2008

Cis-regulatory Analysis Of The Pigment Cell Differentiation Gene Polyketide Synthase

David Rogers
University of Central Florida

CIS-REGULATORY ANALYSIS OF THE PIGMENT CELL DIFFERENTIATION GENE POLYKETIDE SYNTHASE

by

DAVID JOSEPH ROGERS
B.S. Truman State University, 2001
B.A. Truman State University, 2001

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Biology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2008

ABSTRACT

The analysis of Gene Regulatory Networks (GRNs) is essential to understanding the complete process of embryo development. Elucidating every gene regulatory circuit from maternal regulatory inputs all the way to the activation of differentiation gene batteries is an important step in increasing our understanding of developmental biology. In this work I study the *cis*-regulatory architecture of a pigment cell differentiation gene, *polyketide synthase* (*SpPks*) in the sea urchin *Strongylocentrotus purpuratus*. *SpPks* encodes an enzyme that is responsible for the biosynthesis of the sea urchin pigment echinochrome in larval pigment cells. The analysis of the promoter of a differentiation gene will lead to identifying the direct upstream regulators and ultimately to elucidating the structure of the upstream gene regulatory network, which is mostly uncharacterized.

From previous studies the transcription factors *SpGcm* and *SpGatae* are predicted to be positive regulators of *SpPks*. Here, I identify a minimal 1kb promoter region containing putative DNA-binding sites for both GCM and GATAE that is able to recapitulate the expression of *SpPks*. I further show by mutagenesis that a putative DNA-binding site for GCM located 1,179 base pairs upstream of the start of transcription is a direct target for the positive *cis*-regulation of *SpPks*. Quantitative analysis of the transcriptional regulatory function of the GCM-mutagenized construct suggests that GCM is not necessary for the start of *SpPks* transcription but is required for its maintenance. Several GATA E binding sites have been identified within the minimal promoter for *SpPks* by means of consensus sequence. My analysis suggests that GATA E may be a direct positive regulator and could potentially be required for the onset of transcription of

SpPks, though further experimentation will be necessary to characterize the exact regulatory function of GATA E.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Development of the Sea Urchin

After fertilization, the sea urchin zygote divides equally for the first few cleavages. During the fourth cleavage, the animal half of the embryo divides once more, producing eight equal cells called mesomeres. The vegetal half, however, divides unequally along the equator to produce four large macromeres as well as four smaller micromeres. The sixth cleavage occurs equatorially, producing two animal tiers, two vegetal tiers and a cluster of micromeres at the vegetal pole as seen in Figure 1. During blastula stage the veg2 and part of the veg1 descendants will form the vegetal plate, a thickened group of cells opposite to the animal pole, that constitute the endo-mesoderm territory. Between the 7^{th} and 9^{th} cleavage, a cell signaling from the micromeres to surrounding cells is required for the differential specification of Secondary Mesenchyme Cells (SMCs) and endoderm (Sherwood and McClay, 1999).

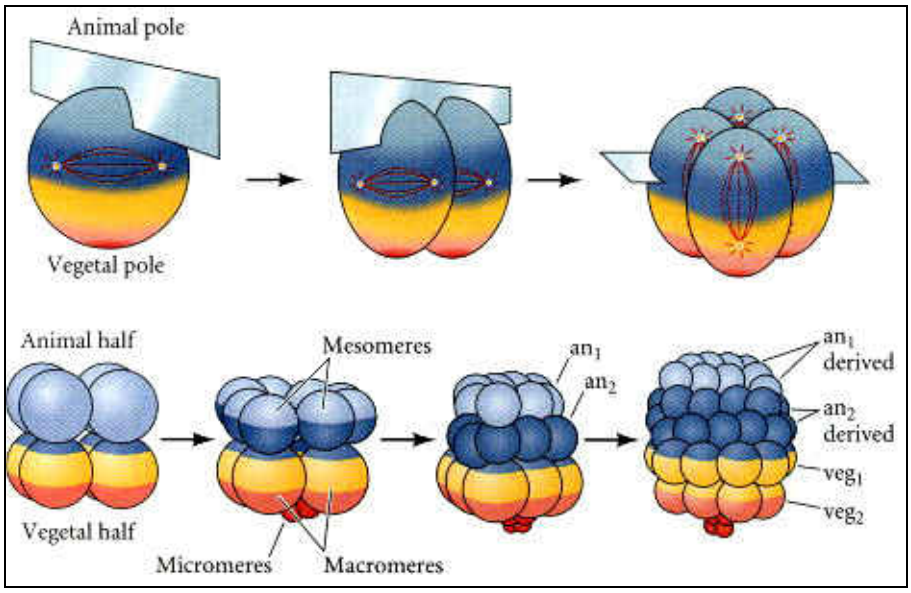


Figure 1: Early development of the sea urchin embryo from first to sixth cleavage. Cells colored in light and dark blue give rise to the ectoderm, cells in yellow become the endoderm and cells in red contribute to the mesoderm (Gilbert, 2000).

Development of the Secondary Mesenchyme Cells (SMCs)

Pigment cells are one of the four cell types that develop from Secondary Mesenchyme Cells (SMCs). The other three cell types deriving from SMCs are blastocoelar cells, coelomic pouches and circumesophageal muscle cells (Cameron *et al.,* 1991). It has been seen that many SMC precursors are already specified before the onset of gastrulation (Ruffins and Ettensohn, 1993). Using a monoclonal antibody for a cell surface protein expressed in pigment cells, it has been shown that pigment cell precursors are localized in the vegetal plate (Ruffins and Ettensohn, 1996). As illustrated in Figure 2, the dorsal (aboral) and ventral (oral) regions of the mesenchyme blastula stage embryo do not equally contribute to SMC derivatives. Pigment cells

are completely excluded from the ventral region whereas blastocoelar cells are not found in the dorsal region (Ruffins and Ettensohn, 1996).

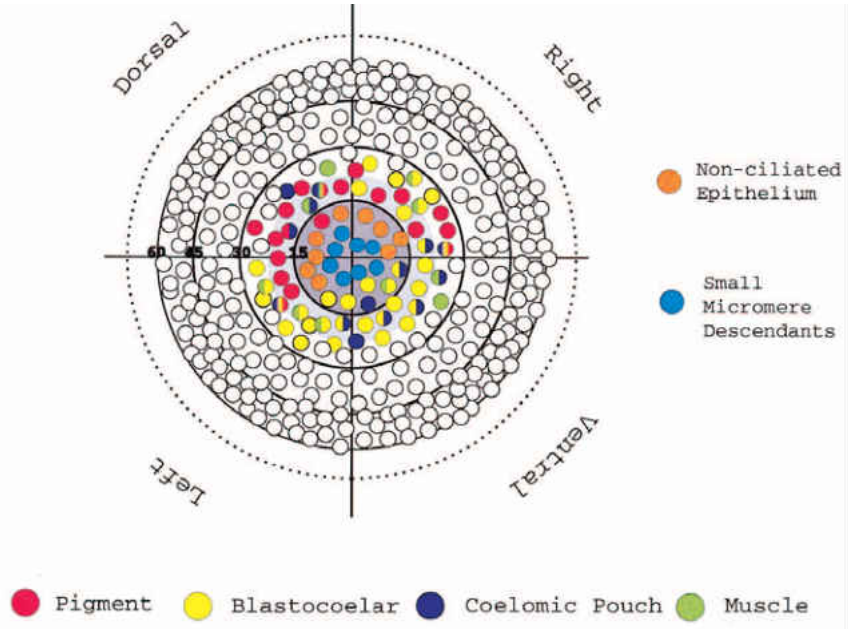


Figure 2: Fate map of Secondary Mesenchyme Cells. At mesenchyme blastula stage, approximate numbers of SMC precursors indicating their presumptive fate are shown (Ruffins and Ettensohn, 1996).

During gastrulation pigment cells detach from the tip of the archenteron and begin to invade the ectoderm. In *S. purpuratus,* at the end of gastrulation, approximately 30 pigment cells have become embedded in the aboral ectoderm and start to develop pigment granules (Gibson and Burke, 1985; Cameron *et al.,* 1991). These pigment cells or echinophores are long, slender, branched cells and seem to be distributed completely randomly throughout the ectoderm (Cameron *et al.,* 1991). Pigment cells might have a role in the immune system of sea urchin larvae. Their morphology and behavior are similar to macrophages. Pigment cells have pseudopodia that are able to rapidly extend and contract, and they are able to migrate within the ectoderm, potentially in response to immune stressors (Gibson and Burke, 1987).

Several genes have been found to be specifically expressed during specification and differentiation of pigment cells. The transcription factor *hmx* is expressed strongly in the blastula stage and continues to be expressed throughout development to the pluteus stage (Martinez and Davidson, 1997). Another transcription factor, *not,* is expressed at the mesenchyme-blastula stage but the expression only lasts through gastrulation (Peterson *et al.,* 1999). *Profilin,* which binds actin, is present from maternal inputs in the unfertilized egg with expression increasing at the onset of gastrulation (Smith *et al.,* 1994). An uncharacterized gene, *S9* is present in late blastula stage embryos and continues to be expressed through development in a pattern coincident with pigment cells (Miller *et al.,* 1996). Three genes, *capk*, a cAMP-dependent protein kinase, *dopt*, a dopachrome tautomerase-like gene and the uncharacterized *PI103,* are present at both 24 and 48 hours post fertilization in pigment cells or their precursors (Rast *et al.,* 2002). Interestingly, the transcription factor, *glial cells missing (SpGcm)* begins to be expressed between 10 and 12 hours post fertilization in pigment cell precursors, a few hours before the onset of expression of a number of pigment cell specific enzymes including *polyketide synthase* (*SpPks), sulfotransferase (SpSult), flavin monoxygenases (SpFmo1, SpFmo2,* and *SpDimethilaniline monooxygenase*, similar to *fmo3*; Ransick *et al.,* 2002; Calestani *et al.,* 2003). Considering that *SpGcm* is known to positively regulate the expression of *SpSult, SpFmo1 and SpPks* (Davidson *et al.*, 2002) and given the close timing of their expression within the same cell type, it is hypothesized that *SpGcm* is a direct regulator of these enzymes. To test this hypothesis, I choose to look at the *SpPks* promoter in order to uncover the architecture of the genetic pathways regulating pigment cell development.

Polyketide Synthase

The *pks* gene family is very large and encodes for enzymes responsible for the synthesis of several polyketide compounds. Mostly bacteria and fungi produce polyketides, with a number also synthesized in plants. These polyketides have a wide variety of biological properties including antibiotic, antitumor, antifungal, immunosuppressive and predator defense (Reviewed in Hopwood, 1997).

The sea urchin *pks,* (as well as *SpFmo1*) are necessary for the synthesis of the sea urchin echinochrome pigment (Calestani et al., 2003). The sea urchin echinochrome is a naphthoquinone that belongs to a class of polyketide compounds (Griffiths, 1965). When synthesis of SpPKS and SpFMO was blocked, pigment cells developed but they did not produce the echinochrome pigment (Calestani *et al.,* 2003).

Given the biological properties of polyketides, it is possible that *SpPks* is involved in the immuno response of sea urchin larvae. Echinochrome A (Figure 3), which is produced in the coelomocytes of adults and pigment cells of larvae, has been shown to have antibiotic properties against many different bacteria (Service and Wardlaw, 1984). Based on phylogenetic analyses, the sea urchin *pks* was found to be most closely related to slime mold and eubacteria *pks* genes, and generally not to other animal or fungal genes (Castoe et al., 2006). The only two related animal *pks* were found in chicken and fish. Interestingly, chicken *pks* was isolated from macrophage and lymphocytes EST libraries (Locus: XP_418587: http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=protein&id=118085656), further supporting the hypothesis that this gene may have a role in the sea urchin immune system.

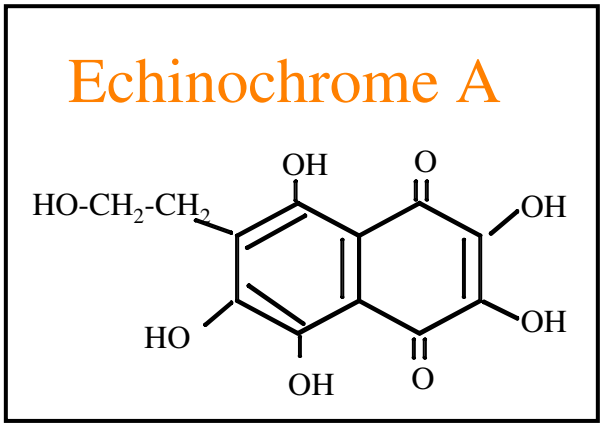
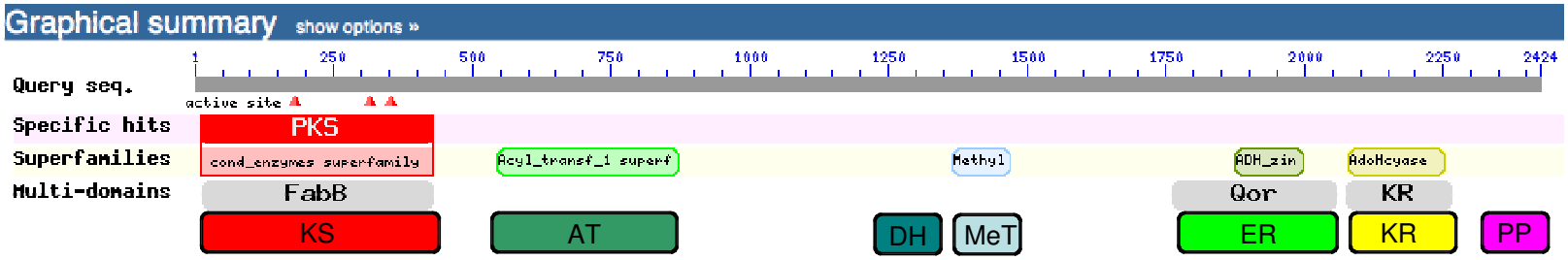


Figure 3: The chemical structure of Echinochrome A, a polyketide compound.

The particular class of PKS found in sea urchins belongs to Type I PKSs. This class typically includes the following conserved domains: a ketoacyl synthase (KS) domain, acyl transferase (AT) domains, and phosphopantetheine attachment sites (PP) as these three domains are necessary for the biosynthesis of the polyketide. The acyl transferase domain positions the next organic acid to be added while the phosphopantetheine attachment site anchors the growing polyketide, finally the ketoacyl synthase domain is what joins the organic acid to the growing chain by condensation (Hopwood, 2004). *SpPks* includes these domains as well as a dehydratase (DH), a methyltransferase (MeT), an enoyl reductase (ER), and a ketoreductase domain (KR) in the following order: KS-AT-DH-MeT-ER-KR-PP (Figure 4). These additional domains make modifications to the polyketide to produce the proper structure. The biochemistry involved in the synthesis of polyketides is diagramed in Figure 5.

A).


Graphical summary
show options »
Query seq.
1
250
500
750
1000
1250
1500
1750
2000
2250
2424
active site
Specific hits
PKS
Superfamilies
cond_enzymes superfamily
Acyl_transf_1 superf
Methyl
ADH_zin
AdoMcyase
Multi-domains
FabB
Qor
KR
KS
AT
DH
MeT
ER
KR
PP

B).

1 mgsnktswgy fpvavvgigt rhacganttd dfwkvlkegk ecildipper waidnfhded
61 qtrqgkmvtk rcgliddleg fdnlffkisp reaasldpqq rhllevnyea fedaginpdn
121 lgescgvfvg igmmdhaiql vdtsttdayt ltgiahsvsa nrisyafnlk gpsfavdtac
181 asgltalhla ctslwnrecs valmsacngi qlpditvgfs algvlspdgr cspfsstang
241 yvrsegwgai vlkplsqala dndhiytvir gsaiaangla nsltmpsppa qeyvmkeaye
301 kfgvsmsdvh yveahgtgtm vgdpleaeai srafnrtkdn plkigsvksn fghtevaagv
361 taaikvalmm enrtipptin fvssnphidp eemkldivtn vqpfptedkh iiglnsfgfa
421 galahcifee apkrpkkelt peqvcgwkfg dsdkegqpii iplsakspea ltavakqwqn
481 ldidqdamsa vswmstrrrv henrltviss sgkqfkaqmk dfvetggaen atsgtvysge
541 pkicmifpgq gqqygnmgrq lyktepvfkn tvdecdaifk kisgwsvlee kslfverphs
601 adykpdtfin dlevsqpsil fmqlglfnlw thwgvkpacv vghslgevsa ayacggmtle
661 eavetiyirs veqgklkgtg smaalrmtle earelcskhe rlyvaainap gstaiagntq
721 aieqiaadnp tiakqlrvqc afhtpdmdpt ektfkekmek vvktpagvrn ipfystltga
781 ryegdfktay wwdnirnave fqsavenvlr dfecdmflec asaatllssv nqivkgsgvk
841 iqlttiasgq rnqddrmcal rglanmhnng vslnwknitk dsaaytklpl ypwqhkpfml
901 epeyrrkrrl glddrtykgq ngqlsletfp fhsdrtakdk lvfpesgyve ymmeatsgen
961 elpvvnkvtf tqslewpeek tvtgtkkatl nldlvrdgnk veisykgdvc ssaeveegia
1021 qdntipvndi iqrcskktta edfysymqem gleygakfqe vnevclgdge svgylkpaqd
1081 nkqriqtthl dacfqlltyt lgarsslyqp amiesirmnv pslpagepll aytsiidcds
1141 walrgnvtit ltngkvlaei qgctckntsg tqtdidinkc lykrefqsvk ahlppikeva
1201 kvfdeenlrk rfpelmesvt raeqvfsnmg aiclayikhg ldqvpvkers dyldpryyrr
1261 leklkrdtsi rqikyedipk vkeemlkvap elkqelsmaq clgehlpttl rnpqsamtll
1321 fkpecmasyf ldsltttfyy kagaemvrqa vlkaletkat vrllevgarm gglthhileh
1381 ledlclegrv eyvftdlsva ffphardhlv dypfvkyqql dietdiesqg fvpgsvdili
1441 cldtlhstgh lqealyfmre licddgwmil yeattvkfia evifgalrlc wvfeddrpec
1501 cwleqnewke alekngfddv valsspkelf hsvligrkag gdgacinpks tpittrkqwl
1561 vvshpdnakf adlvksslsg svtslsydei mkadlgklkk dgsviealfi wnvdhdngfk
1621 vllnflqqig vnvenvcklw mvtfaatsga rpinaagagl vhaaanacqi pfvtvdipee
1681 vtngdkvwas rlvntmlgnk lsdmelvvkd givltprltr mqlpevkvne tpywqltqav
1741 dpfktessve dlgiayqdgl evapgtvlvk vsaaginkrd vdlardstvq kedtssfgme
1801 fcgvvekvge gvttvkpkde vlgfgthcla sytlahadlv vkkpknltps qaattsiafa
1861 tayyslvera nitngeslli qvadpglrda avqianhaga kvicsvddpt tatplkkmga
1921 mivptsssss fvndvnnvtn gagvdvvlns lqgkqmeksl ellaaggrfc sitdsnainf
1981 klqmrllqkn rsliscnies mnqhqkpllq rilrkvtdlm dkgklkpldv tsrpitdypt
2041 lfadesitna gkvaieipsa fkpnkvistt qlfkknatyv vtaaesglsq ifarwlynng
2101 arhiamcylm esgkskasrt vnyltrkgae vfeychqldv rgpdggiaki fgdlkkrnvp
2161 amrgifclgg yrlpgketms dvtfdslqam lsakvrpakl shimsdkmgl eldyfftlss
2221 ddvawgnpsa vasvtgdsyl esfalkrrle gkpalnlqvg alrgidayef ggqttlpvkd
2281 getslhveef lmvlgkllss pdtppcvcit nqdwesvlkf shdhtlkfrh laggeqvais
2341 ecklsledlq kqvknklgdl lcvnpdtidl rqpminygvd slmavemvtw asrelsvvis
2401 qldilggitt gvllekaidn svci

Figure 4: Conserved Domains of SpPKS. A) Results of the SpPKS sequence comparison to the NCBI Conserved Domain Database (http://www.ncbi.nlm.nih.gov/sites/entrez?db=cdd). Below the NCBI Graphical Summary is illustrated a summary of the domain structure employing a commonly used abbreviation nomenclature for each conserved domain found: KS, ketoacyl synthase; AT, acyl transferace; MeT, methyltransferase; ER, enoyl reductase; KR, ketoreductase. In addition putative domains found for dehydratase (DH) and phosphopantetheine

attachment sites (PP) were found by manual search of the SpPKS sequence B) Amino Acid sequence for SpPKS with putative dehydratase domains (hxxxxxxxxp) and putative phosphopantetheine attachment sites (gxxs) highlighted as indicated. GVDS is the most likely PP site.

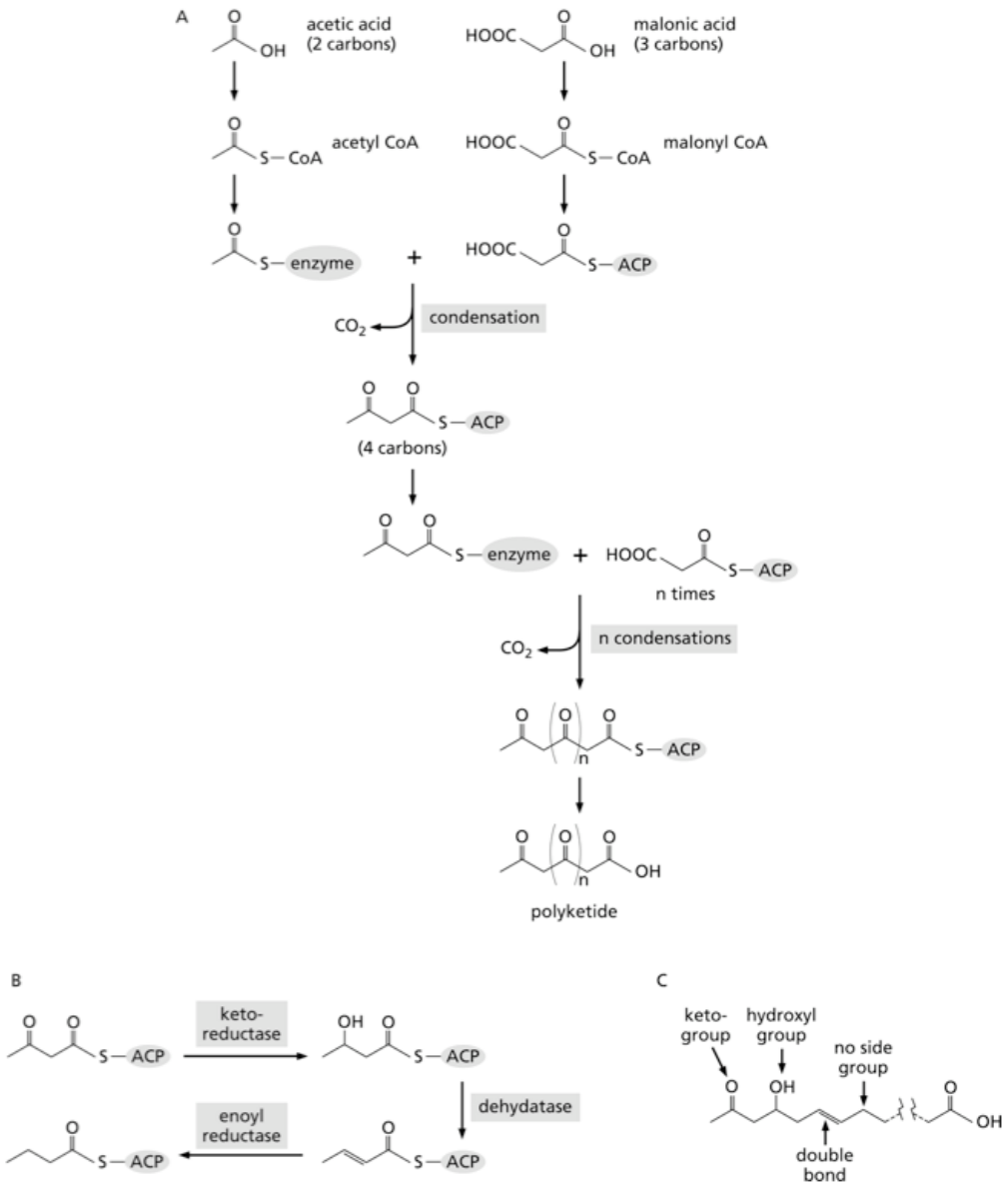


Figure 5: Biochemistry of polyketide synthesis (Hopwood, 2004). A). The growing polyketide is anchored to the phospantetheine attachment site (PP or acyl carrier protein domain, ACP) while new organic acids (positioned by the acyl transferase domain) are added by condensation by the ketoacyl synthase domain. B). Further modifications to the growing chain, including reduction of the ketone group, dehydration of the hydroxyl group and reduction of the enoyl are made by the domains ketoreductase, dehydratase and enoyl reductase. C). A possible complex polyketide.

For over 50 years, most medicines responsible for the treatment of infectious diseases have come from the natural products of microorganisms. Besides penicillins, the most important class of chemical for medicinal use has been polyketides. While this study is not directly applicable, the study of these compounds, the enzymes that produce them, and the regulation of the genes coding these enzymes is one of great importance to the medical field, specifically to biopharmaceutical research.

Gene Regulatory Networks in Development

The information for the development of animal body plan from maternal inputs all the way to differentiated genes, such as *SpPks,* is encoded in the genome. Development occurs as regulatory proteins, known as transcription factors, bind to DNA sequences at the right time and cell type, leading to the correct development of the embryo. The spatial and temporal regulatory states of the cell determine the development of the embryo (Davidson, 2006). The regulatory state of the cell is established by a combination of maternal and zygotic factors that are spatially distributed as cleavage occurs or from intercellular signaling produced by neighboring cells. As development progresses, cells go through specification, a process in which a group of cells in one region of an embryo all express a specific set of genes. These cells, once specified, typically produce the signaling which then leads to the specification of adjacent cells. Cell signaling leads to the expression of genes that encode transcription factors, which activate or repress downstream target genes producing distinct patterns of gene expression (reviewed in Davidson *et al.*, 2002).

The regulatory region of the gene is comprised of *cis*-regulatory modules. These *cis*-regulatory modules are typically 200-500 base pairs in length and can be acted upon by different

transcription factors on upwards of 10 different binding sites. The same transcription factor can be expressed at different times and in different cell-types during development. As each gene can be regulated by multiple transcription factors, and these transcription factors can act on more than one module, gene expression in development can be visualized as an interlocking network (Davidson et al., 2002; reviewed in Levine and Davidson, 2005). The interlocking network that describes the endomesoderm specification to 30 hours is seen in Figure 6.

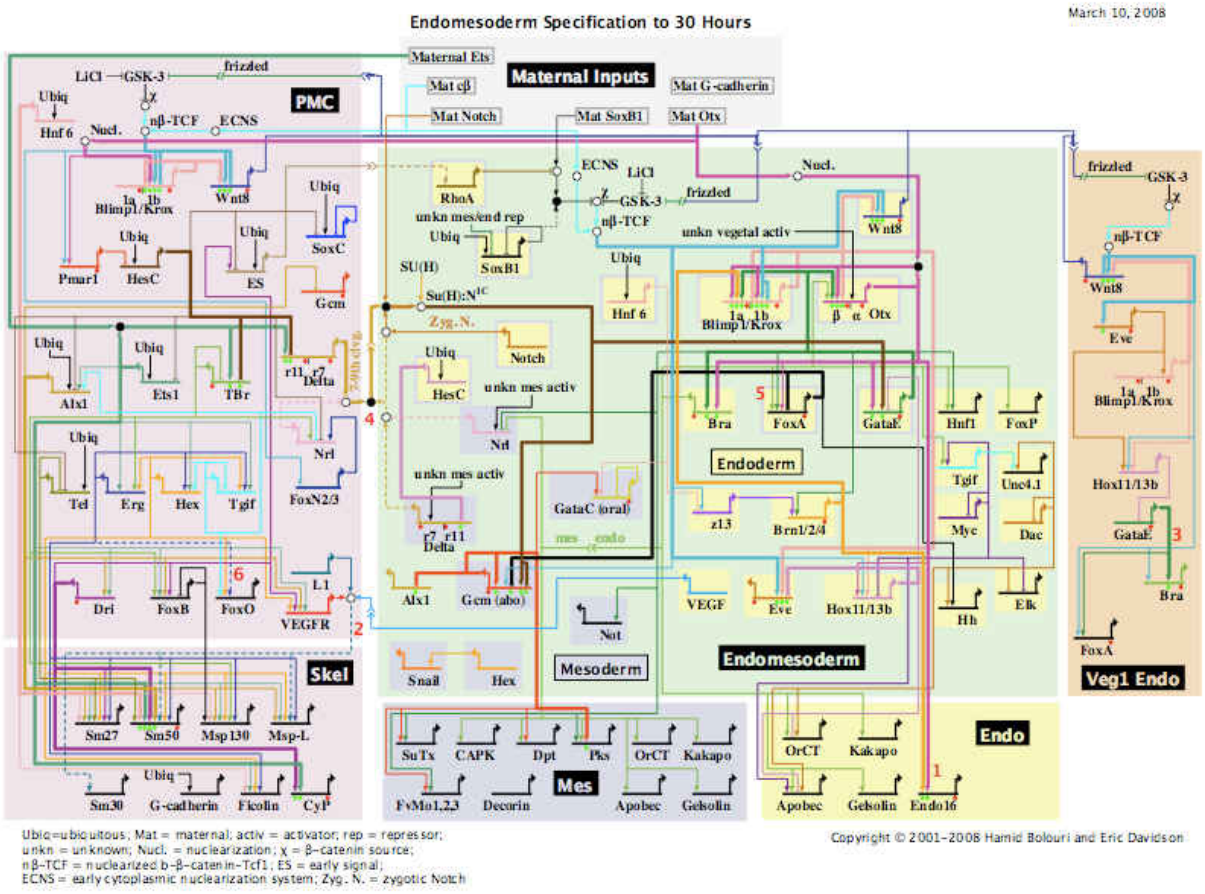


Figure 6: Current knowledge on the Gene Regulatory Network describing endomesoderm specification to 30 hours of development for the sea urchin, visualized as an interlocking network (http://www.biotapestry.org/, Davidson *et al,* 2002).

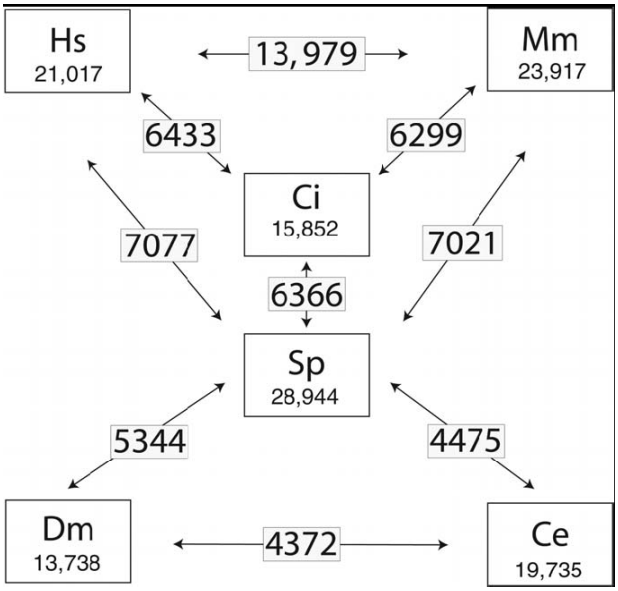
In a gene regulatory network, different logic can be applied to exactly how these transcription factors interact to cause the expression of certain genes. When two or more transcription factors must be present in order for gene expression to occur, this is referred to as "and" logic, which allows for new regulatory states to occur only where expression of different transcription factors overlap (Bolouri and Davidson, 2002; Davidson *et al.,* 2002; Istrail and Davidson, 2005). Also, *cis*-regulatory modules can produce amplifying effects. While one module or transcription factor may be required for gene expression, a separate module can act to amplify that expression when it is present. Another common role of *cis*-regulatory modules is to silence the expression of a gene. Repressors can bind to these modules, which then prevent the start of transcription. Understanding the structure and functions of *cis*-regulatory modules in gene regulatory networks will lead to a better understanding of development as a whole, and the sea urchin, *Strongylocentrotus purpuratus* is an ideal model organism in which to study this.

Gene Regulatory Networks in the Sea Urchin

The sea urchin embryo is relatively simple, given that it is made up of a single layer of differentiated cells. The adult body plan does not form until later, after the larva feeds and metamorphoses. Sea urchins are classified as Type 1 embryos as they develop from a small (less than 200 μm) egg and consist of only a few hundred cells at gastrulation. The GRN regulating the development of a Type I embryo functions to produce differentiated cells at the right time and place in as direct as possible way to form a larva able to feed. By one day after fertilization in the *S. purpuratus*, most of the major cell lineages (the gut, skeleton, mesoderm and endoderm) have already been specified or are well on their way with established regulatory states, and

differentiation gene batteries are already being expressed (Davidson *et al.,* 2002; Davidson, 2006; Levine and Davidson, 2005). In contrast, *Drosophila* and vertebrate model organisms are more complex developmental systems because they go through successive stages of pattern formation to produce a multilayered, juvenile form of the adult.

In addition to its simplicity, the sea urchin embryo, being an invertebrate deuterostome, has a very significant evolutionary placement, as deuterostomes are more closely related to chordates than the protostome model organisms such as fruit flies and nematodes. Recently, the genome of *S. purpuratus* was reported and determined to encode about 23,300 genes, some of which were previously thought to be only in vertebrates as well as others that were thought to only exist outside of deuterostomes (Sea Urchin Genome Sequencing Consortium, 2006). Of these 23,300 genes, it was also found that this sea urchin shared 7,077 genes in common with humans, as seen in Figure 7 (Sea Urchin Genome Sequencing Consortium, 2006).



(Sea Urchin Genome Sequencing Consortium, 2006)

Figure 7: Number of Orthologs among bilateria found by BLAST alignments from sequenced genomes. Numbers on double arrows indicate total number of orthologs between the two species indicated. Numbers under species reflect total number of protein sequences per species. Hs*: Homo sapiens,* Mm, *Mus musculus,* Ci, *Ciona intestinalis,* Sp: *S. purpuratus,* Dm: *Drosophila melanogaster,* Ce: *C. elegans.*

The Gene Regulatory Network of SMCs

The specific signaling pathway that leads to the development of SMCs and therefore pigment cells has been well defined by a series of experiments. Micromeres induce SMC specification by expressing the ligand Delta during the seventh to ninth cleavage, which specifies the surrounding veg_2 cells to become SMC precursors by activating a Notch receptor (Sherwood and McClay, 1999; Sweet *et al.,* 1999; Oliveri *et al.,* 2002; Sweet *et al.,* 2002). In order for the cells to be receptive to this signal, the nuclearization of β-catenin must first occur (McClay *et al.,* 2000). Once the Notch receptor binds to Delta, the intracellular component (NICD) is cleaved allowing for the nuclearization of NICD where it is then able to relieve the repression of the

transcription factor Suppressor of Hairless (Su(H); reviewed in Ehebauer *et al.* 2006). Two experiments performed by Sherwood and McClay (1999) show that Notch regulates the development of SMCs through its intracellular component. In one experiment they used an overexpression of Notch while in the other they overexpressed a dominant negative form of Notch, which contains only the receptor portion, lacking the intracellular component. Overexpression of Notch leads to increased SMCs by changing the fate of presumptive endoderm cells, whereas overexpression of a dominant negative form of Notch leads to no pigment cells and fewer SMC derived cells in general, causing the endoderm to take over territories normally reserved for SMCs (Sherwood and McClay, 1999).

SpGcm, which encodes the transcription factor glial cells missing, was found to be expressed in SMC precursor cells in a pattern much like that of pigment cells (Ransick *et al.,* 2002). When expression of *SpGcm* was blocked, the resulting larvae developed without pigment cells, showing that *SpGcm* is necessary for pigment cell specification. Likewise, overexpression of a dominant negative form of Su(H), in which the DNA binding site of the protein is mutated, leads to less expression of *SpGcm* and therefore a lack of pigment cells (Ransick and Davidson, 2006). This confirms that intracellular Notch acts by releasing the repression of Su(H) to activate *SpGcm* leading to the specification of pigment cells . It is also worth noting that *foxa*, a transcription factor expressed in the endoderm, represses *SpGcm* expression in these cell types as a result of Notch signaling, allowing endodermal cells to not be specified to a mesodermal fate (Oliveri *et al.,* 2006). Putative DNA binding sites for FoxA have been identified within a minimal promoter for *SpGcm* but have yet to be experimentally tested (Ransick and Davidson, 2006).

To isolate a comprehensive pool of genes specifically expressed in pigment cells, a differential macroarray screening was performed (Calestani et al., 2003) using a highly sensitive procedure described by Rast *et al.* (2000). Transcripts from LiCl-treated embryos, which have an excess of endo-mesodermal precursor cells, were compared with transcripts from dnN-expressing embryos, which lack SMCs (Calestani *et al.,* 2003). LiCl-treated embryos result in an excess of SMC cells most likely due to the fact that lithium ions inhibit the enzyme glycogen synthase kinase-3β (GSK3β), which is a part of the Wnt/Wingless pathway. The function of GSK3β is to promote the degradation of β-catenin, thus, by inhibiting this enzyme; there is a build up of β-catenin which leads to its nuclearization, therefore promoting the specification of SMCs as described above (Ransick *et al.,* 2002). The screen resulted in the identification of several genes including *SpPks, SpSult, SpFmo1, SpFmo2, SpDimethlaniline monooxygenase* (similar to *fmo3*), and *SpGcm*. These genes were found to be expressed in patterns typical of SMC precursors and, by pluteus stage, coincident with the distribution of pigment cells, being that they were embedded in the aboral ectoderm (Calestani *et al.,* 2003).

In order to further clarify the potential inputs involved in the *cis*-regulation of these genes, specifically *SpPks*, perturbation analyses have been performed, in which the effect of knocking out certain genes is seen by quantifying the transcript levels of other genes. The transcription factors, *gcm* and *gata E* have been found to be positive regulators of *SpPks, SpSult, SpFmo1, SpFmo2, and SpDimethlaniline monooxygenase* (Davidson *et al.*, 2002). *SpGcm* also appears to have an auto-regulatory loop but has not been shown to be positively regulated by *gata E* (Davidson *et al.*, 2002). These transcription factors are not involved in regulating the expression of each other, but are both required for the expression of pigment cell specific genes.

Therefore, they act in parallel pathways to regulate these genes. Another positive input comes from the genes *kruppel-like* (*Krl*) and *brachyury* (*Bra*). *Brachyury* is expressed in the endomesoderm and *kruppel-like* is expressed in the mesoderm at the hatched blastula stage, but moves into the endoderm by gastrulation (Peterson *et al.,* 1999; Howard *et al.* 2001; Davidson *et al.,* 2002 (2); Rast *et al.,* 2002; Lee and Davidson, 2004; Minokawa *et al.*, 2004; Yamazaki *et al.* 2007). *Kruppel-like* is involved in the Wnt pathway and appears to act independently and parallel to the Delta/Notch pathway, which works through *gcm* (Yamazaki *et al.* 2007). The spatial and temporal pattern of expression of the upstream transcriptional regulators of the pigment cell differentiation genes *SpPks, SpFmo,* and *SpSult* is summarized in Figure 8.

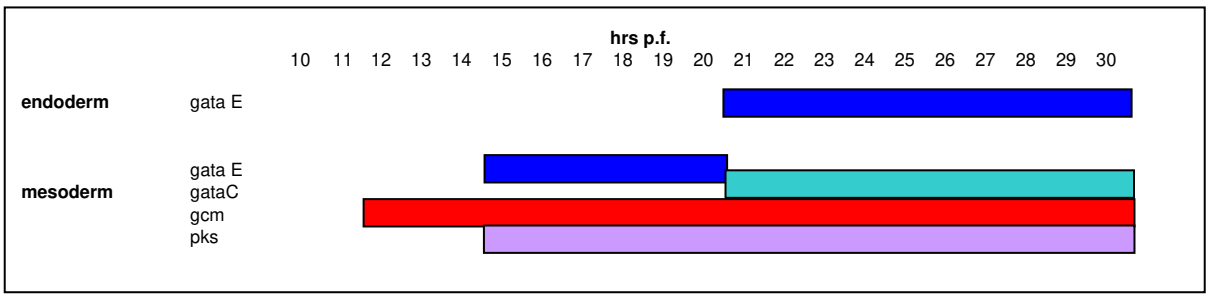


Figure 8: Upstream transcriptional regulators of pigment cell specific genes. Bars indicate temporal expression in hours post fertilization. *gata E* is expressed in the mesoderm, but only until ~21 hours post fertilization. *gcm* is expressed in the mesoderm ~3 hours before the onset of expression for *pks*, which is consistent with it being a direct positive regulator of *pks*. *Gata C* is also positively regulated by *gcm*.

MATERIALS AND METHODS

Computational Analysis of *SpPks* Promoter Region

The software Family Relations (Brown *et al.*, 2002) was used to search for DNA-binding sites for the putative *SpPks* direct regulators, GCM and GATA E. The DNA-binding site sequences for GCM and GATAE in sea urchin are not known. For this reason DNA-binding consensus sequences known from other organisms were used. The GCM DNA-binding site sequence that was used is ATRCGGGY (where R indicates either a G or A and Y indicates T or C; Akiyama *et al.*, 1996) and for GATAE is WGATAR (where W indicates A or T; Evans *et al,* 1988).

Cloning of *SpPks* Promoter Elements

The genomic regions of interest were amplified from a Bacterial Artificial Chromosome (BAC) containing SpPks (NCBI accession NW_001307661) by Polymerase Chain Reaction (PCR). PCR primers to amplify the regions -3kb, -2kb -1.5kb, -2 to -1kb and -1kb were designed using the Primer3 Program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi). Forward and reverse primers were first designed for the -3kb promoter region. Keeping the same reverse primer, new forward primers were designed for the -2kb, -1.5kb and -1kb region. Keeping the same forward primer for the -2kb construct, a new reverse primer was designed for the -2 to -1kb construct. Each primer pair sequence also included restriction digestion sites for SacI (forward primer) and MluI (reverse primer) to facilitate directional cloning into the Green Fluorescent Protein (GFP) reporter vector EpGFPII (Arnone *et al*., 1997). A list of all forward and reverse primers (with restriction sites added) can be found in Table 1.

Table 1. Primer sequences used for PCR amplification of promoter regions.

Construct	Forward Primer	Reverse Primer
6.5kb intron	TACTGAGCTCGCAGGTATTTATACGGAGCA	ATATACGCGTCGCCTTAGGTTGATTTCTCG
-3kb	TACTGAGCTCCACCACTGTGCCAATCTTAAA	ATATACGCGTCCTTCTTGTTGCAGTGGTGA
-2kb	TACTGAGCTCTCCCTCTTTCTCTCCCACTCT	ATATACGCGTCCTTCTTGTTGCAGTGGTGA
-1.5kb	TACTGAGCTCCCGACCGCGTCAAATCT	ATATACGCGTCCTTCTTGTTGCAGTGGTGA
-2 to -1kb	TACTGAGCTCTCCCTCTTTCTCTCCCACTC	ATATACGCGTACCTTTCAATTGCAGACAGGA
-1 kb	TACTGAGCTCGGGGCATAATGACAAATCGT	ATATACGCGTCCTTCTTGTTGCAGTGGTGA
“3 sites”	TACTGAGCTCCCGGAGATTCTCGTCTTTGA	ATATACGCGTCGCCATAATAGTTGCAAAACA
500 bp	TACTGAGCTCCCGACCGCGTCAAATCT	ATATACGCGTACCTTTCAATTGCAGACAGGA
400 bp	TACTGAGCTCTGAAATGCCACTGATTAGTATGATGA	ATATACGCGTACCTTTCAATTGCAGACAGGA
300 bp	TACTGAGCTCTTGAGTGGCTGTTAAGAAACCAT	ATATACGCGTACCTTTCAATTGCAGACAGGA
200 bp	TACTGAGCTCTCGCTTATTCGTATTATACCCGCATC	ATATACGCGTACCTTTCAATTGCAGACAGGA

The -3kb promoter was amplified using the Expand High Fidelity PCR System method according to the manufacturer (Roche, Indianapolis, IN) using the following program: Initial denaturation at 94°C for two minutes, 10 cycles of denaturation at 94°C for 15 seconds, annealing at 60°C for 30 seconds and elongation at 72°C for two minutes, followed by 15 cycles of denaturation at 94°C for 15 seconds, annealing at 58°C for 30 seconds and elongation at 72°C for two minutes + five seconds for each successive cycle, ending with a final elongation cycle at 72°C for seven minutes. The -2kb, -1.5kb, -2 to -2kb and -1kb regions were amplified with Taq DNA Polymerase (Roche, Indianapolis, IN) using the following program: Initial denaturation at 94°C for two minutes, 30 cycles of denaturation at 94°C for 30 seconds, annealing at 58°C for 60 seconds and elongation at 72°C for two minutes, ending with a final elongation cycle at 72°C for seven minutes. PCR products were purified using QIAquick PCR Purification Kit (Qiagen, Valencia, CA), doubly digested with SacI and MluI and purified again with the same kit. The EpGFPII vector was linearized by dual digestion using SacI and MluI and purified with the QIAquick PCR Purification kit. The insert and vector were then ligated using T4 DNA Ligase according to the manufacturer (Promega, Madison, WI). Finally, Fusion-Blue Competent Cells were transformed with 2.5 µl of ligation product according to the manufacturer's protocol (Clontech, Mountain View, CA).

Three aliquots of transformed cells were then spread on LB agar plates containing 100 µg/ml ampicillin (50 µl, 25 µl and all the remaining cells concentrated to a volume of approximately 50 µl). After incubation at 37° C for 14-16 hours, single colonies were cultured in 3 ml of LB broth with 100 µg/ml ampicillin at 37° C for 14-16 hours with shaking.

Plasmids were purified from the bacterial culture using the Qiagen Mini Prep kit according to the manufacturer (Qiagen, Valencia, CA). Verification that the correct DNA insert had been cloned was done by restriction digestion of the plasmids followed by electrophoresis on agarose gel. Plasmids containing the insert were then linearized with SacI (cutting only at the beginning of the promoter construct) and purified using the QIAquick PCR purification kit for later injection.

Site Directed Mutagenesis

To better test the functionality of the GCM site at -1,179bp, the putative binding site was mutagenized within the -2kb construct by a PCR-based site-directed mutagenesis methodology. PCR primers were designed to amplify two overlapping DNA fragments covering the –2Kb region. The first DNA fragment included the region from -2kb to the GCM site to be mutagenized, with the right primer containing a different sequence than the actual binding site (the GCM binding site ACCCGCAT was changed to GTATTAGC). The second DNA fragment was PCR amplified with Taq DNA Polymerase (Roche, Indianapolis, IN) using a similar left primer that overlaps the DNA-binding site with the same changes as the previously mentioned right primer using the following program: Initial denaturation at 94°C for two minutes, 30 cycles of denaturation at 94°C for 30 seconds, annealing at 55°C for 60 seconds and elongation at 72°C for two minutes, ending with a final elongation cycle at 72°C for seven minutes. Five ng of each of these two overlapping fragments were then combined and allowed to anneal to one another (annealing temperature 55°C). Additional PCR cycles (Initial denaturation at 94°C for two minutes, 15 cycles of denaturation at 94°C for 10 seconds, annealing at 55°C for 30 seconds and

elongation at 68°C for seven minutes, ending with a final elongation cycle at 68°C for seven minutes using only the fragments, no primers followed by initial denaturation at 94°C for two minutes, 25 cycles of denaturation at 94°C for 10 seconds, annealing at 60°C for 30 seconds and elongation at 68°C for seven minutes, ending with a final elongation cycle at 68°C for seven minutes adding in the forward and reverse primers for the entire 2kb product) extended the complementary sequence of each side, creating a PCR product the same size as the -2kb construct, with the sequence at the putative GCM binding site mutagenized (Figure 9). This new PCR product with the mutagenized site was then cloned into the EpGFPII reporter vector between SacI and MluI as described above.

PCR 1 and 2:

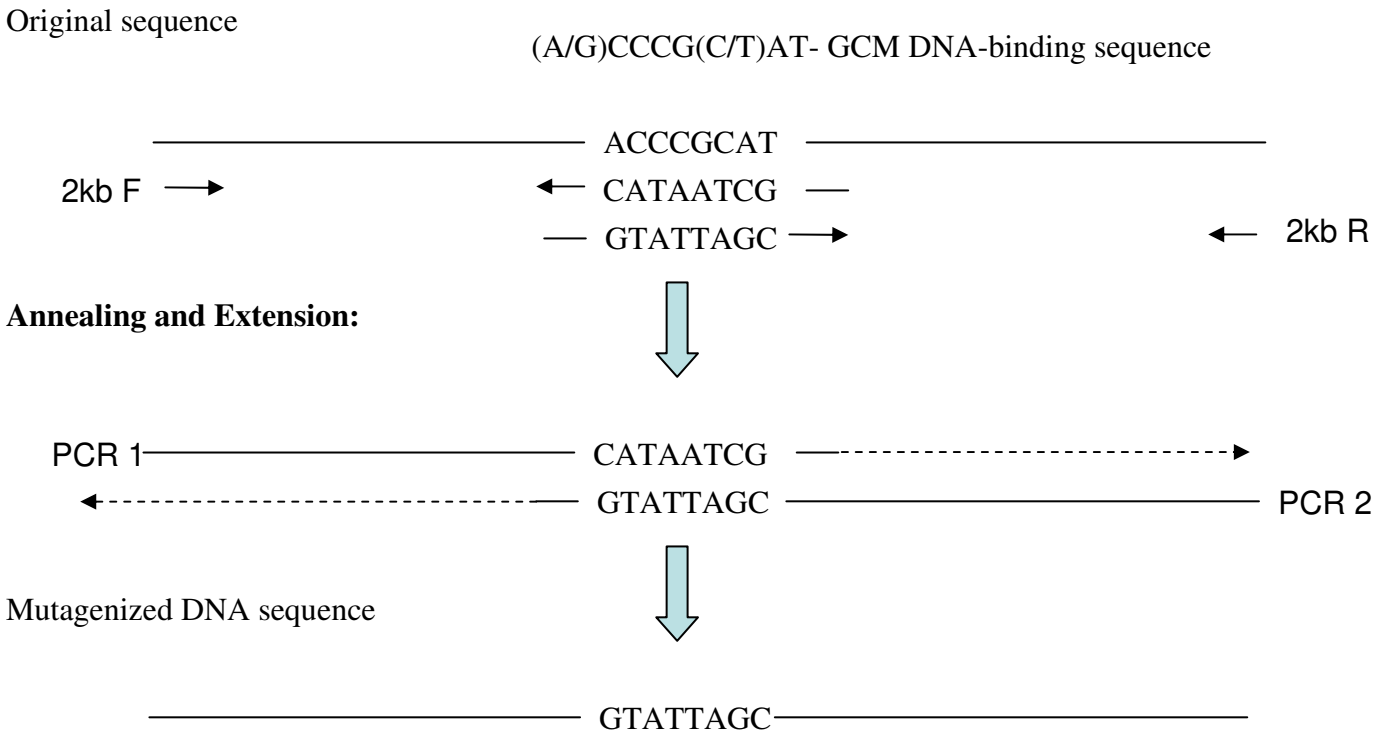


Figure 9: Diagram showing site directed mutagenesis of GCM DNA-binding site. Two overlapping PCR fragments were produced using the primers -2kb F with a right primer with mutagenized sequence as indicated and the -2kb R with a left primer with a mutagenized sequence as indicated.

Procurement of Gametes

Upon arrival, adult sea urchins that had not already spawned were separated from those that had and rinsed with artificial seawater. Spawning was induced in individual urchins by injecting a combined total of approximately 2 ml of 0.5M KCl into 2-3 locations near the mouth of the urchin. The urchin was then placed oral side up on paper towels until sex could be determined. Sex is determined by the color of gametes, eggs being orange and sperm being white. Eggs were then collected by placing the female urchin oral side up into a glass beaker filled with filter-sterilized seawater (FSW) placed on ice. Sperm was collected by placing the urchin oral side down and transferring the sperm by pipetting into a microcentrifuge tube placed on ice.

Microinjections of Constructs

Eggs were transferred with a Pasteur glass pipette from the collection beaker to a Petri dish containing acidic seawater (made by adding 0.3M citric acid to filtered seawater until the pH reached 4.75) and incubated for one minute to remove the egg jelly coat. Eggs were then washed three times for one minute each by transferring them to a new dish filled with FSW. De-jellying allows the eggs to stick to protamine sulfate coated Petri dish lids by removing the outer polysaccharide layer from the eggs. The negatively charged eggs will adhere to the positively charged surface of the treated lids. Petri dish lids were treated by filling them with a 1% solution of protamine sulfate for one minute and immediately rinsing them in double distilled water and then allowing them to air dry.

Approximately 150 eggs were rowed in a line by mouth pipetting on a protamine sulfate coated lid filled with 6 ml of 10mM para-aminobenzoic acid (PABA) seawater. PABA seawater prevents the stabilization of the fertilization envelope by inhibiting ovoperoxidase, an enzyme that chemically cross-links tyrosine residues (Ettensohn, 2004). The envelope forms, but the tyrosine residues do not cross-link, allowing zygotes to be microinjected through the fertilization envelope. Rowing the eggs facilitates rapid injection of zygotes before the fertilization envelope hardens prohibiting injection (about 15 minutes). Rowed eggs were fertilized by sperm diluted (and thus activated) in PABA seawater. Fertilized eggs were then injected with a PicoSpritzer III (Parker Instrumentation, Cleveland, OH), allowing for 2 to 5 picoliters of solutions to be injected using capillary needles pulled with a P-97 flaming micropipette puller (Sutter Instrument Company, Novato, CA). Injection solutions consist of 1000 molecules/pl of linearized plasmid (construct + EPGFPII), 0.12 M KCl and a 5 molar excess of HindIII digested sea urchin genomic DNA (carrier DNA) and water. After injections were completed, penicillin (20 units/ml) and streptomycin (50 ug/ml) were added to each embryo culture. The embryos were incubated at 16° C throughout their development. After the embryos hatched (around 18 hours post-fertilization), they were transferred into new dishes with FSW containing penicillin (20 units/ml) and streptomycin (50 ug/ml) to ensure proper development.

Observation of GFP expression in microinjected embryos

Embryos injected with each construct were observed using fluorescent microscopy (Olympus BX60, Center Valley, PA). Approximately 50 embryos were collected by mouth pipette and placed under a cover slip on a microscope slide. Using the GFP LP 32001 filter for GFP (Chroma Technology, Rockingham, VT), the embryos were exposed to UV light (425 nm)

and GFP expression was observed spatially within the embryos. Initially, embryos were observed at varying developmental time points (15 hours, 26 hours, 32 hours, 45 hours, and 72 hours) to determine if GFP expression was occurring in SMC precursor cells initially, and later specifically in pigment cells. During further experiments, embryos were observed only at 45 hours, a time point where if GFP were being expressed, it would be easily detected and also when pigment cells could be easily identified. To test the statistical significance of the differences between constructs, R version 2.7.0 (http://www.r-project.org/) was used to run a pair wise ANOVA, Tukey multiple comparisons of means using the following script:

"data name" <- read.table ("datatable.txt", header=T)

attach ("data name")

anova (lm(PCT~GROUP))

"model name" <- aov (PCT~GROUP)

TukeyHSD ("model name")

where "data name" and "model name" are user imputed names typed without quotation marks and "datatable.txt" is the name of the data table file, typed with the quotation marks.

RNA Extraction

Approximately 100 injected embryos were collected at various developmental time points (15 hours, 21 hours, 26 hours, 40 hours, 50 hours and 72 hours) for RNA extraction. The embryos were collected by mouth pipette and put into a microcentrifuge tube on ice. The embryos were centrifuged at $\leq$ 2,000 rpm and the seawater was removed from the tube. Immediately, 350 µl of buffer RLT containing β-Mercaptaethanol from the RNAeasy kit

(Qiagen, Valencia, CA) was added to the embryos and they were homogenized by pipetting up and down followed by vortexing for 20 seconds. RNA was then isolated following the manufacturer's procedure. Forty-four µl of the isolated RNA was DNAse treated with DNA-Free (Applied Biosystems/Ambion, Austin, TX) for 15-20 minutes at 37°C. The RNA was then purified following the Qiagen RNAeasy kit protocol. cDNA was then prepared from the total sample of DNAse-treated RNA using the High Capacity cDNA Archive Kit (Applied Biosystems, Foster City, CA). The remaining six µl of isolated RNA was stored at -80°C.

Quantification of transcript accumulation

Quantitative real-time PCR (QPCR) was used to measure the relative amount of the endogenous *SpPks* mRNA, as well as the *gfp* mRNA in injected embryos at the previously mentioned developmental time points. The cDNA made from the isolated RNA of the embryos at these time points was used as template DNA. QPCR reactions were set up using ABI SYBR Green 2X master mix and 15 pmol of each forward and reverse primer for *SpPks, gfp*, in addition to two endogenous control genes, *ubiquitin* and *SpZ12*. *Ubiquitin* is used as a control gene as previous research has indicated that levels of *ubiquitin* expression remain consistent throughout sea urchin development (Nemer et al., 1991; Ransick et al., 2002). *SpZ12* is an internal standard that can be used to quantify the number of mRNA transcripts, as the number of *SpZ12* transcripts in various stages of embryo development is known from RNA titration (Wang *et al.,* 1995). Each reaction was run in triplicate for each combination of developmental stage and primer set on an ABI PRISM 7000 Sequence Detection System (Applied Biosystems, Foster City, CA). A non-template control for each primer set was also included and samples were run at one cycle of

95°C for ten minutes, fifteen seconds followed by 40 cycles of 95°C for fifteen seconds and 60°C for one minute.

SYBR Green is a molecule that emits fluorescence when it binds to double-stranded DNA. At each cycle, the fluorescent dye in each sample is excited by a tungsten-halogen lamp resulting in emissions between 500 and 660 nm, which is then recorded by a charge-coupled device (CCD) camera. Based on the starting number of cDNA transcripts in the sample and the efficiency of the DNA amplification, the level of fluorescence in a sample begins to be detectable after a certain number of cycles have completed. The greater the number of starting transcripts, the earlier the fluorescence becomes detectable by the CCD camera. The PCR cycle at which the fluorescent signal becomes greater than a threshold that is set by the user is called the threshold cycle (Ct). The Ct average for each primer set was subtracted from the Ct for ubiquitin at each developmental time point (dCt), as ubiquitin is known to be expressed at approximately the same levels throughout development (Nemer et al., 1991; Ransick et al., 2002). In this way relative gene expression can be determined at each time point, normalized for other factors such as the actual number of embryos collected, efficacy of the cDNA RT reaction, and any developmental differences that may vary between cDNA samples. The standard deviation for each time point was calculated by taking the square root of the sum of the average Ct subtracted from the Ct for each well squared for *ubiquitin* plus the sum of the average Ct subtracted from the Ct for each well squared for the target gene divided by four (i.e. sd = $\sqrt{((\Sigma(\textit{Ubiq}\text{ sample-average})^2 + \Sigma(\textit{pks}\text{ sample-average})^2/4))}$.

RESULTS

Computational Analysis of Promoter Region

Putative DNA-binding sites of known positive regulators of *SpPks* were searched for using the software Family Relations (Brown *et al.*, 2002) and DNA–binding sites for GCM and GATAE were identified at the following locations: GATAE: -1,846, -1,386, -931, -810, -672; GCM: -1,179, +4,004. The GCM DNA-binding site consensus sequence is ATRCGGGY (where R indicates either a G or A and Y indicates T or C; Akiyama *et al.*, 1996) and for GATAE is WGATAR (where W indicates A or T; Evans *et al,* 1988). The actual sequences found within the 3kb upstream of *SpPks* for GATAE were CTATCT at -1,846, TGATAG at -1,386, TGATAA at -931, AGATAA at -810 and TGATAA at -672 (Table 2). The GCM sequences found in the 3kb upstream and within the first intron were ACCCGCAT at -1,179 and at +4,004 (Table 2). Within the 3kb upstream of *SpPks* there are five GATAE sites with distances from the next closest GATAE site ranging from 121 base pairs to 460 base pairs. The distance between the one GCM site and the two most proximal GATAE sites are 207 and 248 base pairs, respectively.

These data lead to the testing of the following twelve GFP reporter constructs (Figure 10): 6.5kb intron, a construct that includes all putative DNA-binding sites, beginning 2kb upstream of the start of transcription and continuing through the first intron; -3kb, containing all upstream putative DNA-binding sites (1 GCM and 5 GATAE); -2kb, also containing all upstream putative DNA-binding sites but with a 1kb 5' deletion compared to the -3kb construct; -1.5kb, containing 4 of the 5 GATAE sites and the GCM site, with a 500 base pair 5' deletion with respect to the -2kb construct; -1kb, containing only 3 of the 5 GATAE sites, omitting the

GCM site, a 500 base pair 5' deletion compared to the -1.5kb construct; -2 to -1kb, with 2 of 5 GATAE sites and the GCM site, a 1kb 3' deletion of the -2kb construct; -2kb gcm mutagenized (gcm mut), containing all 5 GATAE sites with the GCM site mutagenized via site-directed mutagenesis; "3 sites", a construct that only includes the GCM site and its two flanking GATAE sites; 500 bp, a construct 500 base pairs in length with the same 3' boundary as the -2 to -1kb construct, containing only 1 GATAE and 1 GCM site; 400 bp, a construct 400 base pairs in length with the same 3' boundary as the -2 to -1kb construct, containing only 1 GATAE and 1 GCM site; 300 bp and 200 bp, two constructs 300 and 200 base pairs in length respectively, with the same 3' boundary as the -2 to -1kb construct, containing only the 1 GCM site.

Table 2: Consensus sequences and *SpPks* promoter sequences for putative GCM (Akiyama *et al.*, 1996) and GATAE (Evans *et al,* 1988) DNA-binding sites. The listed DNA-binding sites are located from the 3kb upstream through the first intron of *SpPks* with positions relative to the start of transcription as noted. R indicates either a G or A, Y indicates T or C and W indicates A or T.

Transcription Factor	**Consensus Sequence**	***SpPks* Promoter Sequence**	**Position**
GCM	**ATRCGGGY**	ACCCGCAT	-1,179
		ACCCGCAT	+4,004
GATAE	**WGATAR**	CTATCT	-1,846
		TGATAG	-1,386
		TGATAA	-931
		AGATAA	-810
		TGATAA	-672

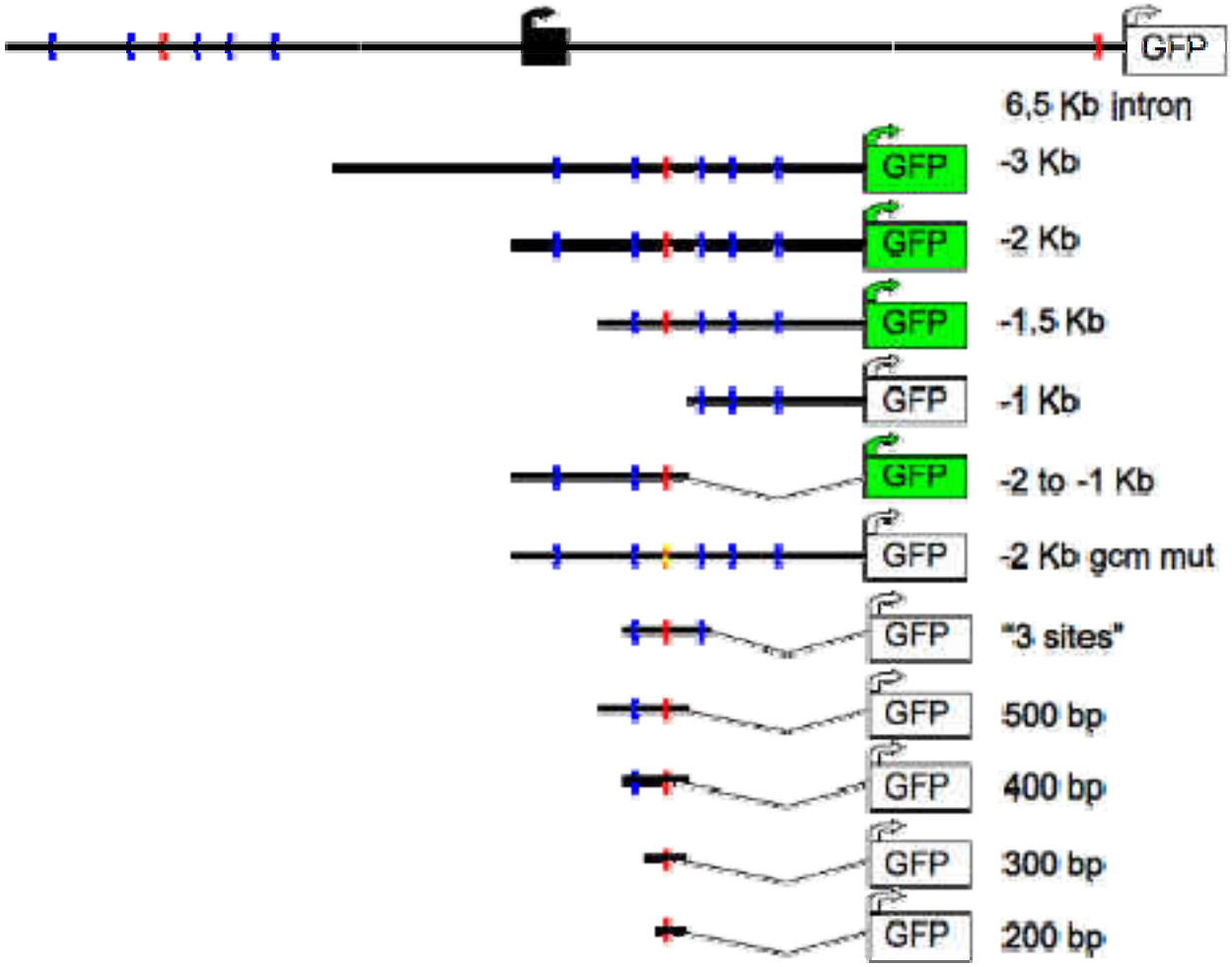


Figure 10**:** GFP Reporter Constructs With Putative Binding Sites Identified**:** GATA E (in blue): -1,846, -1,386, -931, -810, -672; GCM (red): -1,179, +4,004. A 6.5kb construct was made containing each of the identified putative binding sites, first exon is indicated by black box. Serial 5' and 3'deletions to identify the minimal promoter were performed on the -3kb reporter construct resulting in the constructs -2kb, -1.5kb, -1kb, -2 to -1kb, "3 sites", 500 bp, 400 bp, 300 bp and 200 bp. Site-directed mutagenesis of the GCM site at -1,179 within the -2kb construct is shown here as -2kb gcm mut.

In Vivo **Analysis of GFP Reporter Constructs**

Sea urchin embryos injected with the constructs -3kb, -2kb, -1.5kb and -2 to -1kb all showed expression of GFP in pigment cells or their precursors. GFP was observable by 24 hours

post fertilization in the vegetal plate throughout development to 72 hours coincident with pigment cells (Figures 11 and 12). The -1kb construct (which omits the GCM site at -1,179bp) and the -2kb construct with the GCM site mutagenized showed little to no GFP expression. For the -3kb construct, at 26 hours (blastula stage) GFP expression was seen in the vegetal plate of embryos, specifically in the ring of cells that derive from the veg 2 tier (Figure 11 A,D). By 45 hours (late gastrula stage) GFP can be seen in cells that are delaminating from the archenteron, migrating through the blastocoel and embedding in the aboral ectoderm (Figure 11 B,E). At 72 hours (pluteus stage) GFP expression is clearly seen in pigmented cells embedded in the aboral ectoderm (Figure 11 C,F). Constructs such as -2kb showed the same pattern of expression, with GFP being seen in the veg 2 cells at blastula stage (Figure 12 A-C) and in pigment cells by gastrulation (Figure 12 D-G). Other constructs such as the 200 bp construct occasionally showed expression confined to pigment cells, though not at a level, in terms of number of expressing cells, that was considered above background expression (Figure 12 H).

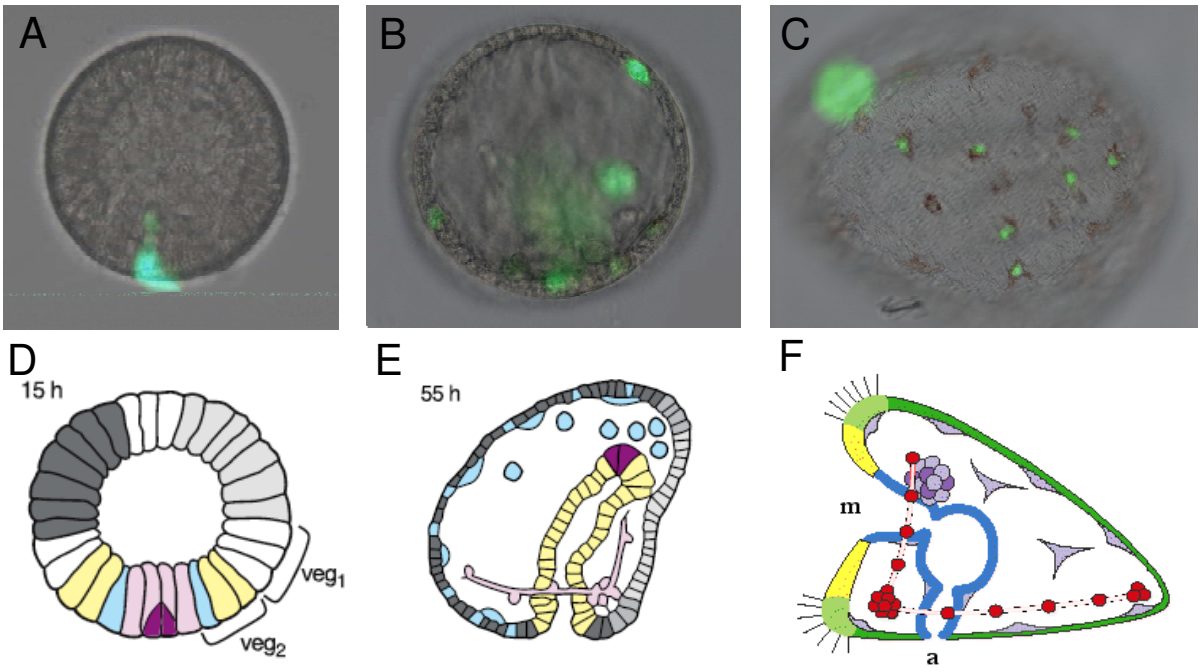


Figure 11: Sea Urchin Embryos Expressing GFP in Pigment Cells. A) Blastula stage embryo (26 hours post fertilization) injected with the -3kb construct. B) Gastrula stage embryo (45 hours) injected with the -3kb construct. C) Pluteus stage embryo (72 hours) injected with the -3kb construct. D) Diagram of blastula stage embryo highlighting the ring of veg2 cells (modified from Davidson *et al.*, 2002). SMC (including pigment cells) precursors are indicated in blue and endoderm precursors in yellow. E) Diagram of gastrula stage embryo with pigment cells delaminating from the archenteron and being embedded in the aboral ectoderm (modified from Davidson *et al.*, 2002.) F) Diagram of pluteus stage embryo with pigment cells embedded in the aboral ectoderm (indicated in violet; modified from Davidson *et al.*, 1998).

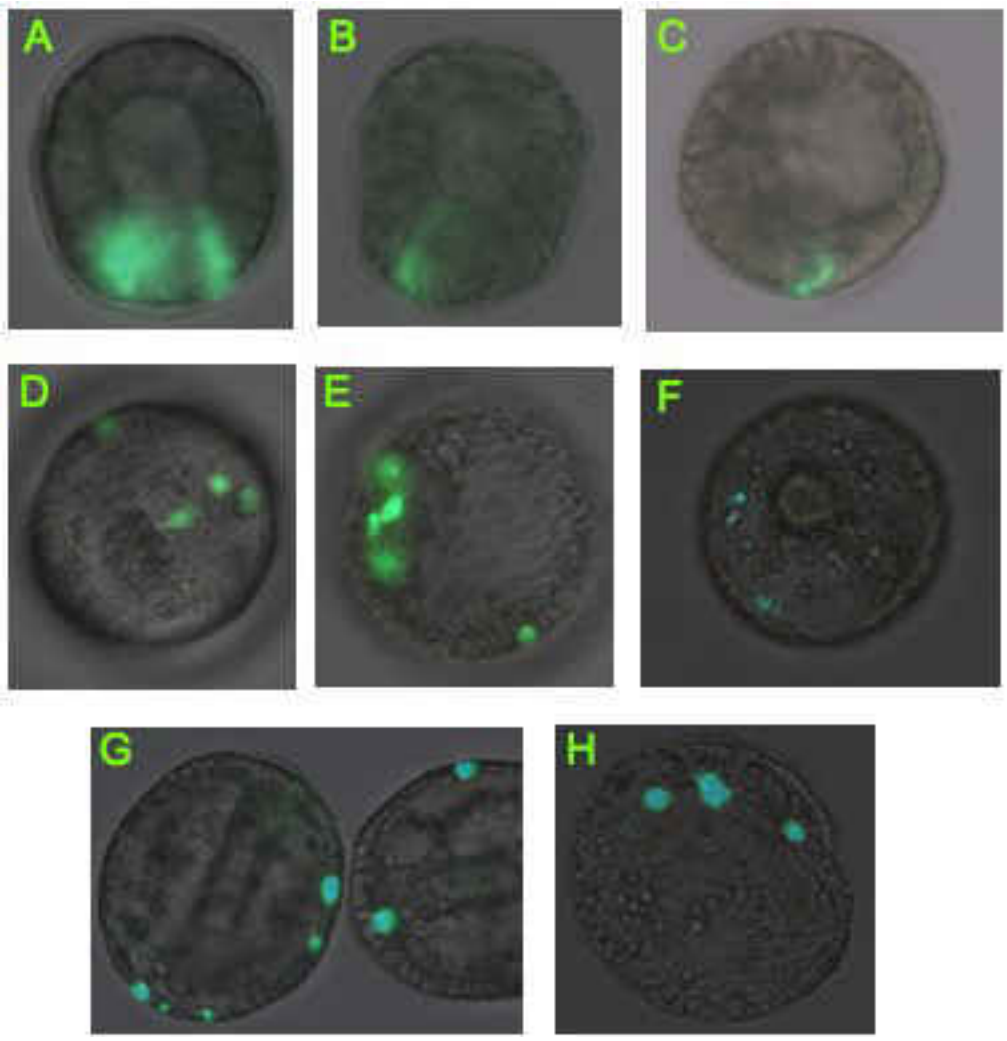


Figure 12: Sea Urchin Embryos Expressing GFP in Pigment Cells. A-C) Blastula stage embryo (25 hours post fertilization) injected with the -2kb construct expressing GFP in veg2 cells. D-G) Late gastrula stage embryo(s) (45 hours) injected with the -2kb construct. H) Late gastrula stage embryo (45 hours) injected with the 200 bp construct.

Data for all *in vivo* observations of GFP reporter constructs are listed in Table 3. In each experiment, approximately 40-50 embryos were observed at a time. Expression of GFP in only one to two cells is considered background expression, that is, random expression of GFP due to the position of integration in the genome and not to the function of the *pks* promoter fragment fused to *gfp*. The 6.5kb intron construct produced no significant GFP expression in the single experiment performed with only 9% of the embryos observed expressing GFP, and their expression was restricted to pigment cells. The -3kb construct produced the most consistent expression of GFP in pigment cells. In the five different experiments with the -3kb construct the percentage of embryos showing GFP expression varied from 26% to 93% of the observed but all embryos were expressing GFP in pigment cells only (no ectopic expression). Interestingly, no construct regardless of its size produced any ectopic expression of GFP. The -2kb construct also showed similar results as the -3kb construct, with results varying between 20% and 66% amongst the five experiments. The -1.5kb construct did produce GFP expression but with more variable results among the four experiments resulting in 57%, 20%, 0% and 46% of embryos showing expression in pigment cells. A 500 base pair 5' deletion of the -1.5kb construct, the -1kb construct, reduced the percentage of expressing embryos to a non-significant amount throughout four experiments, in which only two experiments were able to produce any GFP expression, and at levels of only 4%-5%, however expression was confined to pigment cells. The -2kb to -1kb construct was the smallest construct to produce expression, though at an inconsistent level between three experiments. In the first experiment, the construct produced expression in 57% of embryos, however two further experiments produced only 8% and 0% of embryos respectively. The shorter constructs typically were not able to produce significant

expression of GFP in pigment cells in the single experiments performed. The “3 sites” construct produced expression in 0% of embryos, the 500 bp, 400 bp and 300 bp constructs all had 0% of embryos expressing GFP. Two products of slightly different sizes resulted from the amplification of the 200 bp construct, resulting in two roughly 200 bp promoter regions being tested. 200 bp #1 showed 3% of 59 embryos expressing GFP in pigment cells, while the 200 bp #2 had 0% of 45 embryos expressing. The -2kb construct with the GCM site mutagenized (gcm mut) showed no expression of GFP in pigment cells throughout three experiments.

Pooling all the data from each experiment with replication, the -3kb construct showed 65% of 223 embryos expressing GFP in pigment cells. Likewise, the -2kb construct recapitulated expression in 40% of 201 embryos. As the promoter region was serially deleted, the percentage of embryos showing correct expression decreased: the -1.5kb construct produced correct expression in 30% of 223 embryos while the -2 to -1kb construct had 16% of 135 embryos expressing GFP in pigment cells. In the experiment with the -1kb construct (lacking the putative GCM site) only 3% of 178 embryos expressed GFP in pigment cells. The -2kb construct with the putative GCM site mutagenized expressed GFP in 0% of 167 embryos (Table 4).

Table 3. : *In vivo* observations of *gfp* reporter constructs. Data for each experiment are listed. Columns indicate: Construct used; date of the experiment; total number of embryos observed; the percentage of embryos showing GFP expression in pigment cells (SMCs); the number of embryos with ectopic expression (other than SMCs).

Construct	**Date of Experiment**	**# of embryos observed**	**% of embryos expressing in pigment cells (actual #)**	**# of embryos with ectopic expression**
6.5 kb intron	6/3/08	53	9 (5)	0
-3 kb	8/15/06	56	84 (47)	0
-3 kb	9/19/06	55	93 (51)	0
-3 kb	3/26/08	50	52 (26)	0
-3 kb	4/8/08	42	26 (11)	0
-3 kb	6/3/08	20	45 (9)	0
-2 kb	9/19/06	29	66 (19)	0
-2 kb	2/21/07	62	53 (33)	0
-2 kb	4/4/07	12	50 (6)	0
-2 kb	5/31/07	37	27 (10)	0
-2 kb	6/19/07	61	20 (12)	0
-1.5 kb	4/4/07	60	57 (34)	0
-1.5 kb	6/12/07	49	20 (10)	0
-1.5 kb	6/19/07	64	0 (0)	0
-1.5 kb	4/8/08	50	46 (23)	0
-1 kb	9/19/06	30	0 (0)	0
-1 kb	6/12/07	47	4 (2)	0
-1 kb	6/19/07	55	5 (3)	0
-1 kb	4/8/08	46	0 (0)	0
-2 to -1 kb	4/4/07	30	57 (17)	0
-2 to -1 kb	6/12/07	48	8 (4)	0
-2 to -1 kb	6/19/07	57	0 (0)	0
“3 sites”	3/26/08	49	0 (0)	0
500 bp	5/31/07	63	0 (0)	0
400 bp	5/31/07	56	0 (0)	0
300 bp	5/31/07	41	0 (0)	0
200 bp #1	5/31/07	59	3 (2)	0
200 bp #2	5/31/07	45	0 (0)	0
gcm mut	2/21/07	67	0 (0)	0
gcm mut	6/12/07	46	0 (0)	0
gcm mut	6/19/07	54	0 (0)	0

Table 4. : *In vivo* observations of *gfp* reporter constructs. Combined data for all experiments are listed (replicated a minimum of 3 times). Columns indicate: Construct used and the number of replicate experiments; total number of embryos observed; the percentage of embryos showing GFP expression in pigment cells (SMCs); the number of embryos with ectopic expression (other than SMCs).

Construct (# of exp)	**# of embryos observed**	**% of embryos expressing in pigment cells (actual #)**	**# of embryos with ectopic expression**
-3 Kb (5)	223	65 (144)	0
-2 Kb (5)	201	40 (80)	0
-1.5 Kb (4)	223	30 (67)	0
-1 Kb (4)	178	3 (5)	0
-2 to -1 Kb (3)	135	16 (21)	0
gcm mut (3)	167	0 (0)	0

To test if the observed percentages of GFP expressing embryos among constructs were significantly different, a pair wise ANOVA test was performed on all constructs that had replication. Based on this analysis, the only constructs that were significantly different were the -3kb construct and the -1kb construct with a p-value of 0.0099 as well as the -3kb construct and the gcm mut construct with a p-value of 0.014. All other constructs that appeared to show high levels of expression of GFP in pigment cells had a variance between experiments (different set of parents) too high to be significantly different at a 95% confidence level (Table 5).

Removing the experiments that resulted in most of the variation (5/31/07, 6/12/07 and 6/19/07) due to poor gamete quality, the observed differences between the constructs appears to diminish. Although there are less replicates to be compared, the -1.5kb, -2kb, and -3kb constructs all produce similar averages of 51.5, 56.33 and 60.0 respectively. Given the standard deviation of these experiments the results cannot be considered significantly different from one another

(Table 6). Performing a pair wise ANOVA on this modified data set produces no significant differences, most likely due to the reduced power of the test when used on less than three replicates (Table 7).

Table 5. : Pair wise ANOVA (Tukey multiple comparisons of means) of the different GFP reporter construct data. Columns indicate: Construct pairs being compared; p-value for 95% confidence level. Significance is noted with ** for p-values less than 0.01 and * for p-values less than 0.05 (R version 2.7.0).

Construct Pairs	**P-Value**
-3kb and -2kb	0.82
-3kb and -1.5kb	0.38
-3kb and -1kb	0.0099**
-3kb and -2 to -1kb	0.20
-3kb and gcm mut	0.014*
-2kb and -1.5kb	0.95
-2kb and -1kb	0.10
-2kb and -2 to -1kb	0.75
-2kb and gcm mut	0.11
-1.5kb and -1kb	0.46
-1.5kb and -2 to -1kb	0.99
-1.5kb and gcm mut	0.46
-1kb and -2 to -1kb	0.84
-1kb and gcm mut	1.0
-2 to -1kb and gcm mut	0.82

Table 6. : *In vivo* observations of *gfp* reporter constructs with 5/31/07, 6/12/07 and 6/19/07 data removed. Columns indicate: constructs used with at least one replicate; average percentage of embryos showing GFP expression in pigment cells (SMCs); standard deviation of the average.

Construct (# of exp)	**Average % of embryos expressing in pigment cells**	**Standard Deviation**
-3 Kb (5)	60.0	27.9
-2 Kb (3)	56.3	8.5
-1.5 Kb (2)	51.5	7.8
-1 Kb (2)	0	0

Table 7. : Pair wise ANOVA (Tukey multiple comparisons of means) of the different GFP reporter construct data with 5/31/07, 6/12/07 and 6/19/07 data removed. Columns indicate: Construct pairs being compared; p-value for 95% confidence level (R version 2.7.0).

Construct Pairs	**P-Value**
-3kb and -2kb	1.0
-3kb and -1.5kb	1.0
-3kb and -1kb	0.059
-3kb and -2 to -1kb	1.0
-3kb and gcm mut	0.18
-2kb and -1.5kb	1.0
-2kb and -1kb	0.11
-2kb and -2 to -1kb	1.0
-2kb and gcm mut	0.26
-1.5kb and -1kb	0.22
-1.5kb and -2 to -1kb	1.0
-1.5kb and gcm mut	0.39
-1kb and -2 to -1kb	0.30
-1kb and gcm mut	1.0
-2 to -1kb and gcm mut	0.42

Quantitative Temporal Analysis of GFP Reporter Constructs

The relative amount of *SpPks* and *gfp* transcripts during the course of development was measured by Quantitative real-time PCR (QPCR). Data shown are derived from one cDNA batch (one set of parents). Cycle threshold (Ct) values are means of triplicates. Average Ct for each primer set were normalized to the average ubiquitin Ct. dCt indicates the difference between the ubiquitin and the gene of interest average Ct values. dCt values below -11 indicate no gene expression.

Data for the -2Kb construct showed a similar trend of expression for *gfp* as for *pks* after 21 hours with some oscillation in expression of for the -2kb-*gpf* construct. During early blastula stages (15 and 21 hours) the expression of -2kb-*gfp* and the endogenous *pks* differ. For the endogenous *pks*, there is a relatively low level of expression at 15 hours followed by a steady incline through 50 hours at which point the expression reaches a plateau. Expression of *gfp* in the -2kb construct follows roughly this same pattern but with notably higher expression at 15 and 21 hours and a drop in expression at 50 hours (Figure 13). This experiment was replicated once with similar results.

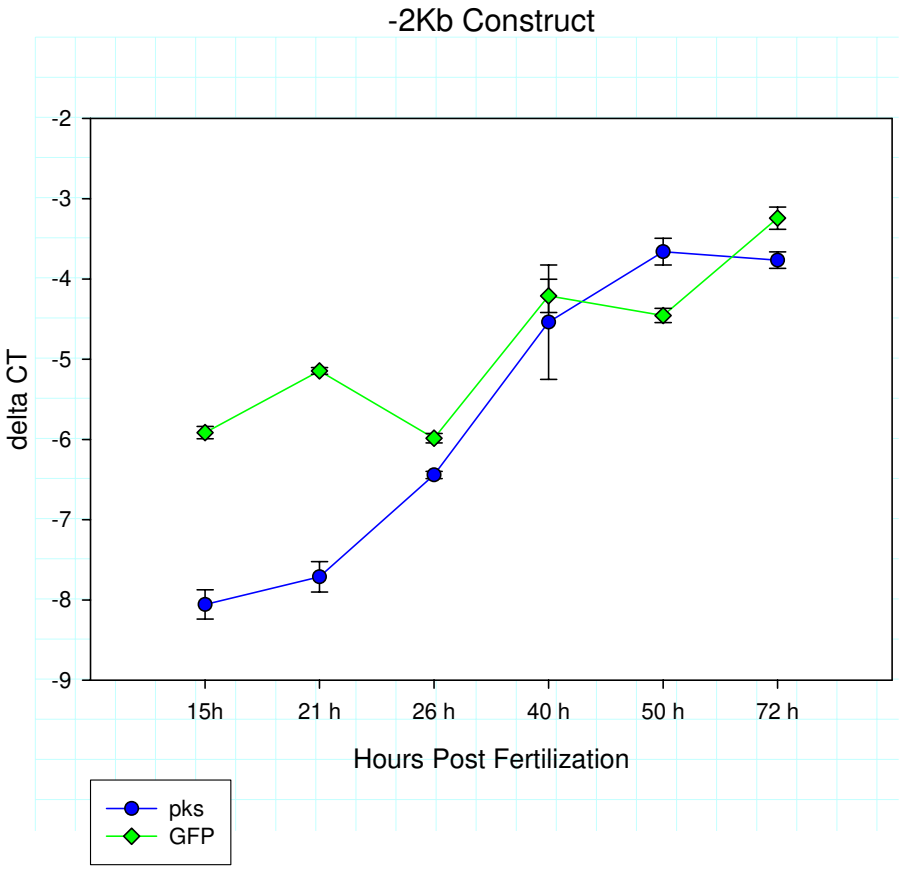


Figure 13: Temporal Expression of *gfp* and *pks* for the -2Kb Construct. On the y-axis is the delta Ct for the endogenous *pks* transcript and for the *gfp* transcripts regulated by the -2Kb DNA region. dCt indicates the difference between the ubiquitin and the gene of interest Ct values. dCt values below -11 indicate no gene expression. On the x-axis the developmental time points are indicated. Error bars indicate the standard deviation for the dCt at each data point calculated as described in Methods.

QPCR data for the gcm mutagenized transcript (gcm mut) again shows endogenous *pks* with little or no expression at 15 hours and a steady increase throughout 40-50 hours at which point the expression begins to decline. Expression of *gfp*, however, increases from 15 to 21

hours, followed by a sharp decline with no expression throughout 72 hours post fertilization (Figure 14).

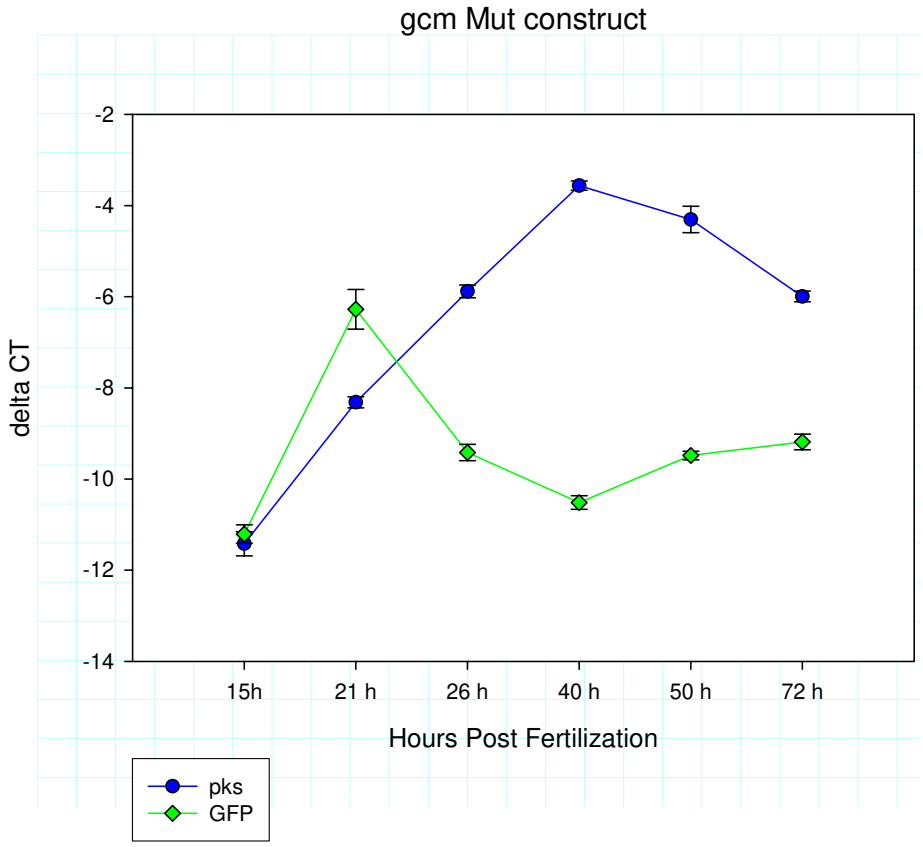


Figure 14: Temporal Expression of *gfp* and *pks* for the gcm mutagenized construct. On the y-axis is delta Ct for the endogenous *pks* transcript and for the *gfp* transcripts regulated by the –2Kb DNA region with the gcm DNA-binding site mutagenized. dCt indicates the difference between the ubiquitin and the gene of interest Ct values. dCt values below -11 indicate no gene expression. On the x-axis the developmental time points are indicated. Error bars indicate the standard deviation for the dCt at each data point calculated as described in Methods.

DISCUSSION

Identification of *SpPks* Minimal Promoter

To identify the minimal promoter for *SpPks*, 500-1000 base pair serial deletions were performed from both the 5' and 3' ends of the -3kb region with fusion to a reporter gene. The minimal promoter that recapitulated expression of *SpPks* in pigment cells was the -2 to -1kb promoter region, containing two GATAE sites and one GCM site. This promoter construct was able to reproduce the correct spatial expression of *SpPks* without showing any ectopic expression. It needs to be noted that this construct only showed expression in 16% of the embryos in the bulked data, however one experiment did produce 57% expression. This seemingly low percentage of GFP expressing embryos could be due to missing some DNA-binding site for positive regulators but it could also be due to the contribution of other factors independent from the regulatory function of the promoter region. Factors that may lead to low expression of GFP include: poor gamete quality (gametes procured outside of the normal breeding season); a low copy number of the construct being incorporated, resulting in an undetectable amount of GFP protein being made (GFP is easily detectable by fluorescence microscopy in sea urchin embryos at concentrations of 1-5x10^6 molecules/embryo, Yuh *et al.,* 1996; Arnone *et al.,* 1997; Damle *et al.,* 2006); incorporation into regions of the genome that lead to silencing GFP expression; incorporation of the DNA construct into the genome late in the cleavage of the developing embryo leading to less cell descendants being able to produce GFP. In the case of the 6.5kb intron construct another factor that may have lead to the low GFP expression, especially in a construct that includes all possible putative DNA-binding sites, is the

positioning of the binding sites relative to the start of transcription. By including the first exon and intron of *SpPks* after the -2kb upstream region, that region was therefore shifted 4.5kb upstream of the start of transcription of *gfp*. If there were any positional relationships or DNA-looping of the promoter region necessary, this shift may have resulted in the loss of expression of GFP in this construct.

Statistically, the only differences seen between constructs were for the -3kb construct and the -1kb construct as well as the -3kb and gcm mut constructs. This is most likely attributed to the high level of variance that was seen between experiments of the same construct. Most notably, the experiments conducted on 5/31/07, 6/12/07 and 6/19/07 (dates that are outside of the normal breeding season resulting in poor gamete quality) created the majority of this high variability. The -2kb construct tested on 5/31 and 6/19 only produced expression in 27% and 20% of the embryos respectively, down from levels of 50-66% from other experiments. Likewise, the -1.5kb construct varied a great deal. Two experiments resulted in expression in 46% and 57% of embryos, while the two experiments conducted on 6/12 and 6/19 reduced these numbers to 20% and 0%. The 6/12 and 6/19 experiments also drove down the percentage of expressing embryos for the -2 to -1kb construct. An experiment conducted on 4/4/07 (with good gametes) produced expression in 57% of the embryos scored. The two June experiments only showed expression in 8% and 0% of embryos. These constructs need to be studied further using higher quality gametes to determine if there are any statistical differences to be seen between them that might be attributed to the difference in number of putative DNA-binding sites. Alternatively if I analyze the data removing the results for each replicate experiment performed outside the normal breeding season, some of the differences in the percentages of expressing

embryos could still be attributed to different levels of DNA incorporation between experiments, although this was mostly controlled for by testing the same construct (usually the -3kb or -2kb construct) in every experiment. Conclusions can then be drawn about the percentages of expressing embryos for other constructs relative to the percentages for the -3kb or -2kb constructs between experiments. Removing these experiments, there appears to be no differences between the constructs with varying numbers of GATA E sites, though these sites may still play a role that could be further explored by QPCR.

Each construct that produced expression of GFP contained two to five putative GATAE sites (two in the -2 to -1kb, four in the -1.5kb construct and 5 in the -2kb and -3kb constructs). It is possible that these additional GATAE sites produce additive effects on GFP expression as an increase from two to four to five sites might result in an increase of the rate of transcription of *gfp*. An increase in the rate of transcription could result in GFP observation in embryos that otherwise would not have produced detectable levels of the protein. Sometimes, though, the contribution of each DNA-binding site to enhance transcription is very small and difficult to detect eliminating one or two sites at a time.

Interestingly, each construct that recapitulated expression in pigment cells contained the putative GCM binding site. The -1kb construct that omitted this DNA sequence showed only 3% expression of GFP in pigment cells. This indicates that the -1kb construct lacks an essential positive regulator. The mutagenesis of the GCM DNA binding site within the -2kb construct showed a complete depletion of GFP expression giving compelling evidence that the putative GCM binding site at -1,179kb is a real transcription factor binding site that is necessary for the expression of *SpPks* in pigment cells.

Quantitative Analysis of *SpPks* Promoter Function

To evaluate the temporal expression produced by the *SpPks* promoter, QPCR was used to measure transcript expression throughout development. The results obtained by QPCR showed that the -2kb region overall recapitulates the endogenous *pks* expression. During the early stages (15-21 hours) expression of *gfp* by the -2kb promoter was notably higher than the expression of endogenous *pks*. This is most likely due to the presence of a higher copy number of injected *gfp* genes than the one copy of endogenous *pks* within the genome resulting in higher early expression. In later stages there is a decrease in expression compared to the early stages with some mild oscillation of *gfp* expression compared to the endogenous *pks*. Taken together with the fact that the -2kb construct did not have as high of a percentage of embryos expressing GFP in pigment cells as the -3kb construct, it is possible that the -2kb construct is missing one or more transcription factor sites that help to stabilize expression throughout development and also maintain the higher earlier expression levels.

Quantitative analysis of transcript accumulation of the gcm mutagenized construct showed drastically different results as compared to the endogenous expression of *SpPks*. While there is a peak of expression seen at 21 hours, for all other time points there is virtually no expression of *gfp*. From these data, it is clear that the mutagenized site corresponds to a functional transcription factor DNA-binding site, specifically for GCM, based on the binding site sequence homology to other GCM homologues. These results strongly suggest that *SpPks* is a direct target of GCM. While GCM is necessary for correct expression, the peak at 21 hours implies that it is not the only transcription factor acting on the -2kb construct and also not entirely necessary for the start of transcription. As there are GATAE sites within the -2kb region,

GATAE is expressed in the mesoderm until 21 hours and GATAE is a known positive regulator of *SpPks*, it is likely that GATAE at least partially contributes to the peak at 21 hours. Alternatively, other unknown or known positive regulators of *SpPks* such as *Brachyury* and *Kruppel-like* may also contribute to this peak through currently unknown pathways.

Analysis of *SpPks* Promoter Structure

Generally, the overall structure of a gene promoter is modular, consisting of several nonoverlapping segments of DNA that each contribute specifically to the overall pattern of gene expression. Given that *cis*-regulatory modules are typically 200-500 base pairs in length, several putative modules can be seen within the -2kb construct. Each pair of GATAE sites may constitute a module as they are all within 500 base pairs of one another (460, 455, 121 and 138 base pairs respectively). More importantly, the GCM site at -1,179 may form a *cis*-regulatory module with both or either of the two flanking GATAE sites. The GATAE site located 5' of the GCM site is 207 base pairs away while the GATAE site flanking the 3' region is 248 base pairs away. A *cis*-regulatory module including both proximal GATAE sites and the GCM site would be 455 base pairs in length.

Cis-regulatory modules for differentiation genes such as *SpPks* are typically unique, in that they are often compact and do not include sites for repressors (Davidson, 2001). Differentiation genes do not need to lay the groundwork for new developmental states or specific patterning, as specification has already occurred. Genes that are only active in a specific cell type, such as muscle cells, skeletal cells, or pigment cells are already confined spatially and temporally by the transcription factors combinations that are specifically expressed in such cells. As such, differentiation genes are often controlled by one key positive regulator and may be

assisted by one or more enhancer modules. An interesting example of this is the well-characterized *cis*-regulatory architecture of the sea urchin gene, *Endo16*. *Endo16* is expressed in the vegetal plate and gut during development of *S. purpuratus* embryos and is controlled by six *cis*-regulatory modules that bind many different transcription factors early in development. Later, after the gut starts to differentiate, the majority of this *cis*-regulatory region is not needed, and the terminal differentiation stage of *Endo16* is controlled by a single module lacking any sites for repressor proteins (Yuh *et al.,* 1994; Yuh *et al.,* 1996; Yuh and Davidson, 1996). *Endo16* belongs to a differentiation gene battery for endoderm specific genes that also includes *CyIIa*. The one key positive regulator (a currently unknown protein) that binds to the differentiation stage module of *Endo16* is the same positive regulator for the differentiation stage of *CyIIa*. A 440 base pair reporter construct that was found to be necessary and sufficient to recapitulate *CyIIa* expression contains two of these *Endo16* regulator DNA-binding sites, the mutagenesis of which eliminated expression (Arnone *et al.,* 1998).

Similarly, the promoter region for *SpPks* appears to be largely controlled by a single transcription factor, GCM, as when the binding site is mutagenized, expression is nearly completely eliminated. *SpPks* also does not appear to contain any sites for repressor proteins in its *cis*-regulatory architecture, as serial deletions of the promoter never resulted in ectopic expression. Typically, as a promoter region that reliably recapitulates the correct spatial expression of a non-differentiation gene is serially deleted, an increase in ectopic expression is observed as DNA-binding sites for negative regulators are eliminated. The putative modules containing GATAE binding sites identified above may act as enhancers for *SpPks* expression as has been discussed.

The Gene Regulatory Network of Pigment Cells

Previous works have shown that the differentiation of pigment cells is dependant on *Notch* signaling, acting through *Su(H)* to control the expression of *gcm* (Sherwood and McClay, 1999; Ransick *et al.,* 2002; Ransick and Davidson, 2006). In this work, I hypothesized that *gcm* was a direct, positive regulator of *SpPks*. The analysis of the promoter region performed in this study supports that hypothesis. Based on consensus sequence, a putative GCM DNA-binding site has been identified and experimentally tested to show that GCM is necessary for the expression of *SpPks*. Other positive regulators such as *gata E, brachyury,* and *kruppel-like* may act as enhancers for this expression (Davidson *et al.,* 2002 (2); Yamazaki *et al.* 2007). The pathway through which *brachyury* acts is still unknown, though based on the location of expression it is likely that it acts in parallel to the Delta/Notch pathway (Peterson *et al.,* 1999). *Kruppel-like* is involved in the Wnt pathway and therefore acts independently and parallel to the Delta/Notch pathway (Yamazaki *et al.* 2007). Differentiation genes such as *SpPks,* though often require only one master regulator, are commonly regulated by multiple positive inputs that act in parallel to stabilize the system and/or enhance the level of expression. These multiple positive inputs are locked into a feed forward mechanism that controls the expression of the differentiation gene (Oliveri *et al.,* 2008). Several GATAE binding site-containing modules have been identified within the minimal promoter for *SpPks* by means of consensus sequence. The architecture of this minimal promoter is similar to that of other differentiation genes as *SpPks* appears to be controlled by a single key regulator in a compact module lacking any repressor binding sites.

Future Work

Replication of the gcm mutagenized QPCR data must still be performed in order to draw any meaningful conclusions from the data. Furthermore, co-injection of the gcm mutagenized construct with a Morpholino Anti-Sense Oligonucleotide (MASO) for GATAE would establish if GATAE was responsible for the peak at 21 hours. MASO block the translation of mRNA to protein by binding to the mRNA and physically preventing either the translational machinery to work or blocking proper splicing from occurring, essentially knocking down that protein in the injected individual. If GATAE is binding to the gcm mutagenized construct to create expression at 21 hours, knocking down GATAE with a MASO specific for GATAE would eliminate the expression seen at 21 hours. If GATAE is working additively with other transcription factors, the expression will decrease but may not be completely eliminated. If GATAE is not binding to the construct the results should be the same as in the gcm mutagenized construct alone.

If other transcription factors than GATAE are suggested by the MASO experiment, a yeast one-hybrid (Y1H) methodology may be employed. In a Y1H system, many protein-DNA interactions can be identified without any prior knowledge of possible interactions. This is done by using the promoter sequence for a gene as "bait" DNA sequence and inserting it into a vector containing reporter genes (such as *His3* and *LacZ*). This vector is then incorporated into the genome of yeast. To identify what proteins might interact with this sequence of DNA, cDNA libraries will be made from total RNA of sea urchin embryos at different developmental time points. This cDNA is then inserted into an expression vector in frame with the sequence of an activating domain (AD) necessary for expression of the previously mentioned reporter genes, and again the vector is incorporated into the yeast genome. When these "prey" proteins bind to

the “bait” DNA sequence, the yeast is both able to grow in selective media and express the LacZ reporter gene, allowing for reduced possibilities of false positives. The cDNA encoding the proteins that are identified as interacting with the promoter region will then be sequenced using primers for the AD (Deplancke *et al.* 2004). Proteins that are known to not be expressed in pigment cells will be eliminated and Whole Mount In-Situ Hybridization (WMISH) will be performed to test if previously unknown proteins are actually expressed in pigment cells. Alternatively, given the fact that more genomic sequences from other sea urchin species are now available, comparative genomics may be used to identify areas of conserved DNA sequences, which are more likely to have a cis-regulatory function.

REFERENCES

Akiyama, Y., Hosoya, T., Poole, A.M., and Hotta, Y. (1996). The gcm-motif: A novel DNA-binding motif conserved in *Drosophila* and mammals. *PNAS* **93**: 14912-14916.

Arnone, M.I., Bogarad, L.D., Collazo, A., Kirchhamer, C.V., Cameron, R.A., Rast, J.P., Gregorian, A., and Davidson, E.H. (1997). Green Fluorescent Protein in the sea urchin: new experimental approaches to transcriptional regulatory analysis in embryos and larvae. *Development* **124**: 4649-4659.

Arnone, M.I., Martin, E.L., and Davidson, E.H. (1998). Cis-regulation downstream of cell type specification: a single compact element controls the complex expression of the CyIIa gene in sea urchin embryos. *Development* **125**: 1381-1395.

Balhoff, J.P. and Wray, G.A. (2005). Evolutionary analysis of the well characterized *endo16* promoter reveals substantial variation within functional sites. *PNAS* **102(24)**: 8591-8596.

Bolouri, H. and Davidson, E.H. (2002). Modeling DNA sequnce-based cis-regulatory gene networks. *Developmental Biology* **246(1)**: 2-13.

Brown, C.T., Rust, A.G., Clarke, P.J.C., Pan, Z., Schilstra, M.J., De Buysscher, T., Griffin, G., Wold, B.J., Cameron, R.A., Davidson E.H. and Bolouri, H. (2002). New computational approaches for analysis of *cis*-regulatory networks. *Developmental Biology* **246(1)**: 86-102.

Calestani, C., Rast, J.P. and Davidson, E.H. (2003). Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. *Development* **130**: 4587 – 4596.

Cameron, R.A., Fraser, S.E., Britten, R.J. and Davidson, E.H. (1991). Macromere cell fates during sea urchin development. *Development* **113**: 1085 – 1091.

Castoe, T.A., Stephens, T., Noonan, B.P. and Calestani, C. (2006). A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. *Gene* **392:** 47-58.

Damle, S., Hanser, B., Davidson, E.H., and Fraser, S.E. (2006) Confocal quantification of cis -regulatory reporter gene expression in living sea urchin. *Developmental Biology* **299**: 543-550.

Davidson, E.H., Cameron, R.A. and Ransick, A. (1998). Specification of cell fate in the sea urchin embryo: summary and some proposed mechanisms. *Development* **125(17)**: 3269-3290.

Davidson, E.H. (2001). Genomic Regulatory Systems. San Diego: Academic Press.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z.j., Schilstra, M.J., Clarke, P.J.C., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., and Bolouri, H. (2002). A genomic regulatory network for development. *Science* **295**: 1669-1678.

Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Schilstra, M.J., Clarke, P.J.C., Rust, A.G., Pan, Z.j., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., and Bolouri, H. (2002). A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Developmental Biology* **246**: 162-190.

Davidson, E.H. (2006). The Regulatory Genome. San Diego: Academic Press.

Deplancke, B., Dupuy, D., Vidal, M., and Walhout, A.J.M. (2004). A Gateway-Compatible Yeast One-Hybrid System. *Genome Res.* **14**: 2093-2101.

Ehebauer, M., Hayward, P., and Martinez-Arias, A. (2006). Notch Signaling Pathway. *Science Signaling*. **364**: cm7

Evans, T., Reitman, M, and Felsenfeld, G. (1988). An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *PNAS* **85(16)**: 5976-5980.

Gibson, A.W. and Burke, R.D. (1985). The origin of pigment cells in embryos of the sea urchin *Strongylocentrotus purpuratus*. *Dev*. *Biol*. **107**: 414 – 419.

Gibson, A.W. and Burke, R.D. (1987). Migratory and inasive behavior of pigment cells in normal and animalized sea urchin embryos. *Exp. Cell Res.* **173**: 546-557.

Gilbert, S.F. (2000). Developmental Biology, 6th Edition. Massachusettes: Sinauer Associaties, Inc.

Griffiths, M. (1965). A study of the synthesis of naphthoquinone pigments by the larvae of two species of sea urchins and their reciprocal hybrids. *Dev. Biol.* **11**: 443-447.

Hopwood, D.A. (1997). Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97**:2465 – 2497.

Hopwood, D.A. (2004). Cracking the polyketide code. *PLoS Biology* **2(2):** 0166-0169.

Hörstadius, S. (1939). The mechanics of sea urchin development, studied by operative methods. *Biology Review* **14**: 132 – 179.

Howard, E.W., Newman, L.A., Oleksyn, D.W., Angerer, R.C. and Angerer, L.M. (2001). *SpKrl*: a direct target of β-catenin regulation required for endoderm differentiation in sea urchin embryos. *Development* **128**: 365-375.

Istrail, S., and Davidson, E.H. (2005). Logic functions of the genomic cis-regulatory code. *PNAS* **102(14)**: 4954-4959.

Lee, E.C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D.A., Court, D.L., Jenkins, N.A., Copeland, N.G. (2001). A highly efficient *Escherichia coli*-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**: 56-65.

Lee, P.Y., and Davidson, E.H. (2004). Expression of *Spgatae*, the *Stongylocentrotus purpuratus* otholog of vertebrate GATA4/5/6 factors. *Gene Expr. Patterns* **5**: 161-165.

Levine, M., and Davidson, E.H. (2005). Gene regulatory networks for development. *PNAS* **102(14)**: 4936-4942.

Martinez, P. and Davidson, E.H. (1997). SpHmx, a sea urchin homeobox gene expressed in embryonic pigment cells. *Dev. Bio.* **181**: 213-222.

McClay, D.R., Peterson, R.E., Range, R.C., Winter-Vann, A.M., and Ferkowicz, M.J. (2000). A micromere induction signal is activated by β-catenin and acts through Notch to initiate specification of secondary mesenchyme cells in the sea urchin embryo. *Development* **127**: 5113-5122.

Miller, R.N., Dalamagas, D.G., Kingsley, P.D., and ettensoh, C.A. (1996). Expression of S9 and actin CyIIa mRNAs reveals dorso-ventral polarity and mesodermal sublineages in the vegetal plate of the sea urchin embryo. *Mech. Dev.* **60**: 3-12.

Minokawa, T., Rast, J.P., Arenas-Mena, C., Franco, C.B. and Davidson, E.H. (2004). Expression patterns of four different regulatory genes that function during sea urchin development. *Gene Expr. Patterns* **4(4)**: 449-456.

Nemer, M., Rondinelli, E., Infante, D., Inante, AA. (1991). Polyubiquitin RNA characteristics and conditional induction in sea urchin embryos. *Developmental Biology* **145(2)**: 255-265.

Oliveri, P., Carrick, D.M., and Davidson, E.H. (2002). A regulatory gene network that directs micromere specification in the sea urchin embryo. *Developmental Biology* **246**: 209-228.

Oliveri, P., Walton, K.D., Davidson, E.H. and McClay, D.R. (2006). Repression of mesodermal fate by *foxa*, a key endoderm regulator of the sea urchin embryo. *Development* **133**: 4173-4181.

Oliveri, P., Tu, Q., and Davidson, E.H. (2008). Global regulatory logic for specification of an embryonic cell lineage. *PNAS* **105(16)**: 5955-5962.

Peterson, K.J., Harada, Y., Cameron, R.A., and Davidson, E.H. (1999). Expression Pattern of *Brachyury* and *Not* in the Sea Urchin: Comparative Implications for the Origins of Mesoderm in the Basal Deuterostomes. *Developmental Biology* **207(2)**: 419-431.

Ransick, A. and Davidson, E.H. (1995). Micromeres are required for normal vegetal plate specification in sea urchin embryos. *Development* **121**: 3215-3222.

Ransick, A., Rast, J.P., Minokawa, T., Calestani, C. and Davidson, E.H. (2002). New early zygotic regulators expressed in endomesoderm of sea urchin embryos discovered by differential array hybridization. *Developmental Biology* **246**: 132-147.

Ransick, A., and Davidson, E.H. (2006). *Cis*-regulatory processing of Notch signaling input to the sea urchin *glial cells missing* gene during mesoderm specification. *Developmental Biology* **297**: 587-602.

Rast, J.P., Amore, G., Calestani, C., Livi, C.B., Ransick, A., and Davidson, E.H. (2000). Recovery of developmentally defined gene sets from high-density cDNA macroarrays. *Developmental Biology* **228**: 270-286.

Rast, J.P., Cameron, R.A., Poustka, A.J. and Davidson, E.H. (2002). *brachyury* target genes in the early sea urchin embryo isolated by differential macroarray screening. *Dev. Biol.* **246**: 191-208.

Ruffins, S.W. and Ettensohn, C.A. (1993). A clonal analysis of secondary mesenchyme cell fates in the sea urchin embryo. *Developmental Biology* **160**: 285-288.

Ruffins, S.W. and Ettensohn, C.A. (1996). A fate map of the vegetal plate of the sea urchin (Lytechinus variegatus) mesenchyme blastula. *Development* **122**: 253 – 263.

Sea Urchin Genome Sequencing Consortium (2006). The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**: 941-952.

Service, M., and Wardlaw, A.C., (1984). Echinochrome-A as a bactericidal substance in the coelomic fluid of *Echinus esculentus* (L.). *Comp. Biochem.* **79B**: 161-165.

Sherwood, D.R. and McClay, D.R. (1999). LvNotch signaling mediates secondary mesenchyme specification in the sea urchin embryo. *Development*. **126**: 1703-1713.

Sherwood, D.R. and McClay, D.R. (2001). LvNotch signaling plays a dual role in regulating the position of the ectoderm-endoderm boundary in the sea urchin embryo. *Development* **128**: 2221-2232.

Smith, L.C., Britten, R.J., and Davidson, E.H. (1992). *SpCoel1*: A sea urchin profilin gene expressed specifically in coelomocytes in response to injury. *Mol. Biol. Cell*. **3**: 403-414.

Sweet, H.C., Hodor, P.G., and Ettensohn, C.A. (1999). The role of micromere signaling in Notch activation and mesoderm specification during sea urchin embryogenesis. *Development* **126**: 5255-5265.

Sweet, H.C., Gehring, M., and Ettensoh, C.A. (2002). LvDelta is a mesoderm-inducing signal in the sea urchin embryo and can endow blastomeres with organizer-like properties. *Development* **129**: 1945-1955.

Wittkopp, PJ. (2006). Evolution of *cis*-regulatory sequence and function in Diptera. *Heredity* **97**: 139-147.

Yamazaki, A., Kawabata, R., Shiomi, K., Tsuchimoto, J., Kiyomoto, M., Amemiya, S., and Yamaguchi, M. (2007). Kruppel-like is required for nonskeletogenic mesoderm specification in the sea urchin embryo. *Developmental Biology* **314**: 433-442.

Yuh, C.H., Ransick, A., Martinez, P., Britten, R.J. and Davidson, E.H. (1994). Complexity and organization of DNA-protein interactions in the 5'-regulatory region of an endoderm-specific marker gene in the sea urchin embryo. *Mech. Dev.* **47**: 165-186.

Yuh, C.H. and Davidson, E.H. (1996). Modular cis-regulatory organization of Endo16, a gut-specific gene of the sea urchin embryo. *Development* **122**: 1069-1082.

Yuh, C.H., Moore, J.G. and Davidson, E.H. Quantitative functional interrelations within the cis-regulatory system of the S. purpuratus Endo16 gene. *Development* **122**: 4045-4056.