STRUCTURE FUNCTION ANALYSIS OF THE DEUBIQUITYLATING ENZYME FAM

Poon-Yu Khut, B.Sc

School of Molecular & Biomedical Science (Biochemistry)
University of Adelaide
Adelaide, South Australia 5005

October 2006
CHAPTER 3

CHARACTERISATION OF USP9 AND ESTABLISHMENT OF A DOMINANT-NEGATIVE DOMAIN SCREEN FOR FAM-BINDING PROTEINS IN ZEBRAFISH

Introduction

IDENTIFICATION OF FAM-BINDING PROTEINS IN ZEBRAFISH

At approximately 290 kDa, FAM is the largest member of USP family of deubiquitylating enzymes of which there are predicted to be over 60 members in mammals (Semple 2003). However, very little information is known of FAM’s structure. Of FAM’s 2554 amino acids, only a single 443 amino acid region, termed Fam-CAT (Taya et al. 1998, 1999), had been partially characterised. This region spans the theoretical boundaries of the catalytic core (but does not fully cover the structural definition) and has been shown to bind two bona-fide substrates, AF-6 and β-catenin (Taya et al. 1998, 1999) and more recently, an additional region within the catalytic core (which extensively overlaps Fam-CAT) has been shown to bind Epsin (Chen et al. 2003) (figure 1.13). It was proposed that the remainder of FAM comprised of extensive N and C-terminal extensions form protein binding domains (approximately 1500 and 600 amino acids respectively), as is the case with numerous other USPs (Kato et al. 2000, Lin et al. 2000). Indeed, this supposition was later confirmed for FAM when a region in the C-terminus (2299-2554 aa) was shown to bind Doublecortin, a microtubule-associated protein involved in neuronal migration (Friocourt et al. 2005).

Given FAM’s complex temporal and spatial expression patterns throughout development, it has been proposed that FAM’s function is not confined to a particular developmental or cellular event (Wood et al. 1997). FAM therefore has the potential to deubiquitylate multiple substrates. Identification of these substrates and their corresponding binding regions within FAM is of interest, as it would help elucidate the cellular role of FAM that contributes to particular developmental outcomes.
It is also likely that FAM itself is regulated by its activity and subcellular location. FAM’s ability to reside at varying subcellular localisations implies regulation by other proteins that target FAM, whether it be to the spindle and chromosomes (Pantaleon et al. 2001), the peri-nuclear region (Pantaleon et al. 2001), microtubules (Friocourt et al. 2005), cell-cell contacts (Taya et al. 1998, 1999) or points of protein trafficking (Murray et al. 2004). Identification of these non-substrate FAM-binding proteins would be fundamental in understanding how FAM is regulated. Conversely, identification of the regions within FAM that bind or are modified by these proteins would also aid in assigning identity to the remaining majority of uncharacterised FAM.

It was thus the aim of this project to identify FAM-binding proteins (be it substrate or regulatory) that are involved in early development.

**DOMINANT-NEGATIVE DOMAIN SCREEN FOR FAM-BINDING PROTEINS**

Previous attempts to identify novel FAM-binding proteins have focused on using FAM as bait for extracting binding proteins from Embryonic Stem (ES) and Early Primitive Ectoderm-Like (EPL) cell lysates. Immunoprecipitation of complexes using FAM antibody proved expensive (Diam 1999) and GST fusions with arbitrarily defined regions of FAM were difficult to solubly express in bacteria (Khut 2000). Ultimately these studies did not identify any novel FAM-binding proteins.

In a new approach, a dominant-negative domain screen for FAM-binding proteins was established in the zebrafish system. This method is a reverse approach whereby a developmental role is first established for a FAM/binding protein interaction, followed by its identification. In this way, future research could immediately focus on FAM’s role in the particular developmental process as revealed by the phenotype.

This screen entailed injection of ectopic mRNA encoding regions of FAM fused to a Myc tag, into one half of a two-cell stage Zebrafish embryo. By experimenting on only half of the embryo, changes in morphology can be compared to the other, non-injected half which develops normally. The FAM fragment mRNA is translated into bait protein, which if properly folded, may bind endogenous zebrafish proteins. If the bound zebrafish protein is required for a developmental process, titration of the protein will lead to a developmental
abnormality on the injected side. As the design of the fusion-baits included the Myc immuno-reactive tag, these protein complexes could later be extracted by immunoprecipitation from embryo lysates with a commercially available Myc antibody (figure 3.1). The protein complexes would then be separated by electrophoresis and the binding-partner identified by mass spectrometry. Injection of mRNA into embryos is an established technique used in gain-of-function experiments and has been used to show that human homeobox-containing gene VENT-like homeobox-2 (VENTX2), belongs to a ventralising homeodomain class of genes by its ability to cause anterior truncations and prevent notochord formation in zebrafish embryos (Moretti et al. 2001).

In the absence of structural data, FAM fragments used in this assay were designed based on amino acid conservation between the drosophila and mouse sequences. It was rationalised that higher degrees of amino acid conservation are associated with functional conservation and are more likely to form folding domains than non-conserved inter-domain linker regions. These regions were thus termed Highly Conserved Regions (HCRs).

This developmental screen has a number of advantages over other approaches.

**Sensitivity:** A number of features make this assay sensitive. The nature of an enzymatic reaction is transient i.e. once substrates are modified they must be released to free the active site. In an immunoprecipitation reaction using a functional enzyme as bait, there exists the possibility that little substrate may remain attached to the enzyme after catalysis, particularly after washing steps. This dominant-negative domain screen thus used fragments of FAM that were catalytically inactive.

This assay is sensitive on the protein level. Abnormalities are only detected when the amount of a critical protein drops below a certain threshold level, irrespective of whether a large or small amount of protein is required for the process.

Operating on the whole organism rather than on a population of cultured cells affords rapid and easy identification of abnormalities. As this assay is performed on only half of an embryo, it is possible to compare a structural difference to the control side by visual inspection and is sensitive enough to identify slight aberrations.
Figure 3.1

Schematic of the Zebrafish Developmental Screen

1. Highly Conserved Regions (HCRs) identified by aligning FAM with fly orthologue FAF.
2. HCRs cloned into GW pCS2+MT vector that allows for in vitro transcription of capped, poly-adenylated mRNA with a N-terminal 6x Myc tag, which could be later used to immunoprecipitate and identify the binding complex via mass spectrometry.
3. In vitro transcription of mRNA.
4. Injection of mRNA into one half of a 2-cell stage zebrafish embryo.
5. In vivo translation of mRNA by endogenous cellular machinery.
6. Protein pool is screened by HCR protein for interactions that affects a developmental process.
7. If there are no meaningful interactions, development proceeds normally. If there is an interaction that interferes with a particular proteins' developmental function, the embryo will develop abnormally one the side that was injected.
**Speed:** Experimental results can be obtained extremely quickly due to the rapid rate of zebrafish development. Within the first 24 hours of zebrafish development, embryos have already completed gastrulation, neurulation and somitogenesis, a stage roughly equivalent to a first trimester human embryo. As this experiment occurs within a developing embryo, this assay screens all developing cell types within their normal context as opposed to individual cell populations. Such a high throughput makes tissue screening extremely rapid.

**Simplicity:** Certain properties of zebrafish development that make it a good genetic model are exploited by this assay. Importantly, zebrafish are vertebrates and so their development is more comparable to the mouse than that of fly which up to this point, has been the main source of developmental extrapolation. Zebrafish embryos are large enough to be readily manipulated and injected with mRNA, and their transparent and external development allows for easy monitoring of morphological and internal structures throughout the experiment.

In theory, the experimental readout should be easy to interpret. As opposed to pull-down experiments where the data can be numerous bands on a protein gel, the dominant-negative domain screen simply looks for developmental abnormalities. In recent immunoprecipitation experiments from T84 lysates (a polarised intestinal cell line), FAM pulled down more than 40 bands as visualised by Coomassie Blue staining (Murray *et al.* 2004). Results like these can be overwhelming and would require time and resources to pursue them all. It is also likely that several of these precipitated proteins contribute little to developmental events, or may be false positives. As the primary focus of this laboratory is to better understand the role of FAM in early mouse development, identifying only developmentally relevant binding proteins is paramount and the dominant-negative domain screen directly addresses this issue.

**Unbiased:** one of best applications of a technique like immunoprecipitation is when testing the *in vivo* binding between two known interacting proteins. When it has been established that two proteins have an affinity for each other, binding conditions can be optimised and other variables can be tested that affect the strength of its binding e.g. addition or subtraction of intracellular proteins. It may be misleading however, to use immunoprecipitation to screen an array of proteins for an unknown interaction to an
enzyme, due to their potentially weak affinity post-catalysis. Although possible to “optimise” binding conditions (primarily through alteration of salt concentration), large deviations from physiological conditions will undermine the stringency and validity of the experiment. The dominant-negative domain screen on the other hand is conducted blind. The experiment is carried out in an intracellular environment and any developmental defects that arises, theoretically results from a physiological interaction. A caveat of this approach however, was that the bait protein were artificial definitions of a folding domain. As their biological relevance was uncertain, a large potential error existed however, one presumes that any grossly misfolded protein would be cleared by the ubiquitin/proteasome system, although the overloading of the system may also lead to unforeseen consequences.

**CHAPTER DESCRIPTION**

The developmental screen assumed that FAM and its zebrafish orthologue were sufficiently conserved to allow recognition of binding proteins between species. This assumption is reasonable given that FAM can rescue fully FAF function in drosophila (Chen *et al.* 2000). Before beginning the experiments, it was therefore necessary to identify and characterise the zebrafish orthologue to determine the extent of sequence similarity. Described in this chapter is the characterisation of the zebrafish orthologue of FAM, Usp9. The predicted amino acid sequence of Usp9 was assembled from database searches and its orthology relationships with other FAM proteins were determined by phylogenetic analysis. Its expression pattern during early zebrafish development was analysed by in situ hybridisation.

This chapter further describes the cloning of four HCRs of FAM, which were ectopically expressed in zebrafish embryos along with enhanced green fluorescent protein (eGFP) and distilled water controls, and were subsequently analysed for changes in phenotype. The phenotypes were scored and statistically analysed to determine whether the fragments significantly increased developmental defects.

Also detailed in this chapter are attempts to express these HCRs in bacterial and mammalian expression systems.
**Assembly of the Zebrafish usp9 Sequence**

In the advent of the whole genome sequencing projects, gene orthologues can be readily identified by screening databases with sequences from a gene of interest. Although the quality of the data is becoming more reliable as sequencing technology improves, it is important to note that the assembly of genomic sequence into cDNA and protein sequences are predicted by computer programs.

To define the putative coding sequence of zebrafish *Fam (usp9)*, a predicted mRNA sequence from the Zv4 assembly of zebrafish genomic sequence by the Sanger Institute was aligned with the known *Fam* cDNA sequences from: mouse *Fam (Usp9x)* and *FamY (Usp9y)*, humans *USP9X* and *USP9Y*, and with predicted mRNA sequences from pufferfish, chicken, frog and rat (*table 3.1*). The drosophila sequence (*faf*) was omitted due to the degree of evolutionarily divergence. When positions of variation in the predicted zebrafish sequence were identified relative to both the closely related pufferfish sequence and the other sequences (114 amino acids), or where (in one instance), an intron-exon boundary was not conserved relative to mouse *Fam*, the sequence was checked by manual examination of electrophoretograms from the zebrafish genome project Ensembl Trace Server. For each variation, the electrophoretograms from at least two independent sequencing reactions were inspected for verification.
<table>
<thead>
<tr>
<th>Common Name</th>
<th>Species Name</th>
<th>Gene Name</th>
<th>Accession Number or Sanger Center Number (italics)</th>
<th>Length (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zebrafish</td>
<td>Danio rerio</td>
<td>usp9</td>
<td>DQ086492</td>
<td>2551</td>
</tr>
<tr>
<td>Mouse</td>
<td>Mus musculus</td>
<td>Fam</td>
<td>DQ086491</td>
<td>2554</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Famy</td>
<td>AJ307017</td>
<td>2556</td>
</tr>
<tr>
<td>Human</td>
<td>Homo sapiens</td>
<td>Usp9x</td>
<td>X98296</td>
<td>2547</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Usp9y</td>
<td>Y13618</td>
<td>2555</td>
</tr>
<tr>
<td>Fly</td>
<td>Drosophila melanogaster</td>
<td>faf</td>
<td>L04958</td>
<td>2711</td>
</tr>
<tr>
<td>Chicken</td>
<td>Gallus gallus</td>
<td></td>
<td>ENSGALG00000016236</td>
<td>2555</td>
</tr>
<tr>
<td>Frog</td>
<td>Xenopus tropicalis</td>
<td></td>
<td>ENSXETG00000015489</td>
<td>2545</td>
</tr>
<tr>
<td>Pufferfish</td>
<td>Fugu rubripes</td>
<td></td>
<td>SINFRUG00000128324</td>
<td>2547</td>
</tr>
<tr>
<td>Rat</td>
<td>Rattus norvegicus</td>
<td></td>
<td>ENSRNOG0000003261</td>
<td>2552</td>
</tr>
</tbody>
</table>

Table 3.1: Table of Fam Orthologue Sequences. Fam orthologue data were derived from either published Genbank entries or by Ensembl orthologue predictions from species genome sequencing. Published sequences have both gene names and an accession number while predicted sequences are identified only by a Sanger Centre number. These sequences were subsequently used to identify non-conserved usp9 amino acids and also for phylogenetic analysis.

These checks identified 17 amino acid errors in the zebrafish Usp9 sequence which were corrected (figure 3.2 A) and inspection of the predicted intron-exon boundary structure of usp9 revealed a computer generated 4bp intron that does not abide by normal splicing consensus. Analysis of the electrophoretograms revealed a sequencing error that had added an extra base pair and its removal correctly fuses the predicted seventh and eighth exons into one single exon, consistent with the conserved intron-exon boundaries of mouse Fam (figure 3.2 B). In this way, all uncertainties in the open reading frame were resolved. In keeping with zebrafish gene nomenclature, the zebrafish Fam gene was named usp9 while the protein was named Usp9.

The errors detected in the predicted zebrafish sequence, in particular the exon structure, highlighted the fallibility of computer predictions. While these predictions rapidly advance gene identification, one cannot solely rely on its data to draw conclusions. This data is best used to help guide research, as was used here.
A) Alignment of Usp9 with various other vertebrate FAM orthologues. Where the computer predicted sequence varied from both the closely related pufferfish sequence and the other sequences, the individual amino acids were checked through examination of the sequencing electrophoregrams. Changes that were made are highlighted.

B) The computer predicted intron/exon structure of usp9 inserts an extra intron (intron 7-8) through a sequencing error that includes an extra guanine base pair (highlighted). When this base pair is removed, the predicted exon 7 and 8 fuse to form a single exon and in the process encodes an extra amino acid (glutamic acid E) as seen in the first panel of B. Figure adapted from blast searches of the Ensembl Danio rerio database for the zebrafish orthologue of Fam.
Phylogenetic Analysis of Fam Orthologues

To establish orthology relationships between zebrafish with published and predicted FAM proteins, a Gendoc Pairwise alignment was first performed comparing the mouse and zebrafish sequences (figure 3.3). The level of conservation between mouse and zebrafish FAM was found to be very high, 92.5% amino acid similarity and 89.7% identical. As no other predicted candidate proteins were found that approach this level of identity in the Zv4 assembly (Sanger Institute), this strongly suggested that these two proteins were orthologues. To establish the orthology relationships of zebrafish Usp9 with FAM proteins from other species, a phylogenetic analysis was performed comparing identified and predicted FAM sequences from zebrafish, human (USP9X and USP9Y), mouse (FAMX and FAMY), rat, frog, chicken, pufferfish and fly. A clustalW alignment of these sequences was used to produce the unrooted phylogenetic tree (figure 3.4). If the highly distant drosophila FAF sequence is omitted from this alignment, the clades remain arranged in the illustrated manner, but bootstrap values are all 100 indicating that the drosophila sequence disrupts arrangement of the sequences somewhat.

In Situ Analysis of usp9 in Early Zebrafish Development

Usp9 transcripts were present during early cleavage stages of zebrafish embryogenesis indicating a maternal contribution, similar to that observed in drosophila (Fisher-Vize et al. 1992) and mouse (Pantaleon et al. 2001, Sato et al. 2004) (figure 3.5 A). The distribution of usp9 transcripts appeared ubiquitous until at least 16 hours post-fertilisation (hpf) (figure 3.5 B-H). By 24 hpf, highest level expression was retained in anterior regions of the central nervous system (CNS, figure 3.5 I). Discrete expression was also observed in the central axis and lateral borders of the CNS posterior of the developing brain (figure 3.5 J) and anterior half of somites (figure 3.5 K). Highest expression was observed in the central region of the developing lens primordium (compare figure 3.5 L to M), a region that is largely exempt from the high levels of apoptosis occurring during lens morphogenesis (Cole and Ross, 2001).
Figure 3.3
Alignment of FAM and Usp9

Gendoc Pairwise alignment of mouse and zebrafish FAM protein sequences. The degree of conservation at each amino acid residue position is indicated by shading (100%=Black; 60%=Grey; 0%=White). The number of the right-hand most amino acid residue in each alignment line is indicated to the right.
Figure 3.4

Phylogenetic Analysis of FAM orthologues

Phylogenetic tree of protein sequences from zebrafish (Usp9), human (USP9X and USP9Y), mouse (FAMX and FAMY), rat, frog, chicken, pufferfish and drosophila FAM orthologues. The distance matrix was generated under the Dayhoff PAM matrix method of amino acid substitution, and a tree constructed using the ‘Neighbour Joining Method’. Nodal bootstrap values are shown as a percentage. Sequences are labelled with species (predicted sequences) and protein names (identified genes).
Figure 3.5

Expression of \textit{usp9} in the Early Stages of Zebrafish Development

\textit{In situ} analysis of \textit{usp9} expression in zebrafish embryos. Lateral views with animal pole or anterior to top and dorsal to right unless otherwise indicated. Expression is ubiquitous throughout the zygote (0 hpf 2-cell stage \textbf{A}), blastula (4 hpf sphere stage \textbf{B}) Gastrula (6 hpf shield stage \textbf{C}, anterior view \textbf{D}. 8 hpf 75\%-epiboly stage \textbf{E}, 10 hpf bud stage \textbf{F}) and up to the segmentation period (12 hpf 6-somite stage, \textbf{G}. 16 hpf 14-somite stage, \textbf{H}). By 16 hpf expression begins to restrict to the anterior regions of the central nervous system and is prevalent in the pharyngula (24 hpf embryo, \textbf{I-M}). Expression is also restricted to the central axis and lateral borders of the CNS (dorsal view \textbf{J}) and anterior half of somites (\textbf{K} shows a closer view of somites above the rostral end of the yolk extension at 24 hpf). Highest expression is seen within the eye, in the medial part of ectodermal tissue (lens primordium, delineated by dots) that is enclosed by the retina to form the lens (see transverse view of the developing eyes in \textbf{L}, sense control \textbf{M}). Scale bars: 250\textmu m (\textbf{A-J}), 100\textmu m (\textbf{K}), 50\textmu m (\textbf{L-M}).
Figure 3.5
By 2 days post fertilisation (dpf) expression in the CNS was now strongly restricted to the brain with very low or no expression in the spinal cord (figure 3.6 A, B, L). However, expression was observed in the endoderm posterior to the caudal limit of brain expression (figure 3.6 A, L). Relatively high level expression was also observed in the distal halves of the developing pectoral fins (figure 3.6 B) and throughout the developing lens except for its outermost epithelial layer (figure 3.6 I-K).

From 3 to 5 dpf, expression became less defined but still predominant in the anterior CNS. Expression was also observed in the lateral line and in some, but not all endodermal structures (figure 3.6 C-H). By 5 dpf, expression within the eye was restricted to the ganglion cell layer and the inner most half of the inner cell layer, but absent from the core of the lens (figure 3.6 M, N). Usp9 transcripts were still detected within the brain (figure 3.6 O). Expression was also observed in the swim bladder and intestinal bulb but although the sense controls were negative, trapping cannot be ruled out (figure 3.6 E, G, P).

The developmental expression of usp9 in zebrafish was consistent with patterns observed in mouse. Expression in both cases appears ubiquitous early, followed by progressive restriction to the CNS (Wood et al. 1997). Polarised expression in the pectoral fin (equivalent to anterior limb bud), and eye are also common features.
**Figure 3.6**

Expression of *usp9* in the Later Stages of Zebrafish Development

*In situ* analysis of *usp9* expression in zebrafish hatching period embryos (2-5 dpf). Lateral views: A, C, E, G. Dorsal views: B, D, F, H. Anterior to the left, dorsal to the top. Sections: K, L, N-P. Expression in long-pec stage embryos (2 dpf, A, B, I-L) is restricted to the brain, eyes, distal tips of the pectoral fin buds and endoderm (L, sagittal section medial to lens). Expression within the eye is particularly concentrated throughout the developing lens (I, transverse view of embryo after limited staining; J, sense negative control; K, sagittal section though retina and lens). Expression becomes continually restricted to regions within the brain, eyes and endodermal tissue and lateral line from the protruding-mouth stage to early larva (3 dpf C-D, 4 dpf E-F, 5 dpf G-H). Expression within the 5 dpf embryo eye is now observed in the ganglion cell layer and central regions of the inner nuclear layer (M, N sagittal section though retina and lens), regions of the brain and gills (O, transverse section, anterior to the top) and swim bladder and intestinal blub (P, sagittal section, anterior to the left, dorsal to the top). It should be noted that the staining in panels E, G and P maybe an artifact of trapping. Endoderm (*e*), pectoral fin (*pf*) lateral line (*ll*), retina (*r*), lens (*l*), inner nuclear layer (*inl*) ganglion cell layer (*gcl*), intestinal bulb (*ib*), swim bladder (*sb*). Scale bars: 250μm (A-H), 100μm (I-J, L-O), 50μm (K, P).
Figure 3.6
**Dominant-Negative Domain Screen**

The extraordinary level of sequence conservation between the mouse and zebrafish genes and their consistent expression patterns, coupled with the ability of FAM to rescue FAF protein function (Chen et al. 2000) established a conserved developmental role across species and made zebrafish an appropriate complementary system for the study of *usp9* in development. For these reasons, it was practical to conduct experiments in the zebrafish system using mouse protein as bait.

**CLONING OF FAM’S HIGHLY CONSERVED REGIONS (HCRs)**

Stretches of amino acids that are maintained throughout evolution exert evolutionary pressure for their conservation in order to maintain structural or functional roles. On the other hand, regions that are not critical for function e.g. linker peptides, are not subject to the same selective pressures and are a major source of amino acid sequence divergence. Therefore cloning some of FAM’s HCRs should have increased the chances of identifying a folding functional domain for ectopic expression in zebrafish.

When the mouse and distantly related drosophila FAM sequences were aligned, it was apparent that although virtually collinear over the entire length of the protein (50% identity, 70% similarity, Wood et al. 1997), certain stretches of amino acids were more conserved than surrounding regions. These variations were apparent due to the large evolutionary distance between mouse and drosophila and would have been less pronounced if two close relatives were compared (figure 3.7 A). As such, this alignment was in stark contrast to an alignment between the mouse and zebrafish sequence (figure 3.3), which were too similar to observe major differences. Other than the highly conserved catalytic Cysteine and Histidine boxes characteristic for this class of enzyme, four other regions were identified to be highly conserved. These HCRs were named TDEE, ERL, RKE and NPF, after the first few amino acids of the fragment. Analysis of their amino acid content demonstrated above average conservation (65.4% average identity, 82.3% average similarity, figure 3.7 B). Within the RKE fragment, much of the catalytic core’s conserved motifs were included with the exception of the Cys box. Excluding this motif ensured that this HCR was not catalytically active when translated. Had a functional catalytic core been
Figure 3.7

Design of Four FAM Highly Conserved Regions (HCRs)

A) Position of the four HCRs, overlayed on top of a FAM/FAF alignment. Due to evolutionary divergence between the drosophila and mouse sequences, highly conserved regions (HCRs) were identified and named TDEE, ERL, RKE, NPF based on the first several amino acids of the HCR. The positions of motifs within the catalytic core conserved amongst the USP class of deubiquitylating enzymes are highlighted.

B) Table of the four HCRs analysing amino acid content. Regional position given in base pairs (bps) refers to the Fam Genbank entry (U67874). Number of amino acids that are Identical and Similar are tallied and also given as a percentage. Each HCR’s degree of conservation is above the average of 50% identity, 70% similarity.

(A adapted from Wood et al. 1997)
Figure 3.7

NOTE: This image is included on page 56a in the print copy of the thesis held in the University of Adelaide Library.
included in this study, any developmental abnormalities that arose may have been attributable to non-specific deubiquitylation effects.

The HCRs were amplified by PCR with flanking attB sites (for site-directed recombination) and transferred into the pDONR\textsubscript{201} donor vector via a Gateway BP reaction. The fluorescent protein eGFP was also cloned as a control, to test for non-specific effects of mRNA injection, as well as determining whether the destination vectors were able to express exogenous protein after injection into the embryo. Sequencing of these vectors revealed there were no PCR-generated errors. A Gateway compatible destination vector was constructed for \textit{in vitro} transcription of capped mRNA with an N-terminal 6x Myc tag, by modification of a pCS2\textsuperscript{+}MT vector to create GWpCS2\textsuperscript{+}MT. The four HCRs and eGFP were then transferred into the newly constructed vector and linearised plasmids of the HCRs were used as template for \textit{in vitro} transcription reactions. A stock solution of mRNA was made for each, at an estimated concentration of 500ng/\textmu l.

**VALIDATION OF HCR PROTEIN PRODUCTION**

To determine whether protein would be expressed from the newly constructed vectors in vertebrate (mammalian) cells, the vectors were transiently transfected into 293T fibroblast cells. Western analysis of the soluble fractions of the transfectants’ cellular lysates was performed with an anti-Myc antibody. All the HCRs were solubly expressed in mammalian cells at their predicted sizes; therefore the injection of mRNA derived from these vectors into zebrafish cells was pursued (figure 3.8).

**DOMINANT-NEGATIVE DOMAIN SCREEN PILOT STUDY**

To ascertain whether this approach was viable and able to cause developmental defects, a pilot study was conducted. Clutches of embryos (ranging from 20 to 50) were injected (into one cell of a 2-cell stage embryo) with the four different HCR mRNA at concentrations of either 50ng/\textmu l or 100ng/\textmu l and then observed between 20 to 24 hours later. Injected embryos were compared to un-injected embryos, to confirm the normality of the clutch. As the purpose of the study was merely to determine whether mRNA injection could cause specific developmental abnormalities, and if so, which HCRs cause them, embryos were not scored or statistically analysed however, initial observations were noted.
Westerns Analysis of HCR Transient Transfections into 293T Cells

Vectors that were constructed for *in vitro* transcription of HCR regions with a 6x myc tag were transiently transfected into 293T cells and analysed by Western blot to determine whether soluble protein would be expressed at the correct size (12.5% glycine gel with Benchmark molecular weight markers (*M1*) and RPN800 Rainbow molecular weight markers (*M2*). Numbers indicate molecular weight in kilo-Daltons. Approximately 100µg of soluble cellular lysate was loaded into the other lanes. The membrane was probed with an anti-myc antibody, visualised by ECL and exposed to film for 10 seconds (**A**) and 5 minutes (**B**).

Predicted HCR-Myc fusion sizes:

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Molecular Weight (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDEE</td>
<td>37.9</td>
</tr>
<tr>
<td>ERL</td>
<td>27.9</td>
</tr>
<tr>
<td>RKE</td>
<td>45.7</td>
</tr>
<tr>
<td>NPF</td>
<td>33.8</td>
</tr>
<tr>
<td>eGFP</td>
<td>37.9</td>
</tr>
</tbody>
</table>
Figure 3.8

A

B
Although many TDEE injected embryos at 50ng/μl appeared normal, some displayed retarded growth compared to the un-injected controls with possible problems in epiboly. Of these, some had abnormally oblong yolks or double yolks, where a portion of yolk is retained towards the tail. Coupled with the yolk distortion were skewed body axes. At 100ng/μl fewer embryos developed normally, most displaying retarded growth. Many embryos failed to properly converge an anteroposterior axis, instead developing a wider and less defined body axis (figure 3.9).

Many ERL injected embryos displayed shorter tails and defects in eye and CNS development. The severity of the eye phenotype varied from either the complete absence of an entire eye, a smaller eye, or in some cases, eyes that were missing lenses. The CNS of some of these embryos appeared either skewed or sometimes contained numerous rounded cells, perhaps cells that were either undifferentiated or undergoing apoptosis. Results were similar for both 50ng/μl and 100ng/μl injections (figure 3.10).

Embryos injected with RKE mRNA at both 50ng/μl and 100ng/μl produced embryos that looked relatively normal after 24hrs with the exception of a few that displayed a loss of mesodermal trunk tissue, reduced head size and slightly enlarged pericardial cavities that surround the heart (figure 3.11 B, C at 48 hpf). Interestingly, when allowed to develop a further 24 hours, many of these embryos were less pigmented when compared to un-injected controls. The intensity of the pigmented cells that are normally present by 48 hpf were decreased but were nonetheless present. Apart from the reduction of pigmentation, the embryos appeared normal (figure 3.11).

NPF injections at 50ng/μl also caused eye defects giving rise to smaller, deformed or absent eyes on one side of the embryo. Often these defects were coupled with a malformed CNS, again displaying a lack of definition along the brain and anteroposterior axis. At 100ng/μl, similar eye defects were again observed, but in some cases, was only a single medially positioned eye. This was possibly due to an imbalance in the division rates of cells between the two sides allowing one side to overgrow the other, giving rise to a curved morphology (figure 3.12).
Figure 3.9

Pilot Injections with TDEE mRNA

Embryos injected with mRNA encoding for the FAM highly conserved region TDEE at 24 hpf. Un-injected control embryo (A). Embryos injected at 50μg/μl (B, C) and at 100μg/μl (D, E). Defects were apparent as yolk abnormalities and skewed body axes. Affected embryos were generally less developed than un-injected controls as judged by comparing eye and CNS development.

Bar represents 250μm
Figure 3.10

**Pilot Injections with ERL mRNA**

Embryos injected with mRNA encoding for the FAM highly conserved region ERL at 24 hpf. Un-injected control embryo (A, E). Embryos injected at 50μg/μl (B-D) and at 100μg/μl (F). Defects included: abnormal ball-like cells disrupting CNS development (B), lack of lens (C) or smaller eye (D) in eye development, and skewed body axes (F).

Bar represents 250μm
Figure 3.11

Pilot Injections with RKE mRNA

Embryos injected with mRNA encoding for the FAM highly conserved region RKE at 48 hpf. Un-injected control embryo (A). Embryos injected at 50μg/μl (B-D) and at 100μg/μl (E). Affected embryos appeared to be less pigmented than un-injected controls. Other defects observed were enlarged pericardial cavities that surrounded the heart (B, C) and kinks in the tail (B, D).

Bar represents 250μm
Figure 3.12

Pilot Injections with NPF mRNA

Embryos injected with mRNA encoding for the FAM highly conserved region NPF at 24 hpf. Un-injected control embryo (A). Embryos injected at 50μg/μl (B, C) and at 100μg/μl (D, E). Eye defects observed included missing eyes (B), missing lens (C), medially positioned deformed eyes (D) and smaller eyes (E)

Bar represents 250μm except in (E), where it represents 100μm
Although not conclusive and by no means complete, the results of the pilot study indicated that injection of exogenous HCR mRNA could in fact cause developmental defects in tissues such as the eyes, CNS and perhaps post-migratory differentiated pigment cells. As \textit{in situ} analysis had detected strong expression of \textit{usp9} in these cell types (\textit{figure 3.6}), a more rigorous series of experiments to determine any statistical significance were performed.

**EXPERIMENTAL DESIGN OF THE DOMINANT-NEGATIVE DOMAIN SCREEN**

Although ideally this experiment has only one variable i.e., the presence or absence of exogenous protein, in reality there are numerous variables that can also affect a developmental defect. These variables needed to be addressed to ensure reproducibility. The following variables were considered and appropriate controls were built into the experimental design.

*Embryo Quality:* Zebrafish embryos can be collected daily and at any one harvest, embryos from several different tanks are combined into a collection. By regulating the light cycle of various tanks, it is possible to collect embryos several times a day. However, as there were numerous fish in each tank, the fertilised eggs were not genetically homogenous. A non-uniform source of embryos could have presented problems as the viability and hence developmental capacity of certain clutches of embryos can vary.

For the purpose of expediting experiments, the embryos from four daily time point collections were used for injections. To control for variations between embryo clutches, twenty un-injected embryos were put aside at the time of injection. Only when this control group developed normally were data from the injected embryos collected for analysis.

*Injection Technique:* Another consideration was that the delivery of the mRNA template into the embryo is extremely invasive. The glass needle used to deliver the mRNA must puncture through the yolk cell into one of the blastomeres. This trauma alone may cause developmental defects. Removing the needle following injection may not only draw out part of the cell, but also the injected RNA back into the yolk cell. Diffusion of the RNA to the other cell may occur via the cytoplasmic bridges that interconnect the blastomeres through the yolk cell.
To control for any non-specific effects arising from the physical process of injection, a distilled water control was included. As distilled water is considered inert, it should not cause developmental defects due to RNA translation.

**Non-Specific Effects of mRNA Injection:** It is also possible that injection of RNA from any gene may cause defects. Such non-gene-specific effects may represent a generic reaction to alien RNA and/or expression of exogenous protein. To control against this possibility, the fluorescent protein eGFP was cloned and included in the injections as it is considered inert during development and is used in many developmental biology experiments. Further, as eGFP will fluoresce under UV light, the ability of the vector to express protein was determined by checking for embryos that glowed green.

**Reproducibility:** In order for the experiments to be reproducible, there must be enough data to achieve statistical significance. To reduce variability, the experiment was limited to the number of embryos that could be injected at any one sitting, as opposed to collating data from a number of injections from different clutches. The number of embryos that could be injected was determined by the time it takes for two-cell stage embryos to progress to the four-cell stage. This number could vary between twenty to seventy embryos depending on the quality of the clutch however, an average of forty embryos were injected per sitting.

To assess the reproducibility of the experiments, each RNA sample was independently transcribed twice. By comparing the results of the two independent samples, one can determine whether developmental defects consistently arose as a result of the RNA species and not by another anomaly. As the pilot study showed that developmental defects could result from injection of RNA at 50ng/µl, this concentration was chosen for injection.

To control against experimental bias in the interpretation and scoring of phenotypes, the experiment was conducted blind. All of the injection samples were randomised and assigned a letter of the alphabet by an independent researcher. The sample assignments were not revealed to this researcher until the injections and analysis were completed.
DATA COLLECTION

The mRNA from the two separate transcription reactions and the distilled water control, were each aliquoted into tubes in duplicate (in case a particular injection needed to be repeated), blindly randomised and injected. An RNA gel was later run to verify that the aliquots did in fact contain RNA at 50ng/μl, which had not degraded in the preparatory steps (data not shown). After the full completion of the experiments, embryos were further checked for fluorescence to determine the expression of the control eGFP. Embryos injected with eGFP did fluoresce albeit with low intensity however, expression of the protein did not appear to be confined to only one-half of the embryo (data not shown).

After each blind injection, the embryo clutch was first checked by assessing the viability of the un-injected controls. Where the un-injected controls were not viable at 24 hpf, the results of the injection were discarded and the mRNA sample was reinjected from a fresh aliquot. When the embryo clutch quality was assured, the results of the injections were analysed and scored at 24 hpf and again at 48 hpf to check for any pigmentation abnormalities.

Embryos were described, illustrated and scored. After the entire series of injections were performed, the data was collated and defined phenotype categories were established that best fitted the array of developmental defects observed (Figure 3.13). The categories into which the observed phenotypes were placed were:

**Gross Disruption:** Embryo does not contain any features that resemble a properly developed structure.

**One eye:** Embryo appears to be missing an eye on one side of its head.

**Cyclops:** Embryo has only one centralised eye.

**Fused eyes:** Embryo contains one centralised eye that appears to consist of two fused eyes but share a common lens. The eye resembles a pair of goggles.

**Small eye:** Embryo contains an eye that is slightly smaller than the other.

**Small eyes:** Embryo eyes smaller than wildtype.
Figure 3.13

HCR mRNA Injection Phenotype Categories

Hand sketches (A-L) and light microscope photographs (M-Q) of phenotypes that were observed from the injection of HCR mRNA.

A. Wildtype
B. Gross Disruption
C. One Eye
D. Cyclops
E. Fused Eyes
F. Blunt Head
G. Small Eye
H. Headless
I. Skewed Head
J. Kinky Tail
K. Swollen Hindbrain Ventricles
L. Pericardial Edema
M. Pericardial Edema (Photo of L)
N. Fused Eyes (Photo of E)
O. Small Eyes (Normal to the right)
P. Cyclops (Photo of D)
Q. Stumpy Tail
Skewed Head: The head of the embryo has not developed symmetrically, instead curving around its axis.

Stumpy tail: Embryo has an unusually short tail that has not developed properly.

Kinky tail: The tail of the embryo does not develop straight but kinks.

Headless: Embryo is missing defined head structures.

Blunt head: Embryo appears to have a shorter forebrain that does not protrude past the eyes.

Swollen Hindbrain: Embryo lacks a defined anteroposterior axis, instead appears split and open.

Ventricles: Embryos have enlarged pericardial cavities, a transparent structure that rests above the yolk and encases the heart.

Pericardial Edema: No apparent developmental defects. The yolk is rounded with a yolk extension, and numerous V-shaped somites are visible along the trunk. The eye has developed a lens and lobes of the brain are apparent.

STATISTICAL ANALYSIS

Each embryo was then reassessed to determine whether they carried one or more of the abnormalities and the data for each injected RNA and controls were then compiled into a spreadsheet for statistical analysis. Due to the amount of variation in the raw data between the independent transcriptions (data not shown), the data for each type of injected mRNA was pooled in an effort to observe some statistical significance. For each mRNA, the number and percentage of embryos that carried a particular phenotype were compared to the grouped negative controls of eGFP and distilled water (table 3.2 B). A Pearson Chi-Square test was performed to determine if there was any significant difference between the three sets of data (table 3.2 A). In some cases, the injected mRNA did significantly increase the occurrence of particular phenotypes compared to both negative controls (p-value <0.05). TDEE injections significantly increased the occurrence of Skewed Head, Small Eye and Pericardial Edema phenotypes; ERL increased Small Eye, Kinky Tail and Pericardial Edema phenotypes; RKE increased Blunt Head and Pericardial Edema phenotypes while NPF did not significantly cause any phenotypes. Disturbingly, the eGFP
## Table 3.2

### Statistical Analysis of HCR mRNA Injections

**A)** Data was processed by cross-tabulating embryo phenotype against the type of injected HCR mRNA (including the eGFP and H$_2$O negative controls) and then assessed for significance with a Pearson Chi-Square test. A p-value less than 0.05 was obtained where there was a significant difference between the negative controls eGFP and H$_2$O, and the injected mRNA (*p*-value < 0.05 in italics). Where the significant difference is the result of the injected mRNA the p-value is listed in **bold italics**. Where the significant difference is the result of eGFP the p-value is preceded by †. Where there are no embryos scored for both the injected mRNA and the negative controls, there is no p-value.

**B)** Percentage and counts for the negative controls eGFP and H$_2$O.
### Table 3.2

<table>
<thead>
<tr>
<th>Injected mRNA</th>
<th>Gross Disruption</th>
<th>One Eye</th>
<th>Cyclops</th>
<th>Fused Eyes</th>
<th>Skewed Head</th>
<th>Small Eye</th>
<th>Small Eyes</th>
<th>Stumpy Tail</th>
<th>Kinky Tail</th>
<th>Headless</th>
<th>Blunt Head</th>
<th>Swollen Hindbrain Ventricles</th>
<th>Pericardial Edema</th>
<th>Wildtype</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TDEE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of embryos</td>
<td>3.6%</td>
<td>1.8%</td>
<td>1.8%</td>
<td>1.0%</td>
<td>0.0%</td>
<td>13.4%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>1.8%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>2.7%</td>
<td>19.6%</td>
<td>75.0%</td>
<td>112</td>
</tr>
<tr>
<td>Count</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>9</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>22</td>
<td>4</td>
<td>112</td>
</tr>
<tr>
<td>p-value</td>
<td>0.219</td>
<td>0.003</td>
<td>0.392</td>
<td>0.161</td>
<td>0.013</td>
<td>0.066</td>
<td>0.000</td>
<td>0.985</td>
<td>0.513</td>
<td>0.000</td>
<td>0.013</td>
<td>ta</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ERL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of embryos</td>
<td>13.8%</td>
<td>12.9%</td>
<td>4.0%</td>
<td>1.0%</td>
<td>3.0%</td>
<td>28.7%</td>
<td>4.0%</td>
<td>0.0%</td>
<td>19.8%</td>
<td>4.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>22.8%</td>
<td>48.5%</td>
<td>101</td>
</tr>
<tr>
<td>Count</td>
<td>17</td>
<td>13</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>27</td>
<td>4</td>
<td>0</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>47</td>
<td>101</td>
</tr>
<tr>
<td>p-value</td>
<td>0.059</td>
<td>0.134</td>
<td>0.122</td>
<td>0.568</td>
<td>0.204</td>
<td>0.000</td>
<td>0.122</td>
<td>0.001</td>
<td>0.000</td>
<td>0.122</td>
<td>0.000</td>
<td>ta</td>
<td></td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td><strong>RKE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of embryos</td>
<td>14.0%</td>
<td>5.6%</td>
<td>2.1%</td>
<td>0.7%</td>
<td>0.7%</td>
<td>3.5%</td>
<td>1.4%</td>
<td>2.1%</td>
<td>4.2%</td>
<td>2.8%</td>
<td>5.6%</td>
<td>0.0%</td>
<td>13.3%</td>
<td>67.8%</td>
<td>143</td>
</tr>
<tr>
<td>Count</td>
<td>20</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>0</td>
<td>19</td>
<td>97</td>
<td>143</td>
</tr>
<tr>
<td>p-value</td>
<td>0.218</td>
<td>0.055</td>
<td>0.331</td>
<td>0.694</td>
<td>0.694</td>
<td>0.480</td>
<td>0.480</td>
<td>0.000</td>
<td>0.597</td>
<td>0.238</td>
<td>0.040</td>
<td>0.007</td>
<td>ta</td>
<td>0.008</td>
<td></td>
</tr>
<tr>
<td><strong>NPF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of embryos</td>
<td>8.2%</td>
<td>3.3%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>4.9%</td>
<td>0.0%</td>
<td>4.9%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>8.6%</td>
<td>83.8%</td>
<td>61</td>
</tr>
<tr>
<td>Count</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>51</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>p-value</td>
<td>0.63%</td>
<td>0.039</td>
<td>0.000</td>
<td>0.351</td>
<td>0.000</td>
<td>0.575</td>
<td>0.001</td>
<td>0.506</td>
<td>0.176</td>
<td>0.303</td>
<td>0.000</td>
<td>ta</td>
<td></td>
<td>0.004</td>
<td></td>
</tr>
</tbody>
</table>

† p value = <0.06 refers to H20

---

### B

<table>
<thead>
<tr>
<th>Injected mRNA</th>
<th>Gross Disruption</th>
<th>One Eye</th>
<th>Cyclops</th>
<th>Fused Eyes</th>
<th>Skewed Head</th>
<th>Small Eye</th>
<th>Small Eyes</th>
<th>Stumpy Tail</th>
<th>Kinky Tail</th>
<th>Headless</th>
<th>Blunt Head</th>
<th>Swollen Hindbrain Ventricles</th>
<th>Pericardial Edema</th>
<th>Wildtype</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>cGFP</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of embryos</td>
<td>9.8%</td>
<td>15.1%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>3.3%</td>
<td>0.0%</td>
<td>9.8%</td>
<td>1.6%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>3.3%</td>
<td>1.6%</td>
<td>65.6%</td>
<td>61</td>
</tr>
<tr>
<td>Count</td>
<td>8</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>40</td>
<td>61</td>
</tr>
<tr>
<td><strong>H2O</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% of embryos</td>
<td>4.7%</td>
<td>2.3%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>2.3%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>2.3%</td>
<td>90.7%</td>
</tr>
<tr>
<td>Count</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>39</td>
<td>43</td>
</tr>
</tbody>
</table>
negative control was not entirely negative with only 66% of embryos developing normally (compared to TDEE 75%, ERL 47%, RKE 68%, NPF 84% and distilled water 91% normal). eGFP was originally included in the experiment to control for possible non-specific effects of generic RNA injection. Given that eGFP was able to cause statistically significant developmental phenotypes in relation to the injected mRNA (in particular; one eye, stumpy tail and swollen hindbrain ventricles), it would appear that this system contains high background due to non-specific RNA effects. Indeed this non-specific phenomenon has been observed by others (Harvey P, Key B, Kortschak D, personal communication). These results do not discount the possibility that some of these FAM HCRs may cause an increase in certain developmental effects but due to the noise in the system, observations of these effects were confounded.

In these series of injections, no embryos were observed that displayed a strong lack of pigmentation as observed in the pilot injections of RKE mRNA. The absence of pigmentation deficient embryos may have been due to either chance or an insufficient concentration of RNA. To assess whether the pigmentation phenotype was a reproducible observation, a smaller scale experiment was conducted with a higher concentration of mRNA (100µg/µl). RKE, ERL (which, in the previous experiment, significantly increased certain developmental phenotypes) and eGFP mRNA was injected along with distilled water into 2-cell stage embryos. All embryos were described and scored after 48 hours to determine whether pigmentation had been disrupted. In order to determine whether the high degree of variation in the previous experiment was due to either the particular transcription of the mRNA, the clutch of embryos or experimental technique, two different samples were randomly and blindly selected to be injected twice each into two different clutches of embryos with the same needle and aliquot of mRNA.

After completion of the injections, it was clear from the data that statistical analysis would again be tenuous due to the high degree of variation in the results. The results were not pooled to demonstrate the variability between the injections of the different transcriptions, and even between injections of the aliquot (table 3.3). Although RKE mRNA injections did affect pigmentation, so did the negative controls of eGFP and distilled water. Ironically, due to the degree of variation in the injection of eGFP and distilled water, ERL mRNA was actually more consistently negative than the negative controls. Interestingly, when injecting two different clutches of embryos with the same needle and aliquot of
Table 3.3

Assessment of Pigmentation Phenotype

To determine whether injection of RKE mRNA into 2-cell stage zebrafish embryos could decrease pigmentation, RKE, ERL and eGFP mRNA at 100μg/μl along with distilled water were injected into embryos and scored 2 days later. For each sample, two independently transcribed mRNAs were injected (labelled 1 and 2). RKE 2 and eGFP 1 were blindly selected at random to be injected twice into two different clutches of embryos with the same needle to check consistency. Due to the high degree of variation amongst all the categories (including the negative controls eGFP and distilled water), statistical analysis was not performed.
mRNA, the observed survival rate of the injected embryos was very consistent (±0.4%) however, no strong conclusions can be drawn on the basis of only two data sets.

Due to the amount of variation observed from the injection of mRNA into zebrafish embryos, it was concluded that the system was untenable. The amount of background noise in these experiments was particularly demonstrated by the variation observed in the negative controls. There are a number of potential causes for this variation. Firstly, the embryos themselves came from different populations of stockfish. Although not a major problem, there were instances where even the un-injected embryos died unexplainably. Another problem with the experiments is the injection procedure itself, which is very invasive. Some embryos may have developed poorly solely from the trauma of being punctured (as potentially demonstrated by the distilled water injections). Further, the needles used in the experiments were not consistent. Injection needles were hand made by heating and drawing capillary tubes with a needle-puller and were of varying thicknesses. Before the experiment, the end of the needle was broken off to unseal it and this made for inconsistent needle tips. The amount of injected material was also variable as a foot pump, which did not allow for the delivery of a regulated volume, controlled injection of the mRNA. This volume was not calibrated or checked. There may have also been subjective technical variation. As the technique of injecting needed to be learned and practised, initial injections may not have been as consistent as the later injections. Lastly and perhaps most importantly, there may be general non-specific reactions to the injection of foreign mRNA (as potentially demonstrated by the eGFP injections). As the problems with the experiment were potentially widespread and difficult to fully control against, further studies were not pursued.

**Expression of FAM HCR Protein**

As it had proven difficult to identify FAM-binding proteins by the reverse approach of the dominant-negative domain screen, a forward approach was adopted to first identify a FAM-binding protein, and then assess its developmental relevance. As the FAM HCRs had already been cloned into Gateway entry vectors, these fragments were chosen as the
protein baits. Previous attempts at expressing FAM fragments as protein baits had run into solubility problems due to arbitrary definitions of domain boundaries. As the HCR fragments were designed based on amino acid conservation, it was hoped that the chances of expressing conserved soluble domains would be higher.

**GST Fusions**

The FAM HCRs were cloned into a GST fusion vector for bacterial expression due to the ability of bacteria to inexpensively express large quantities of exogenous protein. All the HCR fragments were relatively small and when fused to GST, their sizes ranged from 47 to 62 kDa, a size that is routinely expressed in bacteria. The IPTG-inducible GST expression system was chosen as the GST tag can aid solubility, acts as an epitope for immuno-detection, and allows for easy purification and pull-downs under non-denaturing conditions by affinity to immobilised glutathione.

The FAM HCRs were all cloned into a Gateway compatible destination vector pDEST15 that generates an N-terminal GST fusion. Curiously, these plasmids were extremely difficult to clone, yielding clones with variable digest patterns despite the efficiency of Gateway (95% recombinants in the correct orientation and reading frame, Invitrogen 2003) however, similar rare problems with gateway cloning have been reported (Invitrogen website forum, www.invitrogen.com). When correct clones were finally isolated, they were transformed into BL21 (DE3) cells, a bacterial line suited to protein production. Protein expression of the HCRs was initially conducted in 50ml cultures at 37°C however it is important to note that the bacteria transformed with these constructs had an unusually long generation time. A typical culture would normally reach an inducing optical density (OD) of 0.6 in 2.5 - 3 hours, a point where the bacteria are in a log growth phase. The GST-HCR fusion cultures however only reached an OD of 0.06 - 0.09 in this time. As such, these cultures were not able to be induced at an OD 0.6 but were instead induced after approximately 3 hours with [0.1mM] IPTG for approximately 4 hours. The pellets from the induced bacteria were then lysed by sonication.

At 37°C the TDEE, ERL and RKE fragments appeared to be expressed relatively strongly as gauged by Coomassie staining but not so for ERL (data not shown). Also evident from the gels is the appearance of strong bands of protein corresponding to the predicted sizes of
the fusions prior to induction, raising the possibility of lax expression regulation. Of more concern was that almost all of the expressed protein was found in the insoluble fraction.

Protein insolubility in bacteria can be caused by a number of reasons. Besides the possibility of a disrupted domain, high expression of foreign proteins can sometimes lead to the formation of inclusion bodies. Aggregation of foreign proteins is generally favoured at higher temperatures due to the strong temperature dependence of the hydrophobic interactions that cause aggregation (Kiefhaber et al. 1991) and so in some cases, solubility can be improved by decreasing the incubation temperature. Its effect on solubility is two fold. Firstly, temperature reduction partially eliminates certain heat shock proteins that are induced under over-expression conditions (Chesshyre and Hipkiss 1989), and secondly, lower temperatures (around 30°C) are conducive to the expression and activity of a number of *E. coli* chaperones (Mogk et al. 2002, Ferrer et al. 2003). Expression of the GST-HCR fusions was therefore attempted at 25°C however there was little improvement in solubility (data not shown).

The GST fusions may also be insoluble due to inefficient lysis, trapping the soluble protein with the insoluble cellular debris. A previous study found that lysing the bacterial pellets by French pressing rather than sonication increased yields of soluble protein (Khut 2000). As it is unlikely the lysis process itself could increase solubility, this observation may be the result of the higher lysis efficiency of French pressing (~95% lysis), liberating more protein into the supernatant.

To accommodate the volumes necessary for French pressing, the cultures were increased to 400ml and grown at both 37°C and 25°C (figure 3.14). Although the results were clearer (less background bands in the soluble and insoluble fractions), the expressed protein was still largely insoluble at 37°C and had not improved at 25°C. The leaky expression of TDEE protein prior to induction did however appear to have been suppressed at 25°C. ERL and NPF were expressed weakly, if at all. Due to both poor expression and a lack of solubility, further studies into expressing GST-HCR fusions in bacteria were not pursued.
**Figure 3.14**

Expression of HCR-GST Fusions in Bacteria

Solubility of the HCR-GST fusions induced at 37°C (A, B) and 25°C (C, D) and lysed by French press, were compared by Coomassie staining (12.5% glycine gel Benchmark molecular weight markers (M). Numbers indicate molecular weight in kilo-Daltons). Amount of samples loaded in the uninduced (UI) lane represents a 1:1200 dilution of the initial 400ml culture while a 1:2400 dilution was loaded in the induced (I) lane. Samples in the soluble (S) and insoluble (IS) lanes represents a 1:12000 dilution of the initial 400ml culture. Arrows indicate the position of the visible expressed protein.

A) Expression of TDEE (left) and ERL (right) GST fusions at 37°C  
B) Expression of RKE (left) and NPF (right) GST fusions at 37°C  
C) Expression of TDEE (left) and ERL (right) GST fusions at 25°C  
D) Expression of RKE (left) and NPF (right) GST fusions at 25°C

Predicted HCR-GST fusion sizes:

<table>
<thead>
<tr>
<th>Protein</th>
<th>Size (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TDEE</td>
<td>56</td>
</tr>
<tr>
<td>ERL</td>
<td>47</td>
</tr>
<tr>
<td>RKE</td>
<td>62</td>
</tr>
<tr>
<td>NPF</td>
<td>52</td>
</tr>
</tbody>
</table>
MAMMALIAN EXPRESSION

Although the HCRs could not be expressed as soluble GST fusions in bacteria, it was previously shown that they could be expressed as 6x Myc tag fusions in 293T cells (figure 3.8). Investigation was thus carried out to determine whether HCR bait protein could be expressed in the same cell line. Although it is possible to purify the HCR-Myc tag fusions on an affinity column, this process requires a large amount of antibodies, which can be expensive. To address this issue, the FAM HCRs were cloned into pEF6-DEST51, a vector that adds a C-terminal V5 6x His tag fusion. The V5 epitope allows for immuno-detection while the His tag allows for purification on a nickel column. Similar to GST, the 6x His tag can immobilise the bait protein to a column, allowing the passing of cellular extracts over it. Although cellular proteins with a stretch of histidines will have affinity to the column, it is unlikely that they will bind with the same strength due to the sheer number of consecutive histidine residues in the His tag. Stringent washing should be able to reduce background binding.

The four HCRs were cloned into the pEF-DEST51 vector and verified by restriction digests. Similar to the previous cloning of the HCRs into the pDEST15 (GST fusion) vector, isolation of individual clones that gave the expected restriction patterns was difficult. The cause of the problems associated with their cloning is not clear as the Gateway LR reaction inserts the fragment in the correct orientation via site directed recombination. As the reaction should not alter the insert nor vector DNA, and the sequence of both the vector and the inserts (which had already been sequenced) were known, the heterogeneity observed in the restriction patterns remains an anomaly. Given the excessive amount of time invested in attempting to isolate clones that gave correct restriction patterns, the clones that were eventually selected were ones that best fit the predicted restriction patterns. These clones contained the predicted sized bands but would also contain extra bands despite long digestion times with ample quantities of restriction enzyme. It is also curious to note that 50ml DNA preparations of the TDEE, ERL and RKE plasmids gave unusually low yields ranging from ~50ng/µl to 115ng/µl. Normally one would expect yields of around 300-600ng/µl. This is most likely due to unusually slow growth of the cultures transformed with these plasmids.
The four HCR vectors were transiently transfected into 293T cells along with full length FAM (SSSRF+) as a positive control, which had been previously cloned into the same vector. Full length FAM was detected by western analysis with a V5 antibody whereas none of the four HCRs expressed (data not shown). Mammalian expression of the FAM HCRs was not pursued further.

It is curious that the FAM HCRs could be expressed as Myc tag fusions but not as V5 His tag fusions in 293T cells. This is not due to size limitations as the V5 His tag (23 amino acids) is smaller than the Myc tag (84 amino acids). It is possible that the Myc tag, but not the V5 His tag, improves the HCR solubility. Alternatively the problem may have been in the initial Gateway cloning. It is important the note that the difficulty in cloning these particular fragments is an isolated case and other Gateway reactions performed by this researcher and others has achieved the type of high fidelity as claimed by the manufacturers.

The issues encountered for both mammalian and bacterial expression could have been further investigated. To ensure that the HCRs were correctly cloned into the His tag vector, the inserts could have been cloned by traditional methods, but this would require a new round of PCR and sequencing to remove the Gateway recombination sites for both the HCRs and the vector. Bacterial expression could have been revisited by cloning the HCRs with various other tags. Maltose binding protein (40 kDa) has proved to be a much more effective solubility partner than GST (26 kDa) and thioredoxin (15 kDa) (Kapust and Waugh 1999). If new fusion tags do not aid solubility, co-overexpression with combinations of molecular chaperones such as the GroEL, DnaK, ClpB and ribosome associated trigger factor families of chaperones may prevent the formation of inclusion bodies however these measures still do not guarantee soluble protein expression (Amrein et al. 1995, Nishihara et al. 1998, reviewed in Held et al. 2003). If the proteins were still insoluble then one might consider extracting the insoluble fusion proteins from inclusion bodies by isolating, denaturing and refolding them however, this is often unfavourable for a number of reasons. Major obstacles include poor recovery yields, the requirement for optimisation of refolding conditions for each target protein and the possibility that the resolubilisation process could affect the integrity and thus function of the refolded proteins (reviewed in Sørensen and Mortensen, 2005). Utilisation of refolded proteins in functional
analysis experiments (such as binding assays) is thus clouded with the uncertainty of whether the bait protein is in a native conformation.

Although these options were available for further pursuit, they are time consuming. Already much effort had been invested in the dominant-negative domain screen yet no positive results were obtained that could justify devotion of more time into experimentation with the FAM HCRs. Most significantly, although the HCRs were designed based on conservation, this in itself is no guarantee that they form a functional domain or are biologically relevant. In the absence of these assurances, studies using the FAM HCRs were not pursued.

Summary

With the aim of identifying developmentally relevant FAM-binding proteins, a dominant-negative domain screen was devised that sought to first identify protein interactions that cause a developmental abnormality, followed by identification of that protein. The zebrafish system was chosen for the screen due to reasons of speed, robustness, sensitivity and readout simplicity. To verify that a FAM orthologue exists in zebrafish and that they are sufficiently conserved to allow for cross-species experimentation, the zebrafish orthologue of *Fam* (*usp9*) was identified from database searches. This computer generated sequence prediction was cross-referenced against other known and predicted orthologues and corrected by examination of electrophoreograms. A phylogenetic tree was generated that showed that the zebrafish and mouse genes are closely related (92.5% amino acid similarity and 89.7% identical) and are most likely true orthologues. *In situ* analysis of *usp9* expression shows extensive correlation with the previously described mouse patterns (Wood *et al.* 1997) that show early ubiquitous expression that becomes neurally confined at mid-gestation.

In the absence of structural data, four regions of FAM were designed based on conservation between the mouse and drosophila amino acid sequences. It was rationalised that sequence conservation is indicative of functional conservation and so represented the
best chance at cloning a functional FAM domain as bait for the dominant-negative domain screen. These FAM HCRs were cloned into a vector that allowed both in vitro transcription of mRNA for injection into zebrafish embryos, and in vivo expression in mammalian cells. Transient transfections of the cloned vectors into 293T fibroblast cells demonstrated soluble, full-length expression of all the constructs. Pilot injections of the FAM HCRs into one side of 2-cell stage embryos indicated that injection of mRNA could perturb development, which gave impetus to conduct a more rigorous set of experiments. The FAM HCRs were reinjected into zebrafish embryos as before, along with an mRNA negative control eGFP, and distilled water to control for injection technique. The results however were not statistically significant due to high background in the negative controls. The eGFP injections in particular were able to generate many phenotypes, which suggested a general reaction to the injection of RNA. A smaller scale version of this experiment showed a high degree of variation when injecting both different transcriptions of the same mRNA, and injections of the same aliquot of mRNA. Due to these factors that could not be controlled, these experiments were abandoned.

Attempts were made at expression of the FAM HCRs as protein baits for immobilisation on an affinity column to allow identification of binding-partners from cellular extracts. A bacterial expression system was initially chosen for its ability to inexpensively produce a large amount of protein. The FAM HCRs were cloned into a GST fusion vector and expressed at 37°C and 25°C, however all the fusions (if expressed) were found in the insoluble fraction. Lysis by sonication and French pressing did not alter this outcome. As transient transfections of the HCRs into 293T fibroblasts previously demonstrated soluble expression of the proteins as Myc tag fusions, further investigations were conducted into their expression in a mammalian system. The HCRs were transferred into a V5-His tag fusion vector to allow for future purification of the proteins on a nickel column. Westerns performed on the soluble lysates of the transient transfections with these vectors did not however detect expression. It is important to note that contrary to the principles of Gateway cloning (high efficiency, correct orientation, no requirement to sequence), these vectors were extremely difficult to clone. Although more could have been done either to solubilise the fusions in bacteria or to construct new vectors, the effort could not be justified given that these experiments had not provided any evidence that these HCRs formed folding functional domains. As such, further experimentation with the FAM HCRs was abandoned. Given the problems and uncertainty associated with non-structurally
defined regions of FAM, the next chapter describes experiments that sought to biochemically define FAM domains.