Chapter 4:

Characterization of alternatively spliced variants of Bmi-1 during early chick development

Jane Khudyakov, Tatjana Sauka-Spengler, and Marianne Bronner-Fraser

INTRODUCTION

The Polycomb group (PcG) of epigenetic repressors is a highly conserved bipartite protein complex that regulates gene expression in a vast number of organisms ranging from plants to mammals (Whitcomb et al., 2007; Kohler and Villar, 2008). Polycomb genes have been implicated in a number of key developmental processes such as maintenance of stem cell pluripotency, prevention of cell senescence, lineage restriction and differentiation, and axial patterning (Schuettengruber et al., 2007). The PcG was first identified in *Drosophila,* and orthologs of the fly Polycomb genes have since been identified in a number of organisms as core components of the repressive complexes. In recent years, a staggering number of other, non-core Polycomb Repressive Complex (PRC) partners have been characterized in vertebrates, many of which share great similarity in structure and function (Fig. 1.2B, Chapter 1, Whitcomb et al., 2007). Given that these genes regulate critical developmental processes and that mutations affecting PcG members often lead to embryonic lethality, carcinogenesis, deregulation of the stem cell state, and/or improper patterning of the body plan, careful regulation of the Polycomb complexes themselves is necessary for viable embryonic development (Pietersen and van Lohuizen, 2008). In addition to the antagonistically functioning Trithorax Group and upstream regulators such as the Asx family of Polycomb enhancers, an intriguing possibility is that alternative splice variants may modulate PcG protein function (Schwartz and Pirrotta, 2007; Baskind et al., 2009).

Alternative splicing has become widely recognized as one of the most prevalent means of generating proteomic diversity and phenotypic complexity in

higher-order eukaryotes, by modulating protein function. Pre-mRNA splicing events most commonly involve exon skipping or shuffling, usage of alternative 5′ and 3′ splice sites and retention of introns; and splicing factors that mediate these processes are expressed in a highly regulated, rapidly inducible, and tissue-specific manner (Maniatis and Tasic, 2002; Lareau et al., 2004; Stamm et al., 2005; Kim et al., 2008). Protein variants produced by alternative splicing commonly exhibit diverse changes in activity due to removal or alteration of functional domains or localization signals and often function as dominant-negatives. Functional modifications include changes in affinity for other proteins, ligands, or DNA, alteration in signaling or transactivation activity, and changes in intracellular localization, protein stability, and post-translational modifications. In addition, alternative splicing within 5′- and 3′-untranslated regions may control RNA expression levels or its localization, stability, and translation efficiency, respectively (Lareau et al., 2004; Stamm et al., 2005).

The prevalence of alternatively spliced isoforms in vertebrate proteomes and their high degree of evolutionary conservation suggest that splice variants may play important roles in vertebrate physiology. Indeed, alternatively spliced isoforms have been shown to regulate many important functions such as apoptosis, cell type specification, organ patterning, and neuronal activity. Accordingly, transgenic mice carrying variant-specific mutations often exhibit significant developmental and functional abnormalities (Venables, 2006; Moroy and Heyd, 2007; Holland and Short, 2008). Not surprisingly, differentially expressed splice isoforms are prevalent within large protein families of the signal transduction and transcription factor categories that regulate development and

differentiation, such as FGF and Pax, and are conserved across chordates (Holland and Short, 2008).

Perturbations of the delicate equilibrium between splice variants often lead to cancer, and many tumors are characterized by over-expression of alternatively spliced isoforms (Venables, 2006). A number of alternative splice variants have also been identified within chromatin-modifying protein families that are often perturbed in cancers; these include histone acetyltranferases, DNA methyltransferases, and Polycomb repressors. For example, a truncated variant of the DNA methyltransferase DNMT3 which lacks the conserved methytransferase motif has been shown to compete with the full-length protein for targeting to chromatin, leading to DNA hypomethylation, instability, and cancer (Saito et al., 2002). Likewise, many variants of PcG protein members such as Mph1, Mph2, Cbx6, Cbx7, and L3mbt1, have been identified and shown to lack key protein interaction and chromatin recognition motifs, although their physiological functions have not yet been analyzed (Alkema et al., 1997a; Yamaki et al., 2002; Tajul-Arifin et al., 2003; Li et al., 2005). Here, we show that that the PcG member Bmi-1 is characterized by five alternative splice variants, which we have examined within the *in vivo* context of the developing chicken embryo.

The Polycomb Repressive Complex 1 (PRC1) member Bmi-1 is the vertebrate homolog of *Drosophila posterior sex combs* (*psc*) which was one of the first PRC1 members to be identified in the mouse as a stem cell factor (Park et al., 2004). Mutant mice harboring null mutations or overexpressing Bmi-1 exhibit defects in proliferation of hematopoietic stem cells or develop lymphomas, respectively (Haupt et al., 1993; Park et al., 2003). Bmi-1 mutants also display defects in neural stem cell maintenance, as well as axial transformations due to

dysregulation of homeotic genes (van der Lugt et al., 1994; van der Lugt et al., 1996; Molofsky et al., 2003). In addition, Bmi-1 has been shown to function in mouse ESC development by directly associating with and repressing a large number of developmental regulator genes, thereby preventing premature differentiation (Bracken et al., 2006; Dietrich et al., 2007). In view of these findings, we hypothesized that Bmi-1 may play a similar role during development of multipotent neural crest progenitors. We have found that the chick Bmi-1 homolog is expressed by neural crest progenitors during early developmental stages and functions in cooperation with other PRC1 partners to negatively regulate members of the neural crest gene regulatory network (Chapter 3).

The biochemical activity and structure of the Bmi-1 protein have been thoroughly described. It is comprised of 326 amino acids and contains several highly conserved protein domains that are necessary its activity as well as interaction with other PRC1 members (Fig. 1.3A, Chapter 1). The N-terminal RING finger domain is characterized by a conserved cysteine-rich zinc finger binding motif and is necessary for interaction with the other RING finger-containing proteins, Ring1A and Ring1B. Two conserved cysteine residues within this domain and a stretch of several downstream amino acids have been shown to be critical for protein interaction (Hemenway et al., 1998). In addition, the RING finger domain and a downstream sequence containing a putative nuclear localization signal (NLS) are necessary for the oncogenic activity of Bmi-1 in transgenic mice and cell transformation in culture, as well as for prevention of replicative senescence in fibroblasts, suggesting that both the presence of the RING domain and subnuclear localization are critical for Bmi-1 function (Cohen

et al., 1996; Alkema et al., 1997b; Itahana et al., 2003). The full-length Bmi-1 protein also contains a centrally located helix-turn-helix-turn-helix-turn (HTHTHT) domain that is necessary for interaction with the mouse *polyhomeotic* homologs Mph1 and Mph2. In addition, presence of the HTHTHT domain is critical for transcriptional repression of Hox genes and skeletal transformation, and, to a lesser extent, for oncogenic potential (Cohen et al., 1996; Alkema et al., 1997a, 1997b). Finally, the C-terminal part of Bmi-1 contains a proline, glutamine, serine, and threonine-rich domain (PEST) that has no repressive or oncogenic function but may be involved in targeting the protein for degradation (Cohen et al., 1996; Alkema et al., 1997b). In addition, a putative MAPK pathway phosphorylation site is found within the PEST domain, which has been shown to regulate association of Bmi-1 with chromatin (Voncken et al., 1999; Voncken et al., 2005).

Bmi-1 deletion studies and *in vitro* interaction assays have demonstrated that truncated portions of the protein containing intact RING and HTHTHT domains can homodimerize with the full-length protein and bind to other PRC1 factors via homologous regions (Hemenway et al., 1998; Satijn and Otte, 1999). In addition, over-expression of mutant Bmi-1 constructs lacking any of the protein interaction domains in fibroblast cell culture causes a dominant-negative phenotype by inducing premature replicative senescence (Itahana et al., 2003). We hypothesized that naturally occurring truncated splice variants of Bmi-1 may function *in vivo* to modulate the activity of the full-length protein and affect interactions between partners of the PRC1, possibly in a dominant-negative manner. We have characterized three such splice variants that were identified in a chick cDNA library screen and found that two of them represent truncated N-

terminal isoforms containing the RING domain, while a third, C-terminal variant, contains the HTHTHT domain but lacks the RING finger. We have performed expression analysis by RT-PRC and QPCR using whole chick embryos and have found that the N-terminal variant V4 and the C-terminal variant V6 are expressed during early chick development from gastrulation to neural crest migration stages. We have also demonstrated by over-expression analysis that V4 likely functions in a dominant-negative manner to inhibit Bmi-1 activity, leading to upregulation of the *Msx1* target gene.

MATERIALS AND METHODS

Sequence analysis

A 600 bp fragment of the chick Bmi-1 homolog, amplified using degenerate PCR approach, was used to probe a 4-12 somite-stage macroarrayed cDNA library (Gammill and Bronner-Fraser, 2002; Fraser and Sauka-Spengler, 2004). Eight positive clones isolated in this screen were sequenced and the resulting sequences were analyzed and aligned using EditSeq and SeqMan applications (DNASTAR Lasergene 8, DNASTAR, Inc).

Genomic analysis and intron sequencing

The Bmi-1 genomic locus is very poorly characterized and, apart from a partial 3'-UTR sequence found within the unassembled random sequence collection, entirely absent from chick genome assembly. To gain further insight into genomic organization of Bmi-1, a macroarrayed chicken BAC library (Chori 261, purchased from BACPAC (http://bacpac.chori.org)) was screened using a chick Bmi-1 fragment and the resultant positive BAC clones were identified within the chicken genome assembly (http://genome.ucsc.edu). Analysis of BAC clones identified in our screen and positioning of BAC ends of clones absent in our screen have enabled us to narrow down the position of the Bmi-1 locus to ~20kb on chromosome 2 (chr2:17,642,600-17,662,000). The Bmi-1 gene is directly flanked by sperm associated antigen 6 and COMM domain-containing 3 genes and further analysis of synteny between chicken and mouse genomes confirms proper assignment of the Bmi-1 gene to this region. Mouse Bmi-1 protein sequence, which shares 93% identity with the chick, was used to predict the

location of intron-exon boundaries. We then designed several primer sets near the putative exon/intron boundaries in order to amplify the intronic regions by PCR using either ~1 μg of chick genomic DNA or 200 ng of BAC DNA template per reaction. The following primers were used: V123578F 5′-CGACCAGGATCAAAATCACC-3′, 2R 5′-GCAGTACTTGCTCGTCTC-3′, 2F 5′-GTCCAAGTGCACAAAACC-3′, 4R 5′-AGCAGCATAGAAATCCCT-3′, 4F 5′-TATGCTGCTCATCCGTCG-3′, 6R 5′- ATTTCCTTTTCGTTCCAGT-3′, 5F 5′-TCCATTGAGTTCTTTGAC-3′, 7R 5′-GCAGCGCAAATATCTTTT-3′, 7F 5′-AGTAAGATGGATATCCCC-3′, 9R 5′-GGGCCGCACGCGGTACTT-3′, RT1568 5′-TGTTTGCTTCCCGGTCCTTT-3′. Primers were designed using the EditSeq application (DNASTAR Lasergene 8, DNASTAR, Inc.). Reactions were carried out as follows: 94°C for 30 sec, 55°C for 30 sec, 72°C for 1 min, 30 cycles. PCR products were resolved on a 2% agarose gel and specific bands were isolated using the QIAquick Gel Extraction Kit (QIAGEN, Cat# 28706). The resulting products were cloned into pCR®2.1 TOPO vector using the TOPO® TA Cloning® Kit (Invitrogen, Cat# K4500-01) and transformed into One® Shot Top10 Chemically Competent Cells (Invitrogen, Cat# C4040-03). Positive colonies were picked and grown in a 96-well mini culture plate. QIAprep 96 Turbo Miniprep Kit (QIAGEN, Cat# 27191) was used to purify DNA from bacterial cultures. DNA sequencing reactions were set up using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Cat# 4337455) with 1 pmol/μL of each SP6 and T7 primers and 1 μL miniprep DNA in each reaction. Sequencing was performed by Miki Yun (Davidson lab, Caltech). Sequences were aligned using the SeqMan application (DNASTAR Lasergene 8, DNASTAR,

Inc). The following BAC clones, used in PCR reactions, were obtained from BACPAC (http://bacpac.chori.org): CH261-130E19, CH261-107L20, CH261-180F13.

V6 protein translation assay

Sequence analysis of the V6 clone identified two possible open reading frames (ORF), both containing a different C-terminal portion of the Bmi-1 protein. The first putative truncated protein, if produced, would contain the NLS but none of the characterized protein domains. Alternatively, a longer putative protein would include the majority of the HTHTHT protein interaction domain, in addition to the NLS. To determine which of the reading frames is used to produce the V6 protein, *Xenopus* oocytes were injected with V6 mRNA in which either the short or long putative protein was myc-tagged. Western blot analysis of protein extracts from injected embryos using anti-myc and anti-Bmi-1 antibodies demonstrates that the protein translated from V6 mRNA corresponds to the second ORF and contains the HTHTHT domain and a NLS (data not shown).

Chick embryo incubation

Fertilized chicken eggs were obtained from AA Enterprises (Ramona, CA) and incubated at 38°C in a humidified incubator (Lyon Electric, Chula Vista, CA). Embryos were staged according to the Hamburger and Hamilton chick staging system (Hamburger and Hamilton, 1992).

RT-PCR

Whole embryos for RT-PCR were collected and trimmed to remove extra-embryonic membranes in Ringer's solution on ice. Several embryos of each approximate stage were pooled together and lysed in RNAqueous® Lysis Buffer (Ambion, Cat# AM1912). Total RNA was isolated using the RNAqueous® Kit according to manufacturer's protocol (Ambion, Cat# AM1912). Reverse transcription was carried out using First Strand cDNA Synthesis Kit for RT-PCR (AMV) (Roche Applied Science, Cat# 11483188001). Specific RT primers were designed within the unique 3'-UTR of V4 (corresponding to intron 4, RT4 5'-AACCGCCAAAGCTGCAAACT-3'), the distal 3'-UTR exon shared by V2 and V3/7 (RT237 5'-TCGACCAAAGCAAAGCACGA-3'), and within the proximal 3'-UTR portion shared by full-length variants and V6 (RT1568 5'-TGTTTGCTTCCCGGTCCTTT-3'). PCR reactions were carried out using 500 ng of cDNA template and Taq DNA polymerase (Invitrogen, Cat# 10342-020). PCR primers for amplifying specific variant sequences from cDNA were designed as follows: full-length variants were amplified using the primer set designed within the coding region of Bmi-1, with a forward primer in exon 1 (V123578F 5'-CGACCAGGATCAAAATCACC-3') and a reverse primer in exon 9 (V135678R 5'-TATGGAGGATTTCCGTGCTC-3'). The following PCR settings were used: 94°C for 30 sec, 65°C for 30 sec, 72°C for 1 min, 30 cycles. V2 was detected using the same forward primer (V123578F) and a reverse primer in the distal 3'-UTR exon (V237R 5'-TCCATCTCATCTCCCTCGAC-3'). Amplification conditions were the same as above, with the exception of 30 sec elongation at 72°C. This primer set also amplified the V3/7 full-length variant, yielding a larger size

fragment, albeit with lower efficiency. V4 was amplified using the primer set designed within its 3'-UTR region, with a forward primer in intron 2 (V46F 5'-AACCTCAGCCCCCGAACTC-3') and a reverse primer in intron 4 (V4R 5'-AAAAGGAAAGAGCGGAGCA-3'). The PCR reaction was carried out for 35 cycles, with a 45 sec elongation step. V6 was amplified using the same V46F forward primer, situated in its 5'-UTR, and the V135678R reverse primer in exon 9, described above. The V6 PCR reaction was performed for 35 cycles with a 1 min 20 sec elongation step. All primers were designed using Primer3 software (http://frodo.wi.mit.edu) and synthesized by IDT. PCR products were resolved on a 1% or 2% agarose gel.

QPCR

cDNA for use in QPCR was synthesized using RNA extracted from single whole chick embryos prepared by Pablo Strobl and Tatjana Sauka-Spengler. The reverse transcription (RT) reaction was probed using random hexamers and each sample was accompanied by a minus RT control. QPCR was performed using the 96-well plate ABI 7000 QPCR machine (Applied Biosciences) with SYBRGreen iTaq Supermix with ROX (Bio-Rad, Cat# 172-5101). Primers were used at a concentration of 450 nM with the exception of *Gapdh*, which was used at 150 nM, in a 20 μL reaction. Variant-specific sets of primers were designed using the Primer3 software (http://frodo.wi.mit.edu) and synthesized by IDT. The sequences of primers used are as follows: V13578spec1F 5'-AGAGAAAGAAAAGTCGAAGGAGG-3', V13578spec2F 5'-TCACGTCGATCTGGAAAGTG-3', V6spec4F 5-GATGCTCCTTTCCAGGTCAG-

3′, V6spec4R 5′-ATTAGAGCCATTGGCAGCAT-3′, V2spec2F 5′-CTCGTACCGGGCCTTTTC-3′, V2spec2R 5′-CACGTCAATGACTTCCATCTC-3′, V4spec6F 5′-GAGTGCCTGCACTCCTTCTG-3′, V4spec6R 5′-TTCACGGCTCCTTTCAGATT-3′. Each sample was run in three replicates to reduce errors created by pipetting. The baseline and threshold levels were set according to the Applied Biosystem software, and gene expression was calculated by the standard curve assay method as described in Applied Biosystems protocols. In detail, the results for different samples were interpolated from a line created by running four point standard curves for each primer set and then normalized against results for the *Gapdh* housekeeping gene. The standard cDNA was prepared from chick embryos collected between stages HH4 and HH10, at which we have shown expression of Bmi-1 variants by RT-PCR. Minus RT controls were tested for each set of primers, and showed no amplification. cDNA from three separate single embryos collected at HH4, HH6, HH8, and HH10 was individually tested in this assay.

Over-expression constructs

Using a high fidelity enzyme (Expand High FidelityPLUS PCR System, Roche, Cat# 03300242001), the open reading frame of the V4 variant, including the endogenous Kozak sequence, was amplified using full-length V4 cDNA obtained from a chick cDNA library screen (see above). The obtained fragment was ligated into several expression vectors: pCIG-IRES-GRP (pCIG-V4-GFP), pCIG-H2B-RFP (pCIG-V4-H2B-RFP), and pCIG-mem-RFP (pCIG-V4-memRFP (RFP with a membrane linker)). The V2 fragment was prepared in a similar manner and cloned into pCIG-H2B-RFP vector. Maxi preps were prepared using the QIAGEN

EndoFree Plasmid Maxi Kit (QIAGEN, Cat# 12362) and DNA was resuspended in Buffer EB (QIAGEN, Cat# 19086). Plasmids were diluted to 2-5 μg/μL concentration with Buffer EB (10 mM Tris pH 8) and 0.01% Blue Vegetable Dye (FD&C Blue 1, Spectra Colors Corp, Cat# 3844-45-9) for injection into chick embryos. Empty vectors were used as electroporation controls.

Electroporation

HH stage 3-5 chick embryos were explanted on Whatman filter paper rings and placed ventral-side up in Ringer's solution in an electroporation dish containing a platinum plate electrode in a shallow well. pCIG-V4-H2B-RFP or pCIG-V4-memRFP (or a combination of the two constructs), pCIG-V2-H2B-RFP, or the control pCIG-GFP empty vector was unilaterally injected into the lumen between the epiblast and vitelline membrane targeting the prospective neural plate border. The embryo was covered with a flattened-tip platinum electrode and five 7-volt, 50-millisecond pulses with 100-millisecond pauses in between were applied using a square-pulse electroporator. Embryos were cultured in thin albumin in a humidified 37°C incubator. After 6-8 hours, embryos were fixed in 4% paraformaldehyde at 4°C overnight and dehydrated to 100% methanol for analysis by *in situ* hybridization.

***In situ* hybridization**

Chick embryos were dissected in Ringer's solution and fixed in 4% paraformaldehyde at 4°C overnight. Whole-mount *in situ* hybridization was performed as described previously (Nieto et al., 1996; Xu and Wilkinson, 1998),

with some modifications involving more extensive washing adapted from a lamprey *in situ* protocol (Sauka-Spengler et al., 2007). The digoxigenin-labeled antisense *Msx1* probe was reverse transcribed from the chick EST template ChEST900p21 (BBSRC ChickEST Database (http://www.chick.umist.ac.uk)) using Promega buffers and RNA polymerases (Promega Corp), and purified with illustra ProbeQuant™ G-50 Micro Columns (GE Healthcare, product code 28-9034-08). Stained embryos were photographed in 50% glycerol on a Zeiss Stemi SV11 microscope using AxioVision software (Release 4.6) and processed using Photoshop 7.0 (Adobe Systems).

RESULTS

Bmi-1 is characterized by five alternatively spliced isoforms

Using the full-length Bmi-1 gene as a probe in a high stringency chick cDNA library screen, we have identified eight positive clones corresponding to Bmi-1 variants. Five clones contain an open reading frame (ORF) of 981 base pairs (bp), encoding the full-length 326-amino acid (aa) Bmi-1 protein. These clones are identical in coding sequence but vary within untranslated regions. In addition, we identified three clones that encode truncated Bmi-1 isoforms with ORFs ranging from approximately 200 to 450 bp. We found that the putative proteins encoded by these transcripts lack some of the functional domains present in the full-length Bmi-1 protein, which is illustrated in Figure 4.1. In order to determine whether Bmi-1 clones represent naturally occurring splicing isoforms of the Bmi-1 gene, we performed thorough characterization of Bmi-1 genomic locus and structural analysis of Bmi-1 variants.

In order to identify splicing events that generated Bmi-1 variants, we mapped and characterized the genomic locus of Bmi-1 and obtained intronic sequences by PCR using chick genomic DNA and BAC templates (see Materials and Methods for details). The full-length Bmi-1 gene contains nine coding exons, intercepted by eight introns and at least two non-coding exons contributing to 5'- and 3'-untranslated regions (UTR). The RING finger protein interaction motif is generated by the first two exons. One of two putative nuclear localization signals (NLS) lies within the fourth exon similarly to the mouse Bmi-1 protein, while the other one is found in the ninth exon. The helix-turn-helix-turn-helix (HTHTHT) domain is generated by the last three exons (Fig. 4.1).

The five full-length Bmi-1 clones (V1, V5, V8, V3, V7) have identical open reading frames; however, they differ in the 3'-UTR as two of the variants (V3, V7) contain a supplementary non-coding exon (Fig. 4.1A). Two distinct N-terminal variants, termed V2 and V4, are truncated in the C-terminal portion and thus lack the HTHTHT protein interaction domain. Variant V2, generated by the first three coding exons and the second 3'-UTR exon, contains the RING domain and a NLS, whereas the short V4 variant contains the first two exons and includes the RING domain only (Fig. 4.1B,C). The putative NLS found in the C-terminal region of V2 is generated when the last exon, which contributes to the 3'-UTR in the full-length V3/V7 clones, is spliced in frame with the third coding exon and is translated as part of the V2 protein. This rearrangement extends the N-terminal portion of V2, encoded by the first three exons, by approximately 60 amino acids and contributes a 21 amino acid-region that encodes an experimentally confirmed NLS motif (http://www.rostlab.org/cgi/var/nair/resonline.pl). A C-terminal variant, V6, is generated by the last three exons, encoding a protein that lacks the RING domain and the first five amino acids of the HTHTHT domain (Fig. 4.1D). Therefore, we have characterized several splice isoforms of Bmi-1, which include two unique full-length variants that are alternatively spliced in untranslated regions and N-terminal and C-terminal truncated variants lacking protein interaction and nuclear localization motifs.

Bmi-1 variants are expressed in the chicken embryo during early development in a stage-specific manner

We used RT-PCR to examine whether Bmi-1 variants are expressed in the chicken embryo during early stages of neural crest development. Total RNA was extracted from a pooled sample of several whole embryos collected at each stage of development from HH4 to HH10. Reverse transcription (RT) was performed using specific variant RT primers to ensure detection of low expression levels of variant mRNA and to exclude the possibility of genomic contamination in the cDNA library. Specific PCR primers for amplification of variant sequences from cDNA were designed as described in Materials and Methods. Each set of primer pairs was first tested for specificity in a PCR reaction using variant clones isolated from the chick library as templates. Stage-specific RT-PCR results demonstrate that full-length *Bmi-1*, which we refer to as *V8* for simplicity, and N-terminal variant *V4* are expressed during each stage of development from HH4 to HH10 (Fig. 4.2A,C). In contrast, a very weak *V2*-specific band is only detectable at HH9 (Fig. 4.2B). Intriguingly, *V6* is present at HH4 and is progressively downregulated at later stages (Fig. 4.2D).

Next, we used quantitative RT-PCR (QPCR) to accurately measure levels of truncated variant transcripts at several developmental stages and to examine how they differ between these time points. In addition, since we hypothesized that the short variants may be acting to modify full-length Bmi-1 function during development, we also wanted to quantitatively compare their expression levels. We analyzed expression at four specific stages – HH4, HH6, HH8, and HH10 – representative of events occurring during chick neural crest development. cDNA used in variant QPCR was synthesized from total RNA extracted from single embryos (Pablo Strobl and Tatjana Sauka-Spengler), and three embryos of each stage were individually tested in our assay. Variant-specific primer sets for

QPCR analysis were designed similarly to those used in RT-PCR reactions and tested for specificity using specific variant DNA templates as well as genomic DNA. Variant expression results were normalized to expression of the housekeeping gene *Gapdh*, allowing for control of the amount of input material.

We found that although expression levels varied slightly between single embryos due to stochastic variation of endogenous transcriptional levels and slight variations in the age of the embryos, trends in expression levels from stage to stage were conserved. Therefore, we present data using one representative embryo of each stage in Figure 4.3. We were unable to amplify *V2* at any stage including HH9, at which we found low expression by RT-PCR from pooled sample cDNA primed with V2-specific primers, suggesting that *V2* may not be transcribed at levels that are detectable in a single embryo. *V4* was expressed unvaryingly at low levels at every stage tested (Fig. 4.3A). In contrast, *V6* was expressed at high levels at HH4 and HH8 and at low levels that were similar to *V4* at HH6 and HH10. Its expression was highest at HH8 (Fig. 4.3A). *V8* transcripts were present in at higher levels than either of the truncated variants. However, similar to *V6*, its expression was highest at HH8, and slightly lower at HH4, HH6 and HH10. Expression levels of *V6* and *V8* were similar at HH8 (Fig. 4.3A).

We next calculated *V4* and *V6* transcript levels as a function of *V8* expression (*V4*/*V8* and *V6*/*V8*). We find that *V4* is expressed at low levels relative to *V8* at each stage, similar to the results described above, suggesting that V4 may function in a dominant-negative manner to moderate *V8* expression during development (Fig. 4.3B). In contrast, *V6* expression relative to *V8* is much higher at stages HH4, HH6, and HH8. There is a fourfold, twofold, and eightfold

difference between *V6*/*V8* and *V4*/*V8* at HH4, HH6, and HH8, respectively. However, the two measurements are similar at HH10 (Fig. 4.3B). Based on these expression data, we hypothesize that while V4 may be acting as a negative regulator, V6 may function in a dominant-active manner to enhance full-length Bmi-1 activity, especially during gastrulation (HH4) and late neurulation (HH8) stages.

V4 over-expression causes upregulation of *Msx1*, mimicking a loss-of-function phenotype

To examine whether the N-terminal truncated variant V4 may function as a dominant-negative regulator of Bmi-1, we performed *in vivo* over-expression experiments in the chick embryo. The V4 fragment was cloned into an expression vector and injected unilaterally into the prospective neural plate border of the chick gastrula in an electroporation procedure that was described for MO experiments. Electroporated embryos were cultured until HH6-8 and analyzed by *in situ* hybridization for *Msx1* expression. We find that *Msx1* is upregulated weakly in a statistically significant proportion of embryos expressing pCIG-V4-memRFP (Fig. 4.4A-D,G, 8/22, $p<0.05$), but not in control embryos injected with empty pCIG vector (Fig. 4.4E,F). The phenotype is similar to that seen with Bmi-1 MO knock-down. The increase in staining intensity is always observed within the normal expression domain of *Msx1* and is usually most obvious in the open neural plate, which suggests that the phenotype preferentially affects neural plate border progenitors at earlier stages of development (Fig. 4.4B-D). This may be due to the ability of neural crest cells to compensate for the weak phenotype elicited by V4 as electroporated embryos develop.

In summary, our experiments demonstrate that V4 over-expression mimics the Bmi-1 MO effect on *Msx1*, suggesting that it may function to negatively regulate Bmi-1 during neural crest development. The effect is weak despite injection of high concentrations of the over-expression construct, suggesting that either V4 is not translated efficiently under these conditions or, more likely, that it does not alter full-length Bmi-1 protein activity to the same extent as the MO. Given that V4 contains the RING domain but lacks a NLS, it may act as dominant-negative regulator by binding to the full-length protein and preventing it from entering the nucleus. Additionally, V4 may bind to the Ring proteins via the RING domain and prevent them from interacting with other members of PRC1. Based on these preliminary studies, we propose that V4 is expressed during embryonic development at low concentrations, which may serve to regulate or tone down Bmi-1 activity to necessary critical levels. Thus controlled, Bmi-1 may restrict expression of neural crest network genes to levels that may be below the threshold for differentiation or lineage restriction.

DISCUSSION

We have characterized five distinct Bmi-1 variants that were identified in a chick macroarrayed cDNA library (Gammill and Bronner-Fraser, 2002). Identification of intron/exon sequences by PCR from chick BAC templates and alignment of variant sequences has enabled us to make some predictions about splicing events that may have generated these alternative isoforms. We have found that the N-terminal truncated variant V4 contains the RING protein interaction domain generated by the first two exons, but no functional NLS, suggesting a possible role as a dominant-negative regulator of the full-length protein. Another N-terminal variant, V2, contains the RING domain and a potential NLS formed by splicing in and translation of a 3'-UTR exon containing a stretch of amino acids contributing to this domain. Although we saw very weak expression of this variant in stage HH9 embryos by RT-PCR, we were unable to amplify it by QPCR from single embryo cDNA. In addition, in preliminary over-expression experiments with a pCIG-V2-memRFP construct, expression of *Msx1* was not affected (n=4, data not shown). Therefore, although this variant was isolated from a chick neural crest cDNA library and appears to be a true splice isoform based on its sequence, it may not be expressed during the stages that we examined at detectable levels, or may be the result of a very rare splicing event or splicing error amplified in the library. Therefore, we did not pursue this variant further. Finally, we have also characterized a C-terminal truncated variant, V6, and identified the functional translation site, which is found within the HTHTHT domain in the seventh exon. This variant lacks the

RING domain and the first five amino acids of the HTHTHT domain, but contains an alternate putative NLS near the C-terminus.

Variant expression analysis was performed by RT-PCR and QPCR using whole chick embryos collected at several distinct stages of development. We found that *V4* was continuously expressed throughout early chick development from HH4 until HH10 at low levels compared with full-length *Bmi-1*. *V6* expression was quantitatively higher and was characterized by peaks at HH4 and HH8. Full-length *Bmi-1* was expressed at higher levels than the truncated variants with a peak at HH8, similarly to *V6*. The fact that *V4* is expressed at low levels throughout development, while *V6* was detected at similar levels and temporal pattern as the full-length protein, suggests that these variants may be functioning to modify Bmi-1 activity in a dominant-negative and dominant-active manner, respectively. In addition, the peaks of *V6* and full-length variant expression at HH4 and HH8 suggest that Bmi-1 function may be more critical at gastrulation and late neurulation, as opposed to other stages. It is possible that higher Bmi-1 activity in the neural plate border at gastrulation functions to maintain neural crest genes at levels that are sufficient for commitment, but too low for specification and differentiation. At HH8, Bmi-1 activity may be high in the anterior neural folds in order to repress neural crest genes in this region.

In order to obtain some insight into Bmi-1 variant function in the context of the developing neural crest, we electroporated an overexpression construct containing V4 into the prospective neural plate of the chick gastrula. We find that V4 over-expression recapitulates the effect of Bmi-1 MO knock-down on *Msx1*, although the effect is much weaker and primarily observed at younger stages. This suggests that V4 may function to partially inhibit the repressive activity of

Bmi-1 during development, possibly by dimerizing with the full-length protein and preventing it from binding to other PRC1 members and/or from entering the nucleus, and we intend to pursue biochemical studies to elucidate this mechanism. We also plan to examine the function of V6 during neural crest progenitor development using similar *in vivo* over-expression technique. Based on the similarity in expression levels of V6 and full-length Bmi-1, as well as the presence of a putative NLS and the HTHTHT repressive domain in this variant, we predict that over-expression would result in a gain-of-function phenotype similar to that observed with Bmi-1 and Ring1B co-electroporation.

In conclusion, we have isolated and characterized several truncated splice isoforms of Bmi-1, which are expressed in the chick embryo, and have demonstrated that the N-terminal variant V4 negatively regulates Bmi-1 during early neural crest development. However, a large amount of work remains in order to characterize the spatiotemporal expression pattern of Bmi-1 variants during chick neural crest development and to elucidate the functional mechanism by which they modulate Bmi-1 function.

ACKNOWLEDGEMENTS

I am grateful to my co-advisor, Tatjana Sauka-Spengler, for helpful discussion and technical assistance and advice. I would also like to thank Matt Jones and Mary Flowers for their assistance. I am grateful to Miki Yun for help with intron sequencing.

Figure 4.1: Bmi-1 genomic locus and splicing of variants

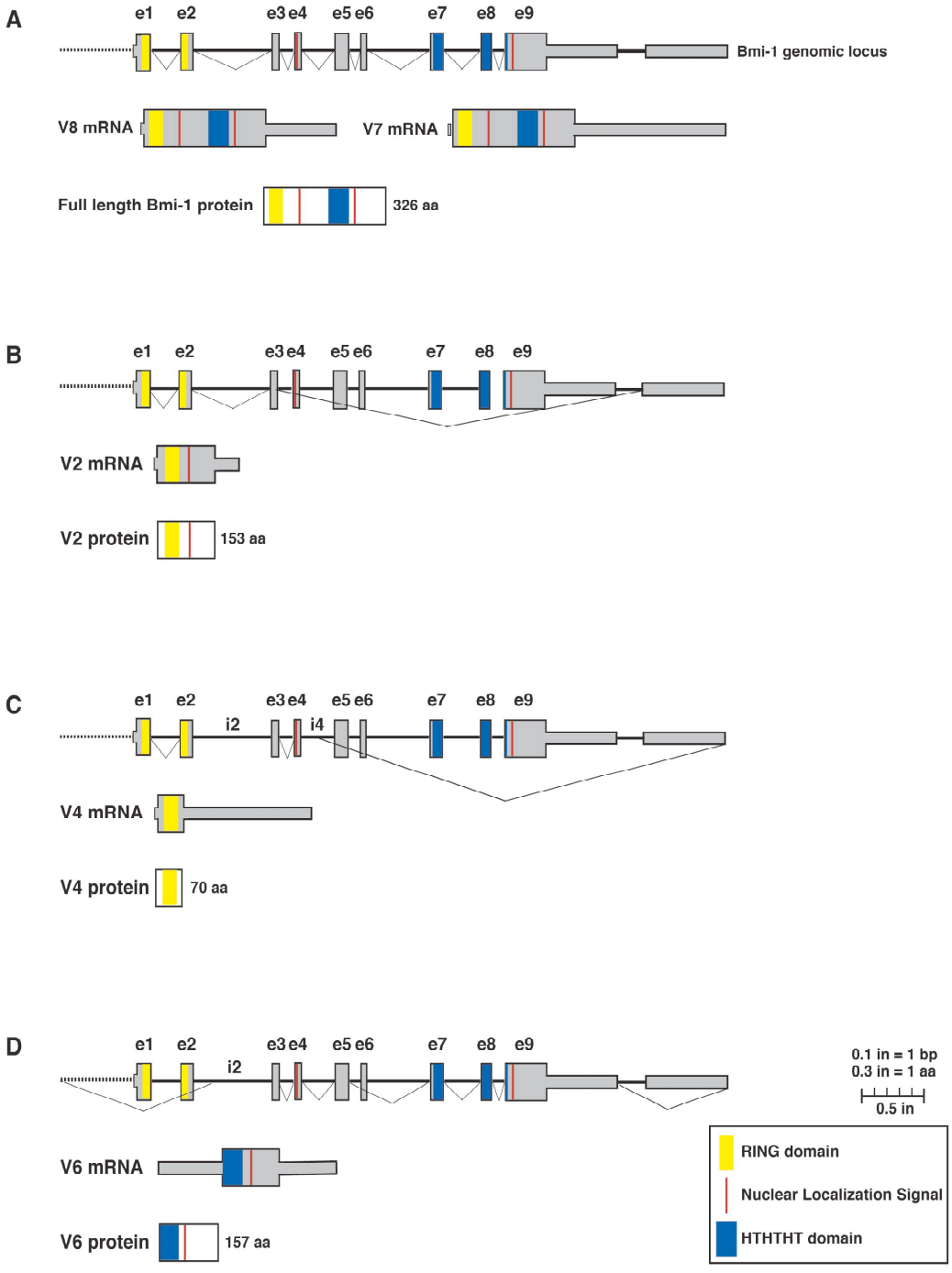

Figure 4.1. Genomic structure of the Bmi-1 locus and predictions of splicing events that yield Bmi-1 isoforms. The chick Bmi-1 gene is characterized by nine coding and at least two non-coding exons. **A.** Full-length Bmi-1 variants V8 and V7 have identical coding sequences and vary only in the 3'-untranslated region (UTR) of the mRNA. V7 contains a supplementary 3'-UTR exon. V7 and V8 encode the full-length 326-amino acid (aa) long Bmi-1 protein containing the RING (yellow box) and HTHTHT (blue box) protein interaction domains and two nuclear localization signals (NLS, red stripe). **B.** A 153-aa long protein encoded by V2 is a result of splicing events assembling the first three exons that contribute the RING domain and the distal 3'-UTR exon, which, when spliced in frame, contributes a putative NLS with an amino acid composition that has been previously experimentally confirmed as functional in other nuclear proteins. **C.** V4 mRNA includes the first four exons as well as the second and fourth introns, which have been retained. Only the first two exons are translated, giving rise to a 70-aa long protein that only contains the RING domain. **D.** V6 is comprised of the second intron and the third, fourth, fifth, seventh, and ninth exons. The functional translation initiation site of V6 lies within the HTHTHT domain in the seventh exon, giving rise to a 175-aa long protein that lacks the RING domain and first five amino acids of the HTHTHT domain. Schematic is drawn to scale.

Figure 4.2: Variant expression during chick development

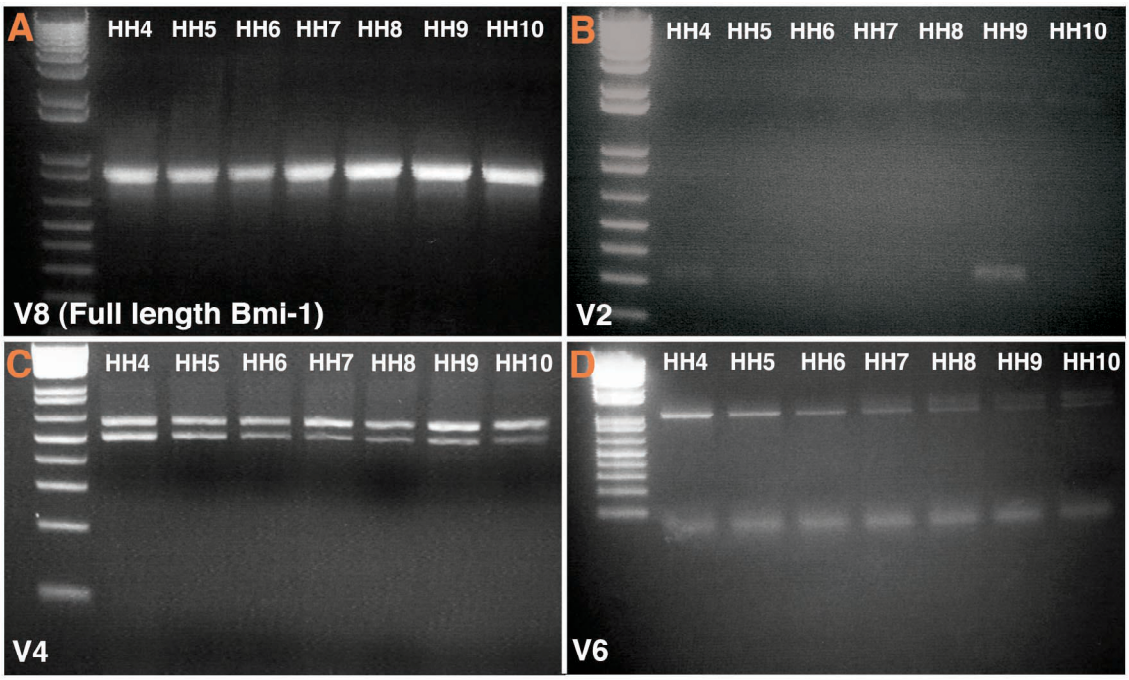

Figure 4.2. Bmi-1 variants are differentially expressed in the chick embryo during early neural crest development. RT-PCR reactions were carried out using material from whole chick embryos collected at HH4-10 and specific variant primer sets described in Materials and Methods. **A.** Full-length Bmi-1 variant *V8* is expressed at all developmental stages tested. **B.** *V2* is amplified at very low levels at HH9 only. **C.** *V4* is expressed throughout HH4-10. **D.** *V6* is expressed at early developmental stages and is progressively downregulated at later stages.

Figure 4.3: Quantification of Bmi-1 variant expression

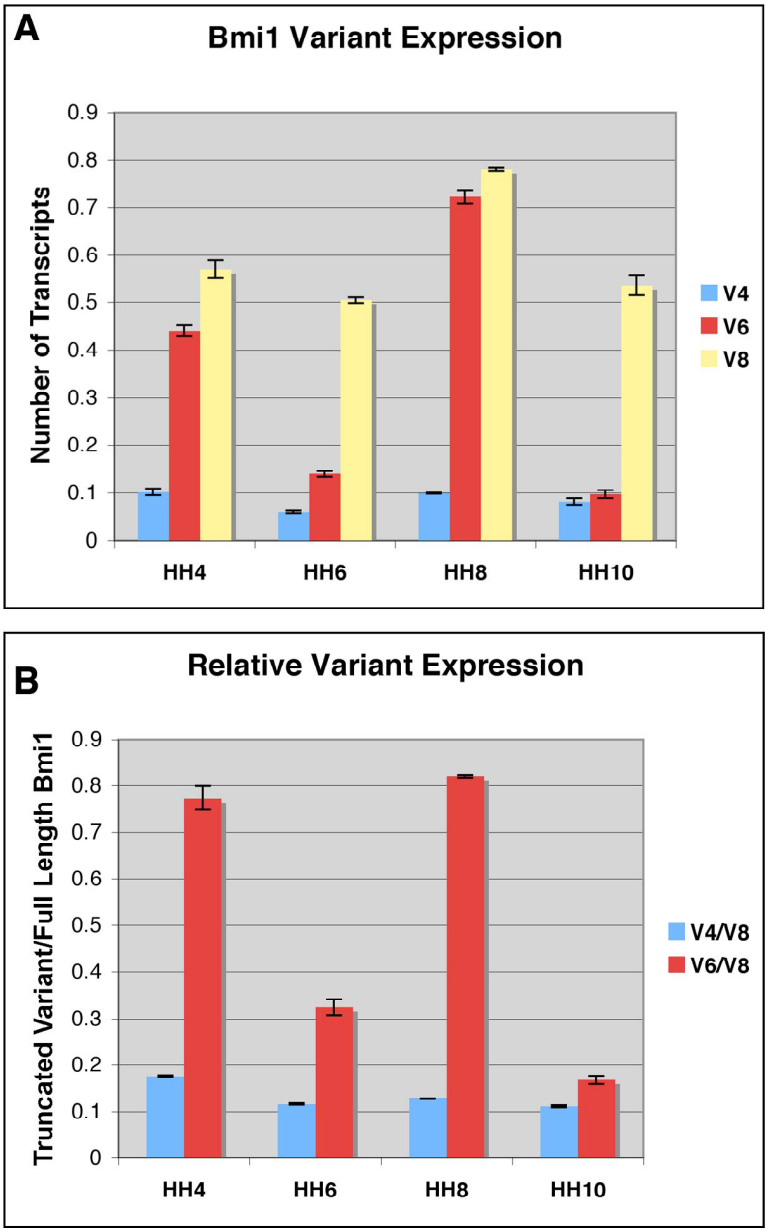

Figure 4.3. Quantitative analysis of Bmi-1 variant expression during early chick development. RT-QPCR was carried out using material from single whole chick embryos staged HH4, HH6, HH8, and HH10 using specific Bmi-1 variant primer pairs. Data shown is from one representative embryo out of three analyzed at

each stage. QPCR reactions were performed in triplicate and relative abundance of transcript was calculated by standard curve method and normalized to *Gapdh*. **A.** The full-length variant *V8* is expressed at high levels at all stages examined with a peak at HH8. In contrast, *V4* levels remain relatively low throughout development. *V6* expression is biphasic: it is expressed at levels comparable to *V8* at HH4 and HH8 but at low levels similarly to *V4* at HH6 and HH10. **B.** Expression levels of truncated variants were compared to that of full-length Bmi-1 variant *V8*. *V4* transcripts are present at low levels compared to *V8* at all stages tested. In contrast, *V6* expression is high compared to *V8* with large peaks at HH4 and HH10.

Figure 4.4: V4 over-expression elicits a dominant-negative phenotype

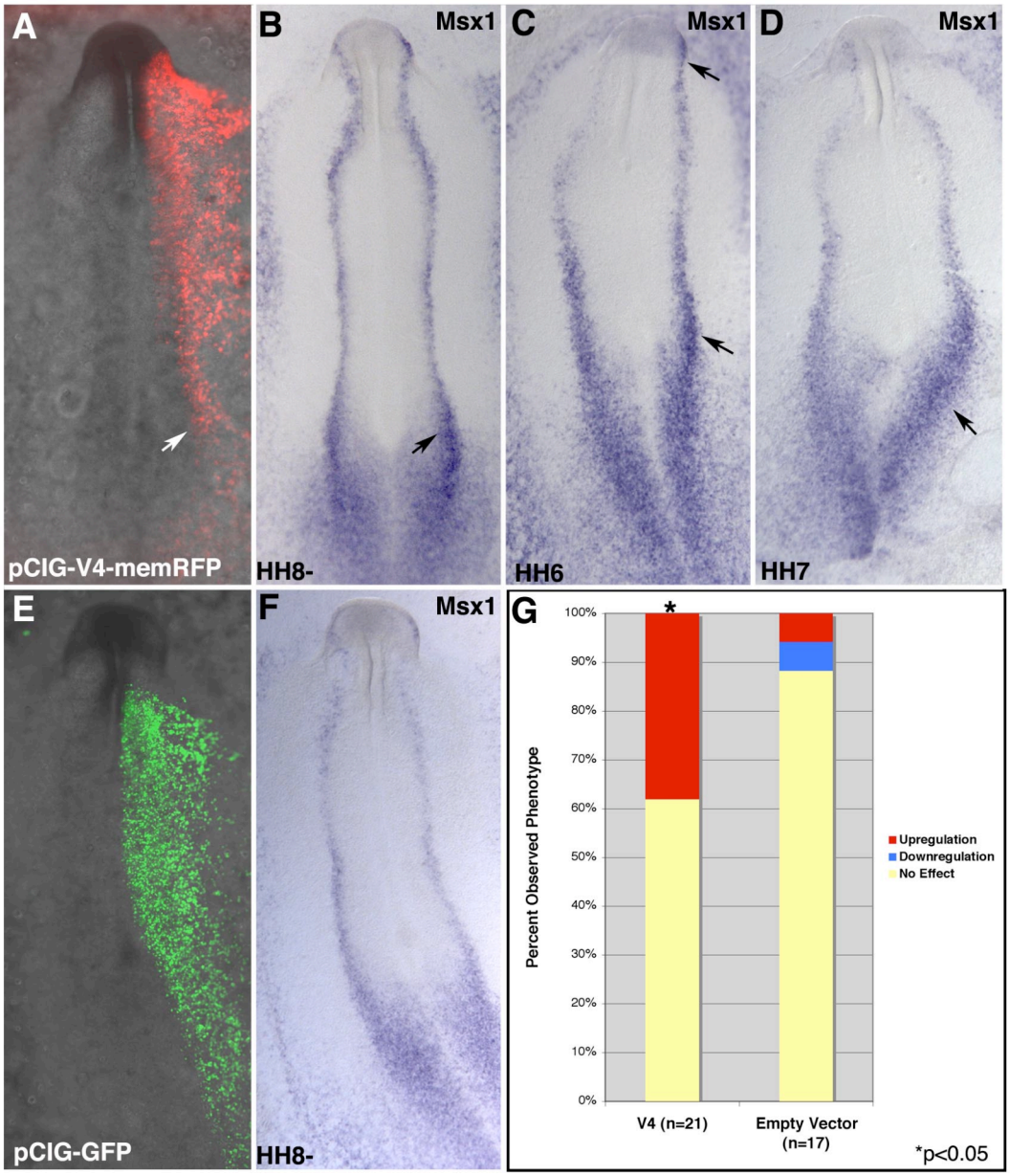

Figure 4.4. Over-expression of pCIG-V4-memRFP results in weak upregulation of *Msx1* during neurulation. **A** and **B.** An embryo that was electroporated with pCIG-V4-memRFP overexpression construct (**A**) exhibits weak upregulation of *Msx1* in the open neural plate at HH8- (**B**, arrow). **C** and **D.** pCIG-V4-memRFP-

electroporated embryos demonstrate more obvious phenotypes when analyzed at HH6 (**C**) or HH7 (**D**). **E** and **F**. In a control embryo electroporated with empty pCIG-GFP vector (**E**), *Msx1* expression is not affected (**F**). **G.** Quantification of phenotypes demonstrates that the effect observed in V4-electroporated embryos is statistically significant and not due to chance (8/21, $p<0.05$).