Biologie Aujourd'hui, 205 (2), 75-85 (2011)

DOI: 10.1051/jbio/2011012

How do Hox transcription factors find their target genes in the nucleus of living cells?

Walter J. Gehring

Growth and Development, Biozentrum University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland
Corresponding author: Walter J. Gehring, walter.gehring@unibas.ch

Received 11 January 2011

Abstract – Homeotic mutations first found in Drosophila led to the identification of Hox genes in all bilateria. These genes are exceptional in that they are arranged in an ordered cluster, in which they are positioned in the same order along the chromosome as they are expressed along the antero-posterior axis to specify the corresponding body regions. They share a highly conserved DNA sequence of 180 bp, the homeobox which encodes the homeodomain, a 60 amino acid polypeptide involved in specific DNA and RNA binding and in protein-protein interactions. The discovery of the homeobox has uncovered for the first time a universal principle of specification of the body plan along the antero-posterior axis. The structure of the homeodomain has been determined by NMR spectroscopy and by X-ray crystallography. However, the mechanism by which the Hox proteins find their target genes in the nucleus of a living cell has been enigmatic. Transcriptome analysis indicates that there are hundreds of target genes to be regulated, both positively and negatively to ensure normal development. In the following, we show by Fluorescence Correlation Spectroscopy (FCS) and single molecule imaging in live salivary gland cells, that the mechanism of recognition is purely stochastic. The homeodomain associates and dissociates rapidly (in the ms range) with chromatin all along the chromosomes. If, however, it associates with a specific binding site in a puffed chromosome region, it remains bound for seconds or minutes to exert its function, by forming a complex with co-activators or co-repressors respectively. These direct measurements solve an old enigma of how Hox transcription factors find their target genes in the nucleus of live cells.

Key words: Hox genes / homeodomain / NMR spectroscopy / Fluorescence Correlation Spectroscopy (FCS) / Single Molecule Imaging / target genes

Résumé – Comment les facteurs de transcription Hox trouvent-ils leurs cibles dans le noyau des cellules vivantes ?

La découverte des mutations homéotiques chez la drosophile a mené à l'identification des gènes Hox chez tous les animaux bilatériens. La particularité de ces gènes tient à leur organisation en *cluster* au sein duquel leur ordre sur le chromosome est identique à celui de leur expression le long de l'axe antéro-postérieur, où ils déterminent l'identité des différentes parties du corps. Les gènes Hox partagent une séquence d'ADN hautement conservée de 180 pb, l'homéoboîte, qui code l'homéodomaine, un polypeptide de 60 acides aminés capable non seulement de se lier spécifiquement à l'ADN et à l'ARN, mais aussi d'interagir avec d'autres protéines. La découverte de l'homéoboîte a revélé un principe universel de spécification de l'identité cellulaire le long de l'axe antéro-postérieur. Bien que la structure de l'homéodomaine ait été résolue par spectroscopie RMN et par cristallographie aux rayons X, le mécanisme permettant aux protéines Hox de trouver leurs gènes cibles dans le noyau cellulaire restait inconnu jusqu'ici. L'analyse du transcriptome indique que l'expression de centaines de gènes est modulée au cours du développement. En combinant la spectroscopie de corrélation de fluorescence (FCS) et l'imagerie moléculaire *in vivo* dans les cellules

Article published by EDP Sciences

de glandes salivaires, nous avons pu montrer que le mécanisme de reconnaissance est aléatoire dans un premier temps : l'homéodomaine s'associe et se dissocie rapidement (le processus est de l'ordre de la milliseconde) avec la chromatine tout au long des chromosomes. Par contre, lorsqu'il atteint un site de liaison spécifique dans une région où la transcription est active (région *puff*), il reste lié à l'ADN pendant plusieurs secondes, voire plusieurs minutes, afin d'exercer sa fonction en formant un complexe avec des co-activateurs ou des co-répresseurs, selon le cas. Ces mesures effectuées directement dans les cellules vivantes permettent de répondre à une vieille interrogation : comment les facteurs de transcription Hox trouvent-ils leurs gènes cibles ?

Mots clés : Gènes Hox / homéodomaine / spectroscopie RMN / spectroscopie de corrélation de fluorescence (FCS) / imagerie moléculaire *in vivo* / gènes cibles

Introduction

As a graduate student, I found a spectacular homeotic mutation in Drosophila which transforms the antennae on the head of the fly into a pair of middle legs (Gehring, 1966). Originally, I called this mutation *Nasobemia*, a term coined by the German poet, Christian Morgenstern. In one of his poems he described an imaginary animal capable of walking on its nose, which he called a "Nasobem". Later the molecular genetic analysis of this mutation indicated that it is a dominant gain-of-function mutation at the *Antennapedia* (*Antp*) locus and provided a key to the understanding of the genetic control mechanisms of development at the molecular level.

Homeotic genes can mutate in two ways giving rise to segmental transformations in opposite directions. In *Antp*, for example, recessive loss-of-function mutations lead to transformations of the second thoracic segment of the larvae into a first thoracic segment (T2 → T1); whereas strong dominant gain-of-function mutations have the opposite effect and convert T1 and some head segments into T2 (Fig. 1). Since T2 is missing in loss-of-function mutants and additional T2 segments are formed in gain-of-function mutants, these phenotypic effects suggest that *Antp* primarily specifies the second thoracic segment (T2), and not the antennae. The same rule applies to the adult phenotypes; loss-of-function mutations (analyzed in genetic mosaics) lead from posterior to anterior transformations, from middle legs (T2) to antennae, whereas gain-of-function mutations cause opposite transformations of antennae to middle legs. Later studies suggested that T2 represents the evolutionary and developmental ground state, which is supported by the observation that in all of the *Hox* genes, loss-of-function mutations lead towards the ground state, whereas gain-of-function mutations lead away from the ground state (Gehring *et al.*, 2009). The nature of these dominant gain-of-function mutations remained enigmatic until the *Antp* gene was cloned and analyzed at the molecular level.

Cloning of the Antennapedia gene and discovery of the homeobox

The *Nasobemia* ($Antp^{Ns}$) mutation was not associated with a large chromosome rearrangement and therefore, it was possible to map it by recombination to locus 48 (Gehring, 1966), whereas all of the other known *Antp* mutations were either inversions or translocations which suppress recombination. These inversions all shared one breakpoint in band 84A on the third chromosome in giant polytene chromosomes, which roughly corresponds to map position 48, the second breakpoints being way outside at different positions. This indicated that *Antp* is located in band 84A, but this was all the information available at that time. Therefore, the isolation of gene products (RNA or protein) by biochemical means was extremely difficult, if not impossible.

With the advent of gene cloning (Cohen *et al.*, 1973) this situation changed radically. David Hogness and his group developed a method for positional cloning or "walking along the chromosome" as they termed it (Bender *et al.*, 1983a) and used it for the first time to clone a homeotic *Drosophila* gene *Ultrabithorax* (*Ubx*) (Bender *et al.*, 1983b). The method consists essentially of the stepwise isolation of overlapping DNA segments from a phage library and mapping them in the giant polytene chromosomes by *in situ* hybridization.

My group (Garber *et al.*, 1983) and Scott *et al.* (1983) "walked" across the *Antp* locus. We isolated some 200 kb of DNA around the various inversion breakpoints and mapped the gene within 160 kb by using deletion mutants. In order to map the *Antp* gene molecularly we first isolated some *Antp* c-DNA clones. The *Antp* gene turned out to be very large, over 100 kb, which made it difficult to prove that we had indeed cloned the *Antp* gene. This was usually done by introducing the cloned gene into loss-of-function mutations and demonstrating that the transgene can complement the loss-of-function mutation and restore the wild phenotype. However, at this time

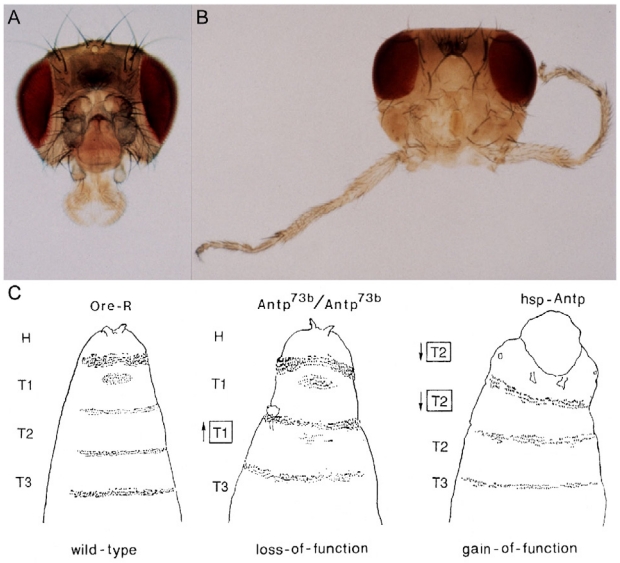

Fig. 1. Gain- and loss-of-function mutants of *Antennapedia* (*Antp*). A. Wild-type $Antp^{+}$ head. B. Gain-of-function $Antp^{Ns}/Antp^{73b}$. The antennae are transformed into second legs. C. Oregon-R (wild-type larva) H, Head; T1-3, thoracic segments. $Antp^{73b}/Antp^{73b}$ (loss-of-function). Transformation of T2 → T1 with second beard (anterior transformation). *Hsp-Antp* (gain-of-function), induced by heat shock. Transformation of T1 → T2 and H → T2 (posterior transformation).

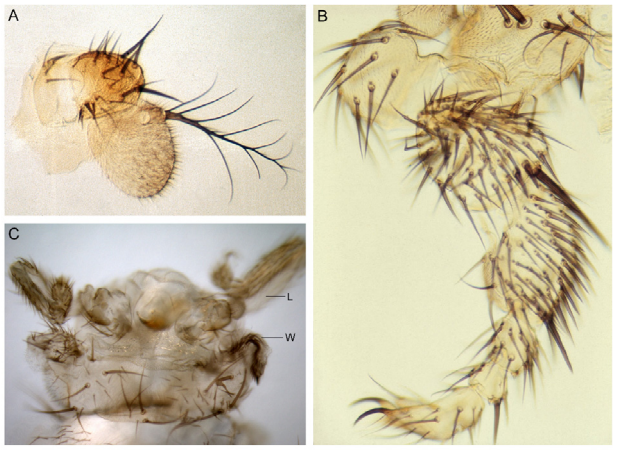

Fig. 2. Homeotic transformations induced by ectopic *Antp* expression. A. Normal antennae. B. Transformation of the antenna into a second leg by heat-induced ectopic expression of *Antp* C-DNA. C. Transformation of the head towards the second thoracic segment (T2) with a pair of second legs (L) and a pair of wings (W) in a fly carrying a hypomorphic *twin of eyeless* ($toy^{D3.3}/toy^{D3.3}$) mutation which switches the antennal program off and a dominant gain-of-function mutation ($Antp^{73b}/Antp^{+}$) which turns on the leg and wing (T2) program.

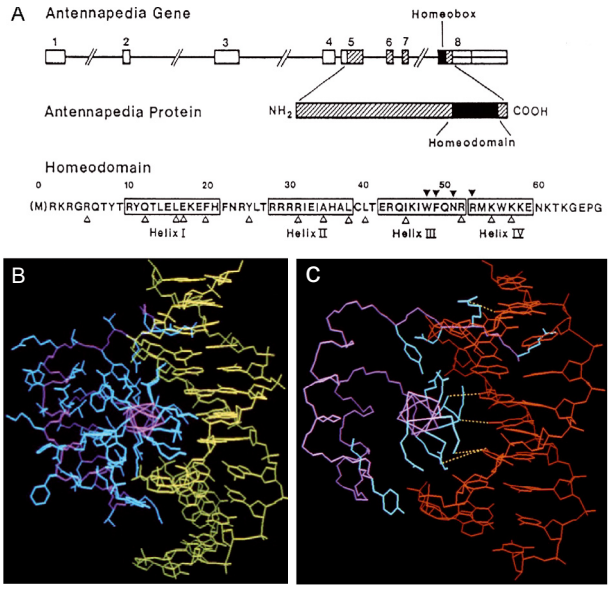

Fig. 3. The homeobox and homeodomain of *Antennapedia*. A. *Antp* gene with exons 1–8 (boxes) and introns (not drawn to scale). The protein-coding exons are crosshatched, the homeobox is indicated in black (located in exon 8). The Antp protein with its amino (NH_2) and carboxy terminus (COOH) is indicated (crosshatched), the homeodomain is shown in black. The amino sequence of the homeodomain is shown in the one-letter-code. Invariant amino-acid residues are marked by black triangles, highly conserved amino acids by open triangles. The four α-helices are boxed in. B. Structure of the Antp homeodomain-DNA complex as determined by NMR spectroscopy (coutesy of Wüthrich, Billeter, 1999). 3D-structure with all amino acid side chains (blue). The backbone of the polypeptide chain is shown in magenta, the DNA in yellow. C. Contacts between the specific base pairs in the DNA (red) and specific amino acid side chains (blue) are indicated by yellow dotted lines. They reach from the recognition helix (Rh) into the major groove and from arginine 5 into the minor groove.

it was impossible to transform flies with DNA fragments of this size. Therefore, we used a different approach based on the gain-of-function phenotype.

The molecular analysis of the inversion mutants giving rise to a gain-of-function phenotype, indicated that one of the breakpoints was located in a large intron in the middle of the gene, leaving us with the question of how a gene can be split in half without losing its function? This apparent paradox was resolved when the structure of the gene was determined in more detail. It turned out that the protein-coding exons of these inversion mutations remained intact, but they became fused to foreign enhancer and promoter sequences located at the other breakpoint of the inversion, leading to abnormal ectopic expression of the

normal *Antp* gene products. The molecular basis of the *Nasobemia* ($Antp^{Ns}$) was analyzed by P. Talbert in R. Garber's laboratory. It turned out to be a 25 kb duplication of the exon 3 region including the second *Antp* promoter (P2) and a *roo* transposon, which leads to misexpression of the ANTP protein (Jorgensen & Garber, 1987).

If ectopic expression of *Antp* indeed leads to a dominant phenotype, *i.e.* the formation of second legs in place of the antennae, then it should be possible to construct an artificial gain-of-function mutation and induce the formation of T2 structures ectopically. Such a mutant was constructed by inserting a normal *Antp* c-DNA into a heat-inducible vector with a heat shock (*hsp70*) promoter, and introducing the construct into transgenic flies (Schneuwly *et al.*, 1987). By application of a heat shock, the expression of *Antp* can be induced at defined developmental stages in all cells of transgenic animals. As predicted, the application of a heat shock at early embryonic stages leads to a transformation of T1 and the head segments to T2 (Fig. 1C *hsp-Antp*). Heat induction at the early third larval stage results in the transformation of the antennae into middle legs (Fig. 2B), and the top of the head into dorsal thoracic structures (T2), which are normally derived from the wing imaginal disc. This result clearly indicates that *Antp* specifies segmental identity, *i.e.*, the second thoracic segment. This was the first successful attempt to redesign the body plan of an animal in a predictable way (Schneuwly *et al.*, 1987).

However, at that time we had placed emphasis on the ectopic expression of the *Antp* gene in the eye-antennal disc at a defined developmental stage, but we largely ignored the fact that the eye-antennal program had to be switched off before the T2 program (wing and second leg) could be switched on. How could the eye-antennal program be switched off? The *twin of eyeless* (*toy*) gene is expressed from very early stages in the primordia of the eye-antennal disc of the embryo (Czerny *et al.*, 1999) and loss-of-function mutants are essentially headless and lethal. Hypomorphic mutants survive until late pupal stages and form adults with reduced heads. By combining a hypomorphic *toy* mutant, which switches off the head and antennal program, with a dominant gain-of-function $Antp^{73b}$ mutant, the formation of a pair of middle legs and a pair of wings on the head of the mutant flies appears. This corresponds to an induction of a T2 segment in the head region, indicating that *Antp* specifies the entire T2 segment (Fig. 2C).

The cloning of the *Antp* gene led to the discovery of the homeobox as described in my book of the homeobox story (Gehring, 1998, 1999a). The *Antp* gene was cloned independently in my group and by Scott *et al.* (1983). *Antp* was found to be a very large gene of over 100 kb with large introns, two promotors, two polyadenylation sites, and multiple, differentially spliced transcripts. When mapping the exons by hybridization of the cDNA clones to the chromosomal DNA isolated on the "chromosome walk", Richard Garber detected crosshybridization between *Antp* cDNA and a neighbouring gene, which later turned out to be the *fushi tarazu* (*ftz*). It was a band at the edge of the gel, which most people would have disregarded, but Richard and I were expecting some crosshybridization, since E.B. Lewis had proposed that the *Bithorax* Gene Complex (BX-C) had arisen by tandem duplication and the duplicated genes were expected to share certain sequences. Tom Kaufman (Kaufman *et al.*, 1980) had extended this hypothesis to the *Antp* gene complex (ANT-C). As Pasteur pointed out, "*In experimental science chance favors only the prepared mind*", and therefore this observation was followed up. My personal view of the homeobox discovery has been challenged (McGinnis and Lawrence, 1999), but after my rebuttal (Gehring, 1999b) these authors had to admit that "*it is always possible that Gehring sensed or knew in 1982 that Garber's band was a crucial clue which should be the basis for further investigation*". This is in fact true and I followed up this idea by asking Atsushi Kuroiwa to clone the gene which crosshybridized with *Antp* cDNA clones, which turned out to be *fushi tarazu* (*ftz*) (Kuroiwa *et al.*, 1984). Using the crosshybridizing *ftz* sequences and fragments of the *Antp* cDNA, Bill McGinnis then identified the homeobox, which is shared by the two genes. Initially, I had expected that the sequence homology would extend all along the coding region and possibly also into the regulatory regions, and that it would be relatively low. However, our data indicated that the homology was confined to a short 180 bp box which we called the homeobox which encodes a defined protein domain, the homeodomain (Fig. 3A) (McGinnis *et al.*, 1984a, 1984b). When McGinnis probed whole genome DNA with homeobox sequences under conditions of low stringency, a ladder of bands of hybridization was found representing different homeobox genes, which allowed us to clone many homeobox-containing genes in a short period of time. Several of these genes mapped to either one of the two homeotic gene clusters, the BX-C or the ANT-C.

The homeobox homology was found independently by Scott & Weiner (1983). However, nobody would have considered this discovery of much importance, if the homeobox would have been found in insects only. Surprisingly, in an evolutionary survey, using DNA from various species, Bill McGinnis showed that the homeobox is not confined to insects, but it is also found in vertebrates, including chickens, mice and humans (McGinnis *et al.*, 1984b). The crosshybridization

of bands formed a similar ladder to that found in Drosophila. In order to prove this point, the respective genes had to be cloned. In collaboration with Eddy De Robertis and Andres Carrasco, the first vertebrate homeobox was cloned from the frog, *Xenopus laevis*, (Carrasco *et al.*, 1984), and it was extremely similar to the *Antp* homeobox. The first mouse homeobox genes were cloned in collaboration with Frank Ruddle who happened to be on a sabbatical leave in my laboratory at that time (McGinnis *et al.*, 1984c). Again the sequence similarity to the Drosophila homeobox was astounding, whereas the flanking sequences showed hardly any conservation, with the exception of a short hexapeptide motif, with the essential sequence YPWM, which is associated with most Hox genes.

Later work showed that the vertebrate Hox genes are also clustered. There are four Hox gene clusters in both mice and man. The colinearity rule first described by Lewis (1978) for the BX-C, indicating that the genes are arranged in the same order along the chromosomes as they are expressed along the antero-posterior axis, also applies to mammals. Since in both insects and mammals, Hox genes are involved in specifying the bodyplan along the antero-posterior axis, the homeobox has uncovered a universal principle of the genetic control of animal development, despite the very different modes of development.

The structure and function of the homeodomain

In my first publication I boldly postulated that *Antp (Nasobemia)* was a regulatory gene as first found in bacteria by Jacob & Monod (1961a, 1961b), and that it would activate all genes required for leg development in the antennal imaginal disc. Finally, this hypothesis turned out to be largely true. Comparative DNA sequence analysis revealed weak but significant sequence homology between the homeobox and the mating-type genes MATa1 and α2 (Shepherd *et al.*, 1984). Since MATα2 had been shown to be a transcriptional repressor, this suggested that the homeodomain proteins might also be transcriptional regulators. Sequence comparisons led to the further speculation that the homeodomain might contain a helix-turn-helix motif similar to the one found in prokaryotic repressors and activators (Laughon & Scott, 1984; Shepherd *et al.*, 1984).

In order to test this hypothesis we undertook an in depth analysis of the structure and function of the homeodomain. Using the isolated, purified homeodomain polypeptide of *Antp*, we found that the isolated homeodomain peptide is sufficient to bind specific oligodeoxynucleotide sequences from putative binding sites. In contrast to prokaryotic repressors the ANTP homeodomain was shown to bind as a monomer with high affinity with a K_D in the range of 10^{-9} M to its DNA binding site and the half-life of the homeodomain-DNA complex *in vitro* was estimated to be one-and-half hours (Affolter *et al.*, 1990).

In collaboration with Kurt Wüthrich and his group, the three-dimensional structure of the ANTP homeodomain peptide (68 amino acids) was determined by NMR spectroscopy in solution (Qian *et al.*, 1989). The core homeodomain consists of three α-helices. Helix 1 is connected by a loop with helix 2, which is separated by a tight turn from a well-defined third helix, which is elongated by a more flexible helix 4. Helix 2, the connecting turn, and helices 3/4 constitute a helix-turn-helix motif, as first found in prokaryotic gene regulatory proteins. The three-dimensional structure of this motif is very similar to that of prokaryotic repressors even though there is hardly any amino acid sequence similarity.

The mode of DNA binding was analyzed by solving the structure of the ANTP homeodomain complexed with a 14 bp consensus binding site (5-6AAAGCCATTAGAG) (Otting *et al.*, 1990; Billeter *et al.*, 1993). Helix 3/4 contacts the DNA specifically in the major groove as in other helix-turn-helix proteins (Fig. 3) and is, therefore, designated as the recognition helix. However, it is shifted away from the DNA by about 7 Å as compared to prokaryotic repressor-DNA complexes. In addition to the contacts of the recognition helix in the major groove, the N-terminal arm preceeding the first α-helix, which is flexibly disordered in solution, reaches into the minor groove of the DNA thus positioning the homeodomain on the DNA. This DNA contact is mediated by arginine 5 which is highly conserved among most homeodomains (Fig. 3A, open triangle). The direct contacts between the homeodomain and the base pairs in the DNA are indicated by yellow-dotted lines in Figure 3C. These involve arginine 5 in the minor groove, isoleucine 47, glutamine 50 and methionine 54 in the major groove. It should be pointed out that one amino acid can contact two base pairs, and that one base pair can be in contact with two amino acids. This indicates that the binding is somewhat flexible. These results which are summarized in Gehring *et al.* (1994) were confirmed by studies using X-ray crystallography (Kissinger *et al.*, 1990; Wolberger *et al.*, 1991). The highly conserved ELEKEF element was shown to be involved in protein-protein interactions between different homeodomains (Plaza *et al.*, 2001, 2008).

The essential features of the protein-DNA contacts have been verified by second site suppression experiments *in vivo* (Schier & Gehring, 1992). However, the mechanism of "recognition" of the target sites by the homeodomain, among the thousands of genes *in vivo*, in live cells has remained an enigma.

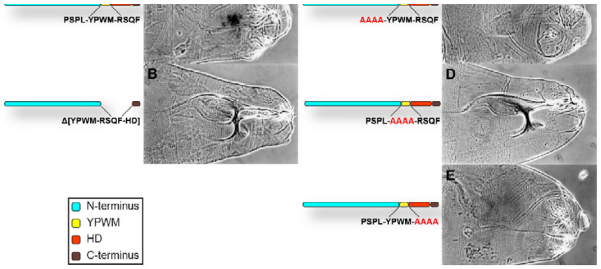

Fig. 4. The YPWM motif is essential for homeotic transformations. The UAS *Antp* constructs depicted on the left were driven by the constitutive expression of hsp70-gal4 and analyzed for head involution defects caused by homeotic transformations of the head segments towards T2. If the YPWM and the adjacent 4 amino acids as well as the HD are deleted, the construct is inactive and fails to induce head involution. Similarly, the substitution of the YPWM motif by AAAA results in a failure to induce homeotic transformations, whereas constructs with a substitution of the four amino acids upstream (PSPL) or downstream (RSQF) of the YPWM motif by AAAA retain their homeotic function (after Papadopoulos *et al.*, 2011).

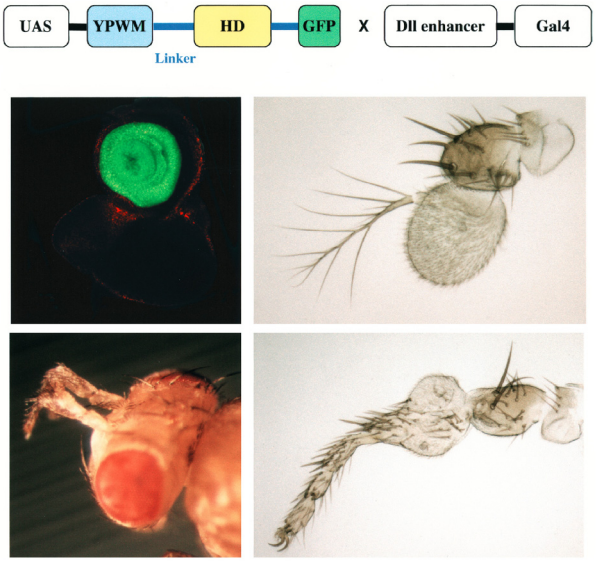

Fig. 5. Synthetic *Antp* genes are capable of inducing antennae-to-leg transformation in transgenic flies. Flies carrying a UAS-YPWM-linker-HD-C terminus construct fused to GFP are crossed to a distal-less enhancer-gal4 driver and express GFP in a circular area of the antennal imaginal disc (green circle). The synthetic *Antp* genes consisting essentially of the YPWM motif and the HD are capable of transforming the antennae into tarsal leg structures with claws. Stocks established from this cross show antenna-to-leg transformations with full penetrance in all of their progeny.

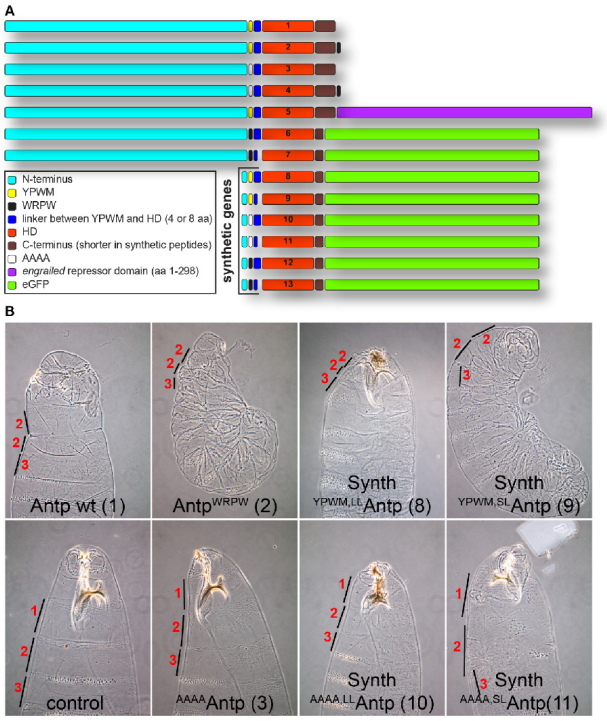

Fig. 6. Homeotic transformation of full-length and synthetic *Antp* constructs in the late embryo. A. Design of the constructs. B. Embryonic phenotypes: Full-length *Antp* wt and *Antp* wt to which a WRPW motif, which interacts with the Groucho repressor protein, attached at the C terminus, induce homeotic transformations of T1 into T2 (2, 2, 3). Synthetic constructs in which the YPWM motif is connected by either the long linker (LL) or the short linker (SL) to the HD are fully active. By contrast full-length and synthetic constructs in which the YPWM motif is substituted by AAAA are inactive and have normal T1, 2, 3 segments like the wild-type control (after Papadopoulos *et al.*, 2011).

Construction of synthetic Hox genes

The functional dissection of the *Antp* gene revealed that a distinct portion of the ANTP protein, including the homeodomain and residues adjacent to both ends of the homeodomain, almost entirely determines its functional specificity (Gibson *et al.*, 1990). However the homeodomain alone has little if any homeotic effect when ectopically expressed in heat-inducible constructs. Subsequently, it was shown that besides the homeodomain, the highly conserved hexapeptide sequences connected to the 5' end of the homeodomain by either a long or short linker-arm, are required for inducing homeotic transformations. As shown in Figure 4, deletion of the YPWM motif and the homeodomain (ΔYPWM-RSQF-HD) inactivates the heat-inducible construct and gives no homeotic transformation, whereas the full-length construct transforms

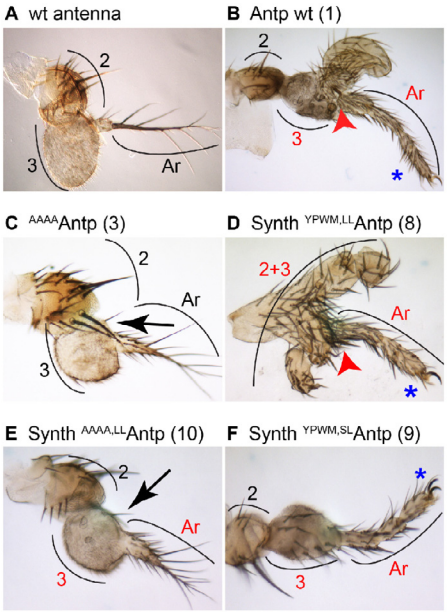

Fig. 7. Antennae-to-leg transformations in transgenic flies carrying a full-length or synthetic *Antp* construct. A. Normal antenna. B., D., F. The wt construct and the synthetic constructs carrying the YPWM motif connected by either the short or the long linker to the HD induce strong transformation of distal antennal segments (2 and/or 3) and the arista (Ar) into leg structures with claws (*). The full-length and the synthetic AAAA mutants show hardly any transformations, mostly a thickening of the arista and some isolated leg bristles (arrows in C. and E.). The red arrow head indicates the position of the apical bristle which is typical for the tibia of the second leg (after Papadopoulos *et al.*, 2011).

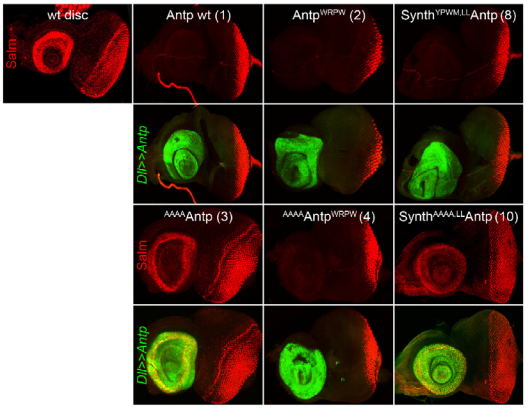

Fig. 8. Synthetic *Antp* constructs are capable of repressing target genes required for normal antennal development. Repression of *Spalt major* (*Salm*) in the antennal disc is mediated by all full-length and synthetic *Antp* genes carrying a YPWM motif or a WRPW repression motif. The AAAA-substituted constructs (lower lanes), with the exception of ^{AAAA} *Antp* ^{WRPW}, do not repress *Salm*. Salm antibody (red), Antp antibody (green) (after Papadopoulos *et al.*, 2011).

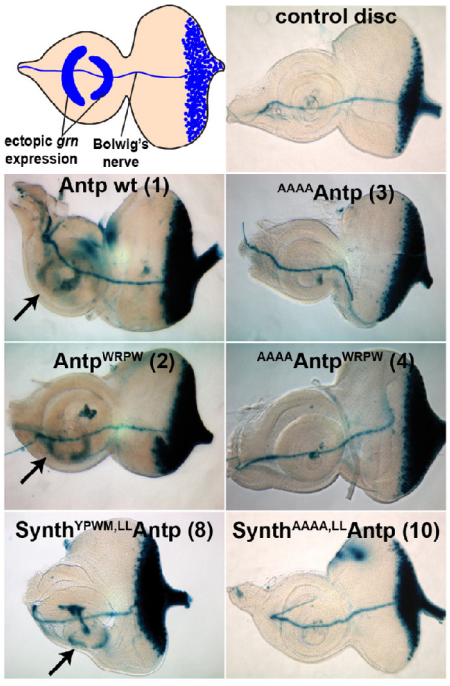

Fig. 9. Synthetic *Antp* constructs can activate target genes in the antennal disc. The full-length wt construct and the synthetic YPWM construct can activate the *grain* (*grn*) gene in the antennal disc (arrows). Addition of a WRPW repression motif at the 3' end of the polypeptide does not interfere with activation. The full-length and the synthetic AAAA constructs are inactive, as is the AAAA construct carrying an additional WRPW repression motif. (after Papadopoulos *et al.*, 2010).

T1 and the head segments towards T2. When the four amino acids YPWM are substituted by four alanines (AAAA) the homeotic activity is lost, whereas the substitution of the four amino acids upstream or downstream of the YPWM motif has little if any effect.

These observations led us to design synthetic genes consisting of the homeobox and the YPWM motif connected by the short (SL) or the long linker (LL). These synthetic genes, like their full-length counterparts, cause antenna-to-leg transformations if expressed under a distal-less gal4 driver (Fig. 5) and transformations of T1 and head segments towards T2. These homeotic transformations require the presence of the YPWM motif whereas the AAAA mutants lead to a thickening of the arista only (Figs. 6 and 7). The synthetic genes can serve as transcriptional repressors of antennal genes as in the case of *Spalt* (Fig. 8) suggesting that the YPWM motif can recruit a co-repressor, whereas AAAA mutants are not capable of

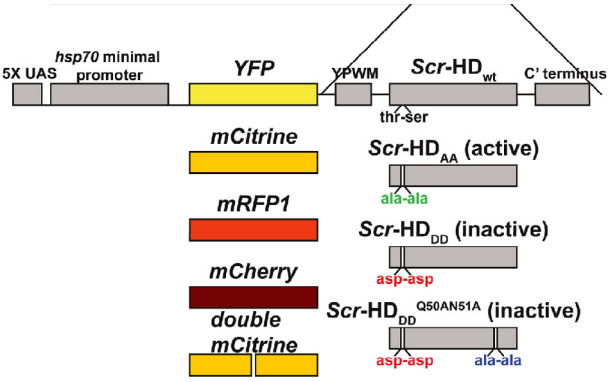

Fig. 10. Synthetic *Sex combs reduced* (*Scr*) constructs. Design of the constructs carrying different fluorescent proteins. The highest photon count was obtained with the mCitrine-YPWM-HD construct (after Vucojevic *et al.*, 2010).

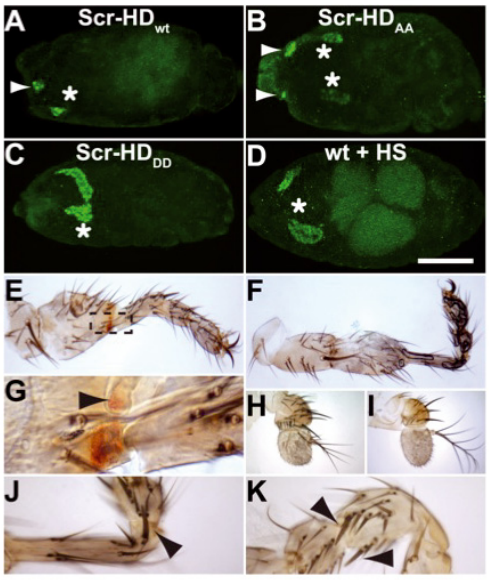

Fig. 11. Functional analysis of synthetic *Scr* genes. A–D The wild type construct (Scr-HD wt) (A) and the constitutively active construct Scr-HD AA (B) induce the formation of a second pair of salivary glands, ectopically in a more anterior head segment (arrowheads) in addition to the normal glands located in the labial segment (labelled by asterisks). The inactive Scr-HD DD construct (C) and heat shocked normal larvae (D) form only one pair of salivary glands in the labial segment. All constructs were induced by a heat shock gal 4 driver and stained for dCREB A, a salivary gland marker. E–I Homeotic transformations generated by using a Dll gal4 driver in the antennal discs. Expression of Scr-HD wt (E) and constitutively active Scr-HD AA (F) result in complete transformation of the distal antenna into tarsus. G. Formation of a few ommatidia at the base of the tarsus due to incomplete repression of the *distal antenna (dan)* gene. H. The inactive Scr-HD dd construct induces a small reduction of the arista only, as compared to the wild type antenna (I). The occasional formation of sex comb teeth in male discs indicates that the tarsal structures induced by Scr-HD wt (J) and Scr-HD AA (K) represent first legs.

repressing *Spalt* expression unless provided with an additional repressor motif WRPW which recruits the Groucho repressor. The YPWM motif is also capable of recruiting co-activators and to activate *grain* transcription (Fig. 9). Again, the AAAA mutants are not capable of activation. Addition of a WRPW repression motif at the 3' end of the synthetic gene does not interfere with transcriptional activation. This rules out the possibility that the activation is caused by repression of a repressor (double negative regulation). Therefore, synthetic *Hox* genes provide powerful tools for functional analysis of *Hox* gene function (Papadopoulos *et al.*, 2011).

Fluorescence correlation spectroscopy and single molecule imaging in live cells

Scientific progress is frequently dependent on new technological developments. Using confocal laser scanning microscopy with avalanche photodiodes (APDs), so-called APD imaging (Vukojevic *et al.*, 2008) and Fluorescence Correlation Spectroscopy (FCS) (Ehrenberg & Rigler, 1974; Magde *et al.*, 1972), we have established a method for nondestructive observation of transcription factor molecules and quantitative measurements of their molecular interactions in live cells with single molecule sensitivity. In a collaborative study with Rudolf Rigler and his group, we have used APD imaging to visualize synthetic Sex combs reduced (Scr)-HD molecules at low expression levels and FCS to study quantitatively the number and mobility of Scr-HD molecules. This is accomplished by monitoring the fluorescence intensity fluctuations in a selected spot in the salivary gland nuclei that is generated by focusing the incident laser light through the objective of the microscope. Statistical analysis of the recorded data is applied in order to derive molecular numbers and macroscopic diffusion constants.

For this purpose we used *Scr* instead of *Antp*. Since *Scr*, in contrast to *Antp,* is normally expressed in salivary glands, the *Scr* gain-of-function mutants are capable of inducing a second pair of salivary glands ectopically (Panzer *et al.*, 1992; Berry & Gehring, 2000). We generated a number of UAS gal4 lines with synthetic *Scr* genes consisting of the HD and YPWM motif fused to various fluorescent proteins (Fig. 10). The homeodomain of *Scr* is inactivated by phosphorylation of a threonine and a serine residue. Substitution

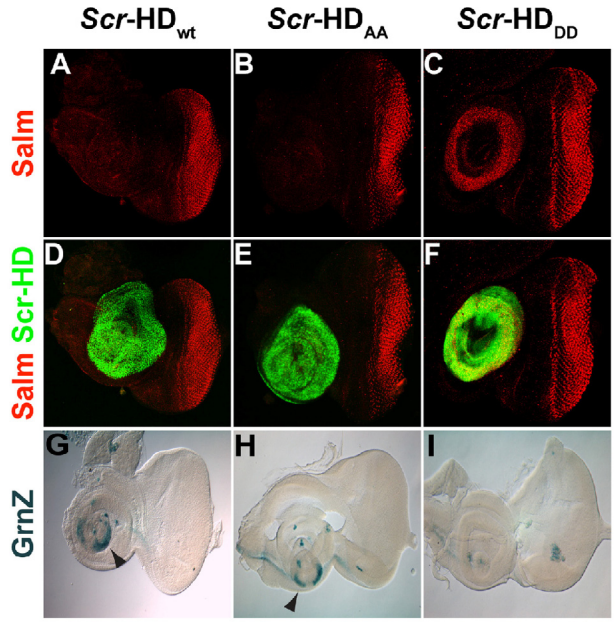

Fig. 12. Synthetic *Scr* constructs are capable of both repression and activation. *Scr* HD_{wt} and the constitutively active Scr-HD_{AA} form can repress *Salm* expression and activate *grain* (*grn*). The inactive HD_{DD} form serves as a negative control.

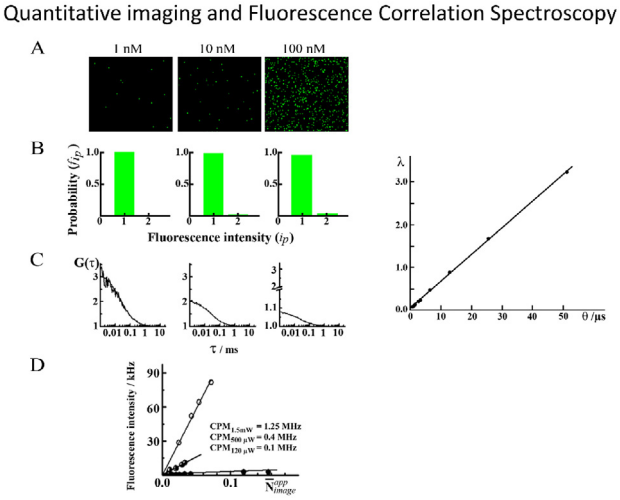

Fig. 13. Comparison of quantitative single molecule Avalanche Photon Diode (APD) imaging and Fluorescence Correlation Spectroscopy (FCS). A. Quantitative APD images of 1, 10 and 100 nM aqueous solutions of Rhodamine 6G recorded fast scanning at 0.64 μs/pixel. The total number of detector-related counts ~50 pixels per image frame, as compared to 337, 1059 and 17069 for 1, 10 and 100 nM Rhodamine 6 G. B. Fluorescence intensity distribution histograms. C. Apparent molecular brightness determined as the average fluorescence count rate versus the apparent number of molecules. (After Vukojevic *et al.*, 2008).

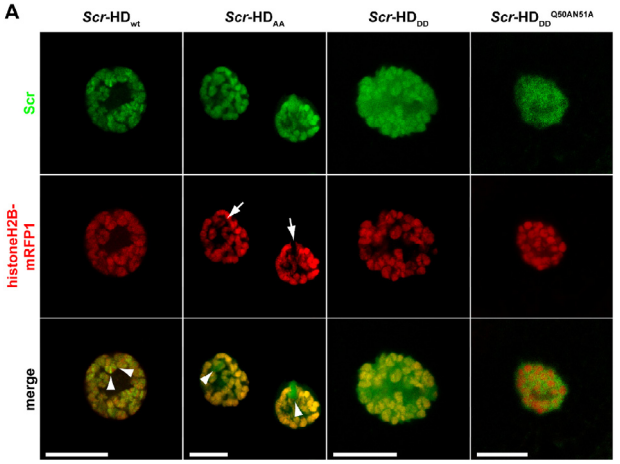

Fig. 14. High resolution APD imaging of DNA. *Scr*-HD interactions in live salivary gland cells. Third instar salivary gland polytene nuclei expressing *Scr*-HD_{wt}, *Scr*-HD_{AA}, *Scr*-HD_{DD}, and *Scr*-$HD_{DD}^{Q50A,\,N51A}$ under the control of the dpp^{blink}-Gal4 driver. Ubiquitously expressed histone H2B tagged with mRFP (red) was used to visualize chromatin. *Scr*-HD_{wt} and *Scr*-HD_{AA} associate with the chromosome (as shown in the green channel), but also show sites of accumulation (puffs) where chromatin is decompacted as shown by the low histone signal (arrows). Arrowheads point at sites of high accumulation of the active forms, *Scr*-HD_{wt} and *Scr*-HD_{AA}. The nucleus expressing the inactive *Scr*-HD_{DD} shows a reduced association of the transcription factor with chromatin and it is also dispersed in the nucleoplasm. The *Scr*-$HD_{DD}^{Q50A,\,N51A}$ which has lost all DNA binding affinity appears almost completely excluded from the chromosomes (staining red in the merged picture), and resides mainly in the nucleoplasm (after Papadopoulos *et al.*, 2010).

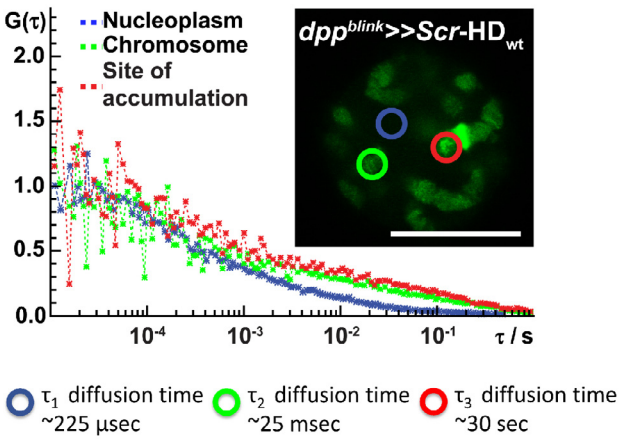

Fig. 15. Measurement of the diffusion times at different sites in live salivary gland nuclei by FCS in the nucleoplasm T_1 is ~225 μs (~free diffusion), in chromosomes (chromatin) T_2 ~ 25 ms (non-specific binding), and at sites of accumulation (puffs) the diffusion T_3 is ~30 s.

of these two residues by two alanines (AA) generates a constitutively active form, whereas two aspartates (DD) mimic the phosphorylated form and are inactive, presumably because the negative charges of the two aspartates lead to repulsion of the N-terminal arm of the HD from the minor groove of the DNA. Additional substitution of Q50 and N51 in the recognition helix abolishes DNA binding completely. The mCitrine-YPWM-HD fusion shows the highest photon count per molecule and was chosen for further analysis.

The mCitrine-YPWM-HD construct was first tested in transgenic embryos and shown to be capable of inducing an additional ectopic pair of salivary glands, both in the wild-type HD construct and in the constitutively active AA form, whereas the DD form was inactive (Fig. 11A–11D). In the adults the wt and AA forms were capable of inducing antenna-to-leg transformations as in *Antp*. However, *Scr* induces first legs and not middle legs, as shown by the presence of sex comb teeth on the antennal legs in male flies (Fig. 11K). Again the DD form was inactive (Fig. 11H and 11I). Like the synthetic *Antp* constructs, *Scr* constructs are also capable of repressing and activating target genes as shown for *Spalt* (*Salm*) and *grain* (*grn*), respectively (Fig. 12).

FCS and APD imaging allows the detection of single GFP molecules in solution (Fig. 13) and in live giant salivary gland nuclei, where approximately 1000 chromosomal DNA molecules are lined up to form giant polytene chromosomes. In active loci the chromatin is decondensed and forms a puff (Fig. 14), whereas the inactive regions form tightly compacted bands. Using a histone H2B-mRFP1 marker the chromosomes are labelled all along their length with the exception of large puffs where the chromatin is decondensed (arrows in Fig. 14) both in the wt and AA constructs. Synthetic *Scr* (green) shows a similar distribution, except that it accumulates strongly in the puffs (arrow heads in the merged pictures in Fig. 14). The Scr-HD_{DD} shows much more diffuse chromosomal staining and can also be detected in the nucleoplasm, indicating that it binds less tightly to chromatin. In the Scr-$HD_{DD}^{Q50A,N51A}$ mutant DNA binding is abolished and it is excluded from the chromosomes, which still stain red with the histone chromatin marker (red areas in the merged picture in Fig. 14).

Measurements of the diffusion times of the Scr-HD_{wt} construct at three different spots; in the nucleoplasm (T_1) condensed chromatin (T_2) and in an active puffed region (T_3) (Fig. 15), where *Scr* accumulates, gave values of ∼225 μs in the nucleoplasm, ∼25 ms in inactive chromatin and ∼30 s in the active puff region. These findings suggest that the synthetic Scr transcription factors find their specific target sites primarily by multiple stochastic association/dissociation events, the rapidity of which is largely due to electrostatic interactions with chromatin. These experiments have been corroborated by Fluorescence Recovery After Photobleaching (FRAP) experiments in which recovery times were measured at a well defined target puff (Vukojevic *et al.*, in prep.). These measurements yielded recovery times in the order of minutes.

These findings resolve the old question of how transcription factors find their target genes among thousands of other genes. The answer is by a purely Darwinian mechanism (random association and selection), *i.e.*, in a very dynamic association/dissociation process and extended association with their target sites.

References

Affolter M., Percival-Smith A., Müller M., Leupin W., Gehring W.J., DNA binding properties of the purified *Antennapedia* homeodomain. *Proc Natl Acad Sci USA*, 1990, 87, 4093–4097.

Bender W., Spierer P., Hogness D.S., Chromosomal walking and jumping to isolate DNA from the Ace and rosy loci and the bithorax complex in *Drosophila melanogaster*. *J Mol Biol*, 1983a, 168, 17–33.

Bender W., Akam M., Karch F., Beachy P.A., Peifer M., Spierer P., Lewis E.B., Hogness D.S., Molecular genetics of the bithorax complex in *Drosophila melanogaster*. *Science*, 1983b, 221, 23–29.

Berry M., Gehring W.J., Phosphorylation status of the SCR homeodomain determines its functional activity: essential role for protein phosphatase 2A, B. *EMBO J*, 2000, 19, 2946–2957.

Billeter M., Qian Y.Q., Otting G., Müller M., Gehring W., Wüthrich K., Determination of the nuclear magnetic resonance solution structure of an *Antennapedia* homeodomain-DNA complex. *J Mol Biol*, 1993, 234, 1084–1093.

Carrasco A.E., McGinnis W., Gehring W.J., De Robertis E.M., Cloning of an *X. laevis* gene expressed during early embryogenesis coding for a peptide region homologous to *Drosophila* homeotic genes. *Cell*, 1984, 37, 409–414.

Cohen S.N., Chang A.C., Boyer R.W., Helling R.B., Construction of biologically functional bacterial plasmids *in vitro*. *PNAS*, 1973, 70, 3240–3244.

Czerny T., Halder G., Callaerts P., Kloter U., Souabni A. Gehring W.J., Busslinger M., Twin of eyeless, a second Pax-6 gene of *Drosophila*, acts upstream of eyeless in the control of eye development. *Mol Cell*, 1999, 3, 297–307.

Ehrenberg M., Rigler R., Rotational brownian motion and fluorescence intensity fluctuations. *Chemical Phys*, 1973, 4, 390–401.

Elson E., Magde D., Fluorescence Correlation Spectroscopy. I. Conceptual basis and theory. *Biopolymers*, 1974, 13, 1–27.

Garber R.L., Kuroiwa A., Gehring W.J., Genomic and cDNA clones of the homeotic locus *Antennapedia* in *Drosophila*. *EMBO J*, 1983, 2, 2027–2036.

Gehring W., Bildung eines vollständigen Mittelbeines mit Sternopleura in der Antennenregion bei der Mutante Nasobemia (Ns) von *Drosophila melanogaster*. *Jul Klaus Arch*, 1966, 41, 44–54.

Gehring W.J., *Master Control Genes in Development and Evolution: The Homeobox Story*. Yale University Press, New Haven, USA, 1998, p. 236.

Gehring W.J., *La Drosophile aux yeux rouges: gènes et développement*. Éditions Odile Jacob, Paris, 1999a, pp. 39.

Gehring W.J., Lifting the lid on the homeobox discovery. *Nature*, 1999b, 399, 521–522.

Gehring W.J., Qian Y.Q., Billeter M., Furukubo-Tokunaga K., Schier A.F., Resendez-Perez D., Affolter M., Otting G., Wüthrich K., Homeodomain-DNA Recognition. *Cell*, 1994, 78, 211–223.

Gehring W.J., Kloter U., Suga H., Evolution of the Hox Gene Complex from an Evolutionary Ground State. *In* Pourquié O. (Ed.) *A contribution of Hox Genes, a volume of Current Topics in Developmental Biology*, 2009, 88, 35–61.

Gibson G., Schier A., LeMotte P., Gehring W.J., The specificities of Sex combs reduced and Antennapedia are defined by a distinct portion of each protein that includes the homeodomain. *Cell*, 1990, 62, 1087–1103.

Jacob F., Monod J., Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, 1961a, 3, 318–356.

Jacob F., Monod J., On the Regulation of Gene Activity. *Cold Spring Harb Symp Quant Biol,* 1961b, 26, 193–211.

Jorgensen E.M., Garber R.L., Function and misfunction of the two promoters of the *Drosophila Antennapedia* gene. *Genes Dev*, 1987, 1, 544–555.

Kaufman, T.C., Lewis, R., Wakimoto, B., Cytogenetic Analysis of Chromosome 3 in *Drosophila melanogaster*: The homoeotic gene complex in polytene chromosome interval 84a-B. *Genetics*, 1980, 94, 115–133.

Kissinger C.R., Liu B.S., Martin-Blanco E., Kornberg T.B., Pabo C.O., Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Ångström resolution: a framework for understanding homeodomain-DNA interactions. *Cell*, 1990, 63, 579–590.

Kuroiwa A., Hafen E., Gehring W.J., Cloning and transcriptional analysis of the segmentation gene *fushi tarazu* of *Drosophila*. *Cell*, 1984, 37, 825–831.

Laughon A., Scott M.P., Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. *Nature*, 1984, 310, 25–31.

Lewis, E.B., A gene complex controlling segmentation in *Drosophila*. *Nature*, 1978, 276, 565–570.

McGinnis W., Lawrence P.A., Master control genes in development and evolution: The Homeobox story, *Nature*, 1999, 399, 301–302.

McGinnis W., Levine M.S., Hafen E., Kuroiwa A., Gehring W.J., A conserved DNA sequence in homeotic genes of the *Drosophila Antennapedia* and bithorax complex. *Nature*, 1984a, 308, 428–433.

McGinnis W., Garber R.L., Wirz J., Kuroiwa A., Gehring W.J., A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. *Cell*, 1984b, 37, 403–408.

McGinnis W., Hart C.P., Gehring W.J., Ruddle F.H., Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of *Drosophila*. *Cell*, 1984c, 38, 675–680.

Otting G., Qian Y.Q., Billeter M., Müller M., Affolter M., Gehring W.J., Wüthrich, K., Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J*, 1990, 9, 3085–3092.

Panzer S., Weigel D., Beckendorf S.K., Organogenesis in *Drosophila melanogaster*: embryonic salivary gland determination is controlled by homeotic and dorsoventral patterning genes. *Development*, 1992, 114, 49–57.

Papadopoulos D.K., Vukojevic V., Adachi Y., Terenius L., Rigler R., Gehring W.J., Function and specificity of synthetic Hox transcription factors *in vivo*. *PNAS*, 2010, 107, 4087–4092.

Plaza S., Prince F., Jaeger J., Kloter U., Flister S., Benassayag C., Cribbs D., Gehring W.J., Molecular basis for the inhibition of *Drosophila* eye development by *Antennapedia*. *EMBO J*, 2001, 20, 802–811.

Plaza S., Prince F., Adachi Y., Punzo C., Cribbs D., Gehring W.J., Cross-regulatory Protein-Protein Interactions between Hox and Pax Transcription factors. *PNAS*, 2008, 105, 13439–13444.

Qian Y.Q., Billeter M., Otting G., Müller M., Gehring W.J., Wüthrich K., The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, 1989, 59, 573–580.

Schneuwly S., Klemenz R., Gehring W.J., Redesigning the body plan of *Drosophila* by ectopic expression of the homeotic gene *Antennapedia*. *Nature*, 1987, 325, 816–818.

Scott M.P., Wiener A.J., Hazelrigg T.I., Polisky B.A., Pirrotta V., Scalenghe F., Kaufman T.C., The molecular organization of the *Antennapedia* locus of *Drosophila*. *Cell*, 1983, 35, 763–776.

Shepherd J.C.W., McGinnis W., Carrasco A.E., De Robertis E.M., Gehring W.J., Fly and frog homeo domains show homologies with yeast mating type regulatory proteins. *Nature*, 1984, 310, 70–71.

Vukojevic V., Heidkamp M., Ming Y., Johansson B., Terenius L., Rigler R., Quantitative single-molecule imaging by confocal laser scanning microscopy. *Proc Natl Acad Sci USA*, 2008, 105, 18176–18181.

Vukojevic V., Papadopoulos D.K., Terenius L., Gehring W.J., Rigler R., Quantitative study of synthetic Hox transcription factor-DNA interactions in live cells. *Proc Natl Acad Sci USA*, 2010, 107, 4093–4098

Wolberger C., Vershon A.K., Liu B., Johnson A.D., Pabo C.O., Crystal structure of a MAT 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell*, 1991, 67, 517–528.

Wüthrich K., Billeter M. *in* Gehring W.J., « la Drosophile aux yeux rouges », Éditions Odile Jacob, Paris, 1999, Fig. 5, p. 150.