UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA

Identification of the RNA regulators operating in the segmentation clock

Sara Maria Ferreira Fernandes

Mestrado em Bioquímica

2011

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE QUÍMICA E BIOQUÍMICA

Identification of the RNA regulators operating in the segmentation clock

Sara Maria Ferreira Fernandes

Tese de mestrado orientada por:

Professora Doutora Margarida Gama Carvalho

Professora Doutora Leonor Saúde

Mestrado em Bioquímica

2011

Agradecimentos

Gostaria de agradecer às minhas orientadoras, Margarida Gama Carvalho, Leonor Saúde e Rita Fior, por me terem recebido nos seus grupos, pelo conhecimento transmitido, pela paciência, por valorizarem o meu trabalho e pela dedicação e abertura de espírito que demonstráram na orientação.

Aos colegas dos três laboratórios agradeço o bom ambiente e a boa disposição e em especial à Ana Luisa e à Verónica agradeço as conversas, os pequenos momentos de distracção e desabafo, mas acima de tudo, a amizade.

Á Lara e ao Fábio agradeço, não só toda a ajuda com os peixes, mas também a boa disposição e a simpatia. Pelos ensinamentos e orientação na secção de microscopia e tratamento de imágem gostaria de agradecer ao Dr. António Temudo.

Á Rita e á Andreia agradeço a paciência, as conversas, os momentos de desabafo e os momentos de calma. Por me lembrar que a vida não é só trabalho agradeço à Verónica.

Por acreditárem em mim, por me apoiarem em tudo e serem uma perpetua fonte de inspiração e afecto, agradeço aos meus pais.

Abstract

Embryonic development is highly controlled in space and time and in many developmental processes this control relies on mechanisms of post-transcriptional regulation, namely the modulation of mRNA stability and translation efficiency.

This study is focused on the formation of transient mesodermal structures, termed somites, which are the precursors of the vertebrae, skeletal muscles and dermis of the back. Somite formation – Somitogenesis - is governed by two mechanisms: a molecular clock, composed of cyclically expressed genes that need to be unstable at the mRNA level to produce sustained oscillations; and a wavefront of differentiation, formed by a morphogen gradient which is thought to be generated by the slow degradation of the morphogens' mRNA. However, the mechanisms that regulate the stability of these mRNAs have not been studied.

We set out to assess the effects that the 3'untranslated regions (3'UTRs) of the zebrafish cyclic and wavefront genes exert on the stability and translation efficiency of their mRNAs. To achieve this we established an *in vivo* fluorescent reporter system and created a mathematical model to describe and characterize this system. Additionally we conducted a bioinformatic identification of sequence elements present in these UTRs that could modulate the observed effects.

We uncovered the existence of three alternative 3'UTRs for the wavefront gene fgf8a. Our *in vivo* studies revealed that two of these fgf8a alternative 3'UTRs, as well as the 3'UTRs of the cyclic genes *her1* and *her7*, modulate a negative regulation of the reporters' expression.

The bioinformatic approach revealed inter-species conserved sequence elements in these 3'UTRs, as well as predicted binding motifs for heterogeneous ribonucleoproteins, and for members of the microRNA-1 family and microRNA 17~19 clusters. The *in vivo* experimental system implemented will enable the future study of the role these sequence elements play in the post-transcriptional regulation of the cyclic and wavefront genes.

Resumo

O desenvolvimento embrionário é rigorosamente controlado no tempo e no espaço, sendo que, grande parte dos processos envolvidos neste controlo depende de mecanismos de regulação pós-transcripcional, especificamente, da modulação da estabilidade do mRNA e da sua eficiência de tradução. A somitogénese, em particular, é um processo que ocorre no desenvolvimento embrionário de vertebrados e consiste na formação de estruturas transientes a partir da mesoderme pré-somítica, denominadas sómitos, sendo que estes actuam como precursores de estruturas repetitivas, incluindo as vertebras, costelas, músculo esquelético e a derme das costas. A formação dos sómitos dá-se de forma sequencial e rítmica, sendo este controlo espacial e temporal exercido por dois mecanismos: um relógio molecular, constituído por genes com expressão cíclica; e uma frente de diferenciação (wavefront) formada por um gradiente morfogénico. Modelos matemáticos demonstraram que as oscilações de expressão génica potenciadas por este relógio molecular só se observam se estes genes derem origem a mRNAs instáveis. Paralelamente, o gradiente morfogénico associado á wavefront é estabelecido pela lenta degradação do mRNA de um dos morfógeneos - o fgf8. Consequentemente, a somitogénese é fortemente depende da regulação da estabilidade dos mRNAs dos genes cíclicos e da wavefront, no entanto os mecanismos moleculares envolvidos nestes processos de regulação pós-transcripcional não foram até data elucidados.

De forma a desvendar os mecanismos envolvidos na regulação da estabilidade do mRNA destes genes, começamos por proceder ao estudo do efeito que as regiões não traduzidas a 3' (3'UTRs) dos genes cíclicos e da wavefront do peixe zebra, têm na expressão de um sistema repórter. Também construímos um modelo matemático descritivo deste sistema repórter com o intuito de caracterizar a estratégia experimental implementada. Numa segunda abordagem, foi realizada uma análise bioinformática das sequências das 3'UTRs destes genes focada na identificação de elementos regulatórios que possam estar envolvidos na modulação dos efeitos observados.

A análise das sequências não traduzidas, disponíveis nas bases dados, revelou a existência de três 3'UTRs alternativas para o gene da wavefront do peixe zebra fgf8a. Estas 3'UTRs alternativas são produzidas como resultado da utilização diferencial de três sinais de poliadenilação alternativos.

A influência destas fgf8a 3'UTRs e das 3'UTRs dos genes ciclicos *her1* e *her7* na expressão do mRNA foi testada *in vivo*. O procedimento experimental envolveu o estabelecimento de um sistema repórter *in vivo*, em particular a produção, *in vitro*, de mRNAs reporter contendo a sequência codificante de proteína fluorescente (Green Fluorescent Protein) acoplada às sequências 3'UTRs dos genes dos genes de interesse. Estes mRNAs reporter foram microinjectados em embriões de peixe zebra no estadío de uma célula e a análise dos níveis de expressão do reporter foi feita 25 a 33 horas após injecção.

Esta análise permitiu a observação da redução da expressão do repórter associada a duas das 3'UTRs alternativas de fgf8a em relação aos controlos, no entanto, interpretação dos resultados obtidos para a terceira 3'UTR de *fgf8a* necessita de uma optimização adicional do sistema

experimental. Relativamente aos genes cíclicos, observámos que as 3'UTRs de *her1* e *her7* provocam uma regulação negativa da expressão do repórter. Estes resultados sugerem que a destabilização dos mRNAs de *her1* e *her7* pode ser mediada por elementos nas suas 3'UTRs. Esta sugestão foi apoiada pelo nosso modelo matemático, representativo do sistema experimental, e o processamento do transcrito foi simulado em diferentes condições de estabilidade e eficiência de tradução do mRNA. A informação obtida com esta caracterização, juntamente com a dinâmica de expressão observada *in vivo* num período de 8 horas, sugere que os efeitos das 3'UTRs na expressão do gene repórter *in vivo* resultam de mecanismos de regulação negativa da estabilidade do mRNA. No entanto, são necessários estudos adicionais para avaliar um possível envolvimento destas 3'UTRs no controlo da tradução dos mRNAs.

A implementação deste sistema experimental conduziu á observação de uma influência diferencial do método de poliadenilação do mRNA reporter na sua expressão *in vivo*, sendo esta influência especialmente marcante para o reporter fluorescente sem 3'UTRs. Este reporter, quando é poli-adenilado *in vitro*, produz níveis de expressão reduzidos. No entanto, se um sinal de poliadenilação for utilizado a sua expressão é favorecida. Uma explicação possível para este fenómeno pode residir no facto de o sinal de poliadenilação utilizado actuar como uma sequência separadora entre o codão de terminação e a cauda poly(A), que não está presente no transcrito poliadenilado *in vitro*. A necessidade desta sequência separadora para o processamento normal do mRNA foi préviamente considerada na literatura. Neste estudo, propomos a hipótese de que na ausência deste separador, a terminação da tradução não se processa correctamente e consequentemente o mRNA é reconhecido pela maquinaria de controlo de qualidade celular como um transcrito anormal, sendo degradado com uma eficiência acentuada.

Utilizando uma abordagem bioinformática, procedemos á identificação de elementos reguladores que poderiam mediar efeito que as 3'UTRs dos genes cíclicos e da wavefront exercem na expressão de um repórter. Nesta abordagem bioinformática tivemos em conta uma teoria actual que considera que a instabilidade dos vários mRNAs cíclicos poderia ser regulada por uma plataforma pós-transcripcional comum, sendo que esta plataforma poderia incluir microRNAs, proteínas de ligação ao RNA, ou ambos. Neste cenário, os constituintes da plataforma iriam interagir com elementos ou motivos de sequência presentes nas regiões não traduzidas (5'UTRs e 3'UTRs) dos genes cíclicos potenciando a sua degradação coordenada.

Neste estudo, previmos a existência de 3'UTRs alternativas para dois genes ortólogos do fgf8a de peixe zebra utilizando duas ferramentas online de previsão de locais de poliadenilação, polyAH e polyApred. A subsequente identificação de elementos de sequência conservados entre diferentes espécies de vertebrados foi feita recorrendo ao alinhamento de sequências relevantes. Esta análise revelou um elemento fortemente conservado entre as 3'UTRs de fgf8a do peixe zebra e as dos seus ortólogos de galinha e murino. Uma subsequente pesquisa unbiased de motivos nas 3'UTRs de interesse revelou a presença que um motivo TC-rich nas 3'UTRs de vários genes da familia *her*, do peixe zebra, incluindo os genes cíclicos, *her1* e *her7*,Este motivo também foi identificado nas 3'UTRs

de genes cíclicos de galinha e murino.

Numa abordagem focada na identificação de potenciais factores de regulação, procedemos à previsão de locais de ligação de proteínas reguladoras (RNA-binding proteins) e microRNAs nas 3'UTRs dos genes cíclicos e do *fgf8a* do peixe zebra. Para tal recorremos às ferramentas SplicingRanibow, microInspector e MicroCosm. Nas 3'UTRs de todos os genes cíclicos e da wavefront do peixe zebra foram previstos motivos de ligação a membros da família hnRNP. Adicionalmente, nas 3'UTRs de dois genes cíclicos - *her7* e *deltaC* - identificámos motivos de ligação a membros da família miR-1 de microRNAs e a membros dos clusters de microRNAs miR-17~92. Note-se que estes microRNAs e proteínas têm funções documentadas no desenvolvimento embrionário. Adicionalmente, os padrões de expressão destes microRNAs e RNA-binding proteins durante o desenvolvimento embrionário do peixe zebra, documentados na literatura, revelam que estes são expressos durante a somitogénese, em estruturas do embrião onde os seus alvos previstos são transcritos e exercem a sua função. Consequentemente, os microRNAs e RNA-binding proteins identificados neste estudo constituem excelentes candidatos para a regulação pós-transcipcional dos genes cíclicos e da wavefront.

O sistema experimental implementado *in vivo* para a realização deste estudo, poderá ser utilizado em abordagens futuras para determinar a influência que cada um dos elementos identificados bioinformáticamente, têm na regulação da estabilidade e eficiência de tradução dos seus respectivos mRNAs.

Abbreviations

mRNA	messenger ribonucleic acid
UTR	untranslated region
PSM	presomitic mesoderm
hairy/ E(spl) *and* hes	hairy and enhancer of split
her *and* hey	hairy and enhancer of split – related
bHLH	basic-helix-loop-helix
Fgf	Fibroblast growth factor
RBP	RNA-binding protein
miR	microRNA
ARE	AU-rich element
MAPK	Mitogen -activated protein kinase
VEGF	Vascular endothelial growth factor
miRISCs	microRNA-induced silencing complexes
PLAS	Power Law and Simulation
NGD	No-Go decay
hnRNP	messenger ribonucleoprotein
PABP	poly(A) binding protein
eGFP	enhanced Green Fluorescent Protein
PCR	polymerase chain reaction
RACE-PCR	rapid amplification of cDNA ends – PCR
EST	expressed sequence tag

Table of contents

I. INTRODUCTION

The development of a living organism is highly regulated in space and time, with many of the genes involved in developmental process being expressed for short periods of time and/or in well defined spatial domains. During embryonic development, these patterns of gene expression need to be precisely regulated in order to specify cellular identities and correctly direct the complex processes that lead to the formation of a multicelular organism. To achieve this control over protein levels, the modulation of transcription is often insufficient because existing mRNAs and proteins must be removed or inactivated to produce the required patterns of expression. Consequently, many developmental processes rely on post-transcriptional regulation, namely the modulation of mRNA stability and translation efficiency. In this project, we focused our attention on one developmental process in particular, somitogenesis, and set out to study the mechanisms of mRNA stability regulation that contribute to this process.

I.1 Somitogenesis

Somites are epithelial blocks of mesoderm containing the precursors of the vertebrae, skeletal muscles and dermis of the back. Somitogenesis – the process of somite formation from the presomitic mesoderm (PSM) – is tightly regulated, both temporally and spatially. Pairs of somites pinch off synchronously from the PSM with a species-specific frequency until a final number of somites, which is also characteristic of the species, is reached. For example, in Zebrafish embryos, a new pair of somites is formed every 30 minutes, until a total of approximately 30 somites is reached. [1]

According to the Clock-Wavefront model, there are two main interconnected mechanisms responsible for the spatiotemporal regulation of somitogenesis – the segmentation clock and the determination front. The clock is thought to play a role in the frequency of somite formation and the determination front appears to be involved in determining the size of the nascent somites. [2] [1]

The clock is composed of a set of genes – cyclic genes – which display transcriptional oscillations in PSM cells. The majority of these genes belong to the Notch signalling pathway which plays a key role in cell-cell communication and, in this case, ensures the synchronization of the oscillations between neighbouring cells. [3] This synchronization is manifest in the form of striped patterns of expression that sweep across the PSM in a posterior to anterior direction. The frequency of these oscillations is at its' highest in the posterior-most end of the PSM, and diminishes as the cells assume more anterior positions, with the expression of the cyclic gene becoming finally restricted to either the anterior or the posterior domain of the nascent somite. This final expression pattern is thought to contribute to the delineation of the future somites' antero-posterior polarity (**Figure 1)**. [4] [5] [1]

When the frequency of oscillation is at its maximum, it coincides with the overall frequency of somite formation. In other words, the time that it takes to complete one such full oscillation is the same time required to form a new pair of somites (**Figure 1**) [6]

The determination front is a virtual frontier between the posterior PSM, composed of loose mesenchyme, and the anterior PSM, where cells are more epithelialized and already committed to their somitic fate. It is conceptually similar to a wavefront of maturation and is defined as the level at which cells first acquire their somitic identity. It's position is determined by two opposing gradients: an anterior-to-posterior Retinoic acid (RA) gradient – involved in the maintenance of the epithelialized state; and a posterior-to-anterior Fibroblast Growth Factor 8 (Fgf8) and Wnt signalling gradient – involved in the maintenance of the mesenchymal state. [7] [8] [9] It is thought that, once the cells cross this frontier, they become competent to respond to a periodic signal, provided by the clock, that potentiates the definition of the future segmental domain. (**Figure 1**) [8] [2]

Somitogenesis is concomitant with the proliferation of PSM cells and axis elongation (caudal growth) which leads to an anterior-to-posterior displacement of the Fgf8 and Wnt gradient and consequent posterior regression of the determination front. In the time it takes to form one pair of somites the determination front travels a distance equal to the length of one somite. In other words, the group of cells that cross the determination front during one cycle of the clock has the length of one somite. This group of cells will become increasingly epithelialized and give rise to a new pair of somites (**Figure 1**) [8] [2]

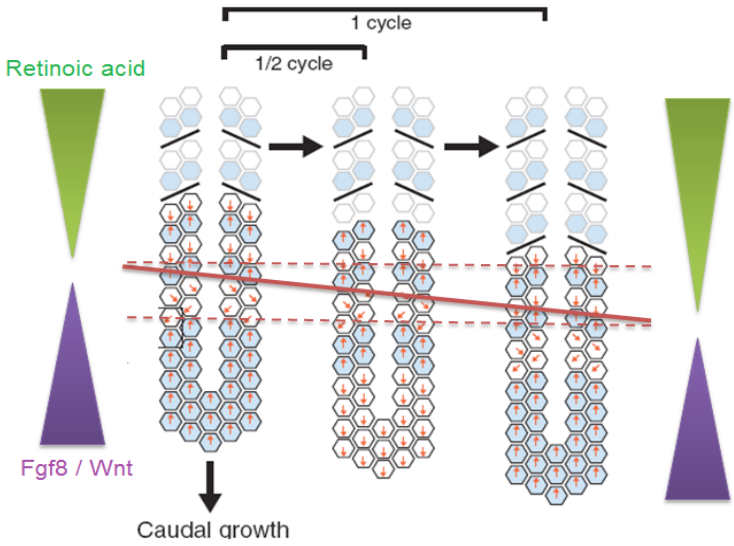


Figure 1- Schematic representation of the clock – wavefront model. Neighbouring cells in the same phase of the clock cycle are depicted in the same colour (blue or white). Note the striped pattern of expression sweeping across the PSM and becoming finally arrested in the nascent somites, where it delineates the nascent somite boundaries (black lines). The position of the determination front (red line) corresponds to the meeting point of the opposing retinoic acid (green) and Fgf8 / Wnt signalling (purple) gradients, and in the course of one cell cycle it travels a distance equal to the length of one somite (dotted red line). Image adapted from J. Lewis et al. (2009) [1]

I.1.1 Somitogenesis in the zebrafish and the importance of mRNA stability regulation

The segmentation clock

In the Zebrafish, the only cyclic genes reported to date are either members or downstream targets of the Notch pathway. An association of this pathway with the segmentation clock was also observed in Mouse and Chick, with the latter two species having cyclic genes belonging not only to the Notch pathway, but also to the Fgf8 and Wnt pathways. [2] This conserved participation of the Notch pathway in vertebrate somitogenesis, made it the object of heightened interest. The zebrafish Notch pathway-associated *her1* (hairy and enhancer of split-related 1), *her7* and *deltaC* genes, in particular, have been the subject of several studies because they oscillate in synchrony with one another (in the posterior PSM) with a period of 30 minutes (at 28ºC), which corresponds to the time taken to generate one pair of somites. [10] [2]

Her1 and Her7 belong to the hairy and enhancer of split [E(spl)] related family of transcriptional repressors and have been shown to repress their own transcription as well as that of *deltaC*, in the posterior PSM [4] . DeltaC is a transmembrane protein that functions as a Notch ligand potentiating the activation of the Notch pathway. In many biological systems, genes of the Her family of transcription factors function as downstream targets of Notch [11] and in the Zebrafish in particular, a constitutively active form of Notch1A results in overexpression of *her1* in the PSM [12] . Taking this into consideration, a two-loop model has been proposed to elucidate the origin of the oscillations and the mechanism of intercellular coordination. The first loop is a negative feedback loop established by direct autorepression of *her1* and/or *her7* and has been proposed to function as the fundamental pacemaker mechanism of the segmentation clock by establishing the intracellular oscillations of *her1*, *her7* and *deltaC* (**Figure 2 -** red loop). While the second loop, generated by the periodic repression of *delta C* would allow the synchronous activation of Notch signalling in neighbouring cells and consequent intercellular coordination of the oscillations (**Figure 2** - green loop). [13]

Mathematical modelling has shown that the *her1*/7 loop would be capable of generating the fore mentioned oscillations provided certain conditions were met. Among which, the sum of the delays associated with transcription (Tm) [1] and translation (Tp)[2] would have to be greater than the lifetimes of the *her7*/*her1* mRNAs (τ_m) and proteins (τ_p). In this model, if oscillations occur, the period of the oscillation would be approximately equal to 2 (Tm + Tp + τ_m + τ_p). Considering that in the zebrafish, the oscillations occur with a period of 30 minutes, this model implies that, sustained oscillations would occur if the mRNA and protein products of these genes are highly unstable. [13] The instability of the *her1* and *her7* mRNAs was confirmed by *in situ* hybridisation experiments, based on which the

1 The time – Tm – that elapses between the beginning of transcription and the arrival of mature mRNA to the cytosol.

2 The time – Tp – that elapses between the beginning of translation and the arrival of the fully functional protein to its site of action.

authors determined that *her1*, *her7* and Delta C have mRNA half-lives between 6.1 and 8.1 minutes. [4] However the mechanisms involved in modulating the instability of these mRNAs have not yet been uncovered.

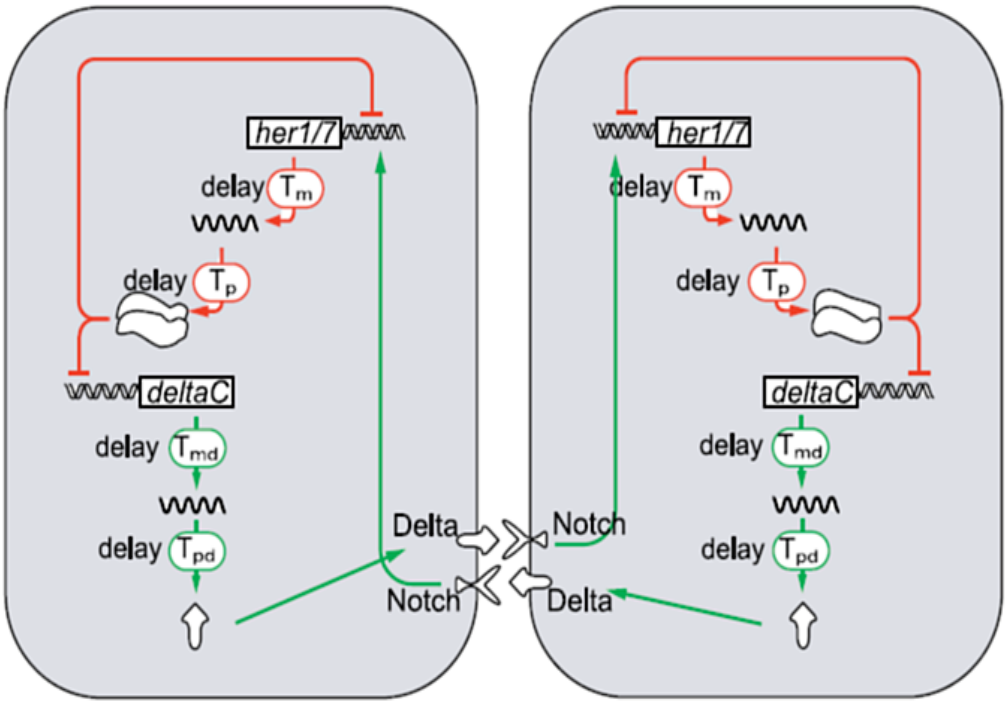


Figure 2 – Schematic representation of the model proposed to elucidate the origin and intercellular coordination of the oscillations. According to this model, an intracellular negative feedback loop (red) is generated by the transcriptional repression of *her1* and *her7* by their own protein products. This feedback loop would potentiate the *her1* and *her7* oscillations and the concomitant periodic transcriptional repression of *deltaC* by Her1 and Her7 would be responsible for the oscillations of *deltaC*. The oscillating levels of *deltaC* would allow a periodic activation of the Notch signalling pathway in neighbouring cells which, in turn would activate the transcription of *her1* and *her7* in these cells thus allowing a coordination of the oscillations between adjacent cells (green). Tm (d) and Tp(d) correspond to the delays associated with the transcription and translation of *her1*/7 and delta C. Arrows denote a positive effect, ⊥ denotes an inhibitory effect. Image adapted from [4]

The hairy and Enhancer of split gene family

The hairy/E(spl) familiy of transcriptional repressors is a subsection of the basic-helix-loop-helix (bHLH) superfamily of transcription factors. Many vertebrate bHLH proteins are effectors of the Notch pathway and have key functions in the regulation, not only of mesoderm segmentation, but also of vasculogenesis, neurogenesis, myogenesis and T lymphocyte development. [11]

In the context of somitogenesis, the zebrafish *her1* and *her7* hairy/ E(spl) family members have well established and largely overlapping cyclic patterns of expression in the PSM. However, an additional set of hairy/ E(spl) family members, with diverse patterns of expression in the zebrafish

PSM, exhibit spatially restricted cyclic or dynamic expression in the PSM. *Her11*, *her12* and the *her15* gene pair belong to this category and will be referred to as pseudo-cyclic genes. (**Figure 3**)

Her11 is a paralog of *her1* and both are, along with *her7* and her5, orthologs of the cyclic mouse gene hes7. *Her11* appears to display a cyclic/dynamic expression in the intermediate to anterior PSM, where it is synexpressed with *her1* and *her7*. However it is not expressed in the posterior portion of the PSM [14] Conversely, the *her15*[3] gene pair and *her12* display posteriorly restricted oscillatory domains, which appear to be largely in phase with *her1* and *her7*. The *her15* and *her12* genes also produce weak dynamic striped patterns of expression in the anterior PSM, located at the future segmental borders (**Figure 3**) [15]

Hey1, also displays cyclic expression and belongs to a subclass of hairy/ E(spl) genes, characterized by the presence of a C-terminal YRPW motif instead of the WRPW motif that is characteristic of the her genes. [14] [16] *Hey1* has a highly dynamic transcription in the anterior PSM and stripes of expression in the posterior halves of the nascent and epithelialized somites. (**Figure 3**) [16]

Her6 and *her1*3.2 are also expressed in the PSM, however they do not show an oscillatory or dynamic behaviour and will be referred to as non-cyclic genes. *Her6* is expressed in the tailbud and in two stripes in the anterior PSM, located in the caudal compartment of the early somitomeres. [17] During somitogenesis *her13.2* expression remains confined to the posterior PSM and tailbud.(**Figure 3**) [18] [19]

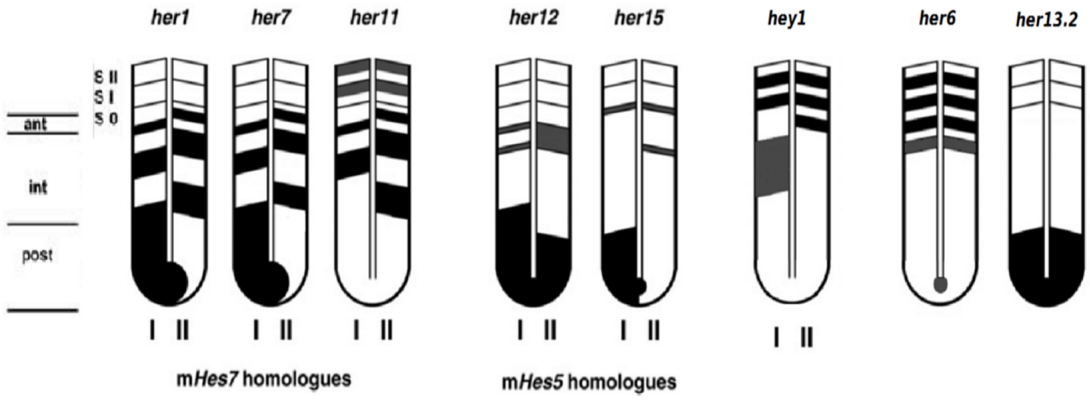


Figure 3 – Reported expression patterns of hairy/E(spl) family members in the zebrafish PSM. *Her1, her7* and *her11* are orthologs of the mouse gene *Hes7*. *Her12* and *her15* are orthologs of the mouse gene *Hes5.* Gray areas indicate low levels of expression, black areas denote regions were the gene is strongly expressed. S0 - somite to be formed next, SI - last formed somite, SII - penultimate formed somite. The left and right halves indicate two different phases of the expression cycles, each, marked by I and II below the drawings. For a description of the expression patterns and references see text. Image adapted from [15]

3 The her15a and her15b genes have a very high degree of homology and the near exact match of the predicted transcripts prevented the distinction of their individual expression patterns by *in situ* hybridization, consequently the authors considered the overall sum of the two expression patterns.

The determination front

As previously stated, the position of the determination front is thought to be determined by two opposing gradients, an Fgf8 and Wnt signalling posterior-to-anterior gradient and a retinoic acid anterior-to-posterior gradient.

The importance of the Fgf8 signaling gradient in particular was demonstrated in zebrafish by experiments that perturbed the slope of the gradient. This was achieved by grafting FGF8-soaked beads[4] next to the PSM and by resorting to a fibroblast growth factor receptor inhibitor (SU5402). This resulted in the formation of abnormally small and abnormally large somites, respectively. These experiments also revealed that MAPK signalling appears to be activated by Fgf signalling and contribute to the maintenance of the mesenchymal state. [7] Even though Fgf8 appears to have a role in the establishment of the determination front, experiments conducted with the zebrafish *acerebellar* (ace) mutant, which does not produce a functional form of Fgf8, did not reveal a somitogenesis phenotype. [20] This phenomenon was later explained by a study that uncovered a redundancy between Fgf8 and Fgf24 in regards to the promotion of posterior PSM formation. [21]

In mouse and chick, *in situ* Hybridization experiments using intronic and exonic probes for *Fgf8* revealed that the FGF8 gradient is an mRNA gradient. In other words, transcription occurs only when the cells are located in the growing posterior tip of the embryo (at the tailbud level). As cells move out of the tailbud and into the PSM, they stop transcribing the *Fgf8* gene, and a slow process of mRNA decay allows the transcripts to remain in the cells while they are progressively displaced to more anterior positions, generating a posterior-to-anterior gradient of Fgf8 mRNA and protein. [22] The presence of an analogous stabilization of the FGF8 mRNA in the Zebrafish is assumed, however the mechanisms involved in regulating this stability have not yet been uncovered.

I.1.2 mRNA stability regulation during somitogenesis – state of the art

In essence, the two mechanisms central to the spatiotemporal regulation of somitogenesis are crucially dependent on the regulation of mRNA stability. While the cyclic mRNAs have to be unstable in order to produce sustained oscillations the fgf8 mRNA has to be stable in order to contribute to the positioning of the determination front. However, very little is known about the mechanisms that regulate these mRNA dynamics.

It is widely accepted that mRNA stability is primarily mediated by motifs present in the transcripts' sequence. The mature mRNA molecule can be divided into four regions: a 5' untranslated region (5'UTR) containig a 5'terminal 7-methyl-guanylate (m7G) cap structure, a coding sequence, a 3' untranslated region (3'UTR) and a 3'-terminal stretch of 100-250 adenine residues termed, the poly(A) tail. [23]. In most cases, the motifs involved in regulating mRNA stability are located in the

4 The authors did not resort to the zebrafish Fgf8a protein, instead mouse-FGF8b was used.

untranslated regions (UTRs) of the mRNA (see section 1.2).

In the case of *her1* and *her7*, studies conducted by Gajewski et al.[48] and A. C. Oates et al,[152] suggest that the 5'terminal 25 nucleotides of the zebrafish *her1* and *her7* 5'UTRs as well as a region located 73 nucleotides downstream of the 5'terminal of the *her7* 5'UTR could be involved in mediating the instability of the transcripts. This suggestion was made based on the observation that morpholinos (MO) targeting these regions induced a stabilization of either the endogenous *her1* and *her7* mRNAs or of a GFP reporter containing the 5'UTR of *her7*. Note that additional stabilizing effects on the *her1* and *her7* endogenous mRNAs were reported for MOs targeting the start of their respective coding sequences. [10] [24]

Evidence regarding the possible existence of motifs involved in the modulation of mRNA stability in the 3'UTRs of the zebrafish *her1* and *her7* comes from Giudicelli et al. [4]. The authors generated stable fish lines carrying a transgene in which a heat-shock promoter was coupled to the coding sequence and 3'UTR of either *her1* or *her7.* After inducing overexpression of the transgenes, the authors observed a rapid and uniform decline of the *her1* and *her7* mRNA levels. These results imply that, even in the absence of their 5'UTRs the transcripts have a short lifetime. In summary, these studies suggest that motifs in the 5'UTRs, coding sequence and possibly the 3'UTRs of these mRNAs could be involved in the destabilization of their transcripts.

In regard to fgf8, V. Hilgers, et al. [25], using and inducible reporter system with the 3'UTR of the chick *FGF8,* observed a slow decline of the reporters' protein levels upon termination of transcription in the chick neural tube. The authors concluded that this slow decline reflects an elevated stability of the reporter mRNA mediated by elements located in the 3'UTR of *FGF8*.

Given the importance of mRNA stability regulation in somitogenesis, it becomes necessary to understand how sequence motifs could influence the stability of a transcript.

I.2 Mechanisms of mRNA stability modulation

In order to understand how mRNA stability is regulated, one must first analyse the mRNA molecule itself. Even though all four regions of an mRNA molecule (5'UTR, coding sequence, 3'UTR and poly(A) tail) can influence both the stability and translatability of the molecule, the 3'UTR is most frequently the main regulator of mRNA stability, being also involved in mRNA transport to the cytoplasm and in the determination of its' subcellular localization. The 5'UTR is the main effector of translation modulation and the coding sequence contains the necessary information for protein synthesis (triplet codons encoding one amino acid each). Lastly, the poly(A) tail has been linked to the enhancement of translation and cytoplasmic mRNA stability. [23]

The influence of the UTRs on post-transcriptional regulation is effected mainly through interactions between regulatory factors and regulatory elements. The main regulatory factors involved in mRNA stability are RNA-binding proteins (RBPs) and non-coding RNAs (primarily microRNAs). Certain sequence motifs and structural elements located within the untranslated regions establish interactions

with regulatory factors and thus act as regulatory elements. These interactions can have different effects on mRNA stability and translational efficiency, depending on the regulatory elements and factors involved in the interaction. [23] [26]

The influence of the poly(A) tail on post-transcriptional regulation is primarily mediated by interactions with poly(A)-binding proteins (PABPs). These proteins, when associated with the poly(A)-tract, protect the transcript, from rapid decay and interact with the translation initiation complex. The latter interaction can not only enhance the recruitment of new ribossomal units, thus stimulating translation, but also inhibit the action of a decapping enzyme, which destabilizes the mRNA by removing the 5'cap. [27] [23] Note that several studies indicate that the effect of the poly(A) tail on the expression of a transcript depends on the length of the tail, with longer tails having a greater positive effect on translation and in some cases on mRNA stability than shorter tails. [28] [29]

I.2.1 Regulatory Elements

As previously stated, mRNA turnover is most frequently regulated by elements located in the 3'UTRs. However, unlike DNA-mediated regulatory signals whose activity is essentially mediated by their primary structure, the biological activity of regulatory motifs at the RNA level can rely either on their primary structure, their secondary structure, or on a combination of both. [23]

AU-rich elements (AREs), and a steem-loop structure termed the iron responsive element (IRE) are possibly the two most studied primary structure and secondary structure-based regulatory elements, respectively. While the first are found mainly in mRNAs encoding nuclear transcription factors such as c-Fos and c-Myc, the second is primarily associated with the regulation of mRNAs encoding proteins associated with the metabolism of iron, such as the transferrin receptor.[23] [30] [26].

While these and other regulatory elements can have an influence on the stability of an mRNA molecule on their own, we frequently come across multiple regulatory elements within a single transcript. In some cases the activities of the individual regulatory elements behave cooperatively (or additively) to promote a certain form of regulation, whereas in other cases different elements induce different forms of regulation, thus generating a finer mechanism of post-transcriptional control.

An example of the first scenario involves the Vascular Endothelial Growth Factor (VEGF) mRNA. This mRNA is intrinsically liable, but is stabilized in response to hypoxia. Under normoxic conditions, the destabilization is effected by the additive action of elements in the 5' UTR, the 3'UTR and the coding sequence. Although each region is independently capable of promoting mRNA degradation, together they act additively to effect rapid degradation. Under hypoxic condition, the observed stabilization is completely dependent on the cooperation between elements present in the 5'UTR, the coding sequence, and the 3'UTR. Combinations of any of two of these three regions are ineffective in responding to hypoxia and only by combining all three regions can the hypoxic stabilization seen with the endogenous VEGF mRNA be reproduced. [31]

An example of the second scenario comes from *Drosophilas*' early embryonic development. Transcripts encoding the RNA-binding protein Nanos or the heat-shock protein Hsp83 are degraded everywhere in the embryo except in the posterior polar plasm, with distinct regulatory elements located in the 3'UTRs of these mRNAs mediating both the degradation in the embryo as a whole and the stabilization at the pole. [32]

Polyadenylation and Alternative 3'UTRs

An additional level of complexity associated with post-transcriptional regulation ties in with the fact that a considerable number of genes give rise to mRNAs with alternative UTRs. Alternative UTRs can be formed from the use of different transcription-start sites, splice donor and/or acceptor sites or polyadenylation sites. Polyadenylation, in particular, is a process concomitant with translation and is triggered by a polyadenylation signal (AAUAAA) located in the 3'terminal region of the 3'UTR. When this signal is recognized by the polyadenylation machinery, the precursor mRNA undergoes endonucleolytic cleavage and subsequent polyadenylation in a polyadenylation site, located 10~30 nucleotides downstream of the polyadenylation signal. The presence of more than one polyadenylation signal can therefore lead to the production of 3'UTRs with different lengths and consequently exclude (or include) subsets of regulatory elements from each alternative 3'UTR. The selection of one polyadenylation signal over another, can be potentiated by different environmental of developmental cues. [23] [33]

In the context of embryonic development, it was recently observed that mouse genes tend to express mRNAs with longer alternative 3'UTRs as embryonic development progresses. In addition, it was reported that, on average, longer alternative 3'UTRs have a greater number of AU-rich sequences and predicted microRNA target sites. Therefore the production of longer 3'UTRs may be a mechanism aimed at augmenting the post-transcriptional control of gene expression. [34] Another relevant study involved four Drosophila *Hox* genes, including Ultrabithorax (*Ubx*). During embryonic development these genes produce mRNAs with variable 3'UTRs in different regions of the embryo. Analysis of the resulting sets of alternative 3'UTRs revealed the presence of different microRNA target sites in each set, making them considerably different substrates for microRNA regulation. It was proposed that this differential production of alternative UTRs is a molecular strategy aimed at ensuring that the transcripts display different levels of visibility to these regulatory factors according to developmental cues. [35]

I.2.2 Regulatory Factors

RNA-Binding Proteins

RNA-binding proteins are important mediators of post-transcriptional regulation of gene expression. These proteins interact with target sites in the mRNA through their RNA-binding domains and posses auxiliary domains that mediate protein-protein interactions and sub-cellular targeting.

An RBP can exert its' function as a regulatory factor in many ways. A destabilizing effect can arise, for example, if the interaction of an RBP with a regulatory element promotes the deadenylation of the transcript and/or the recruitment of members of the mRNA degradation machinery. Conversely, some RBPs, upon interaction, may use their auxiliary domains to protect the poly(A) tail (and consequently the transcript) from degradation. Additionally, the interaction of an RBP with an mRNA molecule can alter the structure of the mRNA and thus facilitate or hinder interactions with other regulatory factors. [36] [26] [37] Many RBPs, like the heterogeneous nuclear ribonucleoproteins (hnRNPs), which are involved in the cytoplasmic post-transcriptional regulation of gene expression can also participate in a wide variety of regulatory processes – like transcription and pre-mRNA processing- within the nucleus [5]. [38] [23]

Of particular importance, in the context of embryonic development, are the members of the Signal Transduction and Activation of RNA (STAR) family of RBPs. Two members of this family, the HOW (Drosophila) and Quaking (mouse) proteins, have been linked to several signaling pathways and developmental defects. The HOW protein is required for mesoderm formation and a mutation of Quaking gives rise to non-viable embryos with abnormal somites and a desorganized antero-posterior axis. [39] The zebrafish orthologues of these genes, quakingA (*qkA*) and quakingB (*qkB*) are expressed in the PSM and functional studies have implicated QkA in the regulation of mRNA stability and translation efficiency in the paraxial mesoderm.[40] [41] More specifically, QkA positively modulates the stability an translation efficiency of *gli2a*, a member of the hedgehog signaling pathway. This pathway, and consequently QkA, are involved in the regulation of muscle cell type specification. [42]

microRNAs

microRNAs (miRs) are small non-coding RNAs that extensively regulate gene expression at the post-transcriptional level [43] [44] miRs inhibit protein synthesis by repressing translation and/or by bringing about deadenylation and subsequent degradation of mRNA targets. Generally, miRs function as part of ribonucleoprotein complexes, termed microRNA-induced silencing complexes (miRISCs). Core components of miRISCs include the AGOfamily of proteins, which directly anchor miRNAs in a deep pocket, and the GW182 family of proteins, that interact directly with AGO proteins and serve as

5 hnRNPs form complexes with RNA polymerase II transcripts and are said to package them into messenger ribonucleoproteins (mRNPs)

the downstream effectors of miRNA-mediated repression. [43] [44] miRs are transcribed from independent non-coding regions, or introns of protein-coding genes as long primary transcripts (pri-miRs). In several cases, transcription is tissue-specific and/or developmental stage-specific and to allow coordinated expression, some miRs are clustered in polycistronic transcripts, termed microRNA clusters. [44] The regulated expression of these clusters may be linked to transcriptional regulation of their host gene promoters. For example, the transcription factor c-Myc upregulates the transcription of the human miR-17~92 microRNA cluster (this cluster contains six miRs: miR-17-5p, miR-18a, miR-19a, miR-20a, miR-19b-1 and miR-92-1). Interestingly, two miRs in this cluster negatively regulate the expression of E2F1, a transcription factor that promotes cell cycle progression. E2F1 is also transcriptionaly activated by c-Myc, thus generating a mechanism through which c-Myc simultaneously activates E2F1 transcription and limits its translation, thus producing a tightly controlled proliferative signal. [45] [44]

Subsequent to transcription, the pri-miRNAs are processed by the successive action of two members of the RNase-III family of enzymes, Drosha and Dicer, giving rise to the mature miR. The 2-8 nucleotides at the 5'end of the mature miR constitute the Seed region. This region nucleates the interaction between the miR and the regulatory elements located in the 3'UTRs of target mRNAs. In the target 3'UTR, often multiple target sites, either for the same or for different miRs, are required for effective repression, and when the sites are close to one another, they tend to act cooperatively. [43] [44]

In addition, the absence of structural motifs near a target site, that could hinder accessibility to the site may improve the efficiency of miR-mediated regulation. [43] Recent studies have shown that, the accessibility of a miR-target site can be influenced by RNA-binding proteins (RBPs) like HuR and dead end 1(Dnd1). In regard to HuR, studies conducted in Hela cells, revealed that HuR appears to inhibits c-Myc expression by recruiting the let-7-loaded RISC to the c-Myc 3'UTR, possibly, not via direct interaction with the complex but, through enhancement of the accessibility of the let-7 binding site to the miRISC complex. [36]. Conversely, in zebrafish primordial germ cells Dnd1 appears to prevent miR-430-mediated repression of nanos1 and TDRD7 mRNAs by binding to U-rich sequences adjacent to the miRs' target site, and thus, interfering with miR binding [46].

RNA operons

Eukaryotes need to coordinate the expression of multiple genes in order to accomplish diverse biological functions required for processes like growth and differentiation. The importance of this coordination gave rise to the hypothesis that multiple mRNAs could be co-regulated by one or more sequence-specific RBPs that would orchestrate their splicing, export, stability, localization and/or trans-

lation. This hypothesis is the basis for the post-transcriptional 'RNA-operon' [6]theory, which has been increasingly supported by experimental data. According to to this model, a set of regulatory factors - primarily RBPs and non-coding RNAs, but also metabolites – could combine to regulate multiple functionally related mRNAs along a coordinated pathway of RNA processing, allowing cells to respond with considerable agility to both intra and extracellular stimuli. The model provides a simplifying principle that helps explain the higher-order organization and dynamics of functionally related mRNAs at several levels of post-transcriptional regulation. [47] [48]

In the context of biological clocks, the oscillation model of circadian rhythms is a good example of coordinated expression events. This model relies on interconnected feedback loops modulated by positive and negative transcriptional regulators. Two post-transcriptional regulators have been proposed to modulate the mRNA stability of multiple cycle-related genes. [49] The first is an RNA deadenylase termed nocturnin which displays high-amplitude circadian expression in the clock-containing photoreceptor cell layer of the *Xenopus laevis'* retina and has been postulated to deadenylate a subset of circadian-cycle-related transcripts and thus potentiate their coordinated decay [50]. The second, termed butyrate response factor 1 (BRF1) is an RBP which also displays rhythmic circadian expression, and is known to interact with AU-rich transcripts and promote their degradation. BRF1 was proposed to function as part of a post-transcriptional operon (that may also include protein kinase B and 14-3-3) which insures the degradation of a subset of transcripts, termed immediate-early genes, during the diurnal phase of the circadian cycle [51].

6 An RNA operon is a ribonucleoprotein structure in which multiple mRNAs are coordinately regulated by RNA-binding proteins and small non-coding RNAs. The combination of multiple mRNAs in an RNA operon can change dynamically following biological perturbations

II. Objectives

In light of our current understanding of somitogenesis it is clear that the spatiotemporal regulation of this process is crucially dependent on what could be called opposing mechanisms of post-transcriptional regulation: on one hand the cyclic genes have to be unstable at the mRNA level in order to produce sustained oscillations and on the other hand the *fgf8* wavefront gene needs to be stable at the mRNA level to position the determination front. However, the molecular mechanisms that regulate mRNA stability in the context of somitogenesis are still unknown.

The work presented in this thesis is part of a long term project aimed at identifying the molecular mechanisms involved in the post-transcriptional regulation of the cyclic genes and the wavefront fgf8 gene, in the zebrafish.

The aim of this study was to perform an initial characterization of the influence that the 3'UTR sequences of the zebrafish cyclic genes *her1* and *her7* and the wavefront gene *fgf8a* have on the post-transcriptional regulation of their respective mRNAs. In this context, our specific aims were:

- To implement an experimental system that would allow the characterization of the effect that each of these 3'UTRs exerts on the expression of a reporter mRNA.

- To identify sequence elements that could potentiate the observed post-transcriptional effects using a bioinformatic approach. This approach was focused on the identification of :
  - cross-species conserved sequence elements within each 3'UTR
  - short sequence motifs present in the 3'UTRs of the cyclic genes
  - known RBP and miR binding motifs in the 3'UTRs

III. Materials and Methods

III.1. Mathematical modelling of the *in vivo* fluorescent reporter system

All the simulations described in the results section **IV.1.1** were done using the PLAS (Power Law Analysis and Simulation) [52] software. Each curve depicted in this section of the results was computed by the software from an input analogous to the one indicated bellow. The rate constant values changed in each individual curve are indicated in section **IV.1.1**. Arguments proceeded by // are ignored by the program and are presented in this section for descriptive purposes only.

```
// Differential Rate Equations representing the experimental system (Figure 5)
// the arguments v1 to v5 indicate the rates of mRNA degradation (v1), translation(v2), maturation (v4)
// and protein degradation (v3, v5)
RNAGFP' = - v1
ProtGFP' = v2 - v3 - v4
ProtMGFP' = v4 – v5
// Individual rate laws of the model, stipulated assuming first-order rate kinetics
v1= kgfp*RNAGFP
v2= kt*RNAGFP
v3= kp*ProtGFP
v4= km*ProtGFP
v5= kp*ProtMGFP
// Initial levels of each molecular species in the system, expressed in number of molecules
// the mRNA levels considered are an approximation of the amount of injected mRNA
RNAGFP= 1.9e8
ProtGFP= 0
ProtMGFP= 0
// Estimated rate constant values, expressed in number of molecules per second
kgfp= 2.75e-5
kp= 7.405e-6
kt= 0.375
km= 1.77e-4
// Initial time point(t0), final time point (tf) and number of time points (Npoints) considered in each simulation
t0 = 0.
Npoints = 60000.
tf = 1000000.0
// Indication of the molecular species to plot as a function of time
!!ProtMGFP
```

The rate constants adopted in this model were adapted from the literature. The GFP mRNA and protein decay rates were calculated from their respective half-lives of 7 and 26 hours, determined by studies conducted in Human Embryonic Kidney 293T cells [53] and mouse LA-9 cells [54] . The

maturation rate constant of 1.77e-4 s^{-1} used in this model was determined for the Enhanced Green Fluorescent Protein (eGFP) *in vitro* by A.G. Evdokimov et al. [55]. The rate of translation of 0.375 s^{-1} was calculated considering that the average ribosome density in eukaryotic mRNAs and the average rate of translation elongation, reported in E coli by H. Bremer et al. [56].

III.2. Cloning of the 3'UTRs of *her1*, *her7* and *fgf8a*

The *fgf8aL* 3'UTR and the 3'UTRs of *her1* and *her7* were amplified by PCR from a sample of genomic DNA. All the forward primers used included a ClaI restriction site and all the reverse primers used included a XhoI restriction site. Note that the 3'terminal of the *her1* 3'UTR consists of a very AT-rich region, consequently the *her1* reverse primer recognizes the stretch of 23 nucleotides in the genomic sequence directly downstream of the 3'terminal of the *her1* 3'UTR. Primer sequences for all the 3'UTRs amplified in this work are listed in **Table 1**.

Table 1 - **Primers used for amplification reactions**. The sequences highlighted in blue correspond to ClaI clevage sites, the sequences highlighted in red correspond to XhoI cleavage sites.

Primer	Sequence
her1_Fw	5'-TGTA**ATCGAT**GCAAAACTGAAGACACTTAGCA-3'
her1_Rev	5'-TGTA**CTCGAG**AGAAACGAGTCATACAAAAAGCA -3'
her7_Fw	5'-TGTA**ATCGAT**CTGAGAGAAAAGACATTCATTGAGA-3'
her7_Rev	5'-TGTA**CTCGAG**AATGGTGAATATTTCACTTTTAATGG-3'
fgf8a_Fw	5'-TGTA**ATCGAT**AGAGTGAAGCCAGAGAAAAG-3'
fgf8aL_Rev	5'-TGTT**CTCGAG**TAGAATCTTAGATCTTCATTTAGGC-3'
fgf8aM_Rev	5'-TGTT**CTCGAG**ATCCTAAGGTAAATTTATTACA -3'
fgf8aS_Rev	5'-TGTT**CTCGAG**TCAAATAAAAATATATTTATATTTGTATAA-3'

A pCS2-eGFP vector had been previously constructed by insertion of the coding sequence of eGFP into a pCS2+ expression vector, between the EcoRI and StuI restriction sites (downstream of an SP6 RNA polymerase promoter sequence) (**Figure 4**). The 3'UTR amplification products were digested with XhoI and introduced into the pCS2-eGFP vector, downstream of the eGFP coding sequence, by ligation between the StuI and XhoI restriction sites. (**Figure 4**) The restriction enzymes used were purchased from Fermentas and the digestions were done according to the manufacturers' instructions. The digestion products were purified with GeneJET™ Gel Extraction Kit (Fermentas_K0691) and the ligations were done using a T4 DNA Ligase (Fermentas_EL0011)

DH5α bacteria (*Escherichia coli*) were transformed with the ligation products. In this process competent bacteria were incubated with each ligation mixture on ice for 20 min, subjected to a heat shock (45 s at 42ºC), incubated on ice for an additional 2 min and grown in antibiotic-free LB medium (1% w/v bacto-tryptone; 0,5% w/v bacto-yeast extract; 1% w/v NaCl) for 45 min at 37ºC and 220 rpm. The transformed bacteria were plated on LB-agar plates supplemented with the appropriate antibiotic

(Ampicilin 100µg/mL) and left to grow at 37ºC overnight. This antibiotic was used because the pCS2+ expression vector contains an Ampicllin resistance gene which allows the selection of transformed colonies. Transformed bacterial colonies were grown in LB medium supplemented with antibiotic and used to extract plasmid DNA. Small scale plasmid DNA extractions and purifications were done with the GeneJET™ Plasmid Miniprep Kit (Fermentas_K0502), according to the manufacturers' protocol.

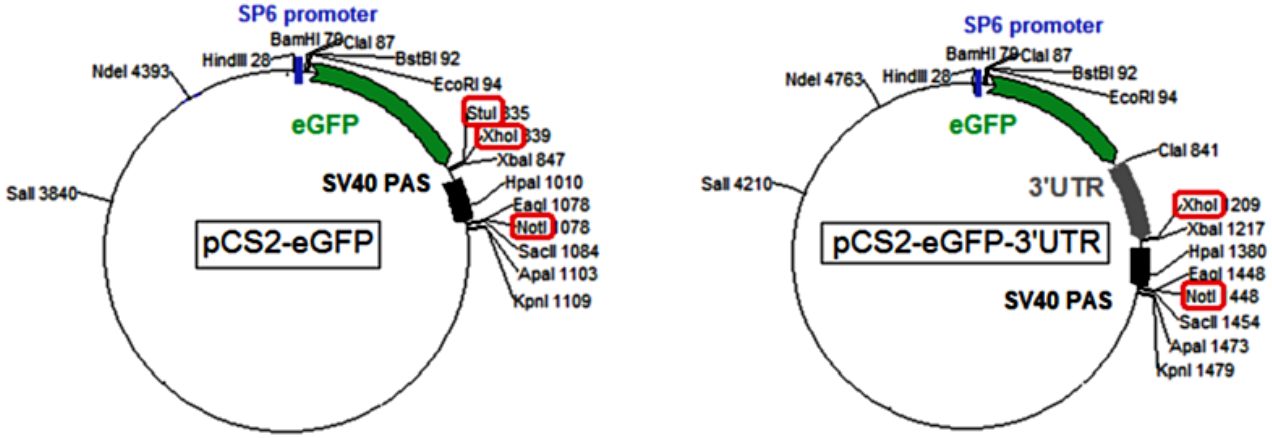


Figure 4 - **pCS2 vectors used in this study**. The eGFP coding sequence (highlighted in green) was inserted into a pCS2+ expression vector between EcoRI and StuI, thus producing the pCS2-eGFP vector. Each 3'UTR amplification product (highlighted in grey) was digested with XhoI and inserted into the pCS2-eGFP vector between StuI and XhoI – highlighted in red – thus producing the pCS2-eGFP-3'UTR vector. The restriction sites XhoI and NotI – highlighted in red – were used to linearise the vectors to produce the templates for *in vitro* transcription (see section III.3). The SV40 late polyadenylation signal (SV40 PAS) is depicted in black. Linearisation of the vectors with XhoI excludes the SV40 signal from the final template, whereas linearisation with NotI leads to the inclusion of the signal in the final template (see section III.3).

Recombinant vectors were identified by restriction digestion and the presence of each 3'UTR was confirmed by sequencing (sequencing reactions were done by the company Stabvida). The sequencing results revealed that the cloned sequences were consistent with the annotations NM_131281 (*fgf8aL*), NM_131609 (*her7*) and OTTDART00000023081 (*her1*), with all the detected single nucleotide polymorphisms being present in either EST data or in the results obtained in two different amplification experiments done with two distinct genomic DNA samples and two different DNA polymerases.

The *fgf8aM* and *fgf8aS* 3'UTRs were amplified by PCR from the purified pCS2-eGFP-*fgf8aL* construct and cloned into the pCS2-eGFP vector as described above. The mCherry coding sequence was cloned into the pCS2+ expression vector between XhoI and XbaI, using the small scale plasmid production protocol described above.

III.3. *In vitro* transcription and polyadenylation

To produce the DNA templates for the SV40-bearing eGFP reporter mRNAs, the pCS2+ constructs described in III.2. were linearized with NotI. To produce the templates for the polyadenylated eGFP reporter mRNAs, the pCS2+ constructs were linearized with XhoI. (**Figure 4**)

The DNA templates were purified with the Zymoclean™ Gel DNA Recovery Kit (Zymo research_D4008) and transcribed *in vitro* with the SP6 mMessage mMachine kit according to manufacturers' instructions (Ambion_AM1340).

In vitro polyadenylation was done using the Ambion Poly(A) Tailing Kit (Ambion_AM1350). In the polyadenylation reactions, two alterations were made to the manufacturers' instructions, namely, only 2U of Poly(A) Polymerase were used, in a final reaction volume of 100µL, and the reaction time was changed to 40 minutes. The average length of the poly(A) tails obtained with this protocol, as assessed by denaturing agarose gel eletrophoresis, was 150 nucleotides. Note that the SV40-bearing transcripts were not polyadenylated *in vitro*.

All the transcripts were purified using illustra™ MicroSpin™ G-50 Columns (GE Healthcare_27-5330-01).

III.4. Microinjection

In the first set of experiments (section **IV.1.2**), wild-type AB zebrafish embryos at the one-cell stage were injected with 1.4nL of a 0.32 fmol/1.4nL solution of each eGFP reporter mRNA. This is the molar equivalent of 0.1ng of *in vitro* polyadenylated eGFP mRNA per embryo. mCherry-SV40 mRNA was co-injected with each of the eGFP reporter mRNAs at 0.1ng/1.4nL.

In the polyadenylation control experiments (section **IV.1.2**), the embryos were injected with 1.4nL of a 0.23 fmol/1.4nL solution of each eGFP reporter mRNA. This is the molar equivalent of 0.07ng of *in vitro* polyadenylated eGFP mRNA per embryo. mCherry-SV40 mRNA was co-injected with each of the eGFP reporter mRNAs at 0.07ng/1,4nL.

In each comparative assay the two reporter mRNAs up for comparison were injected into sibling embryos (different embryos from the same batch) and 2 to 3 batches of embryos were used per assay. The injected embryos were incubated at 28ºC for 23-24h.

III.5. Embryo mounting and image acquisition

At 23-24 hpf (hours post fertilization, or hours post-injection), ten to fifteen embryos injected with each eGFP reporter RNA were dechorionated. The chorions were weakened in a 2 mg/ml pronase solution prepared in Embryo Medium (13,7mM NaCl; 0,5mM KCl; 0,25mM Na_2HPO_4; 0,15mM KH_2; 1,3mM $CaCl_2$; 1mM $MgSO_4•7H_2O$; 4,2mM $NaHCO_3$), for 2 to 5 min, at 28.5°C and removed by gentle rinsing with embryo medium. The dechorionated embryos were transferred to an anaesthetising

0,016% (w/v) ethyl 3-aminobenzoate (tricaine) solution prepared in embryo medium and mounted in 1.5% (w/v) low melting agarose, prepared using embryo medium. In each comparative assay, embryos injected with the two reporter mRNAs up for comparison were mounted in the same glass bottom petri dish.

The fluorescent responses were analysed and photographed with a Zeiss Axiovert 200M widefield fluorescence microscope (ZCA522000494), using a 10x magnification. Each embryo was analysed under two filter sets: a Rhodamine filter set – excitation wavelength (ex.) = 540-552 nm; emission wavelength (em.) > 590 nm – and a GFP filter set – ex. = 450-490 nm; em. > 515 nm. All images were retrieved using a 100ms exposure time. Each embryo was analysed and photographed at 25, 28 and 33hpf. During this 8 hour period of analysis the embryos were kept at 28ºC.

Representative embryos from all the comparative assays were also photographed after 33hpf using a 5x magnification, and in the polyadenylation control experiments an additional set of representative embryos was photographed at 15hpf with a 10x magnification.

III.6. Image processing

Images were processed with ImageJ 1.44p [57]. The average pixel intensities of eGFP and mCherry were measured for each embryo in the same circular section of the embryos' trunk adjacent to the yolk extension.

For each GFP reporter mRNA and each time point, the mean normalized GFP intensities were calculated, according to the following equations.

$$\text{Intensity(I)} = \text{I section} - \text{I background} \qquad \text{(a)}$$

$$\text{Normalized intensity } (I_n) = \text{I (GFP)} / \text{I (mCherry)} \qquad \text{(b)}$$

$$\text{mean In} = (I_{n1} + I_{n2} + ... + I_{n10}) / 10 \qquad \text{(c)}$$

The average background intensity (I background) was measured in an area adjacent to the embryo, for each channel and each time point, using a circular section identical to the one referred previously. The average intensities (I) were calculated by subtracting the background intensity from the total intensity (measured in the embryos' trunk) (equation(a)).

For each embryo and each time point, the average intensity calculated for the mCherry channel was used to normalize the average intensity calculated for the GFP channel (equation(b)).

For each GFP reporter mRNA the mean value of the normalized GFP intensities of the ten embryos considered, was calculated according to equation (c), where I_{n1}, I_{n2} and I_{n10} represent the normalized intensity (I_n) values obtained for each embryo.

The mean I_n values obtained with each RNA reporter sample for each time point were plotted as a

function of time.

In each comparative assay, the fold changes of GFP reporter expression were calculated, for each time point, as the ratio of the mean I_n values obtained with the two reporter mRNAs up for comparison, according to the following equations:

Fold change in intensity= mean I_n (GFP-SV40) / mean I_n (GFP-3'UTR-SV40) (d)

Fold change in intensity= mean I_n (tail) / mean I_n (SV40) (e)

Equation (d) was used in the first set of experiments and equation (e) was used in the polyadenylation control experiments.

T student tests were conducted to assess the statistical significance of the differences observed, both in the mean I_n values measured at 25hpf and in the variation of the fluorescent responses over time (decreasing vs. increasing). In the first set of experiments, a p-value < 0.02 was obtained for all the differences observed between mean I_n values and between the slopes of the curves, with the exception of the difference in mean I_n measured at 25hpf for the eGFP-SV40 vs. eGFP-*fgf8aS*-SV40 assay. In the set of control experiments a p-value <0.04 was obtained for all the differences observed between mean I_n values, however, the only statistically significant differences in the fluorescent responses over time pertain to the eGFP-SV40 vs. eGFP-tail assay (p-value<0.02).

III.7. Bioinformatic identification of 3'UTR sequence elements

All the alignments were done with Clustal X 2.0.12 [58] using the default gap opening and gap extension penalties. All the EST sequences used in this analysis were retrieved from the Unigene database, and their reference numbers are indicated in the alignments.

Polyadenylation signal predictions were conducted using the polyAH [59] and the polyApred [60] online tools, with the default parameters.

The identification of short sequence motifs was done with Improbizer [61], stipulating 2 as the maximum number of motif occurrences per sequence, 6 as the initial motif size and 6 as the number of motifs to find. The option of ignoring motif location was selected. The default settings were adopted for the remaining parameters. The MotifMatcher [61] software was used to complement this search, and once again the option of ignoring motif location was selected and the maximum number of occurrences per sequence was stipulated as 2. The reverse complement search option was not selected.

The search for protein and microRNA target sites was preformed using the online tools SplicingRainbow [62], MicroCosm [63] and MicroInspector [64]. The searches were preformed using

the default parameters of each software. Note, that the MicroInspector analysis was initially conducted on an earlier version of the software. The current version of MicroInspector employs more stringent Gibbs free energy requirements. Consequently, while the results presented in Bold letters (**Figure 19**) refer to miR target sites predicted using the current versions of MicroCosm and/or MicoInspector, the remaining targets were predicted using the earlier version of MicoInspector with the default Gibbs energy cut off value (-20kcal/mol) and using the current version of MicoInspector with a lower cut off value (-16kcal/mol). For more information, and a brief description of each individual software, see section IV.2

The housekeeping control genes, considered in the motif searches and target site predictions were the elongation factor 1-alpha (ef1a), beta-2-microglobulin (b2m) and beta-Actin-1. References for all the annotated 3'UTR sequences used are listed bellow.

*hes7*_human (NM_032580.3)
*hes7*_chipanzee (XM_001167471.1)
*hes7*_orangutan (XM_002826999.1)
*hes7*_macaque (XM_001118139.2)
*hes7*_horse (XM_001504807.1)
*hes7*_mouse (NM_033041.4)
*hairy1*_chick (NM_204472.2)
*hes1*_mouse (OTTMUST00000089350)
*deltaC*_zebrafish (OTTDART00000025961)
*hey1*_zebrafish (OTTDART00000026812)
*her12*_zebrafish (OTTDART00000036935)
*her15*_zebrafish (OTTDART00000025078)
*her11*_zebrafish (ENSDART00000026907)
*her6*_zebrafish (OTTDART00000034578)
*her13.2*_zebrafish (OTTDART00000024892)

*Fgf8*_mouse(NM_001166361.1);
*FGF8*_chick (ENSGALT00000033214)
*fgf24*_zebrafish(ENSDART00000054877)
*fgf24*_zebrafish (NM_182871.3)

Housekeeping controls:
*ef1a*_zebrafish (ENSDART00000023156)
*b2m*_zebrafish (ENSDART00000075127)
*β-actin1*_zebrafish(ENSDART00000054987)

*β-globin*_frog – [65]

IV Results and Discussion

IV.1 Establishment of a reporter system to study the 3'UTRs of *her1*, *her7* and *fgf8a*

To determine the effect that the 3'UTRs, of the cyclic *her1* and *her7* genes and the wavefront gene *fgf8a*, exert on the expression of their respective mRNAs we established an *in vivo* fluorescent reporter system, based on similar previously described approaches [66] [67] [68] [66].The implementation of this system will also enable future studies regarding the identification of specific post-transcriptional regulatory mechanisms involved in the observed effects.

This reporter system relies on fluorescence measurements, which do not directly reflect the mRNA levels but rather the levels of mature fluorescent protein present in the embryo. Therefore, considering that in this study we are interested in characterizing the modulation of mRNA stability and/or translation efficiency by the 3UTRs, it first becomes necessary to understand how changes in these two parameters translate into changes in the levels of mature fluorescent protein. Consequently, we started by designing a mathematical model to describe our experimental system and a series of simulations were carried out to characterize it.

IV.1.1. Mathematical modelling of the *in vivo* fluorescent reporter system

Our experimental approach involves the *in vitro* production of eGFP (enhanced Green Fluorescent Protein) reporter mRNAs with the 3'UTRs of interest, their microinjection into one-cell stage zebrafish embryos and the comparison of the fluorescent responses obtained with and without each 3'UTR. In this mathematical model we considered that the injected mRNA is translated and the resulting protein needs to undergo a post-translational modification termed chromophore maturation, before fluorescence can be detected [55] Therefore we have three reactants in the embryo - mRNA, immature protein and mature protein – and their individual differential rate equations were written as depicted in **Figure 5**.

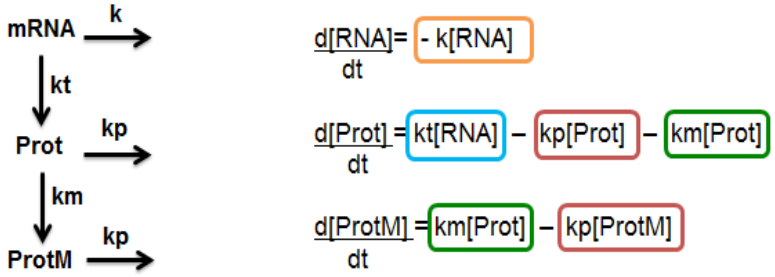


Figure 5 - **Schematic representation of the experimental system and corresponding differential rate equations.** d[N]/dt represents the variation in the levels of species N in the system as a function of time. The terms highlighted in orange, blue, red and green correspond to the contributions of mRNA degradation, translation, protein degradation and maturation respectively, to the levels of the species in the system at the given time. The rate constants of mRNA degradation, translation, maturation and protein degradation are depicted by k, kt, km and kp, respectively.

This model assumes that the embryo behaves like a single cellular entity, with all the considered rate constants being applicable to every cell within the embryo at every stage of development. All the mRNA degradation, translation, maturation and protein degradation events taking place are assumed to display first-order rate kinetics. Additionally, the degradation of the mature and immature proteins is assumed to proceed with the same efficiency (same rate constant). The values adopted for each constant were adapted from the literature (see materials and methods).

In order to translate the kinetic parameters of the model, namely the differential rate equations and the values adopted for each rate constant, into a graphical representation of the predicted levels of mature fluorescent protein in the system as a function of time, we used PLAS (Power Law Analysis and Simulation), an interactive software that supports the dynamical and steady-state analysis of systems of differential equations. The resulting curve is depicted in red in **Figures 6 A** and **B**.

a) Individual impact of mRNA degradation and translation rates in protein levels

In order to assess how the output our model would change in response to changes in mRNA degradation and translation efficiency and whether or not we could use this experimental system to distinguish changes in mRNA stability from changes in translation efficiency, a series of simulations was carried out. In these simulations the rate constants governing mRNA stability and translation efficiency were individually changed by regular amounts, thus producing the curves seen in **Figure 6 A** and **B**, respectively. In each case the arrow indicates the direction of the increase in the respective rate constant. Panels **C** and **D** highlight the predicted behaviour of the system presented in Panels **A** and **B**, respectively, between 25 and 33h. This eight hour interval corresponds to the time frame of analysis during which the in vivo experiments were carried out.

The simulations described were tested using different estimates (five fold higher and five fold lower) of the maturation and protein degradation rate constants and no significant alterations to the overall dynamics described here were detected, from which we conclude that the qualitative behaviour of the system is not conditioned by the values adopted for these constants (see appendix).

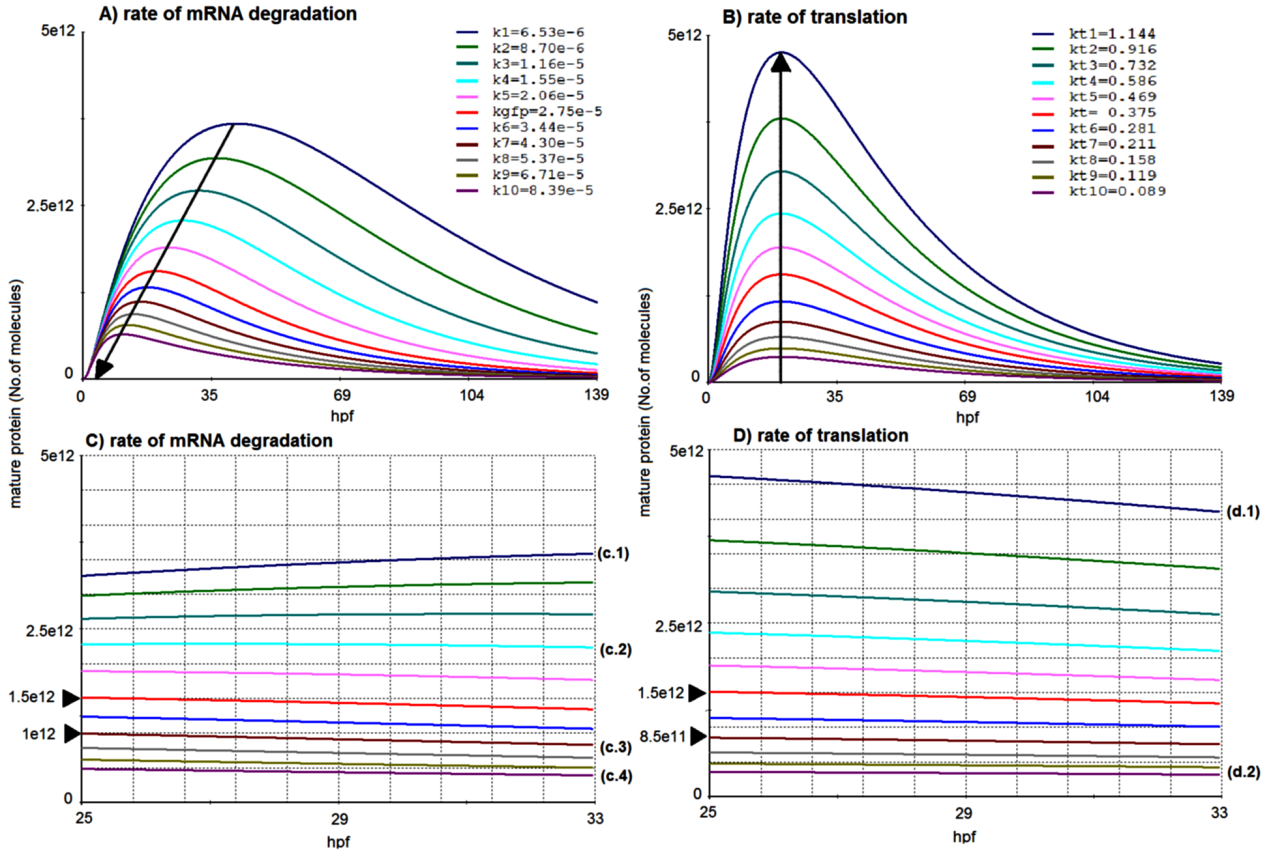


Figure 6 - Impact of the rates of mRNA degradation and translation on reporter protein levels.

A - Effect of varying mRNA degradation rates. Predicted levels of eGFP protein computed using the rate constants adapted from the literature correspond to the red curve. The reported value for the rate of mRNA degradation was altered by gradual 25% increments and decrements to produce the remaining curves. The arrow indicates an increase in the rate constant. The individual rate constant values are indicated in the legend. B - Effect of varying mRNA translation rates. The red curve represents the 'baseline' levels of eGFP protein as before. Values for the rate of mRNA translation were altered by gradual 25% increments and decrements to produce the remaining curves. The arrow indicates the increase in the rate constant and the individual rate constant values are indicated in the legend. C and D - Predicted behaviour of eGFP protein levels in the experimental time frame. Panels C and D provide a focus into the interval between 25h and 33h presented in A and B, respectively. This period corresponds to the time frame of analysis considered in the *in vivo* expression assays. Note that in this time frame, the different curves obtained with different rates of mRNA degradation (panel C) are in different phases of their temporal evolution: (c.1) – ascending phase; (c.2) – peak phase; (c.3) – descending phase; (c.4) – final phase. Conversely, all the curves obtained with different rates of translation (panel D) are in the descending phase of their evolution, with those bearing higher translation rates (d.1) displaying a greater slope than those bearing lower rates of translation (d.2).

As expected, the results presented in **Figure 6** predict that an increase in the rate of mRNA degradation would lead to a decrease in the overall levels of mature eGFP whereas an increase in the rate of translation would produce an increase in the overall levels of mature eGFP.

These predictions indicate that changes in the stability of the mRNA or in the efficiency of translation, have a variable impact on the levels of mature fluorescent protein produced. However, as

a point of reference, an increase in the mRNA degradation rate of approximately 50% leads to a 33% decrease in the levels of mature protein present in the system at 25hpf (**Figure 6 C** - compare red and brown (c.3) curves). A similar decrease (43%) in mature protein levels was observed when we decreased the translation rate by 50% (**Figure 6 D** - compare red and brown curves). Therefore, this experimental system does not appear to allow a distinction between alterations in mRNA stability and alterations in translation efficiency based on the magnitude of the concomitant change in mature protein levels alone.

More interestingly, as shown in **Figure 6 A**, we find that the shape of the curve is altered as a result of changes to the rate of mRNA degradation. Specifically, an increase in the rate of mRNA degradation produces a shift in the peak of mature protein levels toward earlier time points. Strikingly, this shift is not observed when only the rate of translation is altered (**Figure 6 B**). As you can see in **Figure 6 C** the presence of this shift is evident during the time frame of experimental analysis (25 – 33 hpf).

In conclusion, these predictions indicate that, in the context of our experimental system, changes in the rates of mRNA degradation and translation may be primarily distinguishable due to the different impacts they exert on the shape of the resulting mature protein curve.

b) Cooperative effect of altered mRNA degradation and translation rates in protein levels

The simulations presented above focused on the individual impacts that changes in mRNA stability and translation efficiency exert on the system. However, as previously mentioned, the role of 3'UTR sequences in post-transcriptional regulation events isn't always restricted to either a modulation of mRNA stability or an alteration of translation efficiency, as it may simultaneously affect both.

Consequently we proceeded to the study of a scenario in which a 3'UTR contains regulatory elements involved in modulating translation efficiency as well as regulatory elements involved in modulating mRNA stability. In this situation, we can consider two distinct cases:

- a combination of concordant effects – if the regulatory elements associated with stability and translation efficiency individually promote similar effects in the output protein levels (i.e. an increase or a decrease).
- a combination of discordant effects – if the regulatory elements associated with stability and translation efficiency individually promote antagonizing effects in the output protein levels (i.e. an increase and a decrease).

In the first case we expect that the combination of the two concordant effects will enhance the observed change in the output protein levels, whereas in the second case the observed changes in the output protein levels will depend on the relative magnitude of each discordant effect.

Therefore, two additional simulations were done to study the system in the presence of both effects.

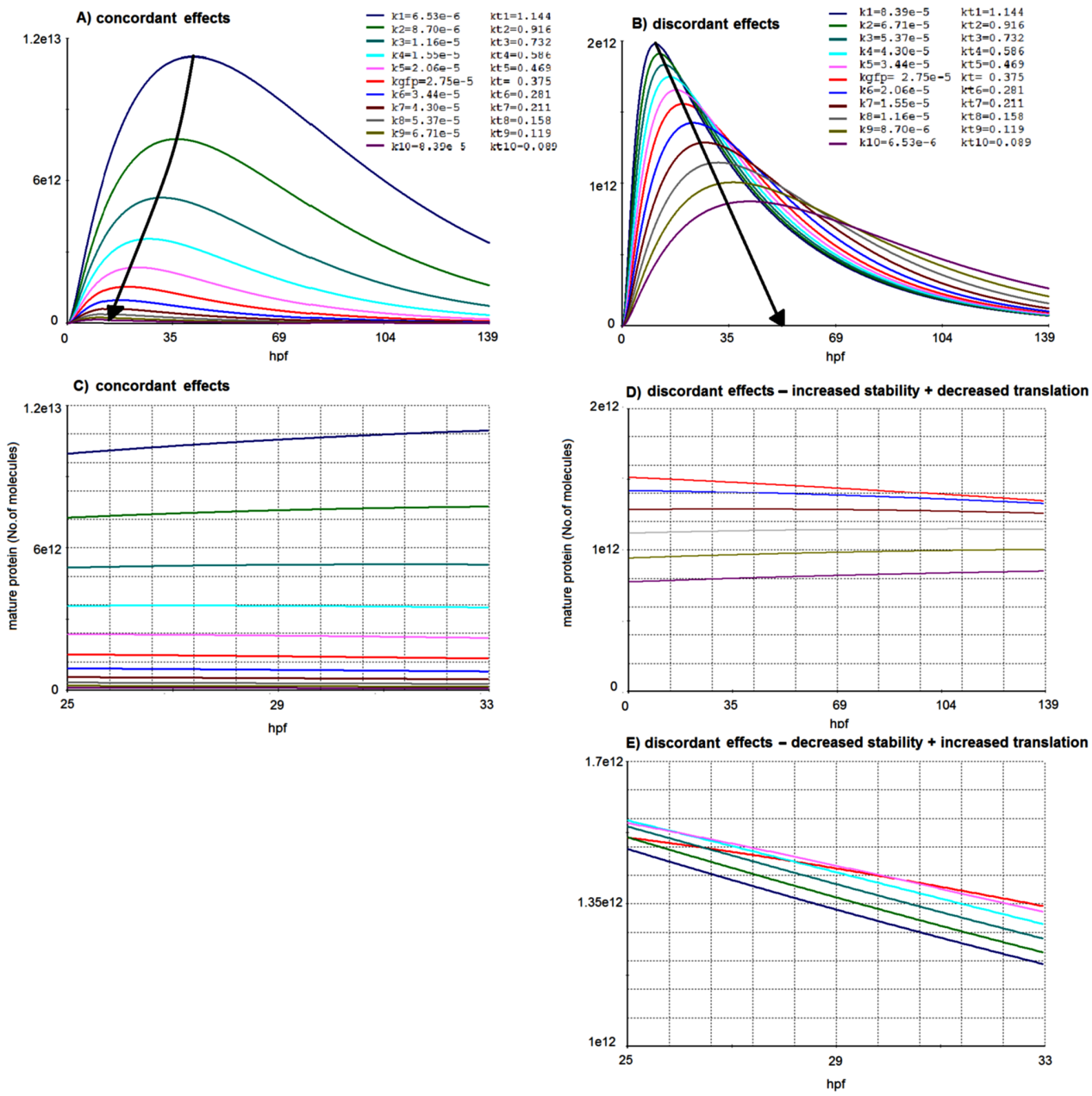


Figure 7 - **Impact of concordant and discordant effects on reporter protein levels.**

A - Effect of concordant variations in mRNA stability and translation efficiency. Predicted levels of eGFP protein computed using the rate constants adapted from the literature correspond to the red curve. The arrow indicates a decrease in the rate of translation and an increase in the rate of mRNA degradation (decreased translation and decreased stability) and the individual rate constant values are indicated in the legend. B - Effect of discordant variations in mRNA stability and translation efficiency. The red curve represents 'baseline' levels of eGFP protein as before. The arrow denotes a decrease in both the rate of translation and the rate of mRNA degradation (decreased translation and increased stability) and the individual rate constant values are indicated in the legend. C, D and E - Predicted behaviour of eGFP protein levels in the experimental time frame. Panels C, D and E provide a focus into the interval between 25h and 33h. C corresponds to the 25-33h interval of the curves presented in A, while D and E correspond to the 25-33h interval of the curves presented in B. D illustrates a decrease in the rates of translation and mRNA degradation (decreased translation and increased stability) and E illustrates an increase in both rates (increased translation and decreased stability) in relation to their original values -red curve.

As shown in **Figure 7 A**, in the presence of concordant effects the output of the system is similar to the one previously obtained by uniquely varying mRNA degradation rates (**Figure 6 A**). As expected in this case the decrease in mature protein levels induced by the increase in the rate of mRNA degradation is heightened by the concomitant decrease in the rate of translation. In contrast, the results obtained in the presence of discordant effects (**Figure 7 B**) reveal that a coupled increase in the rates of translation and mRNA degradation leads to an initial increase and subsequent rapid decline in the levels of mature protein.

In both cases (concordant and discordant), a shift in the peak of mature protein levels toward earlier time points associated with an increase in the rate of mRNA degradation is present (**Figures 7 A**, **B** and **6 A**). Therefore, we conclude that this shift is associated with mRNA instability even in the presence of altered rates of translation.

In the presence of concordant effects, this shift is once again evident in the time frame of experimental interest, with the curves obtained using higher rates of translation and lower rates of mRNA degradation being in the earlier ascending phases or in the peak of their temporal evolution and the curves obtained using lower translation rates and higher mRNA degradation rates being in the later descending phase or in the final phase of their temporal evolution (**Figure 7 C**). In this context, the curves depicted in **Figure 7 C** are relatively similar to the ones depicted in **Figure 6 C**, where only mRNA stability was altered.

Within this time frame we find that the presence of discordant effects (**Figures 7 E** and **D**) does not induce a considerable variation in the levels of mature protein, when compared to the remaining scenarios. However these effects potentiate significant variations in the slope of the curves. Elevated rates of degradation and translation are associated with steeper declines (**Figure 7 E**) and reduced rates of degradation and translation are associated with reduced slopes (**Figure 7 D**). In this context, the curves depicted in **Figure 7 D** are relatively similar to the ones depicted in **Figure 6 D**, where only the translation efficiency was altered.

From these results we conclude that the presence of concordant effects has a greater impact on the levels of mature protein produced than the presence of discordant effects, with the latter scenario affecting primarily the shape of the curve in the time frame of interest.

The fact that a shift in the peak of mature protein levels toward earlier time points was identified in all the scenarios in which the rate of mRNA degradation was increased, but not in the scenario in which only the translation efficiency was changed, leads to the conclusion that this shift is specifically potentiated by a decrease in mRNA stability. This effect can be attributed to the fact that less stable mRNAs disappear faster from the system thus preventing further protein production. The absence of a shift concomitant with alterations in translation efficiency ties in with the fact that mRNA is present in the system, for the same length of time, in all the scenarios in which only the translation efficiency is changed (**Figure 6 B**). Indeed, the only factor that is modulated in the latter scenarios is the ability of the system to use the mRNA to produce protein, and therefore an alteration in the translation efficiency it is not expected to alter the shape of the curve, but instead the levels of accumulated

protein.

In the context of the experimental system we aim to implement in the present study, the association of this shift to a negative regulation of mRNA stability is predicted to determine a different behaviour of stable and unstable mRNAs in the time frame of experimental data acquisition. This phenomenon may therefore assist in the interpretation of our results, more specifically, it may permit the identification of 3'UTRs involved in the modulation of mRNA stability. However, the information provided by this shift appears to be insufficient to distinguish regulatory effects determined exclusively by alterations to the mRNA stability from concordant regulatory effects, affecting both mRNA stability and translation efficiency (compare **Figures 6 C** and **7 C**). Additionally, a mild decrease in translation efficiency may be indistinguishable from a decrease in translation efficiency coupled to an increase in stability (compare **Figures 6 D** and 7 **D**).

In conclusion, based on the results obtained with this model, the experimental approach we proposed to establish seems to be appropriate for the assessment of the effects the 3'UTRs of the cyclic genes and *fgf8a* exert on the expression of their endogenous mRNAs. Additionally, the characterization of these effects can be aided by the information gathered from these simulations. However, any distinction made between changes in mRNA stability and translation efficiency based on the results obtained in these simulations will require experimental validation.

IV.1.2. Determination of the influence of the 3'UTRs of *her1, her7 and fgf8a* on the reporter system

To determine the effects that the 3'UTRs of *her1*, *her7* and *fgf8a* exert on the expression of their corresponding genes, we established an *in vivo* fluorescent reporter system.

As previously referred, this experimental approach involves the *in vitro* production of eGFP reporter mRNAs coupled to the 3'UTRs of interest and subsequent microinjection of the mRNAs into one-cell stage zebrafish embryos, using as a control of injection, a red fluorescent protein (mCherry). In this study, the fluorescent response of each 3'UTR–bearing reporter was examined at 25, 29 and 33 hours after injection and compared to that obtained with the eGFP mRNA without a 3'UTR. This comparison allowed the determination of the influence that the 3'UTRs exert on the expression of the fluorescent reporter.

A requirement for mRNA stability and translation *in vivo* is the presence of a poly(A) tail. Therefore, this experimental approach requires that the eGFP reporter mRNAs produced *in vitro* either undergo *in vitro* polyadenylation or contain a functional polyadenylation signal, like the Simian Virus 40 (SV40) late polyadenylation signal. The second alternative, is the most widely used [67] [68] [66], and was consequently adopted in this protocol. In order to determine whether or not the method of polyadenylation would influence the expression of the individual reporter mRNAs a second set of expression assays was carried out using *in vitro* polyadenylated mRNAs.

a) *Production of eGFP reporter mRNAs in vitro*

The first step in the production of the eGFP-3'UTR reporter mRNAs involved in the identification and analysis of the annotated 3'UTR sequences. This analysis included the alignment of the available EST (Expressed Sequence Tag) data with the corresponding annotated sequences.

As a result of these alignments, the existence of three alternative *fgf8a* 3'UTRs was uncovered **(see section IV.2.1 a)**). The ESTs supporting each alternative 3'UTR displayed a poly(A) tail, from which we concluded that the *fgf8a* alternative 3'UTRs are the result of alternative polyadenylation events. Additionally, the ESTs supporting these 3'UTRs were identified in cDNA libraries obtained from early stages of development, from which we concluded that these alternative 3'UTRs may bare relevance to the development of the embryo during the segmentation period. The smaller alternative UTR is 726 nucleotides long and was termed *fgf8aS*. The intermediate sized UTR is only 71 nucleotides longer than *fgf8aS* and was termed *fgf8aM*. The larger UTR, with a total of 1041 nucleotides is considerably larger than *fgf8aM* and was termed *fgf8aL*.

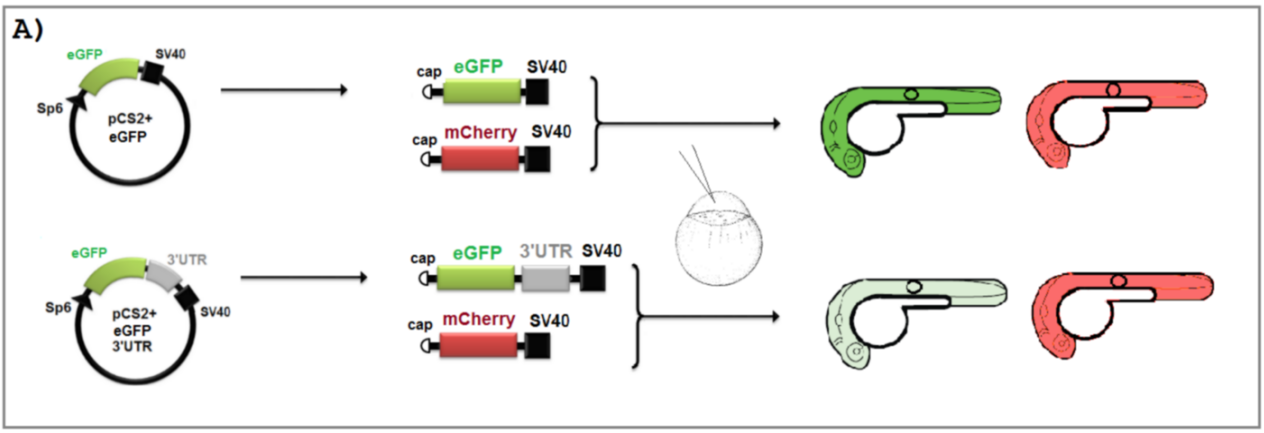
The three *fgf8a* 3'UTRs and the 3'UTRs of *her1* (722 nucleotides in length) and *her7* (365 nucleotides in length) were amplified by PCR from a sample of genomic DNA and subsequently cloned into a PCS2+ expression vector, containing the coding sequence of eGFP. The PCS2+ vector also includes the sequence of the SV40 late polyadenylation signal. To produce the eGFP and eGFP-

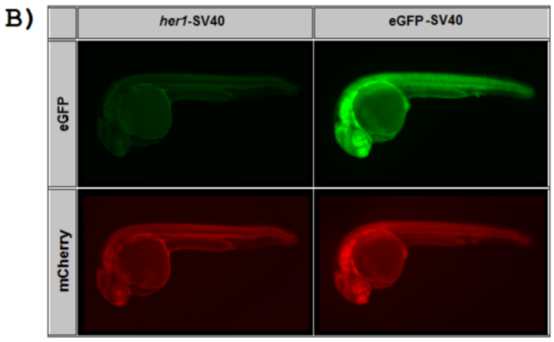
3'UTR reporter mRNAs, the DNA constructs were linearized, transcribed and capped *in vitro* (**Figure 8 A**). Two sets of reporter mRNAs were produced, one set, bearing the SV40 polyadenylation signal and a second set without the signal which was latter polyadenylated *in vitro*.

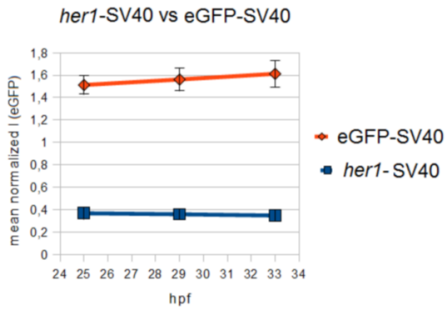
b) *In vivo expression assays*

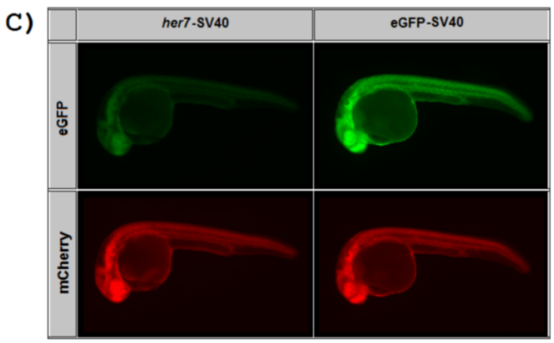
In order to determine if the presence of each 3'UTR influences eGFP mRNA stability and/or translatability, the eGFP reporter mRNAs with and without the 3'UTRs, bearing the SV40 polyadenylation signal, were microinjected into one-cell stage wild type zebrafish embryos. Each reporter mRNA was co-injected with mCherry-SV40 mRNA and the fluorescent responses, normalized to the mCherry signal, were analysed at 25, 29 and 33 hpf. (**Figure 8 A**)The mean normalized eGFP fluorescence intensity values obtained with each eGFP reporter mRNA, in each time point, are presented in **Figures 8 B** to **F.**

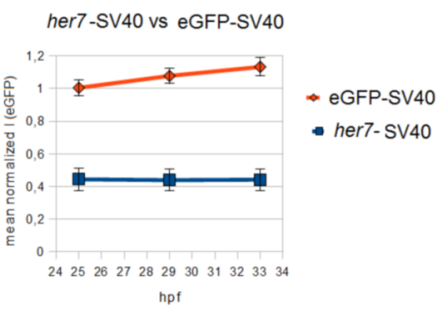
Figure 8 G summarizes the fold differences between the mean normalized fluorescence intensity values observed for eGFP-SV40 and those observed for each eGFP-3'UTR-SV40 reporter.



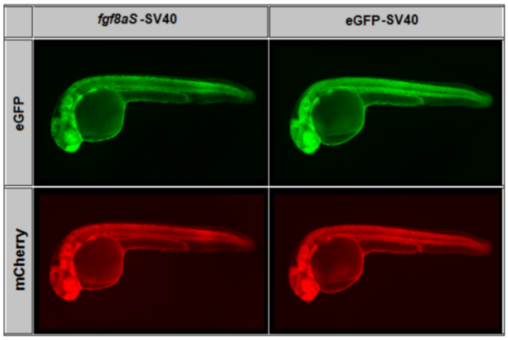


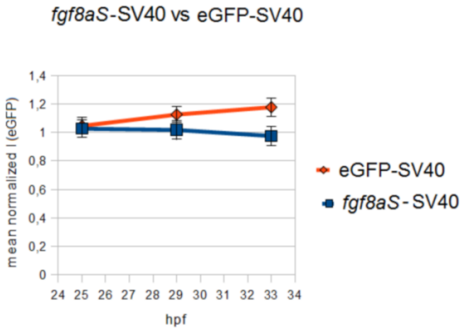




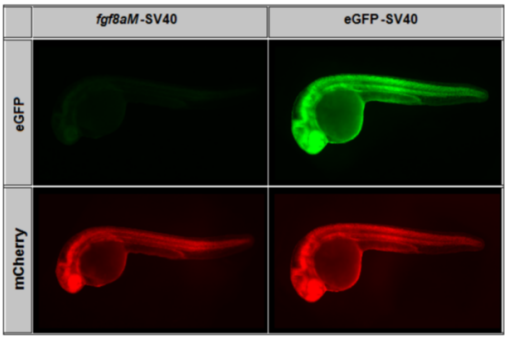


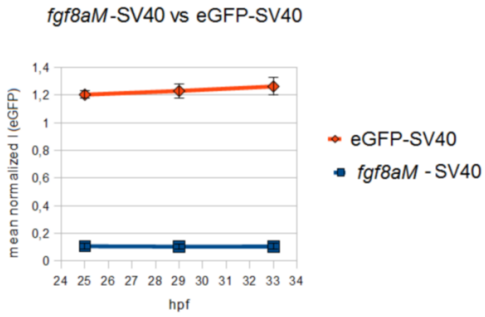
D)



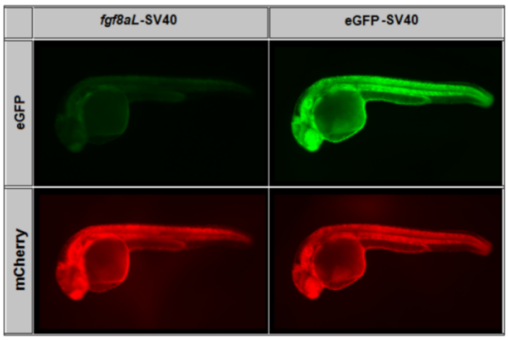


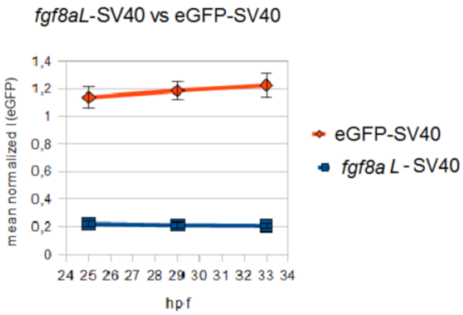
E)





F)





G)

	I (GFP) : I (GFP-UTR)
her1	4
her7	2
fgf8aS	1
fgf8aM	11
fgf8aL	5

Figure 8 – Influence of the 3'UTRs of *her1*, *her7* and *fgf8a* on the expression of an eGFP reporter.

A - Schematic representation of the experimental approach. The PCS2-eGFP vectors with and without a 3'UTR were used to produce eGFP reporter mRNAs bearing the SV40 polyadenylation signal. These mRNAs were co-injected into one-cell stage zebrafish embryos with mCherry-SV40 mRNA. eGFP fluorescence intensity measurements were conducted on 20 embryos from each assay (10 injected with the eGFP-SV40 reporter and 10 injected with the eGFP-3'UTR-SV40 reporter), in a circular area of the embryos' trunk adjacent to the yolk extension. These values were normalized to the mean mCherry fluorescence intensity values measured in the same area of the embryo. B to F – Comparison between the mean normalized eGFP fluorescence intensities obtained with each of the eGFP-3'UTR-SV40 mRNAs and those obtained with the eGFP-SV40 mRNA. The mean normalized eGFP intensities are plotted as a function of time and two representative embryos from each comparative assay are displayed. The error bars indicate ± SD and 10 embryos were considered per reporter sample. G - Fold differences in the fluorescence responses obtained with the eGFP-SV40 reporter and with each of the eGFP-

3'UTR-SV40 reporters, calculated as the ratio between the mean normalized eGFP fluorescence intensity obtained with eGFP-SV40 and that obtained with each eGFP-3'UTR-SV40 construct.

As you can see in **Figure 8**, with the exception of the eGFP reporter mRNA containing the *fgf8aS* 3'UTR (eGFP-*fgf8aS*-SV40), all of the eGFP-3'UTR-SV40 mRNAs studied gave rise to lower mean values of normalized eGFP fluorescence intensity than the control eGFP-SV40 mRNA. The eGFP reporter mRNA containing the 3'UTR of *her1* (eGFP-*her1*-SV40) gave rise to 4 fold lower normalised fluorescence intensity values than the eGFP-SV40 mRNA, while the reporter mRNA containing the 3'UTR of *her7* (eGFP-*her7*-SV40) lead to a value only 2 fold lower than that of the control. The eGFP-*fgf8aS*-SV40 mRNA did not produce a fluorescent response considerably different from that of eGFP-SV40, however, the reporter mRNA bearing the *fgf8aM* 3'UTR (eGFP-*fgf8aM*-SV40), gave rise to a 11 fold lower signal than eGFP-SV40. Finally, the reporter mRNA containing the *fgf8aL* 3'UTR (eGFP-*fgf8aL*-SV40) produced a 5 fold lower signal than the control.

Form these results we conclude that all the 3'UTRs tested have a negative impact on the expression of the reporter, with the exception of the *fgf8aS* 3'UTR, which mediates only a mild decrease in the fluorescence intensity levels over time. Out of the three *fgf8a* alternative 3'UTRs, *fgf8aM* mediates the greatest decrease in the reporters' expression with the *fgf8aL* 3'UTR also affecting the reporters' expression negatively, but to a lesser extent. In addition, the negative impact exerted by the *her1* 3'UTR on the reporters' expression is greater than the impact exerted by the *her7* 3'UTR.

In all five comparative assays, the difference between the signal obtained with the eGFP-3'UTR-SV40 mRNA and that obtained with the eGFP-SV40 mRNA increased over the course of the 8 hour time frame of analysis. This increasing difference results from the fact that the normalized intensity values obtained with the eGFP-SV40 mRNA increase over time while the normalized intensity values obtained with the eGFP-3'UTR-SV40 mRNAs either remain constant or decrease over time.(**Figure 8 B** to **F**)

In light of our mathematical predictions (section **IV.1.1**), the increase in the normalized signal obtained with the eGFP-SV40 reporter is consistent with either the observation of the corresponding curve during the late ascending phase, the peak or the early descending phase of its' temporal evolution. Conversely, the decreasing values or reduced constant values of normalized fluorescence intensity obtained with the eGFP-3'UTR-SV40 reporters are consistent with the observation of the corresponding curves during a late descending phase of their temporal evolution. This behaviour is therefore indicative of a shift in the peak of mature protein levels toward an earlier time point concomitant with the presence of any one of the 3'UTRs tested. As previously stated (section **IV.1.1**), this shift was only identified in the presence of a negative modulation of transcript stability. Therefore, we deduce that all of the 3'UTRs tested may act as negative modulators of mRNA stability.

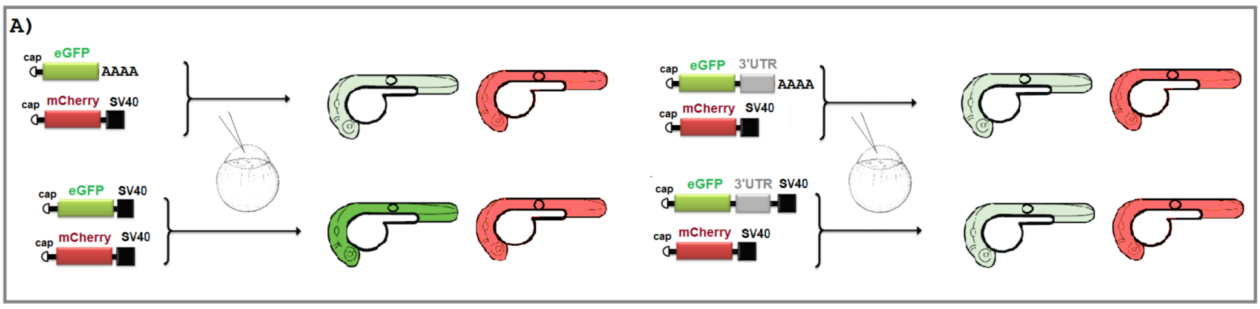
However, in the previous simulations this shift was not only identified in the scenario where only changes to the mRNA stability were tested, but also in the scenarios where concomitant changes to

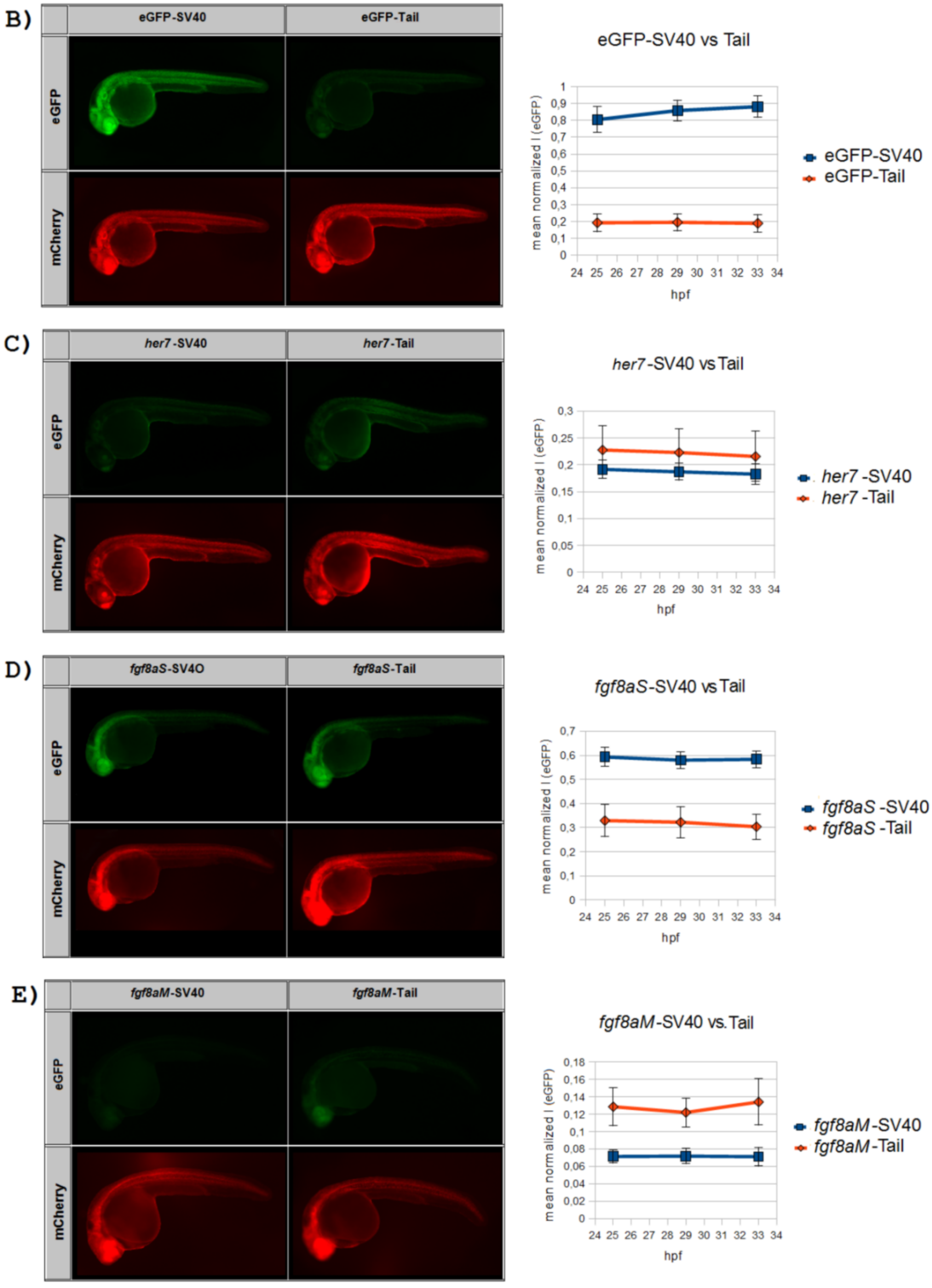
the transcripts' translation efficiency were present (**Figures 6 A** and **7 A** and **B**) . Therefore, we can not, at this time, rule out the possibility that any one of these 3'UTRs could also affect the translatability of the eGFP mRNA. This possibility could explain the results obtained with the *fgf8aS* 3'UTR. These results are consistent with a scenario in which the *fgf8aS* 3'UTR has a negative effect on the stability of the reporter mRNA, which would potentiate the decrease in the fluorescent response over time, and a concomitant positive effect on the translation efficiency of the reporter, which would account for the negligible difference observed between the mean normalized fluorescence intensity values obtained with eGFP-*fgf8aS*-SV40 and with eGFP-SV40. Note that, the simulations depicted in **Figure 7 E** in which this effect was tested on the mathematical model of our experimental system, display similar dynamics to the ones observed in our experimental results.

In summation, with the exception of *fgf8aS*, all of the 3'UTRs tested appear to have a negative impact on the expression of the reporter mRNA, which is probably caused by a decrease in the reporters' mRNA stability. The *fgf8aS* 3'UTR is likely to modulate both an increase in the reporters' translation efficiency and a decrease in its' mRNA stability.

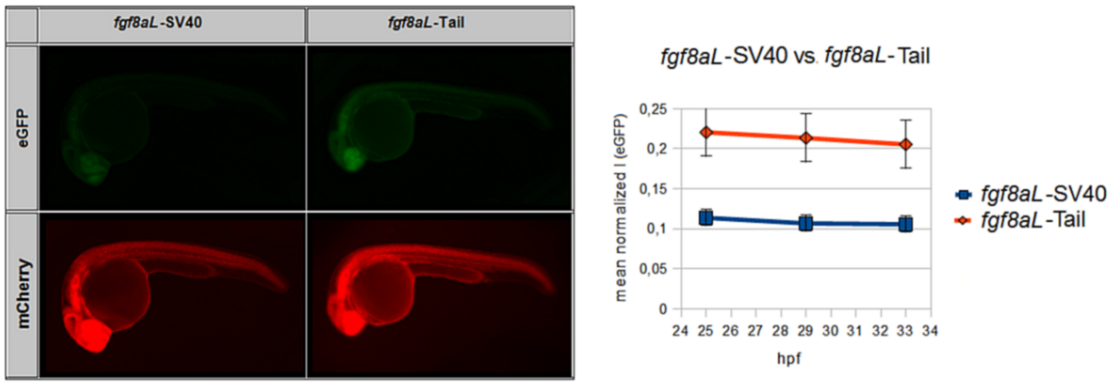
In order to assess if the observed effects are influenced by the SV40 polyadenylation method selected, a second set of *in vivo* expression assays was carried out. In these experiments the fluorescent response obtained with each SV40-bearing reporter RNA was compared to that obtained with the corresponding *in vitro* polyadenylated reporter mRNA (lacking the SV40 signal). The microinjection, data acquisition and data processing protocols described for the first set of experiments were also employed here. The results obtained are presented in **Figures 9 B** to **F**.

In addition, in order to assess whether the changes in fluorescence intensity observed in the previous expression studies and in the current experiments were also visible in the embryo at an earlier stage of development, two representative embryos from each comparative assay carried out in this set of controls were analysed at 15 hpf. These results are shown in **Figure 10**.


A)
cap
eGFP
AAAA
cap
mCherry
SV40
cap
eGFP
SV40
cap
mCherry
SV40
cap
eGFP
3'UTR
AAAA
cap
mCherry
SV40
cap
eGFP
3'UTR
SV40
cap
mCherry
SV40


B)
eGFP-SV40
eGFP-Tail
eGFP
mCherry
eGFP-SV40 vs Tail
mean normalized I (eGFP)
hpf
eGFP-SV40
eGFP-Tail
C)
her7-SV40
her7-Tail
her7-SV40 vs Tail
her7 -SV40
her7 -Tail
D)
fgf8aS-SV4O
fgf8aS-Tail
fgf8aS-SV40 vs Tail
fgf8aS -SV40
fgf8aS -Tail
E)
fgf8aM-SV40
fgf8aM-Tail
fgf8aM-SV40 vs. Tail
fgf8aM-SV40
fgf8aM-Tail

F)



G)

	I (tail) : I (SV40)
eGFP	0,2 (1/4)
her7	1
fgf8aS	0,5 (1/2)
fgf8aM	2
fgf8aL	2

Figure 9 – **Influence of the polyadenylation method on the expression of an eGFP reporter.**

A - Schematic representation of the experimental approach. The PCS2-eGFP vectors with and without a 3'UTR were used to produce eGFP reporter mRNAs lacking a SV40 polyadenylation signal, which were subsequently polyadenylated *in vitro* and co-injected into one-cell stage zebrafish embryos with mCherry-SV40 mRNA. B to F – Comparison between the mean normalized eGFP fluorescence intensities obtained with each SV40-bearing mRNA reporter and those obtained with their *in vitro* polyadenylated counterparts. The mean normalized eGFP intensities are plotted as a function of time and two representative embryos from each comparative assay are displayed. The error bars indicate ± SD and 10 embryos were considered per reporter sample. G- Fold differences in the fluorescence responses obtained with the *in vitro* polyadenylated reporters and with the SV40-bearing reporters. Calculated as the ratio between the mean normalized eGFP fluorescence intensity obtained with the *in vitro* polyadenylated reporter and that obtained with the SV40-bearing reporter.

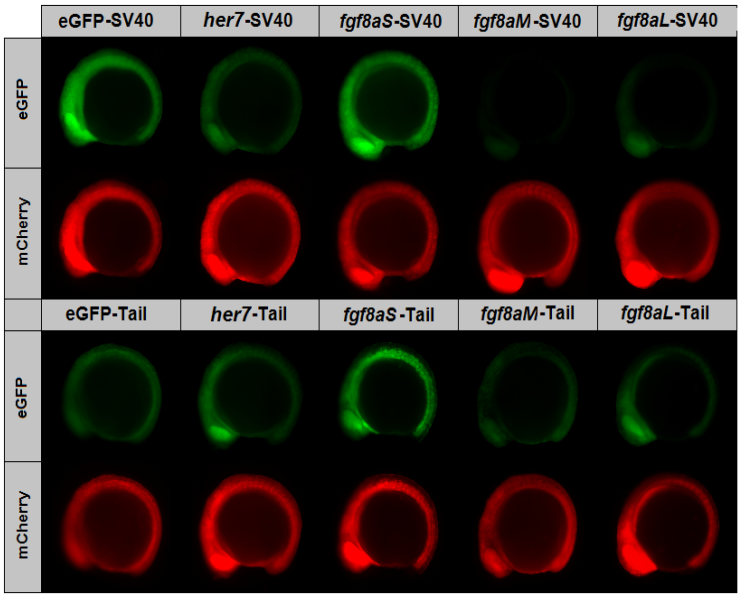


Figure 10 – Fluorescent responses obtained for the reporter mRNAs at 15hpf.

Representative embryos from the second set of expression assays analysed at 15hpf. As observed in the 25-33hpf time frame, with the exception of *fgf8aS*, all the 3'UTRs tested potentiate a decrease in the eGFP-SV40 expression levels at 15hpf. Additionally, out of all the reporter mRNAs tested, the eGFP without a 3'UTR appears to be the most significantly affected by the polyadenylation method at 15hpf, which is once more, consistent with the results obtained at 25-33hpf

As you can see in **Figure 9**, with the exception of the reporter mRNA containing the *her7* 3'UTR, all of the SV40-bearing transcripts tested gave rise to mean values of normalized eGFP fluorescence intensity different from those observed with their *in vitro* polyadenylated counterparts. The eGFP-SV40 mRNA gave rise to a 4 fold higher mean value of fluorescence intensity than the *in vitro* polyadenylated reporter, while the eGFP-*fgf8aS*-SV40 reporter mRNA only potentiated a 2 fold higher signal than the corresponding *in vitro* polyadenylated transcript. Conversely, both the eGFP-*fgf8aM*-SV40 and eGFP-*fgf8aL*-SV40 reporter mRNAs produced 2 fold lower signals than their *in vitro* polyadenylated counterparts, while the eGFP-*her7*-SV40 did not produce a fluorescent response considerably different from that of the *in vitro* polyadenylated transcript.

Additionally, for the *in vitro* polyadenylated eGFP-3'UTR reporters, we observed a considerable variability in the normalized fluorescence intensities of the individual embryos injected with the same reporter mRNA (error bars in **Figures 9 C** to **F**) This increased variability hindered the assessment of the variations in fluorescence intensity over time, associated with these reporter mRNAs and suggests that there is a greater experimental error associated with the *in vitro* polyadenylation method. The only *in vitro* polyadenylated reporter that gave rise to a degree of variability comparable to that of the SV40-bearing transcripts was the in eGFP reporter lacking a 3'UTR (**Figure 9 B**). For this reporter we observed that the normalized fluorescence intensity values remain constant over the 8 hour window of analysis, whereas those obtained with eGFP-SV40, as previously observed, increase over time.

In conclusion, these results indicate that the polyadenylation method influences the expression of all the reporter mRNAs tested, with the possible exception of eGFP-*her7*. Additionally, the magnitude and nature of this influence appears to vary, depending on which 3'UTR is present in the construct. Out of the five reporters tested, the eGFP lacking a 3'UTR is the one associated with the greatest difference in expression in the presence of the SV40 relative to the *in vitro* polyadenylated transcript. The kinetics of these curves indicate that the eGFP-SV40 transcript has an increased mRNA stability in relation to its' polyadenylated counterpart.

In addition, a visual assessment of the fluorescence intensities displayed by the representative embryos analysed at 15hpf (**Figure 10**) indicates that the variations in fluorescence intensity observed in the previous studies and in the current experiments are also visible at this earlier time point, which suggests that the observed effects are not specific to the 25-33hpf developmental stage, being also present in earlier stages of development.

IV.1.3. Discussion and conclusions

Regarding the optimization of the experimental approach, we observed that the polyadenylation strategy employed can influence the expression levels of the eGFP reporter. Specifically, the expression of the eGFP reporter lacking a 3'UTR was greatly increased when an SV40 polyadenylation signal was used as an alternative to *in vitro* polyadenylation. This observation could result from the presence of exogenous elements in the SV40 sequence capable of modulating the reporter post-transcriptionally. However, note that the eGFP-tail RNA was the only construct tested that does not have an untranslated sequence downstream of the eGFP coding sequence. The length of the SV40 signal, after polyadenylation, is 130 nucleotides and consequently this signal provides a spacer between the coding sequence and the poly(A) tail in the case of the eGFP -SV40 construct. No such spacer is present in the eGFP-tail construct.

The importance of a spacer sequence in normal mRNA processing is highlighted in a study carried out by Tanguay, R. L. et al [69]. The authors studied the influence of 3'UTR length on the mRNA stability and translation efficiency of a reporter in chinese hamster ovary cells, and observed that varying the 3'UTR length between 27 and 161 nucleotides had a negligible impact on both translation efficiency and mRNA stability, however, positioning a poly(A) tail close to the stop codon (a 7 nucleotide 3'UTR) greatly reduced translational efficiency *in vivo* (38 fold) and caused a slight decrease in the mRNA half-life (1.4 fold). In addition, *in vitro* studies revealed that this translational repression only occurred in the presence of a poly(A) tail, the same construct without a poly(A) tail exhibited expression levels comparable to those obtained with the constructs bearing longer 3' UTRs. [69]

One theory that could explain this phenomenon involves the formation of a steric block between the poly(A) tail - polyA binding protein complex and the translocating ribosomes approaching the stop codon. This steric block could prevent ribosomes from reaching the termination codon, thus stalling translation and reducing the rate of protein production. Additionally, this block could lead to inefficient translation termination and subsequent production of truncated, non-functional, forms of the protein.

Taking this study into consideration, the difference observed between the fluorescent signals obtained with eGFP-SV40 and those obtained with eGFP-tail, could be due to inefficient translation brought about by the absence of a spacer. However, the kinetics of the curves observed for both constructs (see **Figure 9B**), suggest the presence of an increased mRNA stability associated with the eGFP-SV40 construct. This indicates that, in the zebrafish embryo, additional mechanisms could be involved in disposing of messages without a spacer sequence, or that these disposal mechanisms may have a greater quantitative relevance in the zebrafish than in the Chinese hamster ovary cells.

In this context, the No-Go Decay (NGD) mechanism of mRNA quality control and degradation of aberrant transcripts is particularly relevant. This mechanism was recently discovered in yeast and recognizes mRNAs in which translation elongation has been stalled, for example, as a result of damages to the transcript or to the ribosome. The proteins Dom34 and Hbs1 bind to the transcript

near the stalled ribosome and initiate an endonucleolytic cleavage event near the stall site. This releases the ribosome and generates two mRNA fragments, each with an exposed terminal, thus triggering exosome or exoribonuclease-mediated degradation. [70] [71] It is therefore possible that, in the zebrafish embryo, this mechanism, or an analogous mRNA quality control pathway, recognizes a possible stall in the eGFP-tail mRNA translation as an aberrant phenomena and potentiates the degradation of the transcript.

In conclusion, the difference observed between the fluorescence intensity obtained with eGFP-SV40 and that obtained with eGFP-tail is probably caused by an concordant effect of inefficient translation and enhanced degradation of the eGFP-tail mRNA, potentiated by the absence of a spacer sequence, rather than a specific effect mediated by the SV40 signal. To validate this hypothesis, the experimental approach employed in these studies could be used to compare the signal obtained using the eGFP-SV40 reporter mRNA with that obtained using an *in vitro* polyadenylated eGFP-3'UTR reporter mRNA in which the 3'UTR sequence does not influence the stability or translatability of the transcript. Efforts have already been made to clone an antisense 3'UTRs, which could act as this neutral 3'UTR sequence, and a mutant 3'UTR of the zgc:63829 gene, lacking a miR binding site, kindly provided by Dr. A. Giraldez, will also be tested in this context.

Considering the influence of the polyadenylation method in the fluorescence intensity levels obtained with the 3'UTR-bearing reporter mRNAs (**Figure 9C** to **F**), one technical factor that should be addressed pertains to the variability of the *in vitro* polyadenylation process itself. After an *in vitro* polyadenylation reaction, the poly(A) tail lengths in each pool of polyadenylated messages obey a normal distribution, with the majority of the messages possessing tails of medium length and smaller fractions of mRNAs having either slightly longer, or slightly shorter tails. Knowing that poly(A)tail length can affect a transcripts' expression (introduction), this technical factor could account for the internal variability obtained with the *in vitro* polyadenylated messages. If we consider the potential variability in tail length, not only within each sample, but also between different *in vitro* polyadenylated samples, we could speculate that tail size can also be partially involved in the differences seen between the results obtained for the *in vitro* polyadenylated constructs and those obtained for their SV40-bearing counterparts. Note that all the *in vitro* polyadenylation reactions were done under the same conditions and no considerable variability was detected between the average lengths of the different *in vitro* polyadenylated constructs. However the average tail lengths were determined by their migration in denaturing agarose gels, and a more accurate technique may be required to provide conclusive results on this front. In addition a set of control experiments is required to assess the effect that different tail lengths have on the expression of an eGFP reporter mRNA in the presence of a 3'UTR.

Even though there is no clear explanation at this time, for the differences seen between the results obtained with the eGFP-3'UTR-SV40 reporter mRNAs and the results obtained with the eGFP-3'UTR-tail reporter mRNAs, the magnitude of these differences is relatively small and could simply result from experimental variability associated with the polyadenylation process. Additionally, if we assume

that in the eGFP-SV40 reporter mRNA, the SV40 signal acts mainly as a spacer and has no specific regulatory elements that could influence the stability or translation efficiency of the transcript, the inclusion of an SV40 polyadenylation signal in the *in vitro* transcribed message appears to be the best method of polyadenylation.

In the first set of experiments we concluded that the 3'UTRs of *her1*, *her7* and the *fgf8aL* and *fgf8aM UTRs* have a negative impact on the expression of the reporter. These conclusions were based on results obtained using the SV40 polyadenylation method. In addition, the magnitude of the differences induced by these SV40-bearing 3'UTRs on the fluorescent eGFP signal was greater than the magnitude of the differences observed between the eGFP-3'UTR-SV40 reporters and their *in vitro* polyadenylated counterparts. Consequently, the conclusions drawn from the first set of experimental results appear to be largely unaffected by the variability associated with different polyadenylation strategies. However, in the case of *fgf8aS*, while the lower signal obtained with the *fgf8aS*-tail reporter mRNA, may serve to validate the initial assumption that this UTR acts as negative modulator of mRNA stability, it may contradict the assumption that it also operates as a translational enhancer. Therefore, an unequivocal assessment of the role of the SV40 signal is necessary to verify the predicted effect of this UTR, especially regarding the translation efficiency of the reporter.

The observation that, unlike the *fgf8aS* 3'UTR, the *fgf8aM* 3'UTR modulates a strong repression of the reporters' expression indicates that the 71 nucleotides located between the 3' terminals of the *fgf8aS* and *fgf8aM* 3'UTRs are likely to include regulatory elements that modulate a decrease in mRNA stability and/or an ablation of the translational enhancement predicted to be associated with the *fgf8aS* 3'UTR. Conversely, the smaller impact exerted by the *fgf8aL* 3'UTR on the reporters' expression, relative to that exerted by the *fgf8aM* 3'UTR, suggests that the 244 nucleotides located between the 3'-terminals of *fgf8aM* and *fgf8aL* are likely to be involved in the enhancement of the transcripts' translation efficiency and/or stability. In addition, our results indicate that even though the 3'UTRs of both cyclic genes have a negative effect of transcript expression, the impact exerted by the *her1* 3'UTR on the transcripts expression is more pronounced than the one exerted by the *he7* 3'UTR.

Based on the results obtained with our mathematical model we predicted that all the 3'UTRs tested (*her1*, *her7*, *fgf8aS*, *fgf8aM* and *fgf8aL*) have a negative influence on the stability of the reporter. This prediction can be verified by applying quantitative PCR techniques to total mRNA extracts collected from microinjected embryos at different time points during development, in order to assess the variation of the eGFP mRNA levels over time.

Additional studies that can be carried out to make a more in depth characterization of the effects exerted by these 3'UTRs on the post-transcriptional regulation of their mRNAs include:

- the determination of the developmental stages and embryonic tissues in which each of the alternative *fgf8a* 3'UTRs is expressed;
- a more in depth study of any variations in fluorescent protein levels associated with specific embryonic structures, aimed at identifying any discrete spatial patterns of regulation mediated by these 3'UTRs;

– the study of the variations in fluorescence intensity obtained using the eGFP reporter mRNAs over a larger time frame, aimed not only at providing additional kinetic information to aid the validation or dismissal of the mathematical model created, but also at allowing the identification of any specific regulatory events that could take place during earlier stages of development.

In summary, we established an experimental system which enabled the assessment of the influence the 3'UTRs of the *her1*, *her7* and *fgf8a* have on mRNA expression and will assist future studies regarding the characterization of these effects. We studied this system through mathematical modelling, and experimentally assessed the influence of the polyadenylation strategy employed in the results obtained with this protocol. Additional studies need to be carried out to complete this assessment, however, the results obtained with this system indicate that all of the 3'UTRs tested could have a negative impact on the stability of the reporter mRNA. The possibility that these 3'UTRs could also affect translation efficiency remains viable, especially in the case of the *fgf8aS* 3'UTR, with our initial results suggesting that this 3'UTR could be involved in stimulating the reporters' translation efficiency.

In light of the impact exerted by these 3'UTRs on our reporters' expression, the primary questions raised pertain to the specific mechanism that are involved in this impact. In particular, the identification of the regulatory elements present in these 3'UTRs and the regulatory factors that interact with them to potentiate the observed impact, became the main object of our focus. In this context, a bioinformatic identification of candidate regulatory elements within the 3'UTRs of interest was conducted and is described in section IV.2.

IV.2. Bioinformatic identification of 3'UTR sequence elements

In this section of the study, we conducted a bioinformatic identification of sequence elements in the 3'UTRs of the zebrafish cyclic genes and wavefront gene *fgf8a* that could bare relevance to their post-transcriptional regulation, with a specific interest in identifying:

- cross-species conserved sequence elements in each zebrafish 3'UTR,
- short sequence motifs which could be involved in the regulation of the cyclic zebrafish genes
- known RBP and miR binding motifs that could modulate these genes post-transcriptionaly.

To achieve these objectives, we retrieved the 3'UTR sequences of cyclic genes and *fgf8* from available vertebrate genomes (Materials and Methods) and resorted to standard bioinformatic tools designed to perform sequence alignments and motif searches.

Even though the molecular mechanisms involved in the post-transcriptional regulation of the cyclic genes are largely unknown, one hypothesis suggests that all the Zebrafish cyclic genes can be post-transcriptionaly modulated by the same regulatory platform and that this platform can also be involved in the post-transcriptional regulation of cyclic genes of other vertebrate species. This platform would include a set of regulatory factors (miRs and/or RBPs) that upon binding to regulatory elements in the UTRs of the cyclic genes would potentiate their coordinated post-transcriptional regulation. Therefore, in this study, sequence elements conserved between cyclic genes of different species and/or between Zebrafish cyclic genes were considered especially relevant.

IV.2.1. Cross-species conservation of UTR sequence elements

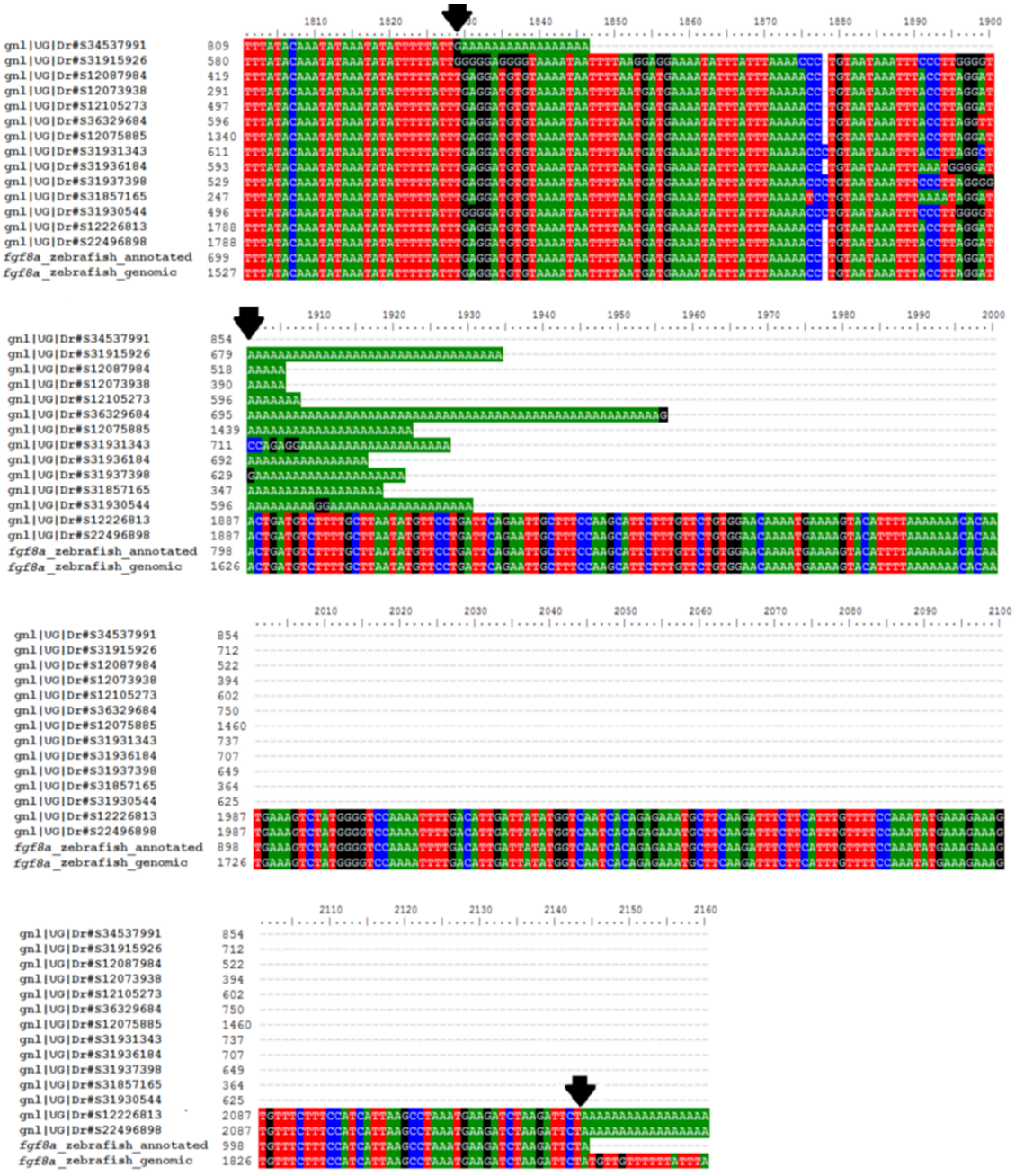
In light of the previous hypothesis and considering that several mechanisms involved in post-transciptional regulation rely on sequence elements conserved between different species [23], we set out to identify elements in the 3'UTRs of the Zebrafish cyclic genes - *her1, her7*- and wavefront gene –*fgf8a* – conserved between the 3'UTRs of ortholog vertebrate genes.

a) The *fgf8* 3'UTR

Analysis of alternative 3'UTRs

In order to identify and characterize the 3'UTR sequences used in this study, we retrieved the annotated 3'UTR sequences and EST data from available databases (materials and methods) and analysed them via multiple alignment, using the ClustalX 2.0.12 software. The results obtained for the 3'UTR of the Zebrafish *fgf8a* gene, in particular, are shown in **Figure 11**.

A



B

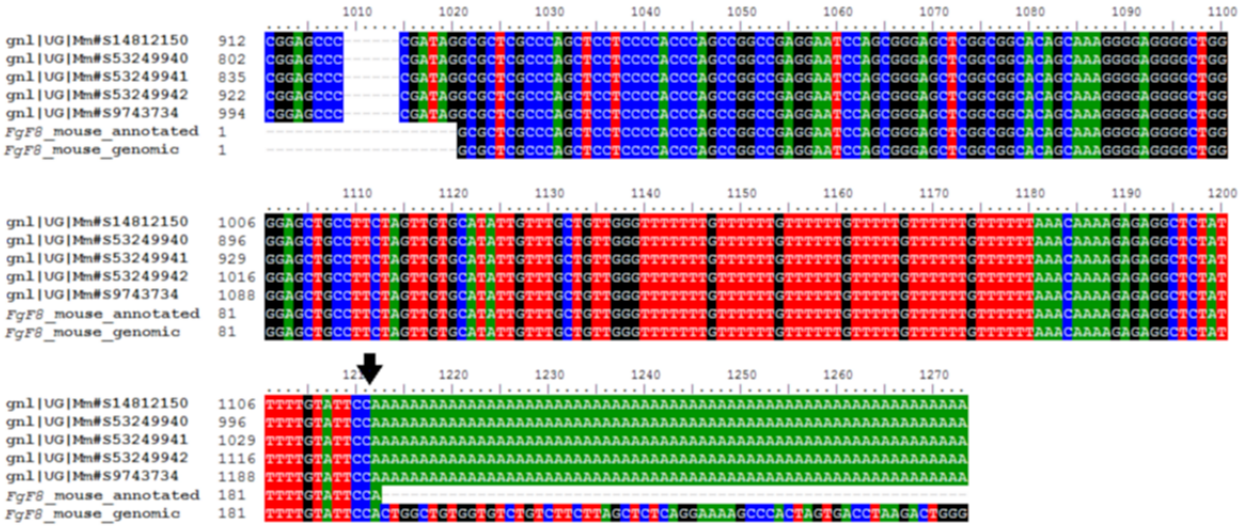
3'UTR	EST sources
fgf8aS	- Myoblast cDNA library (3-9 somite stage)
fgf8aM	- Segmentation stage cDNA library - Myoblast cDNA library (3-9 somite stage) - 26 somite embryos, adult livers, shield stage embryos - Pooled 26 somite embryos - Brain, adult fish 3-4 months old
fgf8aL	-Whole embryo, 2-8hours, mixed wt, cDNA library

Figure 11 – **EST data analysis for the zebrafish *fgf8a* 3'UTRs.** A- Multiple alignment between the annotated zebrafish *fgf8a* 3'UTR, the corresponding genomic sequence and the available EST data. Only the terminal 400 nucleotides of the alignment are shown. Arrows indicate the 3'terminal residues of the three alternative 3'UTRs. B– EST sources, data available at the Unigene database.

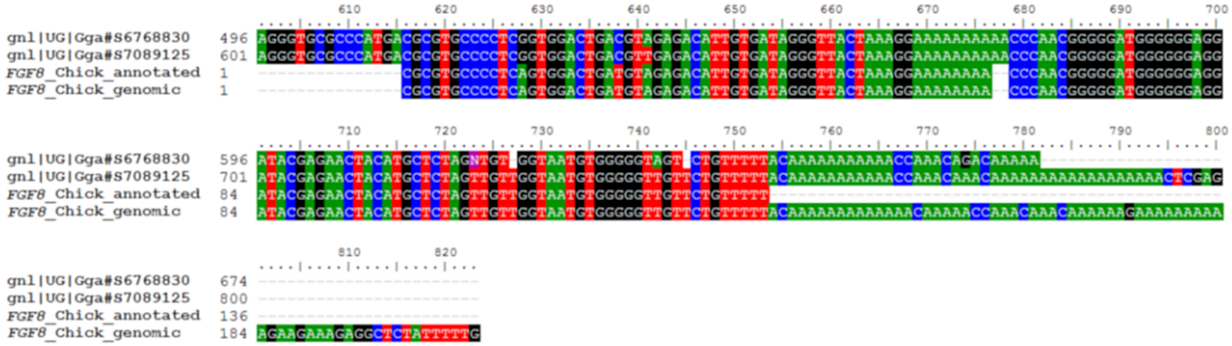
This analysis shows that the annotated *fgf8a* 3'UTR is supported by two ESTs bearing a poly(A) tail which denotes the 3'terminal of the 3'UTR. However eleven additional poly(A)-tail bearing ESTs were identified in support of an alternative 3'UTR, 244 nucleotides smaller than the annotated 3'UTR and one EST was identified in support of a 726 nucleotide long 3'UTR. (**Figure 11 A**) From these results we concluded that the Zebrafish *fgf8a* gene has three alternative 3'UTRs and these 3'UTRs were named *fgf8aL, fgf8aM* and *fgf8aS*, respectively. Additionally, the fore mentioned ESTs were identified in cDNA libraries taken from early stages of development (**Figure 11 B**), from which we conclude that these alternative 3'UTRs may bare relevance to embryonic development during the segmentation period.

The identification of alternative 3'UTRs for this gene raised the possibility that orthologous *FGF8* genes could also undergo alternative polyadenyation events. This possibility was explored through an analogous analysis of the available EST data for *fgf8a* orthologs. We focused our attention on the mouse and chick orthologs, since there is information in the literature regarding the post-transcriptional regulation of these genes (introduction). Additionally, redundant effects were reported, between the zebrafish *fgf8a* and *fgf24* genes in regards to the promotion of posterior PSM formation [21] which suggests that the two genes could be regulated post-transcriptionally by the same mechanisms, and consequently we extended this analysis to include the zebrafish *fgf24* gene. The results obtained are shown in **Figure 12**.

A


gnl|UG|Mm#S14812150
gnl|UG|Mm#S53249940
gnl|UG|Mm#S53249941
gnl|UG|Mm#S53249942
gnl|UG|Mm#S9743734
FgF8_mouse_annotated
FgF8_mouse_genomic

B


gnl|UG|Gga#S6768830
gnl|UG|Gga#S7089125
FGF8_Chick_annotated
FGF8_Chick_genomic

C

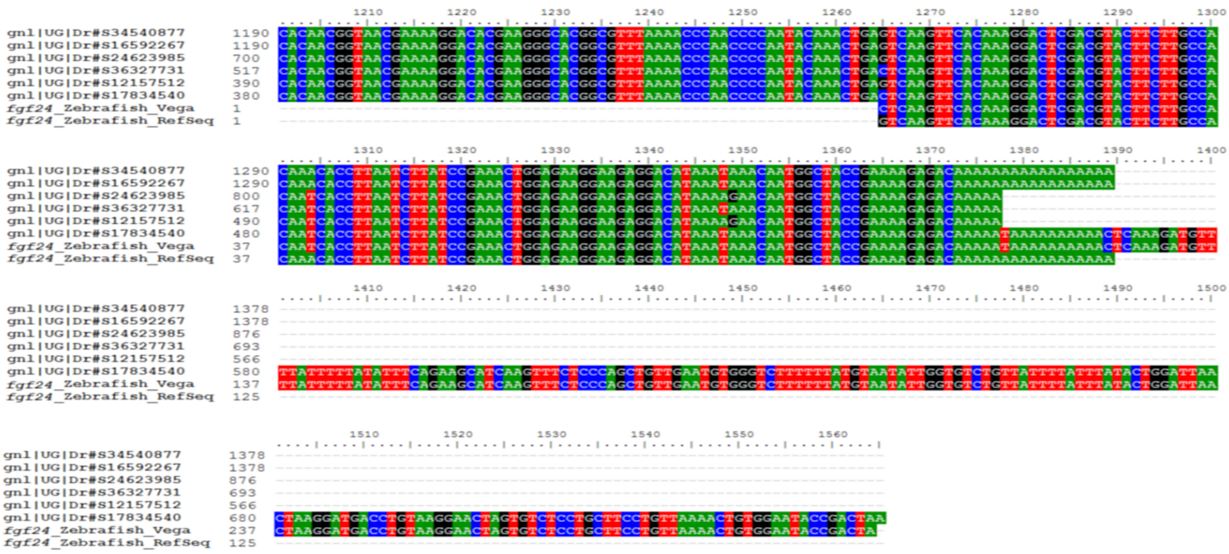

gnl|UG|Dr#S34540877
gnl|UG|Dr#S16592267
gnl|UG|Dr#S24623985
gnl|UG|Dr#S36327731
gnl|UG|Dr#S12157512
gnl|UG|Dr#S17834540
fgf24_Zebrafish_Vega
fgf24_Zebrafish_RefSeq

Figure 12 – EST data analysis for the mouse and chick *Fgf8* 3'UTRs and the Zebrafish *fgf24* 3'UTR. A- Multiple alignment between the annotated mouse *Fgf8* 3'UTR, the corresponding genomic sequence and the available EST data. B - Multiple alignment between the annotated chick *FGF8* 3'UTR, the corresponding genomic sequence and the available EST data C - Multiple alignment between the two annotated zebrafish *fgf24* 3'UTRs, the corresponding genomic sequence and the available EST data.

In regards to the mouse *Fgf8* 3'UTR, as you can see in **Figure 12 A**, the annotated 3'UTR is supported by five ESTs, all of which bearing a poly(A) tail which, once again, denotes the 3'terminal of the 3'UTR (arrow in **Figure 12 A**). However, we detected a considerably greater degree of variability in the ESTs available for the chick *FGF8* 3'UTR and the zebrafish *fgf24*. For the annotated chick 3'UTR, in particular, only two supporting ESTs were available and both display A-rich 3'terminals. However, these A-rich terminals closely resemble the corresponding genomic sequence (**Figure 12 B**). Therefore, these regions could either constitute a poly(A) tail, or be part of the transcribed 3'UTR sequence. A similar scenario was observed for the *fgf24* zebrafish ESTs. As you can see in **Figure 12 C**, five out of the six ESTs available display an A-rich 3'terminal, that once again is mirrored in the genomic sequence. The remaining *fgf24* EST extends further downstream and does not posses an A-rich terminal sequence, thus providing no indication of the 3'UTRs 3'terminal. Therefore, in order to interpret these results, we resorted to our knowledge of the mechanisms underlying the polyadenylation process (introduction), and considered that an annotated 3'UTR sequence is unequivocally supported by EST data only if:

- it has at least one supporting EST with a terminal poly(A) sequence and
- the terminal poly(A) sequence is not present in the corresponding genomic sequence (i.e. is not the product of transcription, but the result of polyadenylation)

Bearing in mind these two criteria, we conclude that, the 3'UTR annotated for the mouse *Fgf8* gene is unequivocally supported by the available EST data, however the 3'UTRs annotated for the chick *FGF8* gene and the zebrafish *fgf24* gene are not.

These results lead us to question the validity of the annotated 3'UTR sequences. In order to address this issue and predict the presence of alternative 3'UTRs in these genes, we used two polyadenylation signal prediction programs – polyApred and POLYAH – to analyse the 3'UTRs and downstream genomic sequences of these three genes. This analysis was manually complemented by searching the sequences for the twelve most frequently occurring polyadenylation signals in eukaryotes [72]

A schematic representation of the results obtained is shown in **Figure 13**. Predicted polyadenylation signals that correspond to the annotated 3'UTRs are depicted in yellow, and alternative polyadenylation signal predictions are depicted in red. The annotated 3'UTR sequences are depicted in black and the alternative 3'UTRs that would result from the use of the alternative polyadenylation signals predicted are depicted in gray. The 3'UTRs unequivocally supported by the available EST data are indicated in this representation by a 3' terminal poly(A) sequence.

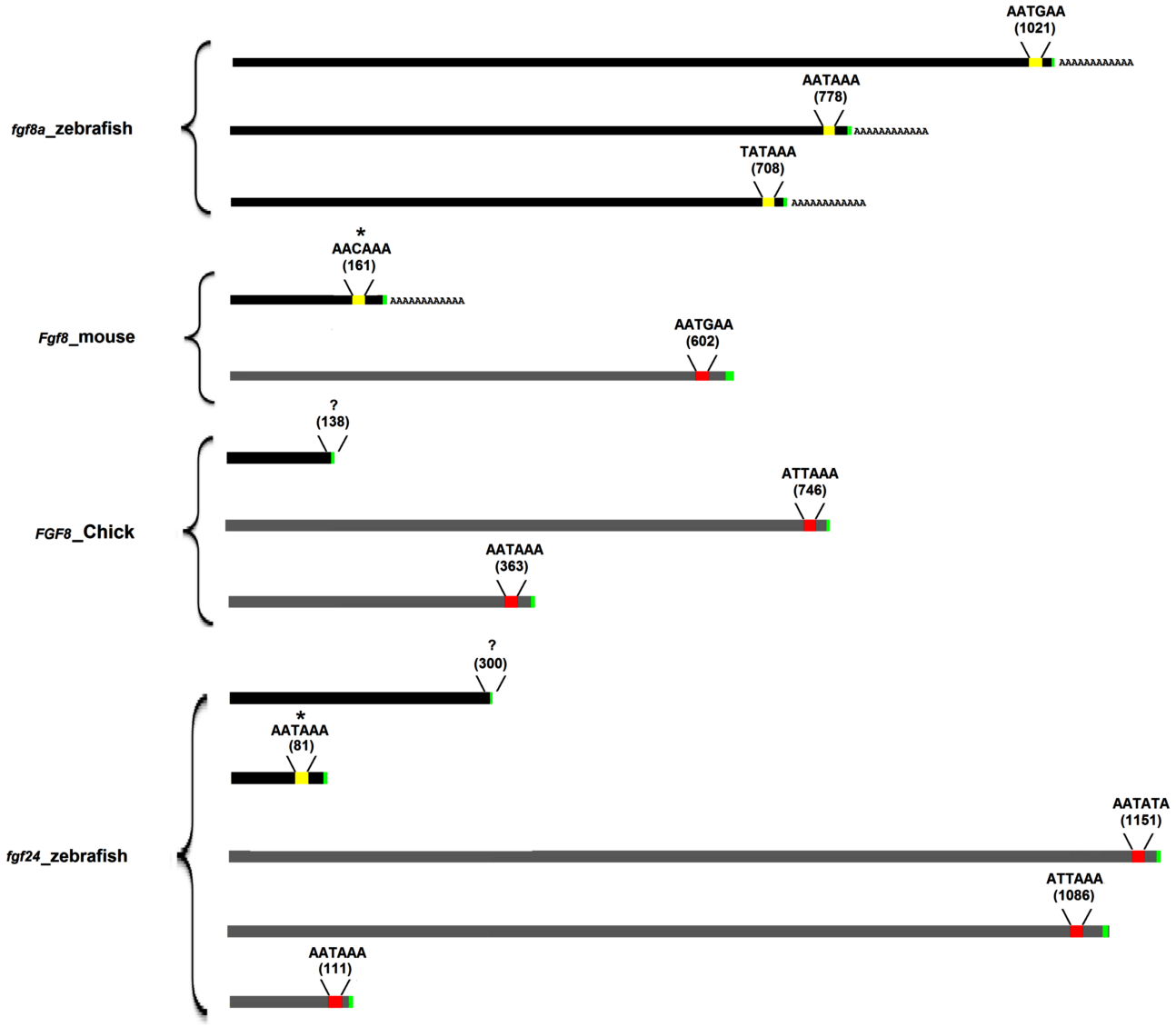


Figure 13 – Schematic representation of the alternative 3'UTRs and polyadenylation signals predicted for the mouse and chick *Fgf8* genes and the zebrafish *fgf24* gene. A schematic representation of the zebrafish *fgf8a* alternative 3'UTRs is also shown. Polyadenylation signals predicted for the annotated 3'UTRs are depicted in yellow and alternative polyadenylation signal predictions are depicted in red. The sequences and relative positions of the polyadenylation signals are indicated (nucleotides were numbered form the 5'terminal of the 3'UTR onward); polyadenylation signals marked with an asterisk were predicted manually. ? marks the absence of an identifiable polyadenylation signal. Green rectangles indicate the polyadenylation sites. Annotated 3'UTR sequences are depicted in black and alternative 3'UTRs that will be produced if the predicted alternative polyadenylation signals are functional, are depicted in grey. The 3'UTRs unequivocally supported by the available EST data are indicated by a 3' terminal poly(A) sequence.

As depicted in **Figure 13**, we predicted two polyadenylation signals for the mouse *Fgf8* gene. One of these signals was identified in the 3'terminal of the annotated 3'UTR while the other was identified 441 nucleotides downstream in the genomic sequence. Considering that the annotated 3'UTR of the mouse *Fgf8* gene was unequivocally supported by the available EST data, from these predictions we conclude that the mouse *Fgf8* may have an additional alternative 3'UTR, resulting from the use of the predicted downstream polyadenylation signal.

n regard to the chick *FGF8*, our alignments did not reveal an unambiguous EST support for the annotated sequence and, as shown in **Figure 13**, no polyadenylation signals were predicted downstream of this annotation. Although the absence of an identifiable polyadenylation signal does not necessarily dismiss the validity of an annotation, it does stress the need to acquire further experimental support for this annotation. Two additional polyadenylation signals were predicted in the chick *FGF8* genomic sequence, respectively 225 and 608 nucleotides downstream of the annotated 3'UTR 3' terminal. From these results we conclude that the chick *FGF8* gene could give rise to at least two alternative 3'UTRs, with the validity of the annotated 3'UTR sequence requiring additional experimental support.

Considering the zebrafish *fgf24* gene, a polyadenylation signal was predicted in the vicinity of the 3' end of one of the annotated 3'UTRs, which could argue for the validity of this annotation. In addition, we predicted three alternative polyadenylation signals in the downstream *fgf24* genomic sequence, one of which located 20 nucleotides apart from one of the annotated 3'UTRs, and the other two located much further downstream (approximately 1000 nucleotides) and in very close proximity to one another (65 nucleotides apart). From this we conclude that further experimental support may be required to validate the annotated 3'UTRs of the *fgf24* gene and that the alternative polyadenylation signals predicted by our study could give rise to alternative 3'UTRs for this gene.

In all three cases, the number of available ESTs was reduced, which raises the possibility that alternative, less abundant 3'UTRs could exist for these genes, having simply avoided detection. Experimental validation, for example via 3'-RACE PCR, is required to assess if any of these signals is functional.

In conclusion, at least two alternative 3'UTRs were predicted for each ortholog *Fgf8* gene and, for the zebrafish *fgf24* four possible alternative polyadenylation events were predicted. Additionally, our results suggest that not only the predicted alternative 3'UTRs but also the majority of the annotated 3'UTRs referred require experimental validation.

Analysis of conserved sequence elements

Next, we sought to identify regions in the *fgf8a* 3'UTR that exhibit a high degree of conservation between orthologous genes as well as sequence elements conserved between the 3'UTRs of the zebrafish *fgf8a* and *fgf24 genes*, which could bare relevance to their post-transcriptional regulation. To accomplish this, we aligned the extended orthologous *Fgf8* 3'UTR sequences, predicted in the previous section, with one another and with the zebrafish *fgf8aL* 3'UTR and we aligned the *fgf8aL* sequence with the extended *fgf24* 3'UTR predicted previously. The results are shown in **Figures 14** and **15** respectively.

These results revealed a high degree of sequence similarity between the *Fgf8* 3'UTRs of the three vertebrate species considered. One highly conserved region in particular is 30 to 35 nucleotides long (**Figure 14** highlighted in black) and includes a 12 nucleotide element which is perfectly conserved in

the three 3'UTRs (**Figure 14**). The alignment between the zebrafish *fgf8a* 3'UTR and the extended *fgf24* 3'UTR also revealed a high degree of sequence similarity between the two 3'UTRs. However, the previously identified 30-35 nucleotide element does not appear to be present in the extended *fgf24* UTR (**Figure 15**). These results suggest that the highly conserved element identified in the 3'UTRs of the zebrafish, chick and mouse *Fgf8 genes* may be involved in an inter-species conserved mechanism of post-transcriptional regulation of these genes, that does not operate on the *fgf24* mRNA.

Additionally, while the mouse and chick extended UTRs align preferentially with the *fgf8aS* and *fgf8aM* zebrafish UTRs, the *fgf24* extended 3'UTR aligns well with the entire zebrafish *fgf8aL* sequence. This observation suggests that there could be additional mechanisms of post-transcriptional regulation specific to the zebrafish *fgf8a* and *fgf24* mRNAs, but absent from the chick and mouse mRNAs, possibly modulated by regulatory elements present in the *fgf8aL* UTR, downstream of the *fgf8aM* 3'terminal.

In summary, we have identified a 12 nucleotide sequence element, highly conserved between the zebrafish *fgf8aS* UTR, the annotated mouse *Fgf8* 3'UTR and the extended chick *FGF8* 3'UTR. The significance of this element is at this time unknown and can be determined in future studies by resorting-for example, to site directed mutagenesis techniques.

Figure 14 – Multiple alignment between the extended mouse *Fgf8*, the extended chick *FGF8* 3'UTR and the zebrafish *fgf8a-L* 3'UTR. Shaded nucleotides denote conservation. The blue triangles denote the 3'terminals of the *fgf8a* 3'UTRs. The green and red triangles denote the 3'terminals of the annotated mouse and chick 3'UTRs, respectively. The green and red stars denote the 3'terminals of the extended mouse and chick 3'UTRs predicted in the previous section, respectively. The black rectangle indicates a highly conserved 30-35 nucleotide element.

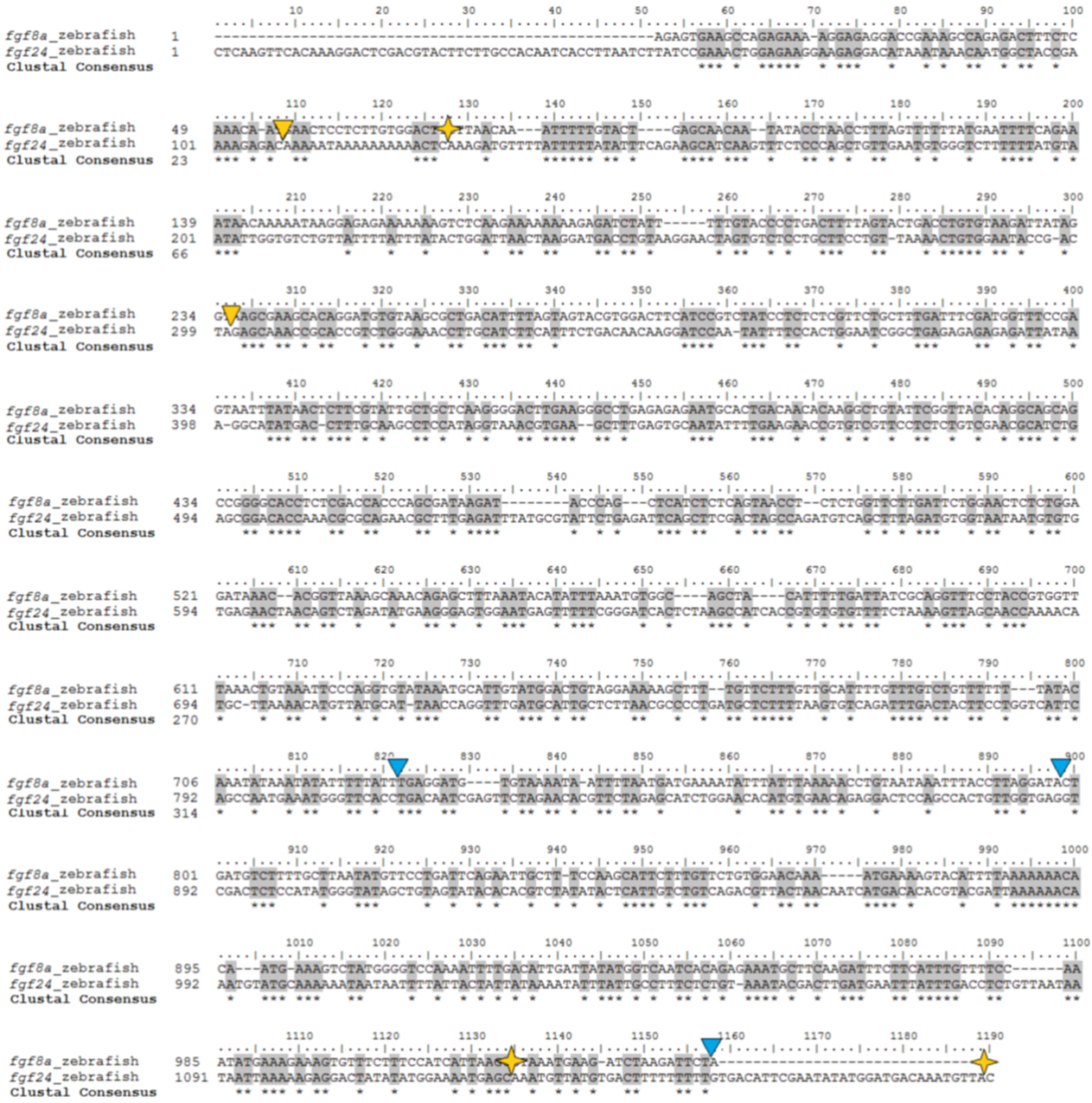


Figure 15 - **Pairwise alignment of the extended zebrafish *fgf24* 3'UTR and *fgf8a-L* 3'UTR.** Shaded nucleotides denote conservation. The blue triangles denote the 3'terminals of the fgf8a 3'UTRs. The yellow triangles denote the 3'terminals of the annotated *fgf24* 3'UTRs, and the yellow stars denote the 3'terminals of the extended *fgf24* 3'UTRs predicted in the previous section.

b) Conserved sequence elements in the cyclic genes

n order to identify cross-species conserved sequence elements with a potential relevance to the post-transcriptional regulation of the Zebrafish *her1* and *her7* genes, we aligned the two zebrafish 3'UTRs with one another and with 3'UTRs of their ortholog - *HES7*- of different mammalian species. The results are shown in **Figures 16** and **17** respectively.

In **Figure 16** we can see that there is a high degree of sequence similarity between the *her7* and

her1 3'UTRs in the aligned region, even though the 3'UTR of *her1* has approximately twice the length of the *her7* 3'UTR. A considerable degree of sequence similarity was also identified between the different *HES7* 3'UTRs (**Figure 17**)**.** However, the *her1* and *her7* 3'UTR sequences are markedly different from the 3'UTR sequences of their *HES7* orthologs, with only a 40-45 nucleotide region of the alignment displaying a moderate degree of homology. (**Figure 17**)

From these results we conclude that the identified 40-45 nucleotide region could constitute an inter-species conserved element**.** However, the overall dissimilarity between the *her1/7* and the *HES7* 3'UTRs suggests that different mechanisms may be involved in the post-transcriptional modulation of these genes in mammals and zebrafish.

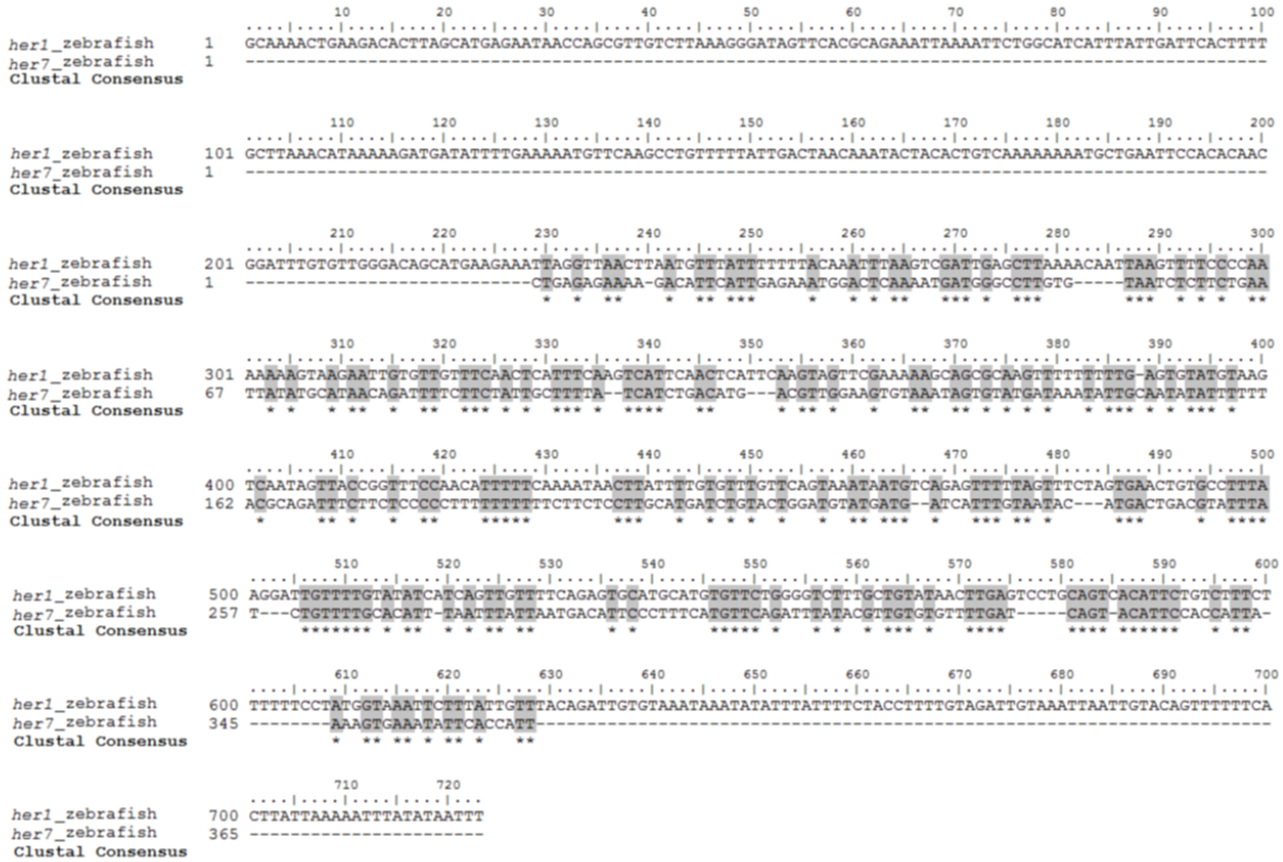


Figure 16 - **Pairwise alignment of the zebrafish *her1* and *her7* 3'UTRs.** Shaded nucleotides denote conservation.

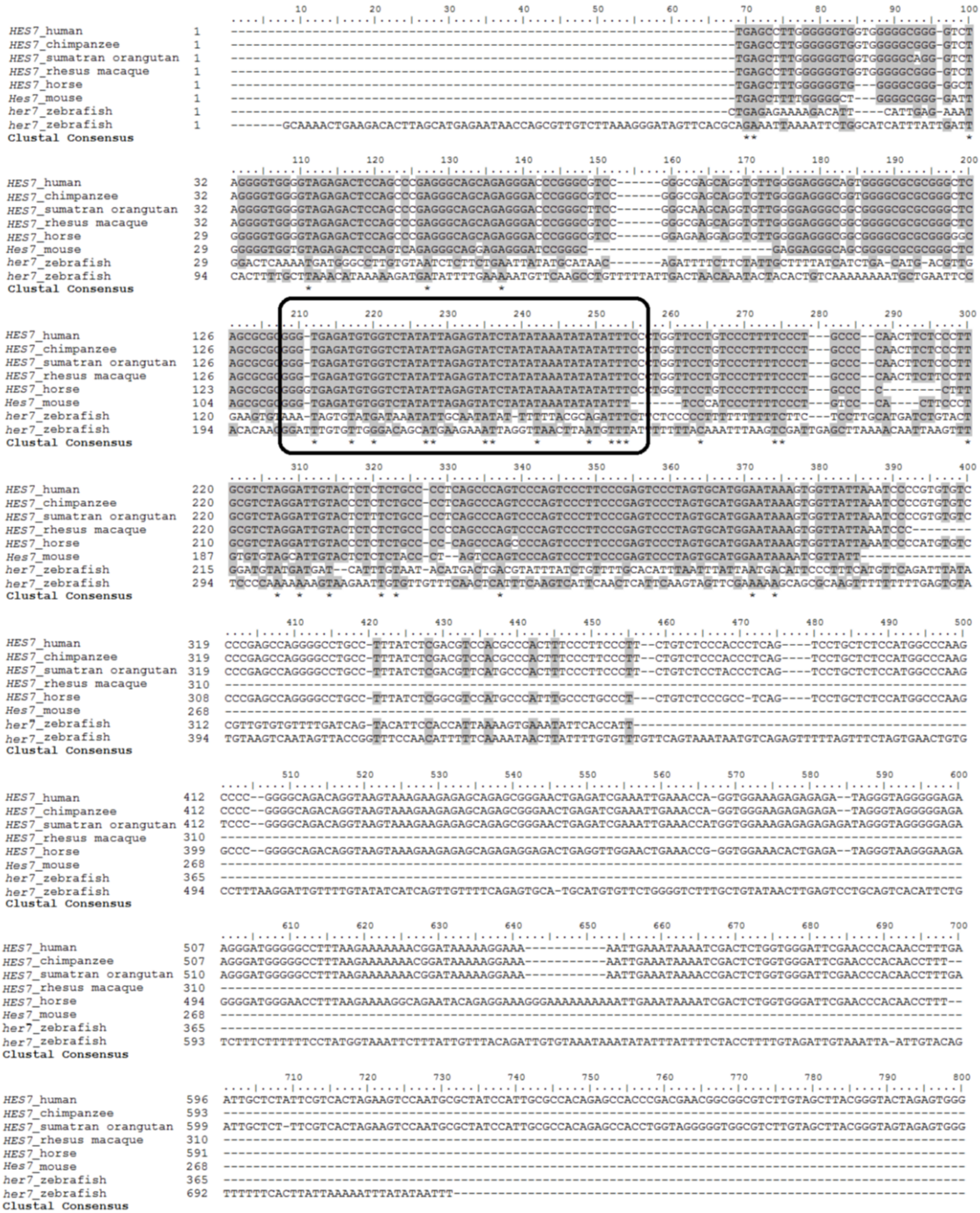


Figure 17 - **Multiple alignment of mammal *HES7* 3'UTRs and zebrafish *her1* and *her7* 3'UTRs**. Black rectangle denotes a mildly conserved region. Shaded nucleotides denote partial conservation.

IV.2.2. Identification of short sequence motifs

In light of the dissimilarity observed between the 3'UTRs of the cyclic *HES7* genes and the zebrafish *her1* and *her7* 3'UTRs, we decided to carry out an unbiased search for short sequence elements – motifs – present in every member of a set of 3'UTRs, in a manner independent of their relative location in the 3'UTR.

To perform a stringent search for motifs with a high probability of being involved in the regulation of cyclic gene expression, two criteria were used:

- We stipulated that the motifs would have to be present in every member of an initial set of sequences containing the 3'UTRs of the zebrafish *her1* and *her7* genes and the 3'UTRs of the cyclic genes Chick_*HAIRY1*, Mouse_*Hes1* and Mouse_*Hes7*. The latter genes were selected, not only because of their phylogenetic connection with the Hairy/[E(spl)]-related family, but also because their cyclic expression in the context of somitogenesis has been reported and studied extensively. [2]
- We stipulated that the motifs would have to be absent from a set of control sequences. This set of controls included the 3'UTRs of three zebrafish housekeeping genes and the 3'UTR of the *Xenopus* beta-globin. The purpose of the housekeeping controls was to account for and discard motifs which were over-represented in the zebrafish 3'UTRs. The *Xenopus* beta-globin 3'UTR has been associated with the production of, stable and translationally efficient mRNA upon injection into zebrafish embryos [73] Consequently, any motifs present in this 3'UTR are, most likely, not involved in mRNA destabilization and/or translational repression. Therefore, considering that the 3'UTRs of *her1* and *her7* have been associated with transcript destabilization, the beta-globin 3'UTR was used as a negative control.

This analysis was conducted using Improbizer a web tool that given an initial set of performs an unbiased search for motifs of a defined size, present in all the provided sequences, that naturally occur with improbable frequency. The motifs identified by Improbizer were then validated using MotifMatcher**,** a web-based tool that given the consensus of a motif determines whether or not the motif is present in a given sequence. MotifMatcher was used to ascertain if the motifs found in the original set of sequences were absent from the control sequences and/or present in the 3'UTRs of other relevant zebrafish cyclic, pseudo-cyclic and non-cyclic genes – *deltaC, hey1, her11, her12, her15, her6 and her13.2* (introduction).

One 8mere motif that satisfied our criteria was found in this analysis, with a consensus sequence of TTTTCCTC (**Figure 18**). This motif was present in the 3'UTRs of all the zebrafish cyclic, pseudo-cyclic and non-cyclic genes considered, and in the 3'UTRs of all the considered cyclic orthologs. Furthermore, the motif occurs twice in each 3'UTR, strongly suggesting that this TC-rich motif could constitute a conserved sequence motif with possible relevance to the post-transcriptional regulation of the cyclic genes. The presence of this motif in the 3'UTRs of the pseudo-cyclic and non-cyclic

Zebrafish *her* genes suggests that any potential regulatory mechanisms mediated by this motif are conserved between *her* – family members.

A)

	T	**T**	**T**	**T**	C	**C**	T	C
a	0.003	0.003	0.003	0.003	0.201	0.003	0.201	0.300
c	0.003	0.003	0.102	0.003	0.596	0.991	0.300	0.398
g	0.003	0.101	0.003	0.003	0.200	0.003	0.003	0.003
t	0.991	0.893	0.892	0.991	0.003	0.003	0.496	0.299

B)

```
                                    10        20
HAIRY1 chick 1          76  -----TTTTCCTA---------- 84
HAIRY1 chick 2          175 -----TTTTACTC---------- 183
Hes1 mouse 1            112 -----TGTTGCTT---------- 120
Hes1 mouse 2            252 -----TTTTACAC---------- 260
Hes7 mouse 1            150 -----TTTTCCCA---------- 158
Hes7 mouse 2            162 -----TTTTCCCT---------- 170
her1 zebrafish 1        290 -----TTTTCCCC---------- 298
her1 zebrafish 2        600 -----TTTTCCTA---------- 608
her7 zebrafish 1        192 -----TTCTCCTT---------- 200
her7 zebrafish 2        260 -----TTTTGCAC---------- 268
deltaC zebrafish 1      636 -----TTTTACTT---------- 644
deltaC zebrafish 2      1049 ----TTTTCCTT---------- 1057
hey1 zebrafish 1        1863 ----TATTCCTC---------- 1871
hey1 zebrafish 2        2173 ----TTTTACAC---------- 2181
her12 zebrafish 1       68  -----ATTTCCAA---------- 76
her12 zebrafish 2       466 -----TTTTACTT---------- 474
her15 zebrafish 1       162 TGTAGTTTTCCTCATGGATGGACATTA 189
her15 zebrafish 2       398 TGTAGTTTTCCTCATGGATGGACATTA 425
her11 zebrafish 1       16  -----TGTTCCCC---------- 24
her11 zebrafish 2       35  -----TGCTCCAA---------- 43
her6 zebrafish 1        80  -----TTGTGCTT---------- 88
her6 zebrafish 2        223 -----TTTTCCAC---------- 231
her13.2 zebrafish 1     289 -----TGTTCCCA---------- 297
her13.2 zebrafish 2     435 -----TTTTACCC---------- 443
```

Figure 18 - **TC-rich motif identified in the 3'UTRs of the zebrafish cyclic, pseudo-cyclic and non-cyclic genes and in the 3'UTRs of the chick_*HAIRY1* and mouse_*Hes1* and *Hes7* genes**. A - Consensus sequence of the motif. The numbers indicate the relative probability of each nucleotide being found in each position of the consensus. The nucleotides highlighted in bold correspond to the most conserved nucleotides in the motif. B – Alignment of the individual motif occurrences identified in the fore mentioned 3'UTRs. Note that the motif occurs twice in each 3'UTR and both occurrences were considered in the alignment. The positions of the motif within each 3'UTR are indicated (nucleotides were numbered from the 5'terminal of the 3'UTR onward). A 27 nucleotide repeat was detected in the *her15* 3'UTR, centred on the TC-rich motif.

n conclusion, a potentially relevant TC-rich motif was identified in the 3'UTRs of all the cyclic, non-cyclic and pseudo-cyclic genes considered. This motif occurs twice in each 3'UTR and its' relevance can be tested in site directed mutagenesis studies.

IV.2.3. Identification of RBP binding motifs

As previously stated (introduction), the two main classes of regulatory factors involved in the modulation of mRNA stability are RBPs and miRs. Consequently, in a more focused approach, we set out to identify known RBP binding motifs in the 3'UTRs of interest, that could be involved in the post-transcriptional regulation of *her1*, *her7*, *deltaC* and *fgf8a*. The web tool used for this purpose, Splicing Rainbow, is associated with a database containing known consensus binding motifs for mammalian splicing factors, including members of the hnRNP class of RBPs. The splicing rainbow algorithm determines whether or not these motifs are present in a user-provided input sequence. The housekeeping controls, used in the previous section, were also used in this analysis to determine the average frequency of occurrence of each motif in the 3'UTRs of zebrafish genes and a negative control was once more provided by the beta-globin 3'UTR.

The results revealed that five out of the thirteen hnRNP binding motifs considered by this software are over-represented in at least two of the 3'UTRs of interest, in relation to the housekeeping control 3'UTRs. These motifs were considered to be the most relevant and their frequencies of occurrence in each 3'UTR are shown in **Table 2 .** Note that no interaction with these five hnRNPs were predicted for the control beta-globin 3'UTR.

Table 2 - hnRNP target sites identified in the 3'UTRs of *her1*, *her7*, *deltaC* and fgf8a. These results were obtained using the SplicingRainbow software and are expressed as the frequency of target site occurrence normalized to UTR length ((n^{er} target sites/UTR length)x100). The values shown for the housekeeping controls (HK) are the mean of the normalized frequencies of occurrence in the three control 3'UTRs. Frequencies highlighted in bold and marked with an asterisk are associated with values ate least 2 fold greater than those associated with the corresponding housekeeping controls. Underlined frequencies indicate the presence of a greater number of target sites in the *fgf8aM* 3'UTR than in the *fgf8aS* 3'UTR.

	her1	*her7*	*delta C*	*fgf8a -L*	*fgf8a-M*	*fgf8a-S*	HK control
hnRNP A1	**0,100***	**-**	**0,100***	**0,300***	**0,376***	**0,274***	0,050
hnRNP B1/A2	**2,600***	**2,500***	1,800	1,600	1,631	1,509	1,250
hnRNP C1/C2	**1,000***	0,500	**0,800***	0,600	**0,753***	**0,823***	0,330
hnRNP A0	**0,600***	**0,800***	**0,600***	**0,500***	**0,627***	0,274	0,210
HuR	**2,100***	**1,900***	1,400	1,300	1,255	1,100	0,960

The motifs identified are associated with the hnRNP A1, hnRNP B1/A2, hnRNP C1/C2, hnRNP A0 and HuR RNA-binding proteins. Note that the binding motifs of the last four proteins are relatively similar, more specifically, all four proteins recognize T-rich motifs including ATTTA motifs and/or poly(T) sequences, with hnRNP A0 binding preferentially to ATTTA motifs, hnRNPs B1/A2 and HuR recognizing both ATTTA and poly(T) motifs and hnRNP C1/C2 recognizing poly(T) motifs. The hnRNP

A1 protein, on the other hand, recognizes TAGGN$^{A}/_{T}$ motifs.[7] [62]

Considering the cyclic genes, this analysis revealed that the 3'UTRs of *her1*, *her7* and *delta C* are enriched in the T-rich motifs associated with the four hnRNPs referred, in particular we identified a 3 to 4 fold enrichment in hnRNP A0 binding motifs – ATTTA – in all three 3'UTRs. The 3'UTRs of *her1* and *deltaC*, have an additional 2 to 3 fold enrichment in hnRNP **A1** binding motifs (**Table 2**). Considering that the four proteins referred (hnRNP B1/A2, hnRNP C1/C2, hnRNP A0 and HuR) bind to very similar motifs, these results indicate that any one of these proteins could modulate these genes post-transcriptionally, with the heightened enrichment of these 3'UTRs in ATTTA motifs suggesting an importance of these motifs in the regulatory mechanism. Additionally, we conclude that hnRNP **A1** could be involved in the post-transcriptional regulation of her1 and deltaC. Interestingly, none of the T-rich motifs identified in this analysis coincides with the TC-rich motif identified in **IV.2.2.**

In the *fgf8a* 3'UTRs, binding motifs for hnRNP **A1** appear to be 6 fold over-represented, and one of these motifs is located in the *fgf8aM* (and consequently *fgf8aL* 3'UTR) but not in the *fgf8aS* 3'UTR. An additional motif, 2 to 3 fold enriched in these 3'UTRs, is ATTTA (hnRNP **A0** binding motif). Note that, five ATTTA motifs were predicted in the *fgf8aM* and *fgf8aL* 3'UTRs, while only two were predicted for the *fgf8aS* 3'UTR. This implies that, in the 71 nucleotides located between the 3'terminals of the *fgf8aM* and *fgf8aS* UTRs we find three ATTTA motifs. (**Table 2**) These results suggest that hnRNP A1 may be a good candidate post-transcriptional regulators for the *fgf8a* gene and the enrichment in ATTTA motifs once again suggests that these motifs could play an important part in the post-transcriptional regulation of *fgf8a*. Note that, if the motifs identified (hnRNP A1 binding motifs and T-rich motifs) are involved in the regulation of the fgf8 mRNA, the presence of additional motifs in the 71 nucleotides between the 3'terminals of the *fgf8aS* and *fgf8aM* UTRs could contribute to the different influences these two 3'UTRs exert on mRNA expression. (see section **IV.1.2**) Interestingly, one of the hnRNP C1/C2 binding motifs identified in this 3'UTR (a poly(T) sequence) partially overlaps the inter-species conserved 12 nucleotide sequence element identified in section **IV.2.1 a)**

In summary, these results indicate that hnRNP A1 might be a good candidate post-transcriptional regulator of the cyclic her1 and deltaC genes and the fgf8a gene and the enrichment of both the cyclic genes' 3'UTRs and the fgf8a 3'UTRs in T-rich motifs (especially ATTTA motifs) suggests that hnRNP B1/A2, hnRNP C1/C2, hnRNP A0 and/or HuR could also be involved in the post-transcriptional regulation of these genes.

7 Note that the software also considers that hnRNP B1/A2 can recognize GTTTG and TTGA motifs and that hnRNP C1/C2 can recognize poly(G) motifs.

IV.2.4. Identification of microRNA target sites

In order to gain further insight into a potential contribution of the miR class of regulatory factors to the post-transcriptional regulation of these genes we scanned the 3'UTRs of interest for potential miR target sites. Two web tools were used for this purpose MicroCosm and MicroInspector.

MicroCosm is a database of miR target sites predicted using the miRanda algorithm. This algorithm detects motifs in the 3'UTRs, complementary to the seed sequences of annotated miRs and rewards miR-target sites conserved between species, but does not exclude non-conserved targets from the pool of predictions. MicroInspector uses a similar approach to MicroCosm, in the sense that it scans a user-provided sequence for motifs complementary to the seed sequences of known miRs. These tools were also used to analyse the previously referred set of housekeeping controls and the *Xenopus* beta-globin 3'UTR.

The results obtained are presented in **Figure 19**. In this figure, the columns correspond to the indicated 3'UTRs and in each column the miRs predicted to interact with a target site in the corresponding 3'UTR are listed. Note that different miRs can have very similar seed sequences (for example miRs from the same cluster and paralogous miRs), and consequently, a single target site can be predicted to interact with more than one miR. All the predicted interactions referred in **Figure 19** involve miRs that have been identified in zebrafish, with the exception of the miRs bearing the prefix “xtr-miR-” which were identified in *Xenopus tropicalis*. The predictions depicted in bold correspond to predicted miR-target site interactions exhibiting a lower theoretical Gibbs free energy value, and thus, more thermodynamically favourable predictions. None of the predicted interactions was over-represented in the set of housekeeping controls and none of the miRs predicted to bind to our 3'UTRs of interest was predicted to interact with the beta-globin 3'UTR.

her1	her7	delta C		her11	her12	her15	her6	her13.2	fgf8aS
miR-202	miR-206 (2x)	miR-129 (2x)	miR-34	miR-137	miR-20b	miR-29b	miR-462	miR-18b	miR-1388
miR-184	miR-2190	miR-138	miR-15b	miR-34	miR-152	miR-2189	xtr-miR-191	miR-18c	miR-2185
miR-24	miR-725	miR-27a	let-7c	miR-365	miR-130b	miR-17a	xtr-miR-33a	miR-27b	miR-727*
miR-449	miR-459*	miR-738	miR-19a*	miR-92a (2x)	miR-148	miR-733	miR-216b	miR-27c	miR-22a
miR-142a-3p	miR-737	miR-2190	miR-723	miR-92b	miR-301b		xtr-miR-106	miR-18b*	miR-22b
miR-132	miR-1 (2x)	miR-27d	miR-206	miR-18b*	miR-130c		miR-18b (x2)	miR-202	
xtr-miR-449	miR-19c	miR-459* (3x)	miR-212	miR-367	miR-430a		miR-17a (x2)		
miR-732	miR-363	miR-128	let-7j	miR-181a	miR-430i		miR-9		
miR-338	miR-133b*	let-7d	let-7f	miR-181a*	miR-130a		miR-20a		
miR-214	miR-19d	miR-181a* (2x)	miR-460-5p	miR-126a	miR-301c		miR-363		
miR-10a*	miR-19b	miR-727* (2x)	miR-15a*	miR-126b	miR-301a (2x)		miR-152		
	miR-144	let-7b	miR-30b		miR-17a (2x)		miR-184 (x2)		
	miR-732	miR-27e			miR-365		miR-124		
	miR-92a	miR-456			miR-155		miR-460-3p		
	miR-124	miR-202			miR-459*		miR-187		
	miR-92b	miR-16a			miR-199*		miR-18c		
	miR-19a	miR-222					miR-126		
	miR-30e*						miR-18a (x2)		
							miR-736		

Figure 19- Predicted miR-target site interactions in the 3'UTRs of the cyclic, pseudo-cyclic, non-cyclic and fgf8a genes. The columns correspond to the indicated 3'UTRs and for each 3'UTR the predicted miR-target site interactions are listed. Numbers in Bold refer to miR target sites predicted using the current versions of MicroCosm and/or MicoInspector. The remaining targets were predicted using an earlier version of MicoInspector with the default Gibbs energy cut off value and by the current version of MicoInspector using a lower cut off value. (see materials and methods for more information). The indication (2x) is used when two binding sites are predicted for the respective miR in the same UTR. Green highlights indicate members of the miR 17~92 and miR 17~92-2 clusters. Blue highlights denote members of the miR-1 family (see text).

As you can see in **Figure 19**, the results obtained for the 3'UTRs of the cyclic zebrafish genes – *her1*, *her7* and *deltaC* – did not reveal miR target sites conserved between the three 3'UTRs. This stems from the fact that the 3'UTRs of *her1* and *her7* have no predicted miR target sites in common. The only predicted target sites shared between zebrafish cyclic genes correspond to dre-miR-206, dre-miR-459* and dre-miR-202, with the first two having predicted target sites in the 3'UTRs of *her7* and *deltaC*, and the third having predicted target sites in the 3'UTRs of *her1* and *deltaC* (**Figure 19**). From these predictions we conclude, post-transcriptional regulation of the three cyclic genes is probably not coordinated by the same miR.

Dre-miR-206, in particular, is in a cluster with dre-miR-133b and they both belong to the miR-1

family of miRs. Note that target sites for the three members of this family were predicted in the 3'UTR of *her7*. (see **Figure 19**, cells highlighted in blue) [74] Therefore, the miR-1 family could be involved in the post-transcriptional regulation of *her7* and *deltaC*.

If we factor in the results obtained for the 3'UTRs of the pseudo-cyclic (*her11*, *her12*, *her15*) and non-cyclic genes (*her6* and *her1*3.2) we see that target sites for members of two miR clusters are over-represented in the 3'UTRs of the cyclic, pseudo-cyclic and non-cylic genes. One of these clusters is designated the miR-17~92 cluster and has six members – dre-miRs 17a-1, 19a, 92a-1, 18a, 20a and 19b. The other, which will be referred to as the miR-17~92-2 cluster, has only three members (dre-miRs 17a-2, 92a-2 and 18b).[74] [8] Target sites for miRs belonging to these two clusters were found in the 3'UTRs of *her7*, *deltaC*, all the pseudo-cyclic genes and all the non-cyclic genes (highlighted in green in **Figure 19**). From which we conclude that these two miR 17~92 clusters could be involved in the post-transcriptional regulation of all the considered her genes, with the exception of *her1*.

In regard to FGF8a, the results obtained with MicroCosm and MicroInspector revealed four miR target sites, one of which capable of interacting with either one of the two paralogous miRs -22a and 22b and the other three located in partially overlapping regions of the 3'UTR. However, all of the predicted miR target sites are located within the *fgf8aS* 3'UTR. Therefore we conclude that the negative effects exerted by the *fgf8aM* and *fgf8aL* 3'UTRs on the expression of our reporter (section **IV.1.2**) probably do not result from the presence of additional miR target sites in these two 3'UTRs.

An additional analysis was done, using the MicroInspector software, to assess whether any additional miR target sites could be identified in the 3'UTR of *her1* if the Gibbs free energy requirements were less stringent. Target sites were identified in this 3'UTR for miR-20a* and miR-206, among others. The theoretical Gibbs free energy values of the predicted interactions were -14 and -12kcal/mol respectively. Note that, in the previous results the interactions with the lowest Gibbs free energy values predicted had a Gibbs free energy of -16kcal/mol. Therefore we conclude that even though *her1* could also be regulated post-transcriptionally by the miR-1 family or the miR-17~92 cluster, it is possible that the resulting miR – target site interactions would not be thermodynamically favourable enough to make this mechanism quantitatively relevant.

IIn summary, we identified two sets of candidate miR regulators of the stability and translation efficiency of the cyclic and fgf8a mRNAs. Namely, the miR-1 family (predicted regulators of *her7* and *deltaC*), and the miR 17~92 clusters (predicted to regulate all the considered *her* genes with the possible exception of *her1*)

8 The mature miR sequence of miR-17a-1 is identical to that of miR-17a-2 and the mature miR sequence of miR-92a-1 is identical to that of miR-92a-2. The miRs 18a and 18b are paralogous miRs, and their mature sequences differ only in one nucleotide.

IV.2.5. Discussion and conclusions

In these studies we identified a highly conserved sequence element in the *fgf8a* 3'UTR, and a short sequence motif conserved in the 3'UTRs of all the cyclic, non-cyclic and pseudo-cyclic vertebrate genes considered. The potential involvement of these sequences in the post-transcriptional regulation of the wavefront gene fgf8a and the cyclic genes *her1* and *her7*, will be explored in future experimental studies.

In a more focused approach, we predicted binding sites in the 3'UTRs of all the cyclic genes considered and *fgf8a* for members of the hnRNP family of RBPs and in the 3'UTRs of Delta C and her 7 we identified target sites for members of the miR-1 family and the miR-17~92 and miR-17~92-2 clusters. Target sites for members of the two latter clusters were also identified in the 3'UTRs of the non-cyclic and pseudo-cyclic zebrafish genes considered. However, in a biological system, regulatory factor–regulatory element interactions can only occur if the regulatory factor and its target are co-expressed. In this context, *in situ* hybridisation data retrieved from the Zfin databank revealed that all of the candidate hnRNPs identified in this study - hnRNP A0, hnRNP A1, hnRNP B1/A2, hnRNP C1/C2 and HuR – are ubiquitously expressed in the embryo during the segmentation period [75]. In addition, information gathered from two independent studies revealed that the miR candidate regulators identified in this study - the miR-1 family and the miR-17~92 clusters - are expressed in the zebrafish embryo during somitogenesis (between 12 and 24hpf) [76] [74] One of these studies conducted a series of *in situ* hybridisation experiments which show that the miR-1 family members are expressed in the embryonic body, more specifically in the muscular system, in the head and in the fin musculature, while the miR-17~92 and miR-17~92-2 clusters are expressed ubiquitously [76] These experiments indicate that the candidate regulatory factors identified in this study are spatially and temporally co-expressed with their predicted targets, making them good candidates for future functional studies.

Additionally, even though the hnRNP family is a family of multi-functional proteins, there is a precedent for the involvement of the candidate regulators identified in this study in the modulation of post-transcriptional events [38]. More specifically, studies conducted on a murine macrophage-like cell line suggest that hnRNP A0 binds to AREs located in the 3'UTRs of a subset of inflammatory mediators (including interleukin-6, cyclooxygenase 2 and tumour necrosis factor-alpha) and may contribute to the stabilization of their mRNAs and/or the stimulation of their translation in a lipopolysaccharide-dependent manner [77]. n regard to the hnRNP A1 RBP several studies suggest that this protein may be involved in modulating mRNA stability, with possible targets of hnRNP A1 including the Xenobiotic-Inducible Gene Cyp2a5 [78] . The hnRNP A2/B1 complexes have been implicated the the stabilization of the Collagen Prolyl 4-Hydroxylase mRNA in human fibrosarcoma cells under iron depletion [79] and hnRNP C was shown to promote the stabilization of urokinase receptor (uPAR) mRNA in human lung epithelial cells and enhance the translation of amyloid precursor protein mRNA by displacing a translational repressor in P bodies [80] [81]. Lastly, HuR has

been associated with mRNA stability regulation and its role in these mechanisms is discussed in greater detail in the introduction.

In addition, the miR candidate regulators identified in this study, have been associated with functions in embryonic development.

More specifically, the miR-1 family has been linked to the development and function of skeletal and cardiac muscle and it has been reported that, in mouse and human stem cell differentiation, miR-1 and miR-133 act comparably to promote mesoderm differentiation while suppressing gene expression of alternative lineages. Additionally, studies done on mouse myoblasts have shown that miR-1 and miR-206 promote muscle differentiation, from the mesoderm precursors, while miR-133 represses differentiation. Lastly, miR-206 has been proposed to take part in a Clock-Myod1-miR-206 feedback loop involved in the timing of the circadian cycle in skeletal muscle. [82] [83]

The miR-17~92 cluster has been primarily linked to cancer pathogenesis and to the regulation of hematopoiesis and immune functions, however it is also involved in the development of the cardiopulmonary system, with miRs-17 and -92a being linked to the control of lung development and postnatal neovascularization, respectively, with mice deficient in the miR-17~92 cluster exhibiting severe developmental defects of the lung and the heart which cause death shortly after birth. [84] [85]

In conclusion, the candidate regulatory factors identified in this study and their binding sites will be explored in future studies, to which the experimental system established in this work will give a valuable contribution.

V. Final remarks

We implemented an experimental system which enabled the assessment of the influence the 3'UTRs of the zebrafish *her1*, *her7* and FGF8a genes have on the expression of a fluorescent reporter. The results obtained indicate the all of the 3'UTRs tested have a negative impact on mRNA expression with the possible exception of one of the fgf8a alternative 3'UTRs (8S), which was postulated to have a negative impact on mRNA stability and a positive impact on translation efficiency. We have also identified a subset of candidate regulatory elements, including target sites for miRs and RBPs, that could be involved in modulating these processes .

Future perspectives include the completion of both the bioinformatic and the *in vivo* studies, the study of the candidate regulatory elements identified in the bioinformatic approach and an analogous study of the influence of the 5'UTRs of these genes in mRNA expression.

The experimental system established in this work, upon optimization, can be used to achieve the two latter goals and, thus contribute to the continued study of the post-transcriptional mechanisms operating in somitogenesis

Appendix

Figure 20 - Impact of the rate of maturation (km) on the system. A -impact of the rate of maturation in a system were the rate of mRNA stability is altered. The alterations to the output of the system under the conditions described in **figure 6A** were tested using five fold higher and five fold lower values of km. B - impact of the rate of maturation in a system were the rate of translation is altered. The alterations to the output of the system under the conditions described in **figure 6B** were tested using five fold higher and five fold lower values of km. The rate constants of mRNA degradation (k), translation (kt) and protein maturation (km) are indicated. Note that the rate of maturation does not significantly affect the output of the system, under the two scenarios tested (altered mRNA stability and. altered translation).

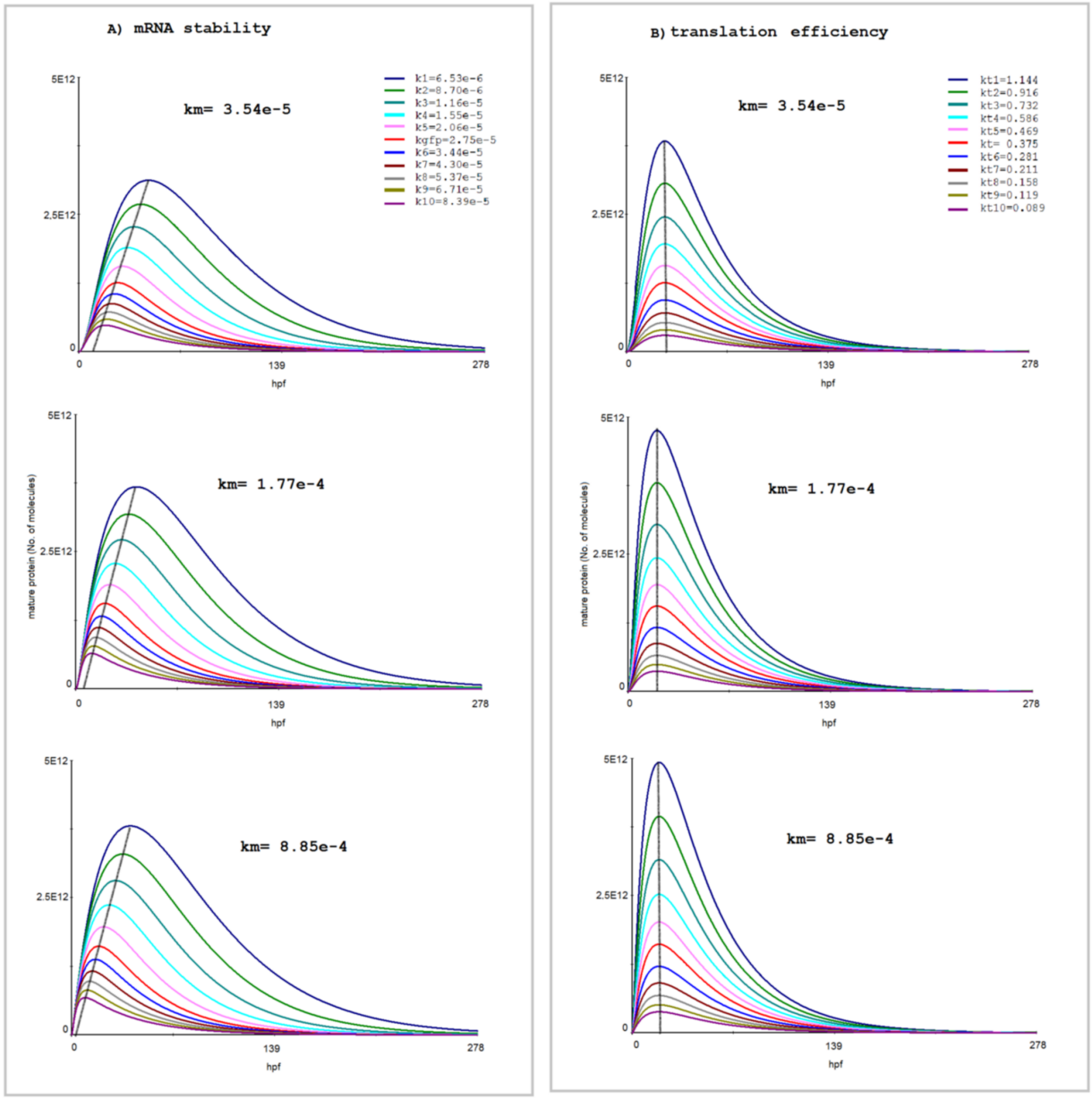
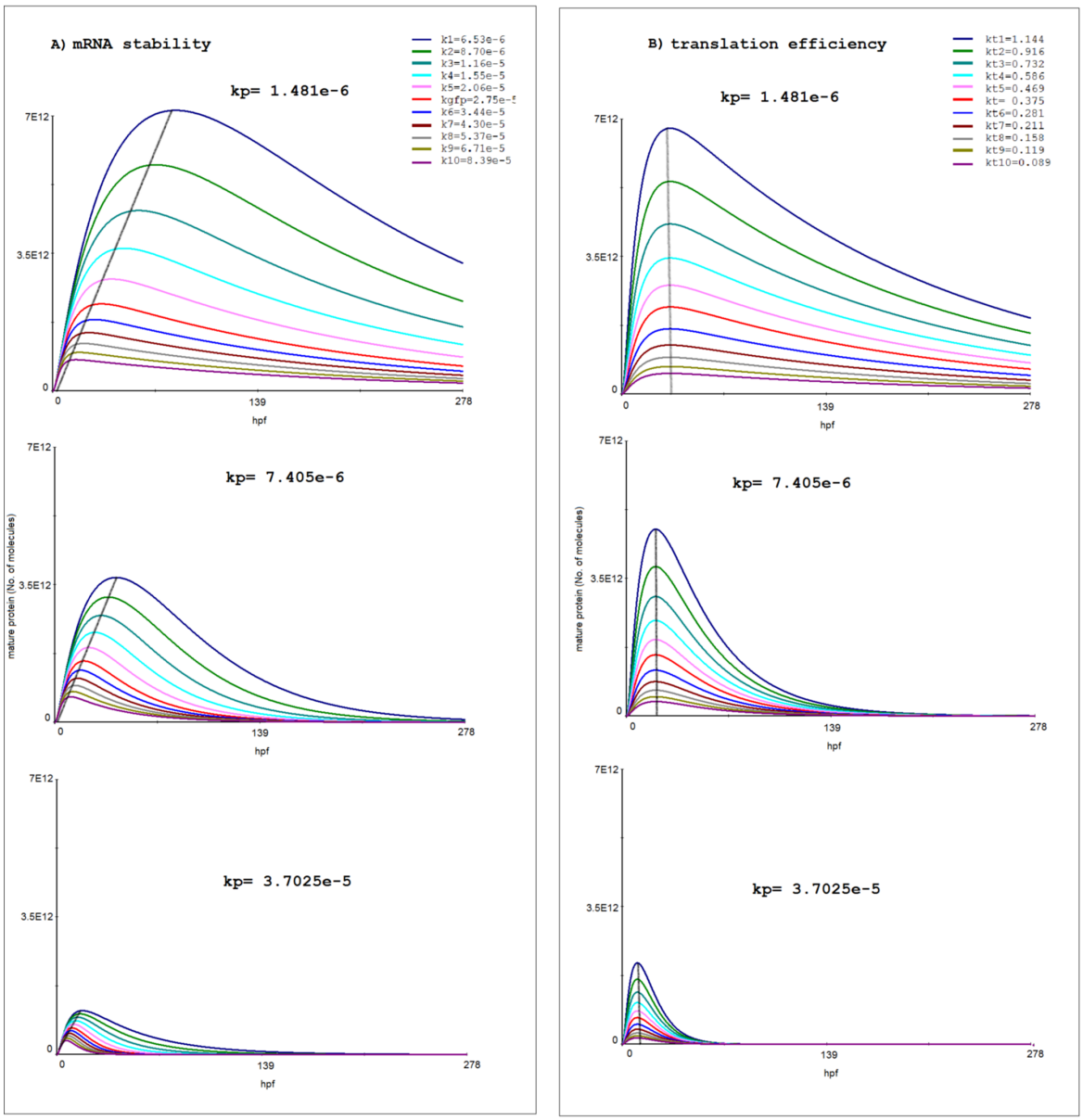


Figure 21 - **Impact of the rate of protein degradation (kp) on the system.** A -impact of the rate of protein degradation in a system were the rate of mRNA stability is altered. The alterations to the output of the system under the conditions described in **figure 6A** were tested using five fold higher and five fold lower values of kp. B - impact of the rate of protein degradation in a system were the rate of translation is altered. The alterations to the output of the system under the conditions described in **figure 6B** were tested using five fold higher and five fold lower values of kp. The rate constants of mRNA degradation (k), translation (kt) and protein degradation (kp) are indicated. Note that the rate of protein degradation primarily alters the amplitude of the curves, having a minor effect on their dynamics. More specifically, a shift of the mature protein peaks toward earlier time points is still associated with a decreased mRNA stability, and the absence of this shift is still associated with variations in the translation efficiency, when the rate of protein degradation is altered.



References

[1] J. Lewis, A. Hanisch, and M. Holder, “Notch signaling, the segmentation clock, and the patterning of vertebrate somites.,” *Journal of biology*, vol. 8, no. 4, p. 44, Jan. 2009.

[2] M.-L. Dequéant and O. Pourquié, “Segmental patterning of the vertebrate embryonic axis.,” *Nature reviews. Genetics*, vol. 9, no. 5, pp. 370-82, May. 2008.

[3] Y. J. Jiang, B. L. Aerne, L. Smithers, C. Haddon, D. Ish-Horowicz, and J. Lewis, “Notch signalling and the synchronization of the somite segmentation clock.,” *Nature*, vol. 408, no. 6811, pp. 475-9, Nov. 2000.

[4] F. Giudicelli, E. M. Ozbudak, G. J. Wright, and J. Lewis, “Setting the tempo in development: an investigation of the zebrafish somite clock mechanism.,” *PLoS biology*, vol. 5, no. 6, p. e150, Jun. 2007.

[5] J. Feller, A. Schneider, K. Schuster-gossler, and A. Gossler, “Noncyclic Notch activity in the presomitic mesoderm demonstrates uncoupling of somite compartmentalization and boundary formation,” *Genes & Development*, pp. 2166-2171, 2008.

[6] I. Palmeirim, D. Henrique, D. Ish-Horowicz, and O. Pourquié, “Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis.,” *Cell*, vol. 91, no. 5, pp. 639-48, Nov. 1997.

[7] a Sawada, M. Shinya, Y. J. Jiang, a Kawakami, a Kuroiwa, and H. Takeda, “Fgf/MAPK signalling is a crucial positional cue in somite boundary formation.,” *Development (Cambridge, England)*, vol. 128, no. 23, pp. 4873-80, Dec. 2001.

[8] J. Dubrulle, M. J. Mcgrew, O. Pourquie, C. D. Luminy, and M. Cedex, “and Regulates Segmentation Clock Control of Spatiotemporal Hox Gene Activation,” vol. 106, pp. 219-232, 2001.

[9] R. Diez del Corral, I. Olivera-Martinez, A. Goriely, E. Gale, M. Maden, and K. Storey, “Opposing FGF and retinoid pathways control ventral neural pattern, neuronal differentiation, and segmentation during body axis extension.,” *Neuron*, vol. 40, no. 1, pp. 65-79, Sep. 2003.

[10] A. C. Oates and R. K. Ho, “Hairy/E(spl)-related (Her) genes are central components of the segmentation oscillator and display redundancy with the Delta/Notch signaling pathway in the formation of anterior segmental boundaries in the zebrafish.,” *Development (Cambridge, England)*, vol. 129, no. 12, pp. 2929-46, Jun. 2002.

[11] R. L. Davis and D. L. Turner, “Vertebrate hairy and Enhancer of split related proteins: transcriptional repressors regulating cellular differentiation and embryonic patterning.,” *Oncogene*, vol. 20, no. 58, pp. 8342-57, Dec. 2001.

[12] C. Takke and J. a Campos-Ortega, “Her1, a Zebrafish Pair-Rule Like Gene, Acts Downstream of Notch Signalling To Control Somite Development.,” *Development (Cambridge, England)*, vol. 126, no. 13, pp. 3005-14, Jul. 1999.

[13] J. Lewis, “Autoinhibition with Transcriptional Delay : A Simple Mechanism for the Zebrafish Somitogenesis Oscillator,” *Current biology*, vol. 13, pp. 1398-1408, 2003.

[14] D. Sieger, D. Tautz, and M. Gajewski, “Her11 Is Involved in the Somitogenesis Clock in Zebrafish.,” *Development genes and evolution*, vol. 214, no. 8, pp. 393-406, Aug. 2004.

[15] S. S. Shankaran et al., “Completing the set of h/E(spl) cyclic genes in zebrafish: her12 and her15 reveal novel modes of expression and contribute to the segmentation clock.,” *Developmental biology*, vol. 304, no. 2, pp. 615-32, Apr. 2007.

[16] C. Winkler, H. Elmasri, B. Klamt, J.-N. Volff, and M. Gessler, “Characterization of hey bHLH genes in teleost fish.,” *Development genes and evolution*, vol. 213, no. 11, pp. 541-53, Nov. 2003.

[17] A. Pasini, Y.-J. Jiang, and D. G. Wilkinson, “Two zebrafish Notch-dependent hairy/Enhancer-of-split-related genes, her6 and her4, are required to maintain the coordination of cyclic gene expression in the presomitic mesoderm.,” *Development (Cambridge, England)*, vol. 131, no. 7, pp. 1529-41, Apr. 2004.

[18] A. Kawamura, S. Koshida, H. Hijikata, T. Sakaguchi, H. Kondoh, and S. Takada, “Zebrafish Hairy / Enhancer of split protein links FGF signaling to cyclic gene expression in the periodic segmentation of somites,” *Genes & Development*, vol. 19, pp. 1156-1161, 2005.

[19] M. Gajewski, H. Elmasri, M. Girschick, D. Sieger, and C. Winkler, “Comparative analysis of her genes during fish somitogenesis suggests a mouse/chick-like mode of oscillation in medaka.,” *Development genes and evolution*, vol. 216, no. 6, pp. 315-32, Jun. 2006.

[20] F. Reifers, H. Böhli, E. C. Walsh, P. H. Crossley, D. Y. Stainier, and M. Brand, “Fgf8 is mutated in zebrafish acerebellar (ace) mutants and is required for maintenance of midbrain-hindbrain boundary development and somitogenesis.,” *Development (Cambridge, England)*, vol. 125, no. 13, pp. 2381-95, Jul. 1998.

[21] B. W. Draper, D. W. Stock, and C. B. Kimmel, “Zebrafish fgf24 functions with fgf8 to promote posterior mesodermal development.,” *Development (Cambridge, England)*, vol. 130, no. 19, pp. 4639-54, Oct. 2003.

[22] J. Dubrulle and O. Pourquié, “fgf8 mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo.,” *Nature*, vol. 427, no. 6973, pp. 419-22, Jan. 2004.

[23] F. Mignone, C. Gissi, S. Liuni, and G. Pesole, “Untranslated regions of mRNAs.,” *Genome biology*, vol. 3, no. 3, p. REVIEWS0004, Jan. 2002.

[24] M. Gajewski, “Anterior and posterior waves of cyclic her1 gene expression are differentially regulated in the presomitic mesoderm of zebrafish,” *Development*, vol. 130, no. 18, pp. 4269-4278, Sep. 2003.

[25] V. Hilgers, O. Pourquié, and J. Dubrulle, “In vivo analysis of mRNA stability using the Tet-Off system in the chicken embryo.,” *Developmental biology*, vol. 284, no. 2, pp. 292-300, Aug. 2005.

[26] C. Barreau, L. Paillard, and H. B. Osborne, “AU-rich elements and associated factors: are there unifying principles?,” *Nucleic acids research*, vol. 33, no. 22, pp. 7138-50, Jan. 2005.

[27] M. Fink, G. Flekna, A. Ludwig, T. Heimbucher, and T. Czerny, “Improved translation efficiency of injected mRNA during early embryonic development.,” *Developmental dynamics : an official publication of the American Association of Anatomists*, vol. 235, no. 12, pp. 3370-8, Dec. 2006.

[28] T. H. Beilharz and T. Preiss, “Widespread use of poly(A) tail length control to accentuate expression of the yeast transcriptome.,” *RNA (New York, N.Y.)*, vol. 13, no. 7, pp. 982-97, Jul. 2007.

[29] A. C. Beckel-Mitchener, A. Miera, R. Keller, and N. I. Perrone-Bizzozero, “Poly(A) tail length-dependent stabilization of GAP-43 mRNA by the RNA-binding protein HuD.,” *The Journal of biological chemistry*, vol. 277, no. 31, pp. 27996-8002, Aug. 2002.

[30] M. W. Hentze, “Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress,” *Proceedings of the National Academy of*

Sciences, vol. 93, no. 16, pp. 8175-8182, Aug. 1996.

[31] J. a Dibbens, D. L. Miller, a Damert, W. Risau, M. a Vadas, and G. J. Goodall, “Hypoxic regulation of vascular endothelial growth factor mRNA stability requires the cooperation of multiple RNA elements.,” *Molecular biology of the cell*, vol. 10, no. 4, pp. 907-19, Apr. 1999.

[32] a Bashirullah, R. L. Cooperstock, and H. D. Lipshitz, “Spatial and temporal control of RNA stability.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 13, pp. 7025-8, Jun. 2001.

[33] C. S. Lutz, “Alternative Polyadenylation: A Twist on mRNA 3’ End Formation,” *ACS Chemical Biology*, vol. 3, no. 10, pp. 609-617, 2008.

[34] Z. Ji, J. Y. Lee, Z. Pan, B. Jiang, and B. Tian, “Progressive lengthening of 3’ untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 17, pp. 7028-33, Apr. 2009.

[35] S. Thomsen, G. Azzam, R. Kaschula, L. S. Williams, and C. R. Alonso, “Developmental RNA processing of 3’UTRs in Hox mRNAs as a context-dependent mechanism modulating visibility to microRNAs.,” *Development (Cambridge, England)*, vol. 137, no. 17, pp. 2951-60, Sep. 2010.

[36] H. H. Kim, Y. Kuwano, S. Srikantan, E. K. Lee, J. L. Martindale, and M. Gorospe, “HuR recruits let-7 / RISC to repress c-Myc expression,” *Genes & Development*, pp. 1743-1748.

[37] W. J. Ma, S. Chung, and H. Furneaux, “The Elav-like proteins bind to AU-rich elements and to the poly(A) tail of mRNA.,” *Nucleic acids research*, vol. 25, no. 18, pp. 3564-9, Sep. 1997.

[38] A. M. Krecic and M. S. Swanson, “hnRNP complexes: composition, structure and function,” *current opinion in cell biology*, vol. 11, pp. 363-371, 1999.

[39] T. Volk, D. Israeli, R. Nir, and H. Toledano-Katchalski, “Tissue development and RNA control: ‘HOW’ is it coordinated?,” *Trends in genetics : TIG*, vol. 24, no. 2, pp. 94-101, Feb. 2008.

[40] C. Thisse, B., Pflumio, S., Fürthauer, M., Loppin, B., Heyer, V., Degrave, A., Woehl, R., Lux, A., Steffan, T., Charbonnier, X.Q. and Thisse, “Expression of the zebrafish genome during embryogenesis,” *(NIH R01 RR15402). ZFIN Direct Data Submission*, 2001.

[41] C. Thisse, B., Thisse, “Fast Release Clones: A High Throughput Expression Analysis.,” *ZFIN Direct Data Submission*.

[42] R. Lobbardi, G. Lambert, J. Zhao, R. Geisler, H. R. Kim, and F. M. Rosa, “Fine-tuning of Hh signaling by the RNA-binding protein Quaking to control muscle development.,” *Development (Cambridge, England)*, vol. 138, no. 9, pp. 1783-94, May. 2011.

[43] M. R. Fabian, N. Sonenberg, and W. Filipowicz, “Regulation of mRNA translation and stability by microRNAs.,” *Annual review of biochemistry*, vol. 79, pp. 351-79, Jan. 2010.

[44] R. S. Pillai, “MicroRNA function: multiple mechanisms for a tiny RNA?,” *RNA (New York, N.Y.)*, vol. 11, no. 12, pp. 1753-61, Dec. 2005.

[45] K. a O’Donnell, E. a Wentzel, K. I. Zeller, C. V. Dang, and J. T. Mendell, “c-Myc-regulated microRNAs modulate E2F1 expression.,” *Nature*, vol. 435, no. 7043, pp. 839-43, Jun. 2005.

[46] M. Kedde et al., “RNA-binding protein Dnd1 inhibits microRNA access to target mRNA.,” *Cell*, vol. 131, no. 7, pp. 1273-86, Dec. 2007.

[47] J. D. Keene and S. a Tenenbaum, “Eukaryotic mRNPs may represent posttranscriptional operons.,” *Molecular cell*, vol. 9, no. 6, pp. 1161-7, Jun. 2002.

[48] J. D. Keene, “RNA regulons: coordination of post-transcriptional events.,” *Nature reviews. Genetics*, vol. 8, no. 7, pp. 533-43, Jul. 2007.

[49] P. L. Lowrey and J. S. Takahashi, “Mammalian circadian biology: elucidating genome-wide levels of temporal organization.,” *Annual review of genomics and human genetics*, vol. 5, pp. 407-41, Jan. 2004.

[50] E. Garbarino-pico, S. Niu, M. D. Rollag, C. A. Strayer, J. C. Besharse, and C. B. Green, “Immediate early response of the circadian polyA ribonuclease nocturnin to two extracellular stimuli,” *Image (Rochester, N.Y.)*, pp. 745-755, 2007.

[51] D. Benjamin, M. Schmidlin, L. Min, B. Gross, and C. Moroni, “BRF1 protein turnover and mRNA decay activity are regulated by protein kinase B at the same phosphorylation sites.,” *Molecular and cellular biology*, vol. 26, no. 24, pp. 9497-507, Dec. 2006.

[52] E. Voit, “PLAS.” 2000.

[53] a Sacchetti, T. El Sewedy, a F. Nasr, and S. Alberti, “Efficient GFP mutations profoundly affect mRNA transcription and translation rates.,” *FEBS letters*, vol. 492, no. 1-2, pp. 151-5, Mar. 2001.

[54] P. Corish and C. Tyler-Smith, “Attenuation of green fluorescent protein half-life in mammalian cells.,” *Protein engineering*, vol. 12, no. 12, pp. 1035-40, Dec. 1999.

[55] A. G. Evdokimov et al., “Structural basis for the fast maturation of Arthropoda green fluorescent protein.,” *EMBO reports*, vol. 7, no. 10, pp. 1006-12, Oct. 2006.

[56] H. Bremer and P. P. Dennis, “Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate,” no. 122.

[57] W. S. Rasband, “Image J", U.S. National Institutes of Health, Bethesda, Maryland, USA, http://imagej.nih.gov/ij/.” .

[58] H. Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., “CLUSTALX.” 2007.

[59] A. A. Salamov and V. V. Solovyev, “Recognition of 3’-processing sites of human mRNA precursors,” *comput appl biosci*, vol. 13, no. 1, pp. 23-28, 1997.

[60] F. Ahmed, M. Kumar, and G. P. S. Raghava, “Prediction of polyadenylation signals in human DNA sequences using nucleotide frequencies,” *In silico biology*, vol. 9, 2009.

[61] W. Ao, J. Caudet, W. J. Kent, S. Muttumu, and S. E. Mango, “Environmentally Induced Foregut Remodeling by PHA-4/FoxA and DAF-12/NHR,” *Science*, vol. 305, no. 5691, 2004.

[62] Varcancel and Morais, “SPLICING RAINBOW.” p. EMBL, 2002.

[63] R. Grocock, S. V. Dongen, and A. J. Enright, “MicroCosm.” .

[64] V. Rusinov, V. Baev, I. N. Minkov, and M. Tabler, “MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence.,” *Nucleic acids research*, vol. 33, no. Web Server issue, pp. W696-700, Jul. 2005.

[65] J. G. Williams et al., “The nucleotide sequence of the major beta-globin mRNA from *Xenopus* laevis,” *Nucleic acids research*, vol. 8, no. 18, pp. 4247-4258, 1980.

[66] W.-Y. Choi, A. J. Giraldez, and A. F. Schier, “Target protectors reveal dampening and balancing of Nodal agonist and antagonist by miR-430.,” *Science (New York, N.Y.)*, vol. 318, no. 5848, pp. 271-4, Oct. 2007.

[67] Y. Mishima et al., “NIH Public Access,” *In Situ*, vol. 16, no. 21, pp. 2135-2142, 2007.

[68] A. J. Giraldez et al., “Deadenylation and Clearance of Maternal mRNAs,” *Science*, vol. 312, no. April, pp. 75-79, 2006.

[69] R. L. Tanguay and D. R. Gallie, “Translational efficiency is regulated by the length of the 3’ untranslated region.,” *Molecular and cellular biology*, vol. 16, no. 1, pp. 146-56, Jan. 1996.

[70] M. K. Doma and R. Parker, “Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation.,” *Nature*, vol. 440, no. 7083, pp. 561-4, Mar. 2006.

[71] N. L. Garneau, J. Wilusz, and C. J. Wilusz, “The highways and byways of mRNA decay.,” *Nature reviews. Molecular cell biology*, vol. 8, no. 2, pp. 113-26, Feb. 2007.

[72] M. Kamasawa and J.-ichi Horiuchi, “Identification and characterization of polyadenylation signal (PAS) variants in human genomic sequences based on modified EST clustering.,” *In silico biology*, vol. 8, no. 3-4, pp. 347-61, Jan. 2008.

[73] H. Ro, K. Soun, and M. Rhee, “Novel vector systems optimized for injecting in vitro-synthesized mRNA into zebrafish embryos,” *Molecular cell*, vol. 17, no. 2, pp. 373-376, 2004.

[74] P. Y. Chen et al., “The developmental miRNA profiles of zebrafish as determined by small RNA cloning,” *Genes & Development*, pp. 1288-1293, 2005.

[75] B. Thisse and C. Thisse, “Fast Release Clones: A High Throughput Expression Analysis,” *ZFIN Direct Data Submission*, 2004.

[76] E. Wienholds et al., “MicroRNA Expression in Zebrafish Embryonic Development,” *Science*, vol. 309, no. July, pp. 310-311, 2005.

[77] S. Rousseau, N. Morrice, M. Peggie, D. G. Campbell, M. Gaestel, and P. Cohen, “Inhibition of SAPK2a/p38 prevents hnRNP A0 phosphorylation by MAPKAP-K2 and its interaction with cytokine mRNAs.,” *The EMBO journal*, vol. 21, no. 23, pp. 6505-14, Dec. 2002.

[78] F. Raffalli-Mathieu, T. Glisovic, Y. Ben-David, and M. a Lang, “Heterogeneous nuclear ribonucleoprotein A1 and regulation of the xenobiotic-inducible gene Cyp2a5.,” *Molecular pharmacology*, vol. 61, no. 4, pp. 795-9, Apr. 2002.

[79] M. Fähling, R. Mrowka, A. Steege, P. Martinka, P. B. Persson, and B. J. Thiele, “Heterogeneous nuclear ribonucleoprotein-A2/B1 modulate collagen prolyl 4-hydroxylase, alpha (I) mRNA stability.,” *The Journal of biological chemistry*, vol. 281, no. 14, pp. 9279-86, Apr. 2006.

[80] S. Shetty, “Regulation of urokinase receptor mRNA stability by hnRNP C in lung epithelial cells.,” *Molecular and cellular biochemistry*, vol. 272, no. 1-2, pp. 107-18, Apr. 2005.

[81] E. K. Lee et al., “hnRNP C promotes APP translation by competing with FMRP for APP mRNA recruitment to P bodies.,” *Nature structural & molecular biology*, vol. 17, no. 6, pp. 732-9, Jun. 2010.

[82] J. J. Mccarthy, “MicroRNA-206: the skeletal muscle-specific myomiR,” *biochim biophys acta*, vol. 1779, no. 11, pp. 682-691, 2008.

[83] K. N. Ivey et al., “MicroRNA Regulation of Cell Lineages in Mouse and Human Embryonic Stem Cells,” *Cell stem cell*, vol. 2, no. 3, pp. 219-229, 2008.

[84] A. Bonauer and S. Dimmeler, “The microRNA-17-92 cluster: still a miRacle?,” *Cell cycle (Georgetown, Tex.)*, vol. 8, no. 23, pp. 3866-73, Dec. 2009.

[85] A. Ventura et al., “Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters.,” *Cell*, vol. 132, no. 5, pp. 875-86, Mar. 2008.