General discussion and summary

Joost M. Woltering

*Hox, microRNAs and evolution; overview and perspectives.*

The patterning of the anterior-posterior axis as done by the *Hox* genes forms one of the best characterized and most investigated regulatory systems within the field of developmental biology. Historically, it probably represents the first developmental problem to be seen with our modern understanding of gene biology; during the early 1980’s the Drosophila *Antenapedia Hox* complex was the first developmental genetic locus to be investigated with the, at that time, brand new toolbox of molecular biology. Despite their long history in biological research the *Hox* genes still receive the unrelenting attention from the scientific community. In Pubmed there are in total 4804 citations for ‘*Hox*’ and the yearly number of papers published on the subject is still steadily on the increase (fig.1).

Although all paradigms in the *Hox* field were set almost 2 decades ago, challenging discoveries are being made each year and novel developments from other fields are steadily incorporated in the framework of understanding on *Hox* functioning and dis-functioning. One of these breakthroughs from recent history is what is now sometimes referred to as the ‘non-coding RNA revolution’. Although it was already known at the time that small non coding RNAs could silence genes (Lee et al. 1993), it was not until the discovery of the mechanism of RNA interference that it was appreciated how widespread this regulatory mechanism is. The term microRNA itself was only introduced in 2001 (Ruvkun 2001). The impact of the discoveries made in 2001 and 2002 was deemed so great that *Science* magazine awarded it the title of ‘Breakthrough of the year 2002’ (Couzin 2002).

Subsequently it turned out that genomes contain hundreds of non coding RNA genes, many of which are microRNAs but novel classes have been discovered more recently (e.g. Girard et al. 2006). From whole library sequencing projects, aimed at identifying all the microRNAs present in the genome, it appeared that two microRNA families are present within the *Hox* clusters, namely *miR-10* and *miR-196* which were both identified in 2003.
As the Hox clusters form such structured and conserved loci where gene functioning is highly interwoven with the surrounding genomic context, there was an immediate interest in the possibility of Hox/AP patterning related functions for these microRNAs. In the year following, the relationship between miR-196 and HoxB8 (Yekta et al. 2004) was one of the first vertebrate microRNA/target relationships to be described. The work on miR-10 described in the first part of this thesis was sparked and inspired by this unexpected presence of these novel components within the Hox clusters. At the time...
General discussion and summary

This work was started (early 2004) the only microRNAs with anything known about their biological roles were lin-4 and let-7 in *C. elegans*.

The complete lack of knowledge about animal microRNA functioning certainly gave the sense of excitement that comes with the entering of completely unexplored scientific territory. The project has always been approached with a classical reverse genetic research strategy, aimed at understanding the developmental role of the *miR-10* microRNA. Also, my point of view has always been from the side of *Hox* and developmental biology and has followed the technical and biochemical advances in the microRNA field rather than having the ambition of developing these myself.

In chapter 2 of this thesis I describe the genomic annotation and expression of the *miR-10* family in Zebrafish. The results of a search for all miR-10 family members and their annotation on the genome turned out to be highly surprising and solved the standing issue of the ‘lost’ Zebrafish *HoxDb* cluster of which only *miR-10d* remains. As much as solving problems for the *Hox* annotations this discovery raises questions about the functioning of *miR-10* and microRNAs in general. It has been a more often noted phenomenon that many microRNA families have been very conserved after genome duplications. This feeling is only confirmed by the status of *miR-10d* as a lone survivor of an estimated 100kb of *HoxDb* cluster that has vanished during evolution. Whether there indeed is a higher genomic survival rate of microRNAs compared to paralogues coding gene groups has to my knowledge never been investigated. This problem also is not as straightforward to answer as it seems; although microRNA paralogues are readily identified by their great sequence homology, there is no good database containing information on paralogues coding genes. Nevertheless this remains an interesting question. If there is indeed a higher conservation ratio the reasons for this are not clear (although chapter 2 contains a few speculations on this issue).

Another interesting observation coming from the structure of the *HoxDb* cluster is that sequences in the *Hox* cluster do not only mutate but also disappear when they are not required anymore. Apparently there are forces in the genome leading to a very strong tendency for compaction. This issue has in the scope of this work however not been looked into further.
Chapter 3 is dedicated to a functional analysis of miR-10 in Zebrafish development. Here I present a loss of function approach by morpholino knockdown and a gain of function by miR-10 siRNA injection to study the effects on Zebrafish development. The main insight coming from this work is that miR-10 represses HoxB1a and HoxB3a. These target genes are located 3’ from the position of the Hox-4 associated miR-10 and are in the embryo expressed more anteriorly. Knockdown of miR-10 leads to an upregulation of the target genes showing that this interaction occurs in the embryo and overexpression of miR-10 leads to specific phenotypic defects associated with the loss of function of the target genes. The phenotypes of the morphant and overexpression embryos are remarkably normal. For microRNA morphant embryos it has now been established in screens that this is more often the case (Kloosterman 2007). Apparently the functions of many microRNA families are so subtle that phenotypic effects are absent or missed upon superficial inspection. For the embryos in which miR-10 is overexpressed the lack of phenotypic defects (besides the nice VIIth nerve and GCM-2 expression defects) is much more puzzling though. As noted in chapter 3 miR-10 is having a total of 1969 predicted targets representing about 10% of the genome and many more than the 738 predicted targets of the average microRNA. This fact seems incompatible with the very specific phenotypic effects observed. The best conclusion to draw from these results probably is that there is still plenty of work to do in the lab in order to understand the rules of microRNA/target recognition and behind the computer for the bio-informaticians developing the target prediction software.

In contrast to previous studies on the Hox related microRNAs we do find a clear loss of function phenotype in the morphant embryos. The migration of the Xth cranial nerve is disrupted and there are obvious problems with the migratory pathways of the motorneurons. Unfortunately, we have not been able to link these phenotypes to a derepression of the identified Hox target genes. The phenotype is consistent with a loss of miR-10 since only motorneurons that are located within the expression domain of miR-10 are affected and motorneurons located more rostrally appear normal. More research will be necessary to understand the precise mechanism by which miR-10 coordinates the normal process of neuronal migration but one of the potential target genes that could be looked at is BDNF (brain derived neurotrophic factor).

‘Genomes regulate gene expression by switching on or off the transcription of genes’ this was and still is the standing paradigm both before and after the non coding RNA revolution.
However, the discovery of a large number of microRNA interactions has taught us that this is only part of the picture. Besides only regulating gene presence at the transcriptional level there appears a need for organisms to do this as well with a second layer of control at the posttranscriptional level. As it appears now, this second layer of control is (at least in vertebrates) subordinate to and less important than the primary level of transcriptional control. Conceptually it is an interesting issue why there is a need for this secondary level at all; what is wrong with the transcriptional machinery that makes it need a helping hand from microRNAs?

In chapter 4 I briefly touch upon this problem after the discovery of the autoregulatory properties of the \( \text{miR-10c}/\text{HoxB3a} \) target combination. I suggest that, at least in the \textit{Hox} clusters, there is a conflict between the need of genomic clustering and clearly specified gene expression domains. On the question ‘which gene will be influenced by an enhancer?’ most biologists would probably just reply ‘the one located behind it’ and this answer probably is the correct one. As we can already understand intuitively, proximity is a strong determinant in the specificity of gene transcription; an enhancer/promoter box is more likely to drive the transcription of an adjacent gene than that of one that is 2 mega-bases away. But how is specificity determined when there is a dense concentration of genes due to a selection on clustering? How easy is it to evolve enhancers that ‘know’ to promote the transcription of one gene but leave others unaffected? To my knowledge this is a very poorly explored field but my guess is that this indeed will be very difficult. The fact that within the \textit{Hox} clusters transcription units do not correspond to traditional genes is now very well established (e.g. Mainguy et al. 2007). The function for this ‘messy’ pattern of transcription is still unknown though. My explanation is that many of these ‘strange transcripts’ represent noise that is inherent to the regulatory mechanisms acting on the clusters. In many cases this noise will have no harmful effects on the biology of the animal. In some instances however, this may result in ‘ectopic’ expression that does interfere with life or development. In these cases a secondary system of microRNA regulation may have evolved to interfere. In chapter 4 this theory is worked out in more detail, also showing the existence of similar relations for \( \text{miR-196} \) and target genes.

\textbf{Chapter 5} describes the axial patterning and expression of \textit{Hox} genes in two organisms with an elongated bodyform. This project was initiated as a spin off from the \( \text{miR-10} \)
General discussion and summary

project. Initially it was aimed at developing the locked nucleic acid hybridization (LNA) technique for use in evo-devo studies by using microRNAs and LNA probes as pan vertebrate marker probes by designing them at ultra conserved regions (UCRs) in mRNAs. Since microRNAs and URCs do not differ in sequence among most vertebrate species, they preclude the time consuming process of picking up probes for each individual species investigated.

The mechanisms behind the evolvement of a serpentine bodyform have since long been subject to intense scientific debate, fueled however for about 95% by theories and very little by experimental facts. The classic paper on this subject is the influential 1999 Cohn and Tickle article published in *Nature* (110 citations according to Google scholar) in which they investigate the expression of 3 Hox genes using immunostaining in Python embryos. (On a personal note I would like to remark here that it was this particular paper that formed the trigger to switch my courses in university from ecology to molecular/developmental biology). The extended bodyform of snakes is characterized by a deregionalization of the axial skeleton and loss of the typical division in cervical, thoracic, lumbar and caudal regions of which only a thorax and caudal region are prominently present. The authors claim that the expression of the investigated Hox genes is spread all over the axis (and follows the deregionalization of the body) and that the neck region has been overpatterned by an anterior shift of HoxC8 and HoxC6 which could possibly also account for the loss of limbs. The actual figures in the article, showing the stained snake embryos, are very difficult to interpret due to a very strong background. The fact that they used only 3 out of the expected 39 Hox genes and that they paid no attention to the changes in the posterior end of the axis made this theory interesting for further investigation, using more Hox genes and producing higher quality data with in situ hybridization. I also aimed at studying the evolutionary event of body elongation/deregionalization in two different species that independently acquired this bodyform, namely a snake and a caecilian species (see chapter 5 for details) to establish whether convergent evolution in bodyform is caused/ accompanied by convergent or divergent shifts in gene expression.

The idea of using LNA probes against normal messengerRNAs unfortunately failed to work. As it is generally recognized now, LNA works very well for microRNA detection but is for unknown reasons virtually useless in the detection of mRNAs. The detection of miR-196 and miR-10 microRNAs worked very well though in both snake and caecilian embryos.
The results were interesting for *miR-196*, which expression pointed towards a convergent shift. Since the LNA probes failed to work I decided to clone a *HoxC8* probe in both to complement the data. The in situ hybridization results obtained with this probe were however completely surprising; instead of being expressed till the rostral end of the trunk as described, *HoxC8* has an anterior boundary of expression located well within the thoracic part of the mesoderm. This lead me to clone a complete panel of *Hox* genes as described in chapter 5.

The conclusion I can draw so far that the idea of complete deregionalization of the thorax on the level of *Hox* gene expression is incorrect; thoracic *Hox* genes are still expressed in a spatial collinear fashion in the somitic mesoderm. The *HoxC6* and *HoxC8* genes investigated by Cohn and Tickle have rostral boundaries within the thorax and can therefore not be responsible for the posteriorization of the snake neck. The *HoxB7* and *HoxB8* genes are expressed much more anteriorly though and the anterior boundary of *HoxB7* is situated two somites posterior from the 1st thoracic vertebra. As *Hox-6* genes are currently believed to be the important genes in rib induction I expect that *HoxB6* is the gene expressed at the cervical/thoracic boundary; the testing of which is currently in progress. In caecilians, using fewer genes I observe about the same pattern with still a high degree of regionalization in the expression of thoracic genes and also more anterior expression of *HoxB* than *HoxC* gene.

Something conceptually novel from these experiments is that in the changes leading to a macro-evolutionary modification not all *Hox* clusters behave the same. *Hox* paralogue group genes are largely redundant in function and show much the same expression patterns. In snakes my results show that the loss of cervical region is due to overpatterning by the *HoxB* but not *HoxC* cluster. Currently I am also investigating the expression of *HoxA6* and *HoxD8* to see how the expression of the *HoxA* and *HoxD* cluster has changed. The specific deregionalization of a single and not all *Hox* clusters may have two explanations. First the change in expression is likely caused by mutation of a regulatory element; simultaneous mutations in all 4 cluster leading to a similar change in expression would be a highly unlikely event. Secondly it could be that changing the expression zones in the somites also affects the expression in other tissues. Besides patterning the axial skeleton, *Hox* genes are also involved in the patterning of the gut, genitals and blood. Expression should probably not be ‘deregionalized’ in these tissues and this could be lethal. In the case of this last
explanation, the deregionalization of one and not another cluster would be a classic example of a developmental constraint at work. At the posterior end of the axis the results are as surprising. The Hox code has been well established in mouse and it is known that both Hox-9 and Hox-10 have rib suppressing activity and are responsible for the induction of the lumbar part of the skeleton, a region that is absent from snakes.

In my experiments I do find though that both HoxB9 and HoxC10 are expressed within the thoracic part of the somitic mesoderm. These results indicate dissociation between the classical Hox code and rib formation in snakes. It is interesting to note that the absence of a lumbar region is quite common within squamates and that this dissociation therefore may be a common theme within evolution, something certainly worth investigating in species showing this type of morphology. This would also mean that alterations in body pattern are not necessarily a reflection of changes in the expression pattern of Hox genes but can also be caused by changes in the interpretation of Hox signals by downstream developmental mechanisms. The expression of HoxC13 is exactly as expected from the mouse; it is expressed in the caudal region of the axis and thereby seems to behave according to the classical Hox code.

The uniting theme within this thesis are the Hox genes and work is presented which extends over a great width of the field from genomics, functional developmental biology to basic evo-devo.

Resuming, the most important contributions to the scientific knowledge described in this thesis are:
- the identification of the Zebrafish HoxDb cluster
- identification of miR-10 as a intrinsic part of the Hox regulatory machinery
- the creation of an insightful model into the rational for posttranscriptional interactions within the Hox clusters
- better understanding of the deregionalization model that is at the basis of the development of a serpentine bodyplan and insight into the snake Hox code.
References

Developmental basis of limblessness and axial patterning in snakes.

Couzin J (2002)
Breakthrough of the year. Small RNAs make big splash
Science 298:2296-7

A germline-specific class of small RNAs binds mammalian Piwi proteins.

Kloosterman WP, Lagendijk AK, Ketting RF, Moulton JD, Plasterk RH. (2007)
Targeted inhibition of miRNA maturation with morpholinos reveals a role for miR-375 in pancreatic islet development.


Extensive polycistronism and antisense transcription in the Mammalian Hox clusters.


MicroRNA-directed cleavage of HOXB8 mRNA.
Science 304:594-6