA had also provided hundreds of signals when it was hybridized to a *Drosophila* Southern blot, and Rick heard that, upon sequencing of that exon of *Ubx*, Akam and coworkers found that the repeat was a long stretch of repeated G/C base pairs, a sequence that, if it encoded any protein fragment, would result in a monotonous stretch of glycine amino acids. This was a boring result in the 1980s, but today we now know such sequences as low complexity regions (LCRs) or intrinsically disordered regions (IDRs) and they are thought to be the driving force for the formation of cellular condensates.

After this, we met with Ernst and Atsushi to plan how to go forward. We decided to cut the *Antp* cDNA into small fragments, reasoning that some fragments might have high-copy repeats and some low copy. Once we used this approach, it was clear that one fragment contained a low-copy repeat, hybridizing to only 8-10 fragments on Southern blots with *D. melanogaster* or two other *Drosophila* species, while another fragment hybridized to hundreds of different sequences in each of the *Drosophila* lanes. We also found cross homology between clones containing the *Antp* low-copy repeat and clones of *Ubx* 3′ RNA coding sequences we obtained from Pierre Spierer, a postdoc from the Hogness lab (Bender et al., 1983). Both homeotic gene clones also hybridized to coding sequences from the *ftz* gene, a *Drosophila* segmentation gene now known to be an evolutionarily derived version of a homeotic gene (Löhr et al., 2001). We cloned a few of the low-copy hybridizing fragments from the unknown genes, and Mike and Ernst used them for *in situ* hybridization experiments on fixed *Drosophila* embryos. The RNAs encoded by a few of these low-copy fragments localized to different regions on the anterior-posterior axis of fly embryos! Mike followed this up with *in situ* hybridizations to fly polytene chromosomes and found that the low copy repeats mapped to the chromosomal regions of the Antennapedia and Bithorax

Development (2024) 151, dev202512. doi:10.1242/dev.202512

clusters! Still not knowing what the repeated sequences encoded, we rushed a paper into *Nature* with these experiments (McGinnis et al., 1984a). In that paper, we labeled the low-copy repeated sequences 'H-repeats' (for Homeotic gene-repeat sequences). Initially, we were puzzled that we could find only three H-repeat sequences in the cloned BX-C DNA that was provided by Welcome Bender and Francois Karch (Bender et al., 1983; Karch et al., 1985; Regulski et al., 1985), even though Lewis had found about eight homeotic genes in that cluster. By chance, we talked to Gines Morata at a meeting in Switzerland, and he told us that he and Ernesto Sanchez-Herrero had found there were only three lethal homeotic genes in the cluster (Sánchez-Herrero et al., 1985). That reduced one considerable source of anxiety about the correlation between H-repeats (homeoboxes) and homeotic genes.

Soon thereafter, we changed the H-repeat label to homeobox once we got the DNA sequence of a few homeobox DNA coding regions, as the sequence homology transitioned so abruptly from near-perfect to non-existent that you could draw boxes around the common amino acid coding sequences in different homeobox genes (Laughon and Scott, 1984; McGinnis et al., 1984b). These proposed homeobox sequences were proposed to encode sequence-specific DNA binding domains (homeodomains) based on their slight similarity to the DNA binding domain of yeast mating-type proteins (Laughon and Scott, 1984; Shepherd et al., 1984), and evidence for this DNA binding function was shown by Claude Desplan and coworkers (Desplan et al., 1985) using filter binding assays.

As described earlier, we used Zoo blots to identify similar homeobox sequences in other animals, including mammals (McGinnis et al., 1984b). Andres Carrasco in Eddy DeRobertis's lab, adjacent to the Gehring lab in Basel, was the first to clone and sequence a vertebrate homeobox sequence, in the frog *Xenopus* (Carrasco et al., 1984), which proved that the hybridizing homeobox sequences in vertebrates encoded homeodomains. In fact, Eddy was so excited to test for homeobox sequences in *Xenopus* that the morning that Walter approved giving out homeobox clones, Eddy rushed to find Bill in a Biozentrum bathroom to get the clones! Frank Ruddle was in Walter's lab on sabbatical at the time, and he and Bill cloned and sequenced a few mouse homeobox genes (McGinnis et al., 1984c). By that time, Mike had departed Basel for Berkeley, where he cloned and sequenced a few human homeobox genes with Bob Tjian and Gerry Rubin (Levine et al., 1984).

Early on, we provided *Drosophila* homeobox clones to hundreds of labs around the world. The subfamily of homeobox genes in the fly clusters and other animals have long been referred to as Hox genes, as the cloning frenzy that started with the Hox homeoboxes led to the discovery of many other subfamilies of animal homeobox genes, almost all of which control developmental patterning and cell type identity. The research on the Hox family and the other homeobox families has led to the realization that animals employ a similar 'toolkit' of genes to control how embryos develop into adults with different morphologies.

This article is part of the collection '40 years of the homeobox'. See related articles in this collection at https://journals.biologists.com/dev/collection/10249/40-years-of-the-homeobox.

Acknowledgements

We thank Matt Scott, Ernst Hafen, Markus Noll, Donald Rio, and Ella Tour for their comments on the manuscript. Our colleagues in Basel in 1983-84 contributed immensely to the homeobox research, with ideas, probes and insightful criticism. Regrettably, we have had to omit the mention of some who made key contributions – our apologies, but that would take a much longer treatise than this abbreviated version. We were notably aided by the atmosphere of the second floor of the Basel Biozentrum, particularly all the wonderful people in the labs of Walter Gehring, Eddy DeRobertis and Markus Noll. What an exciting time it was, full of colorful and interesting characters!

References

Akam, M. E. (1983). The location of Ultrabithorax transcripts in Drosophila tissue sections. *EMBO J.* **2**, 2075-2084. doi:10.1002/j.1460-2075.1983.tb01703

Beachy, P. A., Helfand, S. L. and Hogness, D. S. (1985). Segmental distribution of bithorax complex proteins during Drosophila development. *Nature* **313**, 545-551. doi:10.1038/313545a0

Bender, W., Akam, M., Karch, F., Beachy, P. A., Peifer, M., Spierer, P., Lewis, E. B. and Hogness, D. S. (1983). Molecular Genetics of the Bithorax Complex in Drosophila melanogaster. *Science* **221**, 23-29. doi:10.1126/science

Carrasco, A. E., McGinnis, W., Gehring, W. J. and De Robertis, E. M. (1984). Cloning of an X. laevis gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes. *Cell* **37**, 409-414. doi:10.1016/0092-8674(84)90371-4

Darwin, C. (1859). *The Origin of Species*. London: John Murray Press.

Desplan, C., Theis, J. and O'Farrell, P. H. (1985). The Drosophila developmental gene, engrailed, encodes a sequence-specific DNA binding activity. *Nature* **318**, 630-635. doi:10.1038/318630a0

Garber, R. L., Kuroiwa, A. and Gehring, W. J. (1983). Genomic and cDNA clones of the homeotic locus Antennapedia in Drosophila. *EMBO J.* **2**, 2027-2036. doi:10.1002/j.1460-2075.1983.tb01696.x

Geoffroy Saint-Hilaire, E. (1822). Considerations Generales Sur La Vertebre. *Mem. du Mus. Hist. Nat.* **9**, 89-119.

Gould, S. J. (1985). Geoffroy and the homeobox. *Nat. Hist* **94**, 12-23.

Hafen, E., Levine, M. and Gehring, W. J. (1984). Regulation of Antennapedia transcript distribution by the bithorax complex in Drosophila. *Nature* **307**, 287-289. doi:10.1038/307287a0

Karch, F., Weiffenbach, B., Peifer, M., Bender, W., Duncan, I., Celniker, S., Crosby, M. and Lewis, E. B. (1985). The abdominal region of the bithorax complex. *Cell* **43**, 81-96. doi:10.1016/0092-8674(85)90014-5

Laughon, A. and Scott, M. P. (1984). Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins. *Nature* **310**, 25-31. doi:10.1038/310025a0

Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. and Maniatis, T. (1978). The isolation and characterization of linked delta- and beta-globin genes from a cloned library of human DNA. *Cell* **15**, 1157-1174. doi:10.1016/0092-8674(78)90043-0

Levine, M., Hafen, E., Garber, R. L. and Gehring, W. J. (1983). Spatial distribution of Antennapedia transcripts during Drosophila development. *EMBO J.* **2**, 2037-2046. doi:10.1002/j.1460-2075.1983.tb01697.x

Levine, M., Rubin, G. M. and Tjian, R. (1984). Human DNA sequences homologous to a protein coding region conserved between homeotic genes of Drosophila. *Cell* **38**, 667-673. doi:10.1016/0092-8674(84)90261-7

Lewis, E. B. (1978). A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565-570. doi:10.1038/276565a0

Löhr, U., Yussa, M. and Pick, L. (2001). Drosophila fushi tarazu. a gene on the border of homeotic function. *Curr. Biol.* **11**, 1403-1412. doi:10.1016/s0960-9822(01)00443-2

McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. and Gehring, W. J. (1984a). A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes. *Nature* **308**, 428-433. doi:10.1038/308428a0

McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. and Gehring, W. J. (1984b). A homologous protein-coding sequence in Drosophila homeotic genes and its conservation in other metazoans. *Cell* **37**, 403-408. doi:10.1016/0092-8674(84)90370-2

McGinnis, W., Hart, C. P., Gehring, W. J. and Ruddle, F. H. (1984c). Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of Drosophila. *Cell* **38**, 675-680. doi:10.1016/0092-8674(84)90262-9

Regulski, M., Harding, K., Kostriken, R., Karch, F., Levine, M. and McGinnis, W. (1985). Homeo box genes of the Antennapedia and bithorax complexes of Drosophila. *Cell* **43**, 71-80. doi:10.1016/0092-8674(85)90013-3

Sánchez-Herrero, E., Vernós, I., Marco, R. and Morata, G. (1985). Genetic organization of Drosophila bithorax complex. *Nature* **313**, 108-113. doi:10.1038/313108a0

Scott, M. P. (2024). 40 years of the homeobox: either it is wrong or it is quite interesting. *Development* **151**, dev202776. doi:10.1242/dev.202776

Scott, M. P. and Weiner, A. J. (1984). Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila. *Proc. Natl. Acad. Sci. USA* **81**, 4115-4119. doi:10.1073/pnas.81.13.4115

Scott, M. P., Weiner, A. J., Hazelrigg, T. I., Polisky, B. A., Pirrotta, V., Scalenghe, F. and Kaufman, T. C. (1983). The molecular organization of the Antennapedia locus of Drosophila. *Cell* **35**, 763-776. doi:10.1016/0092-8674(83)90109-5

Shepherd, J. C., McGinnis, W., Carrasco, A. E., De Robertis, E. M. and Gehring, W. J. (1984). Fly and frog homeo domains show homologies with yeast mating type regulatory proteins. *Nature* **310**, 70-71. doi:10.1038/310070a0

Wakimoto, B. T. and Kaufman, T. C. (1981). Analysis of larval segmentation in lethal genotypes associated with the antennapedia gene complex in Drosophila melanogaster. *Dev. Biol.* **81**, 51-64. doi:10.1016/0012-1606(81)90347-x