Regulation of Vg1 biogenesis during mesendoderm induction

P. C. Dave P. Dingal[1,#], Adam N. Carte[1,2,3,*], Tessa G. Montague[1,*], and Alexander F. Schier[1,3,4,#]

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA
[2]Systems, Synthetic, and Quantitative Biology Program, Harvard, University, Cambridge, MA, USA
[3]Biozentrum, University of Basel, Basel, Switzerland
[4]Allen Discovery Center for Cell Lineage Tracing, University of Washington, Seattle, WA, USA

[#]corresponding authors: polimyr_dingal@fas.harvard.edu; alex.schier@unibas.ch
[*]equal contribution

Abstract

The TGF-beta signals Vg1 and Nodal form heterodimers to induce the vertebrate mesendoderm. The Vg1 proprotein is a monomer retained in the endoplasmic reticulum (ER) and is processed and secreted upon heterodimerization with Nodal. Here we investigate the mechanisms underlying Vg1 retention, processing, secretion and signaling in zebrafish. First, using a newly devised Synthetic Processing (SynPro) system, we find that Vg1 can be processed by intra- or extracellular proteases. Second, Vg1 can be processed without Nodal but requires Nodal for secretion and signaling. Third, Vg1-Nodal signaling activity requires Vg1 processing, whereas Nodal can remain unprocessed. Fourth, Vg1 employs exposed cysteines, glycosylated asparagines, and BiP chaperone-binding motifs for monomer retention in the ER. Our results establish SynPro as a new in vivo processing system and define molecular mechanisms and motifs that facilitate the generation of active Vg1-Nodal heterodimers. These observations suggest two strategies for rapid mesendoderm induction: chaperone-binding motifs help store Vg1 as an inactive but ready-to-heterodimerize monomer in the ER, and the flexibility of Vg1 processing location allows efficient generation of active heterodimers both intra- and extracellularly.

Introduction

The TGF-beta signals Nodal and Vg1 (Dvr1/Gdf3) play crucial roles in vertebrate development (Schier 2009; Rogers and Muller 2019), including the induction of mesendoderm and the generation of left-right asymmetry (Dale, Matthews, and Colman 1993; Thomsen and Melton 1993; Zhou et al. 1993; Conlon et al. 1994; Feldman et al. 1998; Birsoy et al. 2006; Bisgrove, Su, and Yost 2017; Montague and Schier 2017; Pelliccia, Jindal, and Burdine 2017; Levin et al. 1995; Collignon, Varlet, and Robertson 1996; Lowe et al. 1996; Pagan-Westphal and Tabin 1998; Long, Ahmad, and Rebagliati 2003; Montague, Gagnon, and Schier 2018). For example in zebrafish, secreted Vg1-Nodal heterodimers induce a gradient of signaling that patterns the embryonic mesendoderm (Montague and Schier 2017). Vg1-Nodal heterodimers exert their effects as ligands for a receptor complex that comprises Activin serine-threonine kinase receptors and an essential co-receptor called Oep (Tdgf1/CRIPTO) (Gritsman et al. 1999; Yeo and Whitman 2001; Cheng et al. 2003). Activated ligand-receptor complexes catalyze phosphorylation of Smad2 (pSmad2), which accumulates in the nucleus to induce the expression of mesendodermal genes (Baker and Harland 1996).

We previously proposed a 4-step model for how Vg1-Nodal heterodimers pattern the mesendoderm of zebrafish embryos (Montague and Schier 2017) (Figure 1A): (1) The maternally ubiquitous Vg1 proprotein is retained as a monomer in the endoplasmic reticulum (ER) of embryonic cells. (2) The expression of the zebrafish Nodal genes, *cyclops* (*cyc*) and squint (*sqt*), initiates at the yolk margin at ~3 hours post-

fertilization (hpf). (3) Nodal forms heterodimers with pre-existing Vg1. (4) Vg1-Nodal heterodimers are processed and secreted to activate signaling. This model explains how Nodal and Vg1 interact during early embryogenesis, but the molecular mechanisms that regulate Vg1 retention, processing, secretion and signaling have remained unclear.

Previous studies of growth factor processing and ER retention provide potential mechanisms for how Vg1 localization and activity might be regulated. TGF-beta ligands are synthesized as pre-proproteins, which comprise an amino-terminal signal sequence, a long prodomain, and a shorter bioactive mature domain. A conserved cysteine in the mature domain is primarily responsible for dimer formation via an intermolecular disulfide bond (Hinck, Mueller, and Springer 2016). Prodomain processing likely occurs in the Golgi apparatus and on the cell surface, where proprotein convertases are found (Miyazono et al. 1991; Beck et al. 2002; Blanchet, Le Good, Mesnard, et al. 2008; Seidah and Prat 2012). Vg1 has an intact prodomain when it is retained in the ER (Montague and Schier 2017), and it was suggested that Vg1 homodimers cannot be secreted because the Vg1 prodomain cannot be processed (Dale, Matthews, and Colman 1993; Thomsen and Melton 1993; Dohrmann, Kessler, and Melton 1996). Alternatively, a suite of protein folding chaperones could potentially interact with specific motifs on the Vg1 proprotein to block its release from the ER and subsequent processing. For example, protein disulfide isomerases (PDIs) that bind to exposed cysteines facilitate the folding of nascent proproteins and promote the retention of unassembled protein complexes in the ER (Reddy and Corley 1998; Wilkinson and Gilbert 2004). Additionally, the lectin chaperones, calnexin and calreticulin, bind to asparagine-linked glycosyl groups on nascent secreted proteins (Helenius and Aebi 2004). Glycosylated proteins are released from the ER after trimming of the N-linked glucose moieties (Hammond, Braakman, and Helenius 1994; Helenius and Aebi 2004). Moreover, the BiP chaperone (also known as Hspa5/Grp78) aids in protein folding by binding to predominantly hydrophobic heptapeptide sequences and can also retain proteins in the ER (Flynn, Chappell, and Rothman 1989; Flynn et al. 1991; Blond-Elguindi et al. 1993). Vg1 has exposed cysteines and asparagines that can be glycosylated (Tannahill and Melton 1989), but it is unclear whether Vg1 or other TGF-beta signals use these chaperone-binding motifs to ensure retention in the ER.

In this study, we investigate the molecular mechanisms that regulate Vg1 retention, processing, secretion and signaling. To control prodomain processing, we created a Synthetic Processing (SynPro) system. Using SynPro, we find that Vg1 can be cleaved and activated either cell-autonomously or non-cell-autonomously. Vg1 can be processed without Nodal but requires Nodal for secretion and signaling. We further show that Vg1, but not Nodal, must be processed for signaling activity. Finally, we identify several chaperone-binding motifs in the prodomain and mature domain of Vg1 that function in ER retention. These molecular mechanisms and sequence motifs regulate Vg1 biogenesis and contribute to the temporal and spatial specificity of mesendoderm formation.

Results

Creation of a synthetic processing system (SynPro)

The processing of Vg1 depends on its dimerization with Nodal (Montague and Schier 2017), but where processing must occur relative to dimerization and secretion has remained unclear. To determine how processing affects the formation of active Vg1-Nodal heterodimers, we first set out to control processing orthogonally by creating SynPro. SynPro employs the ability of a synthetic protease to cleave a specific peptide cleavage site (Figure 1B, Figure 1—figure supplement 1). We screened several proteases that can cleave short peptide sequences and that can be produced in zebrafish embryos without deleterious effects. These included proteases from the tobacco etch virus (TEVp), tobacco vein mottling virus (TVMVp), human rhinovirus 3C, and enterokinase (Waugh 2011). We found that most commercially available

proteases are toxic when produced from mRNAs injected into zebrafish embryos (Figure 1—figure supplement 1A). Codon optimization of the *TEVp* sequence eliminated toxicity. Thus, we synthesized mRNAs of 15 proteases of the *Potyviridae* family that were codon-optimized for zebrafish (Wylie et al. 2017); all of which were non-toxic when expressed in zebrafish embryos (Figure 1—figure supplement 1B).

To test whether the synthetic proteases are functional, we designed a fluorescent reporter of proteolytic cleavage (Figure 1—figure supplement 1C; see Materials and Methods for details). Briefly, a protease-cleavable substrate is initially localized in the cytoplasm. Protease-catalyzed cleavage results in the release into the nucleus and reconstitution of a split fluorescent protein, mNeonGreen2 (Feng et al. 2017). Co-expression of the cleavable substrate and the nuclear reporter in zebrafish embryos did not lead to reconstitution of nuclear mNeonGreen2 fluorescence (Figure 1—figure supplement 1D,i). In contrast, nuclear mNeonGreen2 fluorescence was observed when TEVp was co-expressed with its cognate substrate and the nuclear reporter (Figure 1—figure supplement 1D,ii). We observed protease-induced fluorescence reconstitution for 6 out of 15 proteases tested (Figure 1—figure supplement 1E), indicating that the synthetic proteases can cleave their cognate sequences in zebrafish embryos.

Synthetically processed Vg1 rescues *vg1* mutants

To generate a secreted protease for the SynPro system, we added an amino-terminal signal sequence to TEVp to produce a secreted variant, secTEVp (see Materials and Methods for details). We also introduced five amino acid substitutions in secTEVp that are known to promote solubility (Cabrita et al. 2007) and prevent oxidation in the secretory compartments (Cesaratto et al. 2015). To determine whether secTEVp is functional and can replace the endogenous convertases that cleave the Vg1 prodomain, we generated SynPro Vg1. In this Vg1 derivative, the native cleavage sequence, 'RSRRKR', was replaced with the cognate cleavage sequence of secTEVp, 'ENLYFQS' (Figure 1B). While the expression of *SynPro vg1* mRNA alone failed to rescue mesendoderm formation in maternal *vg1* mutant (M*vg1*) embryos, the co-expression of *SynPro vg1* and *secTEVp* rescued M*vg1* mutants (Figure 1C). Notably, co-production of SynPro Vg1 and an ER-resident secTEVp-KDEL also rescued M*vg1* embryos, revealing that intracellular processing of Vg1 can result in normal Vg1 activity. These results show that endogenous enzymes that process Vg1 can be replaced by an orthogonal protease and that the SynPro system can control subcellular processing and protein activity.

Vg1 processing is not sufficient for secretion

The co-expression of *SynPro vg1* and *secTEVp* rescued M*vg1* mutants and did not induce abnormal overexpression phenotypes, indicating that the signaling activity was restricted to domains of co-expression with endogenous Nodal. This result raised two hypotheses: that Vg1 cleavage by SynPro depends on Nodal or alternatively, that Vg1 is independently cleaved by SynPro, but its secretion is still dependent on Nodal. To test these models, we co-expressed *SynPro vg1* and ER-resident *secTEVp* and assessed the location and processing of SynPro Vg1 with or without the zebrafish Nodal *cyc* (Figure 1—figure supplement 2). SynPro Vg1 was processed but not secreted in the absence of Cyc. These results indicate that synthetic processing of Vg1 does not require Nodal but that Nodal is necessary for Vg1 secretion.

Vg1 processing can be non-cell autonomous

The results above show that processing in the secretory compartments is sufficient to produce functional Vg1. By leveraging the versatility of the SynPro system, we next determined whether extracellular

processing of the Vg1 proprotein might be sufficient to generate active Vg1. We performed a series of transplant experiments in M*vg1* embryos that were injected with mRNAs that express components of the SynPro system (Figure 1D-E). First, donor M*vg1* cells that co-express *SynPro vg1*, *secTEVp* and *cyc* mRNAs were transplanted in host M*vg1* embryos (Figure 1D). As expected, we observed Nodal signaling activity based on immunostaining of nuclear pSmad2 in both donor and neighboring host cells (Figure 1E,i). This result suggests that donor cells secreted secTEVp-processed SynPro Vg1 and Cyc. Second, donor cells co-expressing *SynPro vg1* and *cyc* were transplanted into host M*vg1* embryos expressing *secTEVp*. As in the first case, we observed nuclear pSmad2 accumulation in both donor and the surrounding host cells, even though the protease and its substrate were expressed in different cells (Figure 1E,ii). In this scenario, host-secreted secTEVp cleaved and activated the donor-secreted SynPro Vg1 and Cyc. In control experiments, nuclear pSmad2 was never observed when secTEVp was absent, nor when donor embryos did not express SynPro Vg1 and Cyc (Figure 1E,iii; Figure 1E,iv). The application of the SynPro system indicates that Vg1 can be secreted with Nodal as an unprocessed proprotein and can be processed and activated non-cell autonomously.

Processing is not required for secretion of Vg1 and Nodal

The observation that extracellular secTEVp was sufficient to generate active signaling around cells producing SynPro Vg1 and Nodal suggested that the Vg1 proprotein can be secreted without processing. To further test this idea, we generated non-cleavable mutants of Vg1 and Nodal. We inactivated the cleavage site of Vg1 and of the zebrafish Nodal Squint (Sqt) and inserted a superfolder green fluorescent protein (sfGFP) to generate the non-cleavable variants, *vg1-NC-sfGFP* and *sqt-NC-sfGFP*. We could not generate a non-cleavable variant of Cyc due to additional cryptic cleavage sites (Figure 2—figure supplement 1). When sfGFP-tagged Vg1 (*vg1-sfGFP*) or non-cleavable Vg1 (*vg1-NC-sfGFP*) was co-produced with Sqt or non-cleavable Sqt (*sqt-NC*) in M*vg1* embryos, Vg1-sfGFP and Vg1-NC-sfGFP were secreted (Figure 2A), indicating that processing of Vg1 and Sqt is not required for secretion.

To test whether secreted, non-cleavable Vg1 and Nodal are physiologically active, we injected mRNAs of the non-cleavable constructs into wild-type and M*vg1* embryos and assessed the induction of Nodal target genes. Overexpression of *sqt* with or without *vg1* was able to induce target gene expression (Figure 2B,i-ii) (Montague and Schier 2017). By contrast, non-cleavable Sqt was only able to activate target gene expression in the presence of wild-type (cleavable) Vg1 (Figure 2B,iii, iv). To determine if cleavage of the Vg1 prodomain is required for activity, we co-expressed *vg1-NC* and *sqt-NC* in M*vg1* embryos and found that they were unable to induce Nodal target gene expression (Figure 2C). Similary, *cyc* was only able to activate Nodal target gene expression when co-expressed with *vg1* but not with *vg1-NC* (Figure 2B,v-vi). These results indicate that the Vg1 prodomain – but not the Nodal prodomain – must be cleaved for active heterodimer signaling.

Cysteine thiol and N-glycosyl groups retain the Vg1 prodomain in the ER

Previous studies have suggested that the Vg1 prodomain prevents Vg1 activity, because fusion of other TGF-beta prodomains to the Vg1 mature domain resulted in its processing and mesoderm-inducing activity (Dale et al. 1989; Dale, Matthews, and Colman 1993; Thomsen and Melton 1993; Dohrmann, Kessler, and Melton 1996; Chen et al. 2006). To determine how the Vg1 prodomain might function, we replaced the prodomain of zebrafish Nodals with the prodomain of Vg1. The secretion and activity of these chimeric Nodals were inhibited (Figure 3—figure supplement 1A). Conversely, a chimeric Vg1 protein that fused the Nodal prodomain to the Vg1 mature domain resulted in a secreted and active Vg1 (Figure 3—figure supplement 1B). These results suggest that the Vg1 prodomain has features that promote ER retention.

To determine how the Vg1 prodomain might mediate ER retention, we searched for putative sequence motifs that might mediate retention. We did not find any KDEL sequences in Vg1; these motifs are found at the C-termini of ER-resident proteins and are recognized by KDEL receptors that trigger Golgi-to-ER retrograde transport (Munro and Pelham 1987). However, we found putative sequence motifs in Vg1 that bind to ER-resident chaperones, including exposed cysteines that bind to PDIs (Reddy and Corley 1998; Wilkinson and Gilbert 2004) and glycosylated asparagines in NX[S/T] motifs that bind to calnexin and calreticulin (Helenius and Aebi 2004). In particular, the prodomain of Vg1 (proVg1) has one exposed cysteine (C100) and two potential N-glycosylation sites (N108, N179). Notably, the mature domain of Vg1 (matVg1) also has a free cysteine (C319) and a potential N-glycosylation site (N296), raising the possibility that the mature domain also plays a role in ER retention (Figure 3A).

To elucidate whether these residues promote Vg1 retention in the ER, we systematically mutated the cysteine and N-glycosylation residues in sfGFP-tagged Vg1, proVg1 and matVg1 constructs and visualized their localization. We injected M*vg1* embryos with *sfGFP*-tagged *vg1*, *proVg1* or *matVg1* mRNAs and performed fluorescence microscopy to determine the subcellular localization of the resulting proteins. Similar to full-length Vg1, we observed ER localization of sfGFP-tagged proVg1 and matVg1 (Figure 3B-C). Mutants for the prodomain cysteine (C100S) or glycosylation sites (N108Q, N179Q) were retained in the ER, but the triple mutant proVg1(C100S, N108Q, N179Q) was secreted to the extracellular space (Figure 3B). By contrast, the matVg1 cysteine and glycosylation mutants as well as the quintuple mutant Vg1(C100S, N108Q, N179Q, C319S, N296R) were still retained in the ER (Figure 3C, D). We also found that the loss of N- glycosylation was more deleterious to signaling activity than the loss of cysteines (Figure 3—figure supplement 2). Taken together, our mutagenesis results show that N-linked glycosyl groups and an exposed cysteine thiol retain the Vg1 prodomain in the ER.

BiP-binding motifs retain the Vg1 mature domain in the ER

The observation that the free cysteine and N-glycosylation mutants of matVg1 are still retained in the ER suggests that matVg1 possesses additional ER-retention motifs. We hypothesized that a third chaperone, BiP (Hspa5/Grp78), might promote ER retention. To elucidate which sequence features of matVg1 may bind to BiP, we utilized the Gething-Sambrook scoring system (Blond-Elguindi et al. 1993). Since mature Nodals (matSqt and matCyc) are secreted (Figure 4—figure supplement 1), we compared the sequences of matVg1 and matCyc and scored for all possible BiP-binding heptapeptides (Figure 4A-B, Table 1; see Materials and Methods for details). To test whether the high-scoring features of matVg1 promote ER retention, we systematically mutated three high-scoring matVg1 sequences to their corresponding low-scoring matCyc sequences (labeled as *m1*, *m2*, and *m3* in Figure 4A, C). While all single mutants and double mutants were found in the ER, the triple mutant matVg1(*m1*, *m2*, *m3*) exhibited extracellular localization (Figure 4D). Conversely, a mutant matCyc possessing the high-scoring features of matVg1 was retained in the ER (Figure 4—figure supplement 2, Table 1). These results indicate that the BiP-binding sequences retain the Vg1 mature domain in the ER.

Discussion

This study reveals features of Vg1 that regulate its retention, processing, secretion, and signaling during early zebrafish embryogenesis (Figure 5): (1) Vg1 ER retention is mediated by exposed cysteines, glycosylated asparagines, and BiP chaperone-binding motifs. (2) Vg1 can be processed without Nodal but requires Nodal for secretion and signaling. (3) Vg1 can be processed cell-autonomously or non-cell autonomously. (4) Vg1-Nodal signaling activity requires processing of Vg1 but not of Nodal. These conclusions unify and extend several previous observations about Vg1-Nodal signaling:

First, our study defines Vg1 sequence motifs that have the hallmarks of binding sites for ER-resident chaperones. Our mutational analyses indicate that these motifs help retain monomeric Vg1 in the ER until Nodal is expressed and heterodimerizes with Vg1. We find that the Vg1 prodomain is specifically retained in the ER via its N-glycosylation sites and exposed cysteine. Thus, Vg1 is a member of a class of secreted proteins that are transiently retained in the ER until their cysteines form intermolecular disulfide bonds (Haas and Wabl 1983; Bole, Hendershot, and Kearney 1986; Tu, Sun, and Kreibich 2002; Hu et al. 2008). A classic example of cysteine-mediated retention in the ER are the light and heavy chains of immunoglobulin M antibodies (Sitia et al. 1990; Alberini et al. 1990). In this case, heavy chains are retained in the ER until they are covalently bonded to light chains. Mutating the exposed cysteine in the heavy chain leads to the secretion of monomers. Our results suggest a similar role for the exposed cysteine in Vg1, but in addition, Vg1 employs N-glycosylation sites and BiP-binding regions as additional chaperone motifs for ER retention. Mutating the BiP-binding regions (*m1*, *m2*, *m3*) of the Vg1 mature domain leads to secretion. Interestingly, the *m3* region is juxtaposed to cysteine C319, which mediates heterodimerization with Nodal. It is conceivable that BiP binding blocks the formation of mature Vg1 homodimers.

We found no evidence for KDEL-mediated ER retention of Vg1. This mode of ER retention differs from chaperone-mediated retention in that it mediates retrograde transport from the Golgi apparatus to the ER lumen (Munro and Pelham 1987). In contrast, chaperone-mediated retention prevents the escape of proteins from the ER. We speculate that the Vg1 system has evolved to trap Vg1 in the ER until Nodal is imported and displaces the chaperones as a heteromeric partner.

The ER-resident chaperones might not only retain Vg1 but also support its folding and maturation. For example, a large multiprotein complex of chaperones is associated with unassembled immunoglobulin heavy chains (Meunier et al. 2002) and immature thyroglobulin (Kuznetsov, Chen, and Nigam 1997). During antibody assembly, the BiP chaperone can stably bind to unstructured heavy chains (Vanhove, Usherwood, and Hendershot 2001; Feige et al. 2009) and cooperates with PDI and calnexin to properly assemble the light and heavy chains (Mayer et al. 2000; Shusta et al. 1998). A similar mechanism might control Vg1-Nodal heterodimer assembly. In this scenario, the Vg1 monomer remains in an unstructured, immature state through interaction with various chaperones in the ER and might fold only upon interaction with Nodal. How the chaperone network hands over the Vg1 monomer to Nodal remains to be elucidated.

Second, our study clarifies the roles of Vg1 processing in secretion and signaling. Early studies suggested that Vg1 cannot be secreted because its prodomain cannot be processed (Dale, Matthews, and Colman 1993; Thomsen and Melton 1993; Dohrmann, Kessler, and Melton 1996). However, our application of the SynPro system shows that prodomain processing in the ER is not sufficient to induce the secretion of Vg1 (Figure 1—figure supplement 2). In fact, unprocessed Vg1 can be secreted upon dimerizing with Nodal (Figure 2). These results are consistent with the observations that mouse Nodal and its proprotein convertases are not co-expressed in the same cell (Blanchet, Le Good, Mesnard, et al. 2008; Blanchet, Le Good, Oorschot, et al. 2008; Beck et al. 2002; Tessadori et al. 2015). In this case, prodomain processing of Gdf1/3 (the orthologs of Vg1) and Nodal dimers occurs after they are released in the extracellular space or bound to the target cell surface and early endosomes. Using the SynPro system in zebrafish embryos, we show that Vg1 can also be cleaved after its Nodal-induced release in the extracellular space (Figure 1). These findings establish that prodomain processing of Vg1 can be location-independent and separable from Nodal-induced secretion.

In addition, we find that Vg1-Nodal heterodimers with non-cleavable prodomains are secreted but cannot signal, whereas unprocessed Nodal in combination with processed Vg1 remains partially active. This result demonstrates that Nodal processing is not required for secretion or partial activity, consistent with previous studies that found that non-processed mouse Nodal still promotes mesoderm formation but cannot position or maintain the primitive streak (Ben-Haim et al. 2006). Our finding that the Vg1 prodomain has to be

processed for signaling suggests the possibility that the activity of non-processed mouse Nodal also relies on processed Vg1 (Gdf1/3).

It has been suggested that the Vg1-Nodal heterodimeric system allows for the rapid activation of the Nodal signaling pathway (Montague and Schier 2017). Instead of a time delay caused by the accumulation of sufficient levels of Nodal needed to form homodimers, the preformed pool of Vg1 allows for rapid heterodimer formation as soon as Nodal is synthesized. Our study suggests that two features of Vg1 facilitate this strategy of rapid signaling: chaperone-mediated monomer retention in the ER stores Vg1 in an inactive but ready-to-heterodimerize form, and the flexibility of processing location allows efficient generation of active heterodimers both intra- and extracellularly.

Last, the SynPro system introduced here has broad applications in the targeted processing and regulation of proteins. Our results show that endogenous Vg1 can be fully replaced by a Vg1 derivative containing an engineered cleavage sequence in combination with the corresponding SynPro protease. Thus, the SynPro system can be used for the orthogonal regulation of TGF-beta signals. In addition, endogenous proprotein convertases cleave the same polybasic motif in multiple secreted proteins, whereas each SynPro protease recognizes a unique cleavage sequence. Thus, different SynPro proteases could be combined to separately control the activity of multiple secreted signals. For example, different TGF-beta signals could be processed independently to reveal distinct spatial and temporal requirements. Beyond TGF-beta proteins, the SynPro system could be used to independently process several bioactive peptides from hormone or neuropeptide precursors. By releasing individual peptides one at a time from a polyprotein, the effects of each peptide could be analyzed. Thus, the SynPro system has the potential to accelerate the functional assignment of bioactive peptides generated by secreted polyproteins.

Acknowledgements
The authors would like to thank Richard Losick, Nathan Lord, and Madalena Pinto for helpful comments on the manuscript, and Stephen Hansen for programming and data visualization. We thank the members of the Schier lab for the advice, expertise, insights and discussions. This project was supported by NIH DP1-HD094764 and the Allen Discovery Center for Cell Lineage Tracing.

Ethics
Animal experimentation: All vertebrate animal works was performed at the facilities of Harvard University, Faculty of Arts & Science (HU/FAS). The HU/FAS animal care and use program maintains full AAALAC accreditation, is assured with OLAW (A3593-01) and is currently registered with the USDA. This study was approved by the HU/FAS Standing Committee on the Use of Animals in Research & Teaching under Protocol No. 25-08.

FIGURE LEGENDS

Figure 1. The SynPro system shows that prodomain cleavage can be non-cell autonomous. (**A**) Model for Vg1-Nodal heterodimer formation in zebrafish: Vg1 is maternally deposited and is abundantly found in the ER in the early embryo; at the maternal-zygotic transition, Nodal is produced and form obligate heterodimers with Vg1; Vg1-Nodal heterodimers proceed to the Golgi for prodomain cleavage and are eventually secreted to activate the Nodal signaling pathway. (**B**) The Synthetic Processing system comprises an orthogonal secreted protease derived from tobacco etch virus (secTEVp) and a cognate sequence that replaces the endogenous cleavage site of Vg1 (SynPro Vg1, RSRRKR → ENLYFQS). (**C**) Rescue percentage after 30 hpf of M*vg1* embryos injected with 50 pg of *vg1*, *SynPro vg1*, *SynPro vg1* and *secTEVp*, or *SynPro vg1* and *secTEVp-KDEL* mRNAs. (**D**) Schematic of transplantation assay. M*vg1* embryos were injected with 50 pg mRNA each of: (**i**) DONOR: *cyc*, *SynPro vg1*, and *secTEVp*; HOST: none; (**ii**) DONOR: *cyc* and *SynPro vg1*; HOST: *secTEVp*; (**iii**) DONOR: *cyc* and *SynPro vg1*; HOST: none; (**iv**) DONOR: none; HOST: *secTEVp*. All M*vg1* donor embryos were marked by also injecting 50 pg *sfGFP* mRNA. At high stage, before the onset Nodal signaling, sfGFP-marked DONOR cells were transplanted to the animal pole of HOST M*vg1* embryos. (**E**) At 50-60% epiboly, chimeric embryos were fixed and immunostained for sfGFP and pSmad2. DAPI, nuclei. Scale bar, 20 µm.

Figure 1—figure supplement 1. Developing the SynPro system. (**A**) Viability at 24 hpf of wild-type embryos injected with mRNA of commercially available proteases: tobacco etch virus (TEV) protease (Addgene plasmid #8835), tobacco vein mottling virus (TVMV) protease (Addgene plasmid #8832), enterokinase (Addgene plasmid #49048), and human rhinovirus 3C (HRV 3C) protease (Addgene plasmid #78571). Developmental delays and loss of viability were observed in embryos injected with TEV protease and HRV 3C protease. (**B**) Viability at 24 hpf of wild-type embryos injected with 50 pg of zebrafish codon-optimized mRNA sequences of 15 potyviral proteases. In all conditions, more than 80% of embryos were viable at 24 hpf. (**C**) Schematic of fluorescence reporter assay for protease cleavage, which is composed of three components: Component 1 contains the cleavage site flanked by beta-Arrestin and mNeonGreen2(1-10); Component 2 is the codon-optimized Potyviral protease (e.g. TEV protease recognizes the sequence ENLYFQS); and Component 3 is nuclear histone 2b tagged with mCherry and fused with mNeonGreen2(11). See Methods for construct details. (**D**) Testing the cleavage efficiency of TEV protease. (**i**) Co-expression of Components 1 and 3 does not exhibit mNeonGreen2 fluorescence, whereas (**ii**) co-expression of all 3 components leads to proteolytic cleavage, nuclear translocation, and binding of mNeonGreen2(1-10) to mNeonGreen2(11) – leading to reconstitution of mNeonGreen2 fluorescence in the nucleus. (**E**) Using the reporter assay in zebrafish embryos, six out of fifteen Potyviral proteases exhibited proteolytic activity, based on reconstitution of nuclear mNeonGreen2 fluorescence (bottom row images). Scale bars, 100 µm.

Figure 1—figure supplement 2. Vg1 prodomain processing does not promote secretion. (**A**) Fluorescence images of fixed M*vg1* embryos injected with 25 pg mRNA of sfGFP-tagged *SynPro vg1*, *secTEVp-mCherry-KDEL*, and *cyc* at the indicated combinations. *sfGFP* was inserted upstream of the *vg1* mature domain. Scale bar, 20 µm. (**B**) Anti-GFP non-reducing immunoblot of M*vg1* embryos injected with the indicated constructs as in (**A**). Black arrowhead indicates the position of uncleaved SynPro Vg1, open arrowhead indicates secTEVp-cleaved Vg1. 15 embryos at 3 hpf were loaded per well.

Figure 2. Prodomain cleavage affects Vg1-Nodal signaling but not secretion. (**A**) Live fluorescence imaging of M*vg1* co-injected with 50 pg of *vg1-sfGFP* or non-cleavable *vg1-sfGFP* (*vg1-NC-sfGFP*, RSRRKR → SQNTSN) mRNA and 50 pg of *sqt* or *sqt-NC* (RRHRR → SQNTS) mRNA. Scale bar, 17 µm. (**B**) Nodal target gene (*lefty1*) expression at 50% epiboly in WT and M*vg1* embryos injected with 50 pg of *sqt*, *sqt-NC* or *cyc* mRNA. (**C**) *lefty1* expression in M*vg1* embryos co-injected with 50 pg *sqt*, *sqt-NC* or *cyc* and *vg1* or *vg1-NC* mRNA.

Figure 2—figure supplement 1. Non-cleavable Squint. Anti-GFP reducing immunoblot of M*vg1* embryos injected with 50 pg mRNA of *cyc-sfGFP*, *cyc-NC-sfGFP* (RRGRR → SQNTS), *sqt-sfGFP* or *sqt-NC-sfGFP* mRNA. Black arrowhead indicates the position of full-length protein, open arrowhead indicates processed mature domain.

Figure 3. Cysteine and N-linked glycosylation sites retain the Vg1 prodomain in the ER. (**A**) Schematic and primary amino acid sequence of zebrafish Vg1 pre-proprotein. Cysteines (red) and asparagines (blue) are highlighted. (**B-D**) Fluorescence images of fixed M*vg1* embryos injected with 50 pg mRNA of *sfGFP*-tagged *vg1* prodomain (proVg1) (**B**), *vg1* mature domain (matVg1) (**C**), and full-length *vg1* (**D**), with or without the indicated cysteines and asparagines mutated. *sfGFP* was inserted into *vg1* downstream of the predicted basic cleavage site in all constructs. Scale bar, 20 µm.

Figure 3—figure supplement 1. The Vg1 prodomain inhibits secretion. (A) Immunofluorescence images of fixed M*vg1* embryos injected with 50 pg mRNA of *vg1-sfGFP* (left), *vg1* prodomain fused to *sfGFP*-tagged mature *sqt* domain (middle) or *sfGFP*-tagged mature *cyc* domain (right) or (**B**) *sqt* prodomain (left) or *cyc* prodomain (right) fused to *sfGFP*-tagged mature *vg1* domain, Embryos were immunostained with antibodies against sfGFP and pSmad2. Scale bar, 20 µm.

Figure 3—figure supplement 2. Vg1 mRNA rescue of M*vg1* embryos. Rescue percentage after 30 hpf of M*vg1* embryos injected with 50 pg of *vg1* mRNAs with the indicated mutations. Full or partial (with cyclopia) rescue percentages are respectively indicated as bars that are blue or blue with orange edges.

Figure 4. Binding motifs for BiP promote ER retention of the Vg1 mature domain. (**A**) Amino acid sequence alignment of the mature domains of Vg1 (matVg1) and Cyc (matCyc) and mutant matVg1. Cysteines (red) and asparagines (blue) are highlighted. (**B-C**) Difference charts of the BiP binding specificity scores (Blond-Elguindi et al. 1993) between two protein sequences along a sliding window of seven amino acids. Orange fills indicate the matVg1 or matVg1(*m1*, *m2*, *m3*) score (black line) > matCyc score (blue line). Conversely, blue fills indicate matCyc score > matVg1 or matVg1(*m1*, *m2*, *m3*) score. Arrows denote regions where matVg1 score > 0 and matCyc score < 0. Red arrows specifically denote regions that contain cysteines involved in cystine-knot formation. (**B**) Differences in BiP binding specificity scores between matVg1 (black line) and matCyc (blue line). (**C**) Differences in BiP binding specificity scores between mutant matVg1(*m1*, *m2*, *m3*) (black line) and matCyc (blue line). Note the loss of orange fills in *m1*, *m2*, and *m3* regions (pink arrows) when compared to (**B**). (**D**) Fluorescence images of fixed M*vg1* embryos injected with 50 pg mRNA of sfGFP-tagged *vg1* mature domain variants. *sfGFP* was inserted upstream of the *vg1* mature domain in all constructs.

Figure 4—figure supplement 1. Localization and activity of Vg1 and Nodal mature domains. (A) (Left to right) Immunofluorescence images of fixed M*vg1* embryos injected with 50 pg mRNA of *sfGFP-matVg1*, *sfGFP-matSqt*, *sfGFP-matCyc*, or *sfGFP-matVg1* and *cyc*. Embryos were immunostained with antibodies against sfGFP and pSmad2. DAPI, nuclei. Scale bar, 20 µm.

Figure 4—figure supplement 2. ER retention of matCyc using BiP-binding features of matVg1. (**A**) Amino acid sequence alignment of the mature domains of Vg1 (matVg1) and Cyc (matCyc) and mutant matCyc. Cysteines (red) and asparagines (blue) are highlighted. (**B**) Difference charts of the BiP binding specificity scores (Blond-Elguindi et al. 1993) between wild-type and mutant matCyc sequences along a sliding window of seven amino acids. Orange fills indicate the mutant matCyc (black line) score > matCyc (blue line) score. Conversely, blue fills indicate matCyc score > mutant matCyc score. Note the gain of orange fills in *m1'*, *m2'*, and *m3'* regions (arrows). (**C**) Fluorescence images of fixed M*vg1* embryos injected with 50 pg mRNA of *sfGFP*-tagged *matCyc* and *sfGFP*-tagged mutant *matCyc(m1', m2', m3')*. *sfGFP* was inserted upstream of the mature *cyc* domain in all constructs. DAPI, nuclei. Scale bar, 20 µm.

Figure 5. Model of Vg1-Nodal heterodimer formation and processing. Maternal Vg1 is retained in the ER via chaperone-binding motifs; at the maternal-zygotic transition, Nodal is produced and induces release of Vg1 from the ER via heterodimer formation; processing of the Vg1 prodomain is required for the activity of Vg1-Nodal heterodimers but can be independent from dimer formation and secretion.

Table 1. BiP Binding Specificity Scores of matVg1, matCyc, matVg1(*m1*, *m2*, *m3*), and matCyc(m1', m2', m3') sequences. These scores are visualized in Figure 4B-C and Figure 4—figure supplement 2B.

	matVg1		**matCyc**		**matVg1 (*m1*, *m2*, *m3*)**		**matCyc (*m1'*, *m2'*, *m3'*)**	
Window #	**Sequence**	**Score**	**Sequence**	**Score**	**Sequence**	**Score**	**Sequence**	**Score**
1			GPPVRSP	**-13**			GPPVRSP	**-13**
2			PPVRSPE	**-3**			PPVRSPE	**-3**
3			PVRSPEL	**3**			PVRSPEL	**3**
4			VRSPELQ	**2**			VRSPELQ	**2**
5			RSPELQR	**-7**			RSPELQR	**-7**
6			SPELQRT	**-18**			SPELQRT	**-18**
7	SASYYLP	**-6**	PELQRTP	**1**	SASYYLP	**-6**	PELQRTP	**1**
8	ASYYLPV	**6**	ELQRTPL	**16**	ASYYLPV	**6**	ELQRTPL	**16**
9	SYYLPVT	**-15**	LQRTPLH	**-2**	SYYLPVT	**-15**	LQRTPLH	**-2**
10	YYLPVTP	**-5**	QRTPLHK	**-7**	YYLPVTP	**-5**	QRTPLHK	**-7**
11	YLPVTPS	**1**	RTPLHKS	**-1**	YLPVTPS	**1**	RTPLHKS	**-1**
12	LPVTPSN	**-24**	TPLHKST	**-18**	LPVTPSN	**-24**	TPLHKST	**-18**
13	PVTPSNV	**-7**	PLHKSTT	**-11**	PVTPSNV	**-7**	PLHKSTT	**-11**
14	VTPSNVC	**1**	LHKSTTC	**5**	VTPSNVC	**1**	LHKSTTC	**5**
15	TPSNVCK	**-12**	HKSTTCR	**-4**	TPSNVCK	**-12**	HKSTTCR	**-4**
16	PSNVCKP	**-14**	KSTTCRR	**-9**	PSNVCKP	**-14**	KSTTCRR	**-9**
17	SNVCKPR	**-3**	STTCRRV	**-4**	SNVCKPR	**-3**	STTCRRV	**-4**
18	NVCKPRR	**-13**	TTCRRVD	**2**	NVCKPRR	**-13**	TTCRRVD	**2**
19	VCKPRRL	**5**	TCRRVDM	**0**	VCKPRRL	**5**	TCRRVDM	**0**
20	CKPRRLY	**6**	CRRVDMH	**-7**	CKPRRLY	**6**	CRRVDMH	**-7**
21	KPRRLYI	**-2**	RRVDMHV	**-7**	KPRRLYI	**-2**	RRVDMHV	**-7**
22	PRRLYID	**-5**	RVDMHVD	**-3**	PRRLYID	**-5**	RVDMHVD	**-3**
23	RRLYIDF	**2**	VDMHVDF	**0**	RRLYIDF	**2**	VDMHVDF	**0**
24	RLYIDFK	**-10**	DMHVDFN	**-15**	RLYIDFK	**-10**	DMHVDFN	**-15**
25	LYIDFKD	**3**	MHVDFNQ	**8**	LYIDFKD	**3**	MHVDFNQ	**8**
26	YIDFKDV	**-19**	HVDFNQI	**-3**	YIDFKDV	**-19**	HVDFNQI	**-3**
27	IDFKDVG	**-3**	VDFNQIG	**10**	IDFKDVG	**-3**	VDFNQIG	**10**
28	DFKDVGW	**-7**	DFNQIGW	**-11**	DFKDVGW	**-7**	DFNQIGW	**-11**
29	FKDVGWQ	**8**	FNQIGWG	**8**	FKDVGWQ	**8**	FNQIGWG	**8**
30	KDVGWQD	**1**	NQIGWGS	**1**	KDVGWQD	**1**	NQIGWGS	**1**
31	DVGWQDW	**-1**	QIGWGSW	**2**	DVGWQDW	**-1**	QIGWGSW	**2**
32	VGWQDWI	**9**	IGWGSWI	**1**	VGWQDWI	**9**	IGWGSWI	**1**
33	GWQDWII	**11**	GWGSWIV	**11**	GWQDWII	**11**	GWGSWIV	**11**
34	WQDWIIA	**0**	WGSWIVF	**12**	WQDWIIA	**0**	WGSWIVF	**12**
35	QDWIIAP	**5**	GSWIVFP	**2**	QDWIIAP	**5**	GSWIVFP	**2**
36	DWIIAPQ	**12**	SWIVFPK	**9**	DWIIAPK	**1**	SWIVFPQ	**20**
37	WIIAPQG	**-2**	WIVFPKK	**-12**	WIIAPKK	**-9**	WIVFPQG	**-5**
38	IIAPQGY	**-3**	IVFPKKY	**1**	IIAPKKY	**-12**	IVFPQGY	**10**
39	IAPQGYL	**1**	VFPKKYN	**-21**	IAPKKYN	**-23**	VFPQGYL	**3**
40	APQGYLA	**-26**	FPKKYNA	**-22**	APKKYNA	**-34**	FPQGYLA	**-14**
41	PQGYLAN	**-1**	PKKYNAY	**3**	PKKYNAY	**3**	PQGYLAN	**-1**
42	QGYLANY	**-2**	KKYNAYR	**-5**	KKYNAYR	**-5**	QGYLANY	**-2**
43	GYLANYC	**-2**	KYNAYRC	**-14**	KYNAYRC	**-14**	GYLANYC	**-2**
44	YLANYCH	**-16**	YNAYRCE	**-7**	YNAYRCH	**-8**	YLANYCE	**-15**
45	LANYCHG	**-1**	NAYRCEG	**-7**	NAYRCHG	**-4**	LANYCEG	**-4**
46	ANYCHGE	**-11**	AYRCEGA	**-17**	AYRCHGE	**-6**	ANYCEGA	**-22**
47	NYCHGEC	**-2**	YRCEGAC	**-19**	YRCHGEC	**-14**	NYCEGAC	**-7**

48	YCHGECP	**-10**	RCEGACP	**-13**	RCHGECP	**-3**	YCEGACP	**-20**
49	CHGECPF	**9**	CEGACPN	**6**	CHGECPF	**9**	CEGACPN	**6**
50	HGECPFP	**-13**	EGACPNP	**-11**	HGECPFP	**-13**	EGACPNP	**-11**
51	GECPFPL	**19**	GACPNPL	**16**	GECPFPL	**19**	GACPNPL	**16**
52	ECPFPLS	**-8**	ACPNPLG	**-8**	ECPFPLS	**-8**	ACPNPLG	**-8**
53	CPFPLSE	**5**	CPNPLGE	**-7**	CPFPLSE	**5**	CPNPLGE	**-7**
54	PFPLSES	**-7**	PNPLGEE	**-2**	PFPLSES	**-7**	PNPLGEE	**-2**
55	FPLSESL	**6**	NPLGEEL	**-2**	FPLSESL	**6**	NPLGEEL	**-2**
56	PLSESLN	**-19**	PLGEELR	**-8**	PLSESLR	**-14**	PLGEELR	**-8**
57	LSESLNG	**-5**	LGEELRP	**-11**	LSESLRP	**-7**	LGEELRP	**-11**
58	SESLNGT	**-3**	GEELRPT	**3**	SESLRPT	**11**	GEELRPT	**3**
59	ESLNGTN	**-8**	EELRPTN	**-13**	ESLRPTN	**-14**	EELRPTN	**-13**
60	SLNGTNH	**-13**	ELRPTNH	**-1**	SLRPTNH	**-2**	ELRPTNH	**-1**
61	LNGTNHA	**-9**	LRPTNHA	**-15**	LRPTNHA	**-15**	LRPTNHA	**-15**
62	NGTNHAI	**1**	RPTNHAY	**-5**	RPTNHAI	**-7**	RPTNHAI	**-7**
63	GTNHAIL	**-1**	PTNHAYM	**-7**	PTNHAIL	**0**	PTNHAIL	**0**
64	TNHAILQ	**7**	TNHAYMQ	**-4**	TNHAILQ	**7**	TNHAILQ	**7**
65	NHAILQT	**-1**	NHAYMQS	**-4**	NHAILQT	**-1**	NHAILQT	**-1**
66	HAILQTL	**6**	HAYMQSL	**-4**	HAILQTL	**6**	HAILQTL	**6**
67	AILQTLV	**-3**	AYMQSLL	**-5**	AILQTLV	**-3**	AILQTLL	**3**
68	ILQTLVH	**-2**	YMQSLLK	**-1**	ILQTLVK	**-7**	ILQTLLH	**3**
69	LQTLVHS	**5**	MQSLLKY	**8**	LQTLVKY	**2**	LQTLLHS	**8**
70	QTLVHSF	**-2**	QSLLKYH	**-10**	QTLVKYH	**-9**	QTLLHSF	**1**
71	TLVHSFD	**-9**	SLLKYHH	**-12**	TLVKYHH	**-14**	TLLHSFD	**-6**
72	LVHSFDP	**6**	LLKYHHP	**5**	LVKYHHP	**7**	LLHSFDP	**4**
73	VHSFDPK	**4**	LKYHHPS	**7**	VKYHHPS	**5**	LHSFDPK	**6**
74	HSFDPKG	**-3**	KYHHPSR	**-13**	KYHHPSR	**-13**	HSFDPKG	**-3**
75	SFDPKGT	**-12**	YHHPSRV	**-16**	YHHPSRV	**-16**	SFDPKGT	**-12**
76	FDPKGTP	**-1**	HHPSRVP	**-2**	HHPSRVP	**-2**	FDPKGTP	**-1**
77	DPKGTPQ	**-1**	HPSRVPA	**-6**	HPSRVPA	**-6**	DPKGTPQ	**-1**
78	PKGTPQP	**-10**	PSRVPAS	**-10**	PSRVPAS	**-10**	PKGTPQP	**-10**
79	KGTPQPC	**14**	SRVPASC	**-14**	SRVPASC	**-14**	KGTPQPC	**14**
80	GTPQPCC	**-6**	RVPASCC	**-5**	RVPASCC	**-5**	GTPQPCC	**-6**
81	TPQPCCV	**-6**	VPASCCA	**-20**	VPASCCV	**-9**	TPQPCCA	**-17**
82	PQPCCVP	**-4**	PASCCAP	**-4**	PASCCVP	**-7**	PQPCCAP	**-1**
83	QPCCVPI	**6**	ASCCAPT	**7**	ASCCVPI	**6**	QPCCAPT	**7**
84	PCCVPIK	**-11**	SCCAPTR	**-7**	SCCVPIK	**-13**	PCCAPTR	**-5**
85	CCVPIKL	**6**	CCAPTRT	**-2**	CCVPIKL	**6**	CCAPTRT	**-2**
86	CVPIKLS	**-4**	CAPTRTS	**0**	CVPIKLS	**-4**	CAPTRTS	**0**
87	VPIKLSP	**-13**	APTRTSA	**-27**	VPIKLSP	**-13**	APTRTSA	**-27**
88	PIKLSPI	**5**	PTRTSAL	**0**	PIKLSPI	**5**	PTRTSAL	**0**
89	IKLSPIS	**-2**	TRTSALS	**-7**	IKLSPIS	**-2**	TRTSALS	**-7**
90	KLSPISM	**-1**	RTSALSM	**2**	KLSPISM	**-1**	RTSALSM	**2**
91	LSPISML	**-3**	TSALSML	**-10**	LSPISML	**-3**	TSALSML	**-10**
92	SPISMLY	**-5**	SALSMLY	**0**	SPISMLY	**-5**	SALSMLY	**0**
93	PISMLYY	**3**	ALSMLYY	**-2**	PISMLYY	**3**	ALSMLYY	**-2**
94	ISMLYYD	**-4**	LSMLYYE	**-4**	ISMLYYD	**-4**	LSMLYYE	**-4**
95	SMLYYDN	**-8**	SMLYYEN	**-8**	SMLYYDN	**-8**	SMLYYEN	**-8**
96	MLYYDNN	**-5**	MLYYENG	**1**	MLYYDNN	**-5**	MLYYENG	**1**
97	LYYDNND	**-1**	LYYENGE	**-10**	LYYDNND	**-1**	LYYENGE	**-10**
98	YYDNNDN	**-15**	YYENGEM	**-17**	YYDNNDN	**-15**	YYENGEM	**-17**
99	YDNNDNV	**-13**	YENGEMI	**-19**	YDNNDNV	**-13**	YENGEMI	**-19**
100	DNNDNVV	**-20**	ENGEMIL	**-2**	DNNDNVV	**-20**	ENGEMIL	**-2**
101	NNDNVVL	**2**	NGEMILR	**-5**	NNDNVVL	**2**	NGEMILR	**-5**

102	NDNVVLR	**-9**	GEMILRH	**-3**	NDNVVLR	**-9**	GEMILRH	**-3**
103	DNVVLRH	**-14**			DNVVLRH	**-14**		
104	NVVLRHY	**5**	EMILRHH	**5**	NVVLRHY	**5**	EMILRHH	**5**
105	VVLRHYE	**1**	MILRHHE	**7**	VVLRHYE	**1**	MILRHHE	**7**
106	VLRHYED	**-3**	ILRHHED	**1**	VLRHYED	**-3**	ILRHHED	**1**
107	LRHYEDM	**-1**	LRHHEDM	**-5**	LRHYEDM	**-1**	LRHHEDM	**-5**
108	RHYEDMV	**-14**	RHHEDMQ	**-4**	RHYEDMV	**-14**	RHHEDMQ	**-4**
109	HYEDMVV	**-15**	HHEDMQV	**-12**	HYEDMVV	**-15**	HHEDMQV	**-12**
110	YEDMVVD	**-11**	HEDMQVE	**-6**	YEDMVVD	**-11**	HEDMQVE	**-6**
111	EDMVVDE	**-5**	EDMQVEE	**-3**	EDMVVDE	**-5**	EDMQVEE	**-3**
112	DMVVDEC	**-11**	DMQVEEC	**-7**	DMVVDEC	**-11**	DMQVEEC	**-7**
113	MVVDECG	**2**	MQVEECG	**-4**	MVVDECG	**2**	MQVEECG	**-4**
114	VVDECGC	**-8**	QVEECGC	**-16**	VVDECGC	**-8**	QVEECGC	**-16**
115	VDECGCR	**-14**	VEECGCL	**-7**	VDECGCR	**-14**	VEECGCL	**-7**

MATERIALS AND METHODS

Genotyping of *vg1* mutants

Genomic DNA was isolated via the HotSHOT method from either excised adult caudal fin tissue or individual fixed embryos (Meeker et al. 2007). Genotyping was carried out via PCR using standard conditions followed by 2% gel electrophoresis. Mutant *vg1* fish have the $vg1^{a165}$ allele, which contains a 29 bp deletion in the first exon of *vg1* and was detected as described (Montague and Schier 2017). Allele designation was omitted for brevity in the rest of the text.

Zebrafish husbandry and fish lines

Fish were maintained per standard laboratory conditions (Westerfield 2000). Embryos were raised at 28.5ºC in embryo medium (250 mg/L Instant Ocean salt, 1 mg/L methylene blue in reverse osmosis water adjusted to pH 7 with sodium bicarbonate) and staged according to a standard staging series (Kimmel et al. 1995). Wild-type fish and embryos represent the TLAB strain. The *vg1* mutant fish line was maintained as previously described (Montague and Schier 2017). M*vg1* embryos were generated by crossing zygotic homozygous *vg1* female fish to TLAB wild-type male fish.

Cloning of expression constructs and the Synthetic Processing (SynPro) system

Standard molecular cloning techniques, such as PCR, Gibson Assembly (Gibson et al. 2009), and site-directed mutagenesis, were performed to assemble new constructs used in this study.

The coding sequences (CDS) of *vg1* (Montague and Schier 2017), *sqt*, *cyc* (Muller et al. 2012), and their variants tagged with superfolder GFP (sfGFP) (Pedelacq et al. 2006), were previously assembled into the pCS2(+) vector that contains a β-globin 5' UTR and an SV40 late polyA signal at the 3' UTR. For sfGFP-tagged Vg1 prodomain or mature domain, site-directed mutagenesis (Q5 Kit, New England Biolabs) of pCS2(+)-vg1-sfGFP was performed to truncate the full-length construct into respective domains. All point mutants, indels, and epitope tags were subsequently generated using site-directed mutagenesis. For non-cleavable vg1-NC, vg1-NC-sfGFP and proVg1-sfGFP, the cleavage site 'RSRRKR' was replaced with 'SQNTSN'. For non-cleavable sqt-NC and sqt-NC-sfGFP, the cleavage site 'RRHRR' was replaced with 'SQNTS'. For chimeras of Squint/Cyclops prodomain and sfGFP-tagged Vg1 mature domain (sfGFP-matVg1), the Squint and Cyclops prodomains (including their respective cleavage sites) were PCR-amplified from pCS2(+)-squint and pCS2(+)-cyclops, and sfGFP-matVg1 was PCR-amplified from pCS2(+)-vg1-sfGFP.

Commercially available sequences for tobacco etch virus protease (TEVp, Addgene Plasmid #8835), tobacco vein mottling virus protease (TVMVp; Addgene Plasmid #8832), and human rhinovirus 3C protease (HRV 3Cp; Addgene Plasmid #78571) were gifts from David Waugh (Kapust et al. 2001; Nallamsetty et al. 2004; Raran-Kurussi and Waugh 2016). Commercially available sequence for bovine enterokinase (Addgene Plasmid #49048) was a gift from Hans Brandstetter (Skala, Goettig, and Brandstetter 2013). These sequences were PCR-amplified and subcloned into the pCS2(+) vector for downstream applications *in vivo*.

Zebrafish codon-optimized sequences of 15 Potyviral proteases listed in Supp Fig 3 were synthesized by Integrated DNA Technologies and subcloned into the pCS2(+) vector. For the protease cleavage reporter assay described in Supp Fig 3, Component 1 is a cleavable substrate for Potyviral proteases and was constructed via Gibson assembly of zebrafish beta-Arrestin (*arrb2a* gene PCR-amplified from high-stage cDNA library) and mNeonGreen2(1-10) (a gift from Bo Huang) (Feng et al. 2017) into the pCS2(+) vector. The cognate cleavage site inserted between arrb2a and mNeonGreen2(1-10) was encoded in the primer overhangs. Component 2 is the codon-optimized Potyviral protease described above. Component 3 was constructed via Gibson assembly of four tandem copies of mNeonGreen2(11) (a gift from Bo Huang) (Feng et al. 2017) and mCherry-tagged Histone 2B (a gift from Jeffrey Farrell) into the pCS2(+) vector.

For the SynPro system, the codon-optimized TEV protease was further modified using site-directed mutagenesis to generate secTEVp. The secTEVp construct contains a secretion signal sequence derived from zebrafish toddler (translated signal peptide: MRFFHPLYLLLLLLTVLVLISA) and 5 point mutations (N23Q, C130S, T173G, L56V, S135G). The secTEVp-sfCherry-KDEL construct additionally contains a C-terminal fusion of sfCherry3C (Feng et al. 2017) sequence and an ER-targeting motif '(GSGS)EEKDEL'. SynPro Vg1 was derived from pCS2(+)-vg1-sfGFP using site-directed mutagenesis to replace the 'RSRRKR' cleavage site to '(GSGS)ENLYFQS(GS)'.

Determination of BiP-binding scores

To focus our mutagenesis efforts, we ignored corresponding matVg1 and matCyc heptapeptides that both scored positive in the Gething-Sambrook scoring system because BiP is predicted to bind to both. We also ignored corresponding regions where both heptapeptides scored zero or negative because BiP is predicted to not bind at all. We focused on 8 regions wherein matVg1 heptapeptides scored positive and the corresponding matCyc sequences scored negative (Figure 2B, arrows). We ignored 4 out of the 7 regions because they included six cysteines that participate in cystine-knot formation (Figure 2B, red arrows). BiP-binding scores were visualized using the D3.js chart library (https://d3js.org/) and the code used is available on GitHub (https://github.com/davedingal/BiP_binding_score).

mRNA synthesis and microinjection into embryos

All pCS2(+) plasmids were linearized with NotI and subsequently purified with the E.Z.NA. Cycle Pure Kit (Omega). Capped mRNAs were synthesized with the Sp6 mMessage Machine kit (Invitrogen) using the purified linearized plasmids as templates. Capped mRNAs were then purified with the E.Z.N.A. Total RNA Kit I (Omega). Capped mRNA concentrations were measured using the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). All kits were used according to respective manufacturer's protocols. If not mentioned otherwise, all mRNAs were injected at 50 pg into embryos at the one-cell stage using standard methods (Westerfield 2000).

Transplantation

For transplantation experiments, donor and host M*vg1* embryos at one-cell stage were injected with 50 pg of each mRNA relevant to the experiment and then grown to 1000-cell stage (3 hpf). At 1000-cell stage, cells were transplanted from donor embryos to host embryos, and were grown to shield stage (6 hpf) before fixation for immunostaining.

Live imaging and immunofluorescence imaging

Embryos were injected with 50 pg of each mRNA, grown to sphere stage, and then embedded in 1% low melting temperature agarose (Aquapor) on glass-bottomed dishes (MatTek). Imaging was performed on a Zeiss LSM 880 inverted confocal microscope.

***in situ* hybridization**

Embryos were grown to 50% epiboly then fixed in 4% formaldehyde overnight at 4ºC. Whole mount *in situ* hybridizations were performed according to standard protocols (Thisse and Thisse, 2008). A DIG-labeled antisense RNA probe targeting *lefty1* was synthesized using a DIG Probe Synthesis Kit (Roche). NBT/BCIP/Alkaline phosphatase-stained embryos were dehydrated in methanol before clearing and imaging in 2:1 benzyl benzoate:benzyl alcohol (BBBA) using a Zeiss Axio Imager.Z1 microscope.

Immunoblotting

Embryos were injected at the 1-cell stage with 50 pg of mRNA then allowed to develop to 50% epiboly. 8 embryos per sample were manually deyolked with forceps and frozen in liquid nitrogen. Samples were boiled for 5 min at 95ºC with 2x SDS loading buffer and DTT (150 mM final concentration), then loaded onto Any kD protein gels (Bio-Rad). Samples were transferred to PVDF membranes (GE Healthcare),

blocked in 5% non-fat milk (Bio-Rad) in TBST and incubated in primary antibodies (1:5000 rabbit anti-GFP, ThermoFisher A11122, RRID:AB_221569) at 4°C overnight. Proteins were detected using HRP-coupled secondary antibody (1:15,000 goat anti-rabbit, Jackson ImmunoResearch Labs 111-035-144, RRID:AB_2307391). Chemiluminescence was detected using Amersham ECL reagent (GE Healthcare).

Immunostaining

A previous protocol (Rogers et al. 2017) was modified to improve signal-to-noise ratio. Embryos were fixed in 4% paraformaldehyde overnight at 4ºC in PBSTw (1× phosphate-buffered saline + 0.1% v/v Tween 20), washed three times in PBSTw for 10 min each, dehydrated in a MeOH/PBSTw mixture series (25%, 50%, 75%, and 100% methanol) at 5 min per wash at room temperature (RT), and stored in 100% MeOH at –20ºC for at least 2 h. Embryos were rehydrated in a MeOH/PBSTr (1× PBS + 1% Triton X-100) mixture series (75%, 50%, 25% MeOH) for 5 min each, washed three times in PBSTr for 10 min each at RT, and manually de-yolked. Embryos were then incubated in antibody binding buffer (PBSTr + 1% v/v dimethyl sulfoxide) for 2 h at RT and subsequently stored overnight at 4ºC in antibody binding buffer containing relevant primary antibodies. After primary antibody incubation, embryos were washed six times with PBSTr for 10 min each, before a 30-min incubation in antibody binding buffer at RT. Embryos were then incubated for 2 h at RT in antibody binding buffer containing appropriate fluorescent secondary antibodies. Embryos were then washed six times with PBSTr. To label DNA in cell nuclei, embryos were incubated with 1 µg/mL DAPI in PBSTw for 30 min at RT. Lastly, embryos were washed three times in PBSTw for 10 min each at RT, and then mounted for microscopy.

Primary antibodies were used against GFP (1:1000 chicken IgY, Aves Lab, RRID:AB_2307313) and phosphorylated Smad2 (1:1000 rabbit IgG, Cell Signaling, RRID:AB_2798798). Fluorescent secondary antibodies used were goat anti-chicken Alexa 488 conjugate (1:2000, Thermo Fisher, RRID:AB_2534096) and goat anti-rabbit Alexa 647 conjugate (1:2000, Thermo Fisher, RRID:AB_2633282).

Image processing

All images were processed in Fiji/ImageJ (Schindelin et al. 2012). Brightness, contrast, and color balance were uniformly applied to images.

References

Alberini, C. M., P. Bet, C. Milstein, and R. Sitia. 1990. 'Secretion of immunoglobulin M assembly intermediates in the presence of reducing agents', *Nature*, 347: 485-7.

Baker, J. C., and R. M. Harland. 1996. 'A novel mesoderm inducer, Madr2, functions in the activin signal transduction pathway', *Genes Dev*, 10: 1880-9.

Beck, S., J. A. Le Good, M. Guzman, N. Ben Haim, K. Roy, F. Beermann, and D. B. Constam. 2002. 'Extraembryonic proteases regulate Nodal signalling during gastrulation', *Nat Cell Biol*, 4: 981-5.

Ben-Haim, N., C. Lu, M. Guzman-Ayala, L. Pescatore, D. Mesnard, M. Bischofberger, F. Naef, E. J. Robertson, and D. B. Constam. 2006. 'The nodal precursor acting via activin receptors induces mesoderm by maintaining a source of its convertases and BMP4', *Dev Cell*, 11: 313-23.

Birsoy, B., M. Kofron, K. Schaible, C. Wylie, and J. Heasman. 2006. 'Vg 1 is an essential signaling molecule in Xenopus development', *Development*, 133: 15-20.

Bisgrove, B. W., Y. C. Su, and H. J. Yost. 2017. 'Maternal Gdf3 is an obligatory cofactor in Nodal signaling for embryonic axis formation in zebrafish', *Elife*, 6.

Blanchet, M. H., J. A. Le Good, D. Mesnard, V. Oorschot, S. Baflast, G. Minchiotti, J. Klumperman, and D. B. Constam. 2008. 'Cripto recruits Furin and PACE4 and controls Nodal trafficking during proteolytic maturation', *EMBO J*, 27: 2580-91.

Blanchet, M. H., J. A. Le Good, V. Oorschot, S. Baflast, G. Minchiotti, J. Klumperman, and D. B. Constam. 2008. 'Cripto localizes Nodal at the limiting membrane of early endosomes', *Sci Signal*, 1: ra13.

Blond-Elguindi, S., S. E. Cwirla, W. J. Dower, R. J. Lipshutz, S. R. Sprang, J. F. Sambrook, and M. J. Gething. 1993. 'Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP', *Cell*, 75: 717-28.

Bole, D. G., L. M. Hendershot, and J. F. Kearney. 1986. 'Posttranslational association of immunoglobulin heavy chain binding protein with nascent heavy chains in nonsecreting and secreting hybridomas', *J Cell Biol*, 102: 1558-66.

Cabrita, L. D., D. Gilis, A. L. Robertson, Y. Dehouck, M. Rooman, and S. P. Bottomley. 2007. 'Enhancing the stability and solubility of TEV protease using in silico design', *Protein Sci*, 16: 2360-7.

Cesaratto, F., A. Lopez-Requena, O. R. Burrone, and G. Petris. 2015. 'Engineered tobacco etch virus (TEV) protease active in the secretory pathway of mammalian cells', *J Biotechnol*, 212: 159-66.

Chen, C., S. M. Ware, A. Sato, D. E. Houston-Hawkins, R. Habas, M. M. Matzuk, M. M. Shen, and C. W. Brown. 2006. 'The Vg1-related protein Gdf3 acts in a Nodal signaling pathway in the pre-gastrulation mouse embryo', *Development*, 133: 319-29.

Cheng, S. K., F. Olale, J. T. Bennett, A. H. Brivanlou, and A. F. Schier. 2003. 'EGF-CFC proteins are essential coreceptors for the TGF-beta signals Vg1 and GDF1', *Genes Dev*, 17: 31-6.

Collignon, J., I. Varlet, and E. J. Robertson. 1996. 'Relationship between asymmetric nodal expression and the direction of embryonic turning', *Nature*, 381: 155-8.

Conlon, F. L., K. M. Lyons, N. Takaesu, K. S. Barth, A. Kispert, B. Herrmann, and E. J. Robertson. 1994. 'A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse', *Development*, 120: 1919-28.

Dale, L., G. Matthews, and A. Colman. 1993. 'Secretion and mesoderm-inducing activity of the TGF-beta-related domain of Xenopus Vg1', *EMBO J*, 12: 4471-80.

Dale, L., G. Matthews, L. Tabe, and A. Colman. 1989. 'Developmental expression of the protein product of Vg1, a localized maternal mRNA in the frog Xenopus laevis', *EMBO J*, 8: 1057-65.

Dohrmann, C. E., D. S. Kessler, and D. A. Melton. 1996. 'Induction of axial mesoderm by zDVR-1, the zebrafish orthologue of Xenopus Vg1', *Dev Biol*, 175: 108-17.

Feige, M. J., S. Groscurth, M. Marcinowski, Y. Shimizu, H. Kessler, L. M. Hendershot, and J. Buchner. 2009. 'An unfolded CH1 domain controls the assembly and secretion of IgG antibodies', *Mol Cell*, 34: 569-79.

Feldman, B., M. A. Gates, E. S. Egan, S. T. Dougan, G. Rennebeck, H. I. Sirotkin, A. F. Schier, and W. S. Talbot. 1998. 'Zebrafish organizer development and germ-layer formation require nodal-related signals', *Nature*, 395: 181-5.

Feng, S., S. Sekine, V. Pessino, H. Li, M. D. Leonetti, and B. Huang. 2017. 'Improved split fluorescent proteins for endogenous protein labeling', *Nat Commun*, 8: 370.

Flynn, G. C., T. G. Chappell, and J. E. Rothman. 1989. 'Peptide binding and release by proteins implicated as catalysts of protein assembly', *Science*, 245: 385-90.

Flynn, G. C., J. Pohl, M. T. Flocco, and J. E. Rothman. 1991. 'Peptide-binding specificity of the molecular chaperone BiP', *Nature*, 353: 726-30.

Gibson, D. G., L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, 3rd, and H. O. Smith. 2009. 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nat Methods*, 6: 343-5.

Gritsman, K., J. Zhang, S. Cheng, E. Heckscher, W. S. Talbot, and A. F. Schier. 1999. 'The EGF-CFC protein one-eyed pinhead is essential for nodal signaling', *Cell*, 97: 121-32.

Haas, I. G., and M. Wabl. 1983. 'Immunoglobulin heavy chain binding protein', *Nature*, 306: 387-9.

Hammond, C., I. Braakman, and A. Helenius. 1994. 'Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control', *Proc Natl Acad Sci U S A*, 91: 913-7.

Helenius, A., and M. Aebi. 2004. 'Roles of N-linked glycans in the endoplasmic reticulum', *Annu Rev Biochem*, 73: 1019-49.

Hinck, A. P., T. D. Mueller, and T. A. Springer. 2016. 'Structural Biology and Evolution of the TGF-beta Family', *Cold Spring Harb Perspect Biol*, 8.

Hu, C. C., T. Bachmann, G. Zhou, F. X. Liang, J. Ghiso, G. Kreibich, and T. T. Sun. 2008. 'Assembly of a membrane receptor complex: roles of the uroplakin II prosequence in regulating uroplakin bacterial receptor oligomerization', *Biochem J*, 414: 195-203.

Kapust, R. B., J. Tozser, J. D. Fox, D. E. Anderson, S. Cherry, T. D. Copeland, and D. S. Waugh. 2001. 'Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency', *Protein Eng*, 14: 993-1000.

Kimmel, C. B., W. W. Ballard, S. R. Kimmel, B. Ullmann, and T. F. Schilling. 1995. 'Stages of embryonic development of the zebrafish', *Dev Dyn*, 203: 253-310.

Kuznetsov, G., L. B. Chen, and S. K. Nigam. 1997. 'Multiple molecular chaperones complex with misfolded large oligomeric glycoproteins in the endoplasmic reticulum', *J Biol Chem*, 272: 3057-63.

Levin, M., R. L. Johnson, C. D. Stern, M. Kuehn, and C. Tabin. 1995. 'A molecular pathway determining left-right asymmetry in chick embryogenesis', *Cell*, 82: 803-14.

Long, S., N. Ahmad, and M. Rebagliati. 2003. 'The zebrafish nodal-related gene southpaw is required for visceral and diencephalic left-right asymmetry', *Development*, 130: 2303-16.

Lowe, L. A., D. M. Supp, K. Sampath, T. Yokoyama, C. V. Wright, S. S. Potter, P. Overbeek, and M. R. Kuehn. 1996. 'Conserved left-right asymmetry of nodal expression and alterations in murine situs inversus', *Nature*, 381: 158-61.

Mayer, M., U. Kies, R. Kammermeier, and J. Buchner. 2000. 'BiP and PDI cooperate in the oxidative folding of antibodies in vitro', *J Biol Chem*, 275: 29421-5.

Meeker, N. D., S. A. Hutchinson, L. Ho, and N. S. Trede. 2007. 'Method for isolation of PCR-ready genomic DNA from zebrafish tissues', *Biotechniques*, 43: 610, 12, 14.

Meunier, L., Y. K. Usherwood, K. T. Chung, and L. M. Hendershot. 2002. 'A subset of chaperones and folding enzymes form multiprotein complexes in endoplasmic reticulum to bind nascent proteins', *Mol Biol Cell*, 13: 4456-69.

Miyazono, K., A. Olofsson, P. Colosetti, and C. H. Heldin. 1991. 'A role of the latent TGF-beta 1-binding protein in the assembly and secretion of TGF-beta 1', *EMBO J*, 10: 1091-101.

Montague, T. G., J. A. Gagnon, and A. F. Schier. 2018. 'Conserved regulation of Nodal-mediated left-right patterning in zebrafish and mouse', *Development*, 145.

Montague, T. G., and A. F. Schier. 2017. 'Vg1-Nodal heterodimers are the endogenous inducers of mesendoderm', *Elife*, 6.

Muller, P., K. W. Rogers, B. M. Jordan, J. S. Lee, D. Robson, S. Ramanathan, and A. F. Schier. 2012. 'Differential diffusivity of Nodal and Lefty underlies a reaction-diffusion patterning system', *Science*, 336: 721-4.

Munro, S., and H. R. Pelham. 1987. 'A C-terminal signal prevents secretion of luminal ER proteins', *Cell*, 48: 899-907.

Nallamsetty, S., R. B. Kapust, J. Tozser, S. Cherry, J. E. Tropea, T. D. Copeland, and D. S. Waugh. 2004. 'Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro', *Protein Expr Purif*, 38: 108-15.

Pagan-Westphal, S. M., and C. J. Tabin. 1998. 'The transfer of left-right positional information during chick embryogenesis', *Cell*, 93: 25-35.

Pedelacq, J. D., S. Cabantous, T. Tran, T. C. Terwilliger, and G. S. Waldo. 2006. 'Engineering and characterization of a superfolder green fluorescent protein', *Nat Biotechnol*, 24: 79-88.

Pelliccia, J. L., G. A. Jindal, and R. D. Burdine. 2017. 'Gdf3 is required for robust Nodal signaling during germ layer formation and left-right patterning', *Elife*, 6.

Raran-Kurussi, S., and D. S. Waugh. 2016. 'A dual protease approach for expression and affinity purification of recombinant proteins', *Anal Biochem*, 504: 30-7.

Reddy, P. S., and R. B. Corley. 1998. 'Assembly, sorting, and exit of oligomeric proteins from the endoplasmic reticulum', *Bioessays*, 20: 546-54.

Rogers, K. W., N. D. Lord, J. A. Gagnon, A. Pauli, S. Zimmerman, D. C. Aksel, D. Reyon, S. Q. Tsai, J. K. Joung, and A. F. Schier. 2017. 'Nodal patterning without Lefty inhibitory feedback is functional but fragile', *Elife*, 6.

Rogers, K. W., and P. Muller. 2019. 'Nodal and BMP dispersal during early zebrafish development', *Dev Biol*, 447: 14-23.

Schier, A. F. 2009. 'Nodal morphogens', *Cold Spring Harb Perspect Biol*, 1: a003459.

Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. 2012. 'Fiji: an open-source platform for biological-image analysis', *Nat Methods*, 9: 676-82.

Seidah, N. G., and A. Prat. 2012. 'The biology and therapeutic targeting of the proprotein convertases', *Nat Rev Drug Discov*, 11: 367-83.

Shusta, E. V., R. T. Raines, A. Pluckthun, and K. D. Wittrup. 1998. 'Increasing the secretory capacity of Saccharomyces cerevisiae for production of single-chain antibody fragments', *Nat Biotechnol*, 16: 773-7.

Sitia, R., M. Neuberger, C. Alberini, P. Bet, A. Fra, C. Valetti, G. Williams, and C. Milstein. 1990. 'Developmental regulation of IgM secretion: the role of the carboxy-terminal cysteine', *Cell*, 60: 781-90.

Skala, W., P. Goettig, and H. Brandstetter. 2013. 'Do-it-yourself histidine-tagged bovine enterokinase: a handy member of the protein engineer's toolbox', *J Biotechnol*, 168: 421-5.

Tannahill, D., and D. A. Melton. 1989. 'Localized synthesis of the Vg1 protein during early Xenopus development', *Development*, 106: 775-85.

Tessadori, F., E. S. Noel, E. G. Rens, R. Magliozzi, I. J. Evers-van Gogh, D. Guardavaccaro, R. M. Merks, and J. Bakkers. 2015. 'Nodal signaling range is regulated by proprotein convertase-mediated maturation', *Dev Cell*, 32: 631-9.

Thomsen, G. H., and D. A. Melton. 1993. 'Processed Vg1 protein is an axial mesoderm inducer in Xenopus', *Cell*, 74: 433-41.

Tu, L., T. T. Sun, and G. Kreibich. 2002. 'Specific heterodimer formation is a prerequisite for uroplakins to exit from the endoplasmic reticulum', *Mol Biol Cell*, 13: 4221-30.

Vanhove, M., Y. K. Usherwood, and L. M. Hendershot. 2001. 'Unassembled Ig heavy chains do not cycle from BiP in vivo but require light chains to trigger their release', *Immunity*, 15: 105-14.

Waugh, D. S. 2011. 'An overview of enzymatic reagents for the removal of affinity tags', *Protein Expr Purif*, 80: 283-93.

Westerfield, M. 2000. *The zebrafish book: a guide for the laboratory use of zebrafish (Danio rerio)* (University of Oregon Press, Eugene).

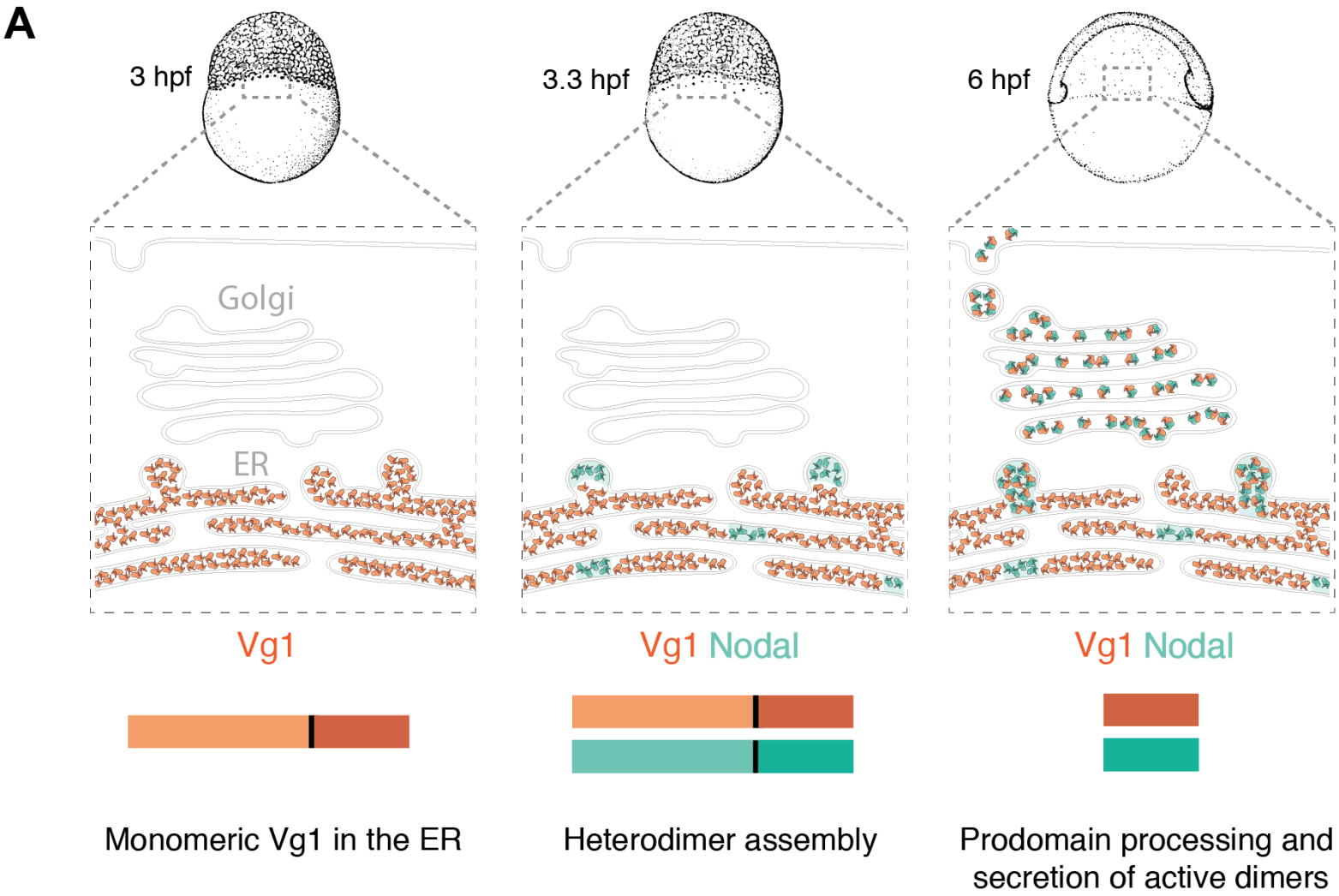
Wilkinson, B., and H. F. Gilbert. 2004. 'Protein disulfide isomerase', *Biochim Biophys Acta*, 1699: 35-44.

Wylie, S. J., M. Adams, C. Chalam, J. Kreuze, J. J. Lopez-Moya, K. Ohshima, S. Praveen, F. Rabenstein, D. Stenger, A. Wang, F. M. Zerbini, and Consortium Ictv Report. 2017. 'ICTV Virus Taxonomy Profile: Potyviridae', *J Gen Virol*, 98: 352-54.

Yeo, C., and M. Whitman. 2001. 'Nodal signals to Smads through Cripto-dependent and Cripto-independent mechanisms', *Mol Cell*, 7: 949-57.

Zhou, X., H. Sasaki, L. Lowe, B. L. Hogan, and M. R. Kuehn. 1993. 'Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation', *Nature*, 361: 543-7.

Figure 1: The Synpro system shows that prodomain cleavage can be non-cell autonomous



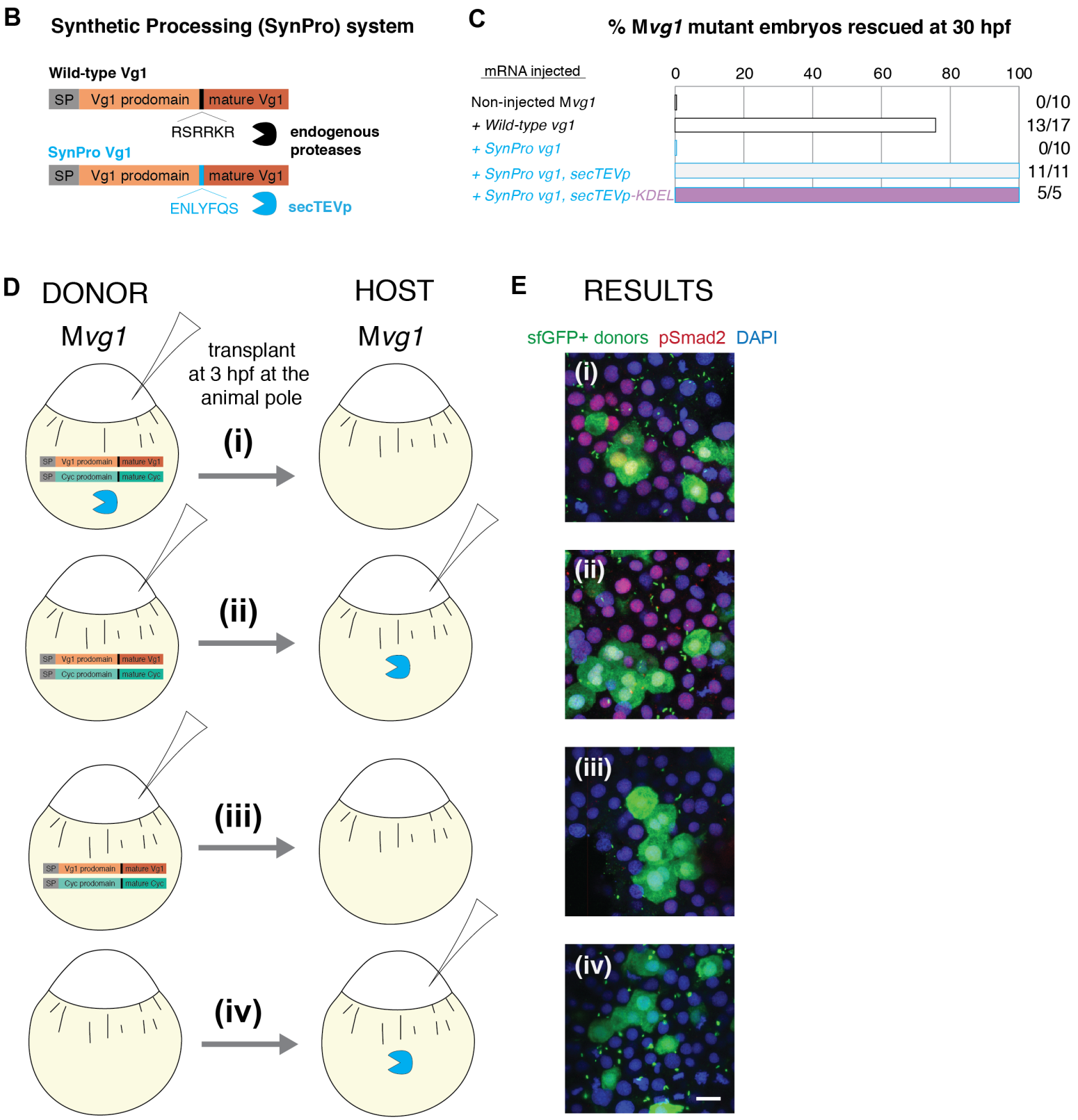

B
Synthetic Processing (SynPro) system
Wild-type Vg1
SP
Vg1 prodomain
mature Vg1
RSRRKR
endogenous proteases
SynPro Vg1
SP
Vg1 prodomain
mature Vg1
ENLYFQS
secTEVp
C
% Mvg1 mutant embryos rescued at 30 hpf
mRNA injected
0
20
40
60
80
100
Non-injected Mvg1
0/10
+ Wild-type vg1
13/17
+ SynPro vg1
0/10
+ SynPro vg1, secTEVp
11/11
+ SynPro vg1, secTEVp-KDEL
5/5
D
DONOR
Mvg1
transplant at 3 hpf at the animal pole
HOST
Mvg1
(i)
(ii)
(iii)
(iv)
SP
Vg1 prodomain
mature Vg1
SP
Cyc prodomain
mature Cyc
E
RESULTS
sfGFP+ donors
pSmad2
DAPI
(i)
(ii)
(iii)
(iv)

Figure 1—figure supplement 1: Developing the SynPro system

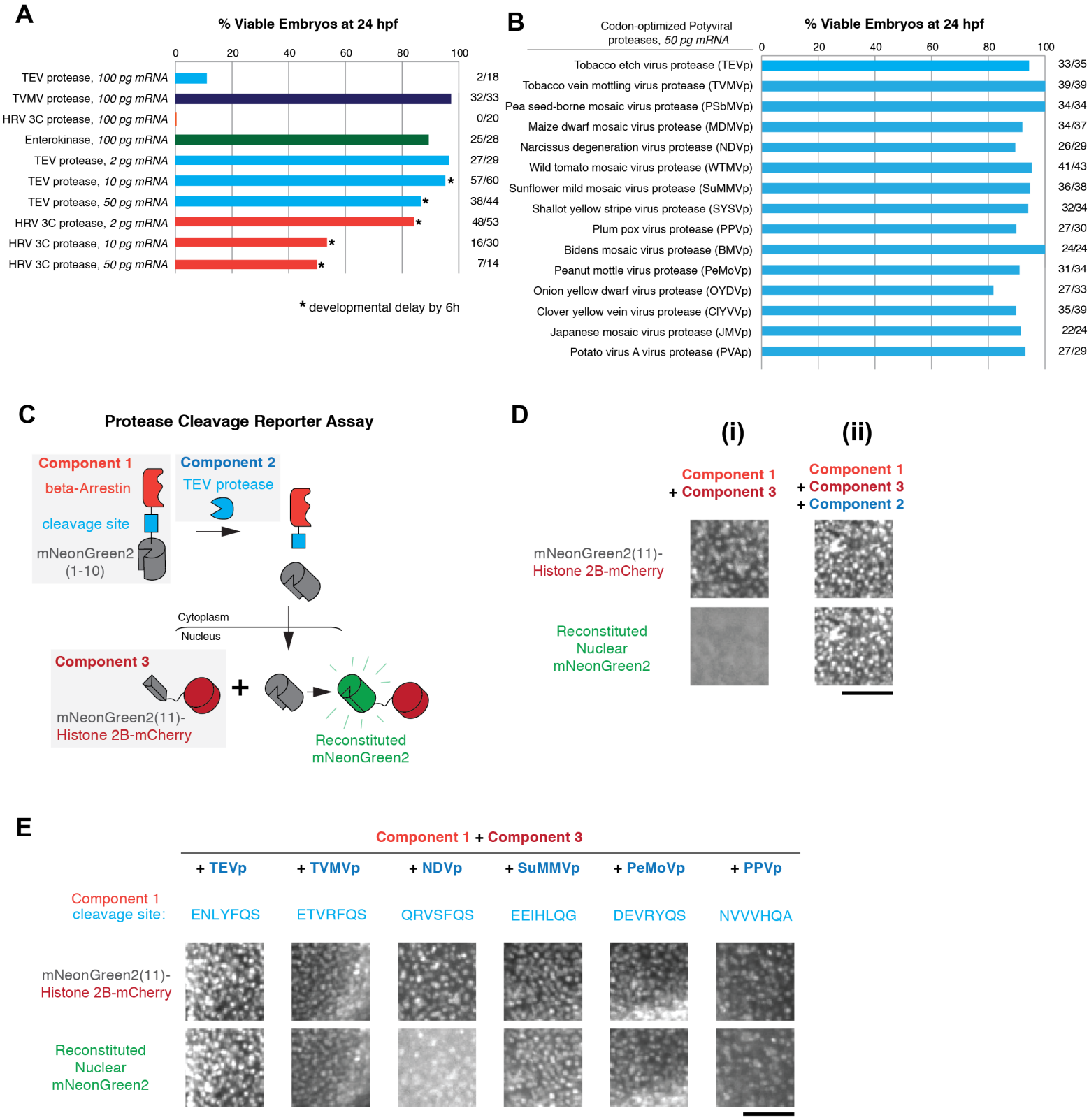


Figure 1—figure supplement 2: Vg1 prodomain processing does not promote secretion.

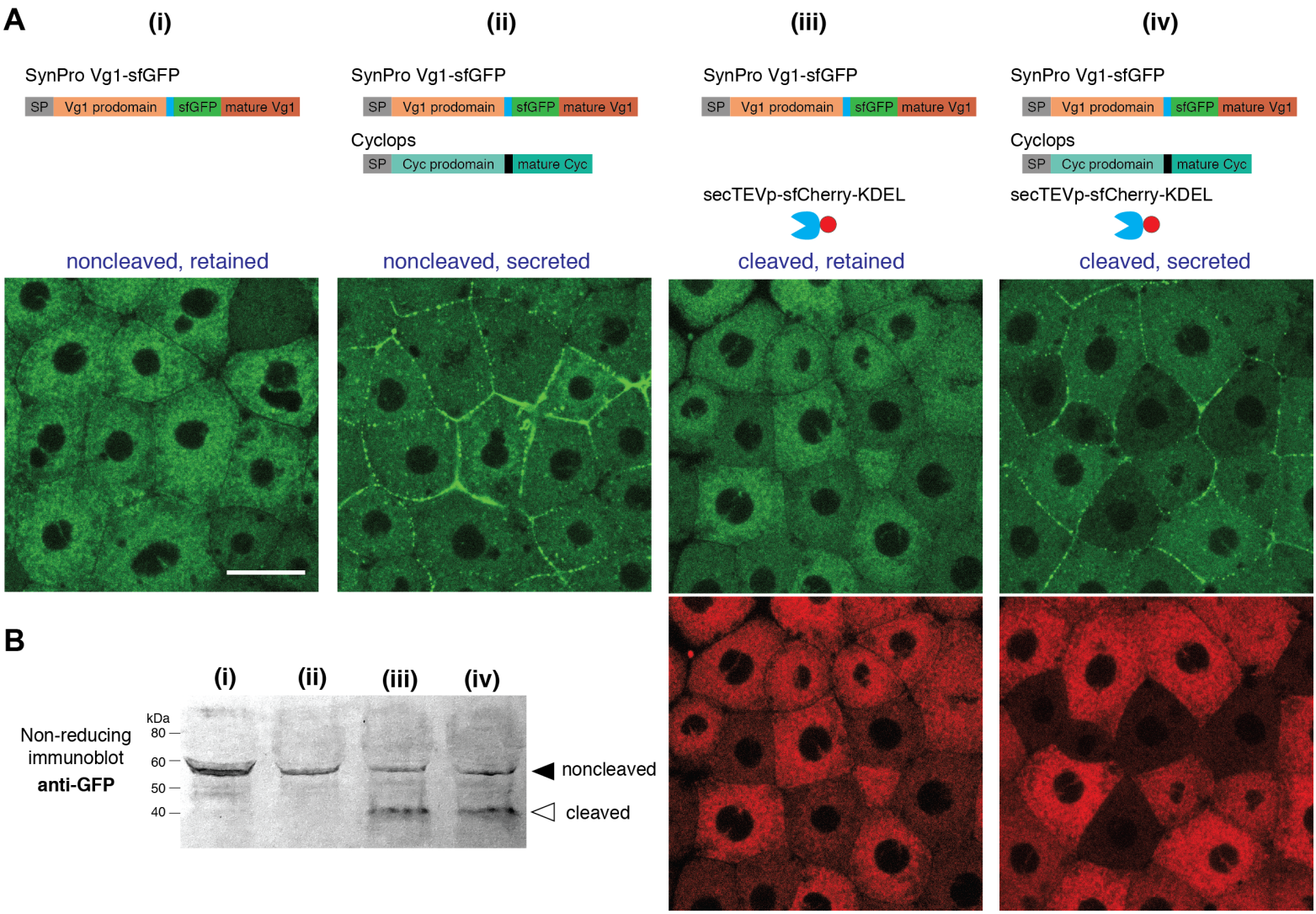


Figure 2: Prodomain cleavage affects Vg1-Nodal signaling but not secretion

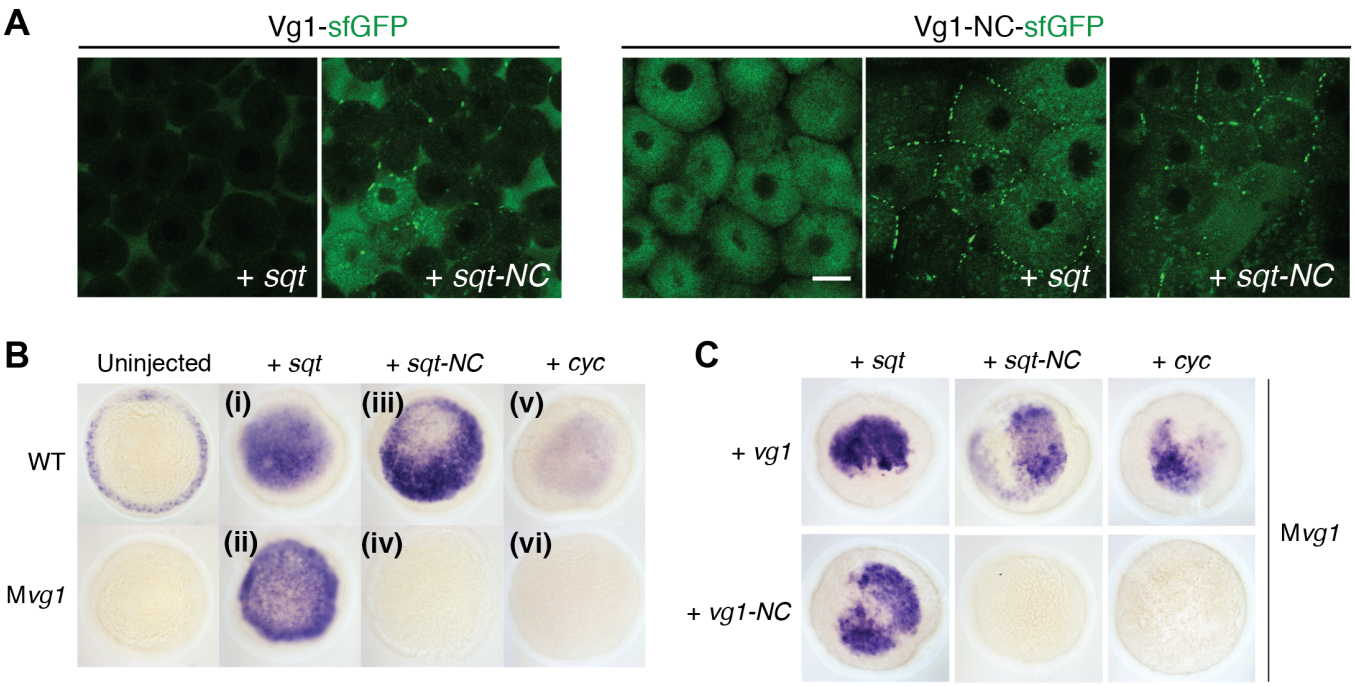


Figure 2—figure supplement 1: Non-cleavable Squint

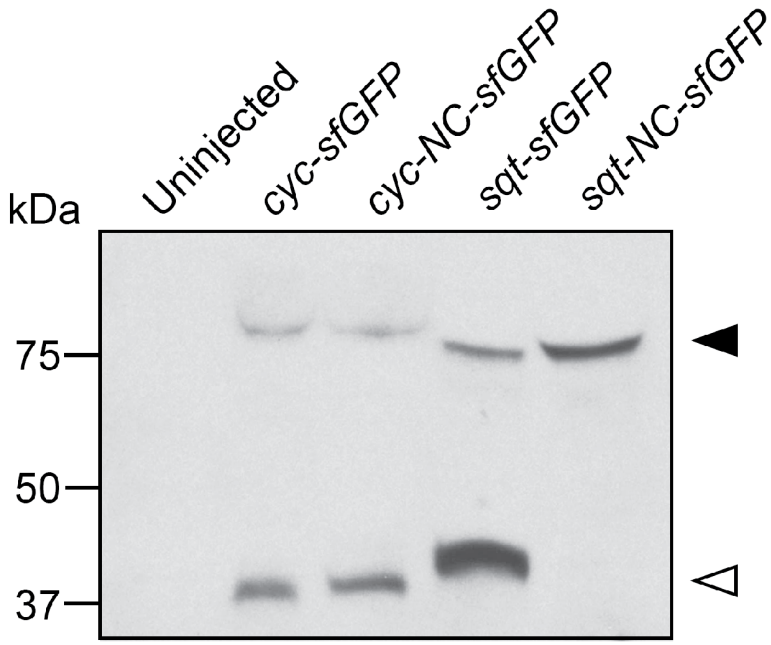
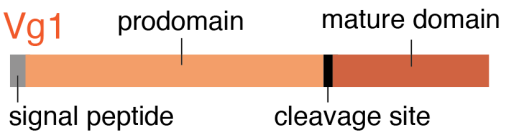


Figure 3: Cysteine and N-linked glycosylation sites retain the Vg1 prodomain in the ER

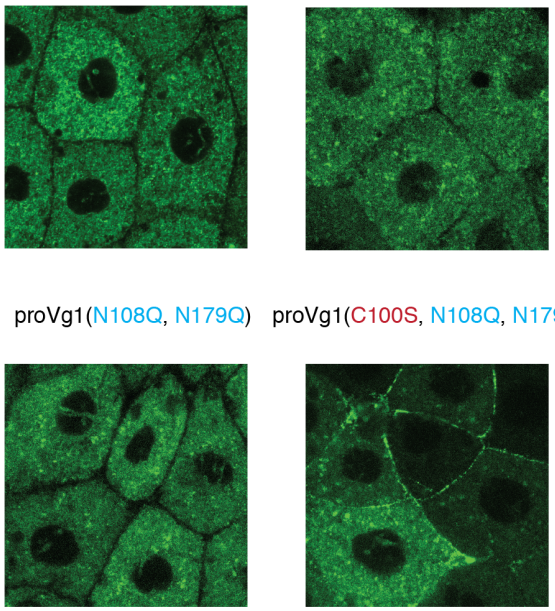
A



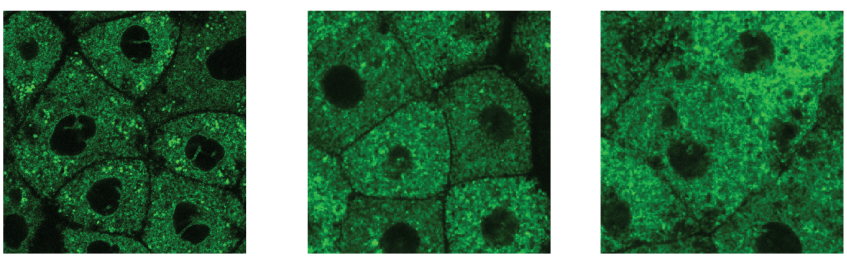

```
MFLVLLRACLLTLSLCSPAEDDGLVQEKLF   30
LSSMGLWSRPKPSHHAAVPSQMWKIFKQAS   60
KQTVNDPCVVSEYGVRGNIVRFMQDQGSLI   90
SAPAVHSFNCVRKHLFFNMSVLEEVEQLSL  120
AQLEMKFKQDLLLLGPHVFSVDLYRVLKTT  150
LKGVTHESSRKLLQSQTLSPGAHASVLVNL  180
TNLAQSWRKPEKNFGMQLELQVMHLNNMLH  210
DHAYVQIPDIHATLVVVSLNPLQCRSRRKR  240
SASYYLPVTPSNVCKPRRLYIDFKDVGWQD  270
WIIAPQGYLANYCHGECPFPLSESLNGTNH  300
AILQTLVHSFDPKGTPQPCCVPIKLSPISM  330
LYYDNNDNVVLRHYEDMVVDECGCR
```

sfGFP insertion site

B



C



D

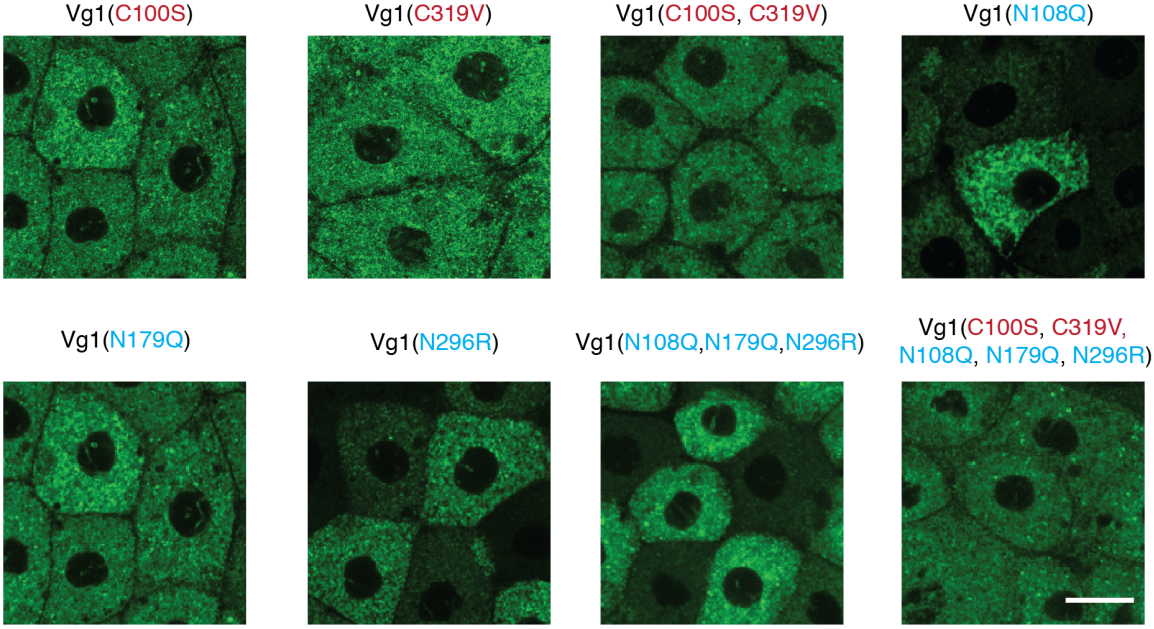


Figure 3—figure supplement 1: The Vg1 prodomain inhibits secretion

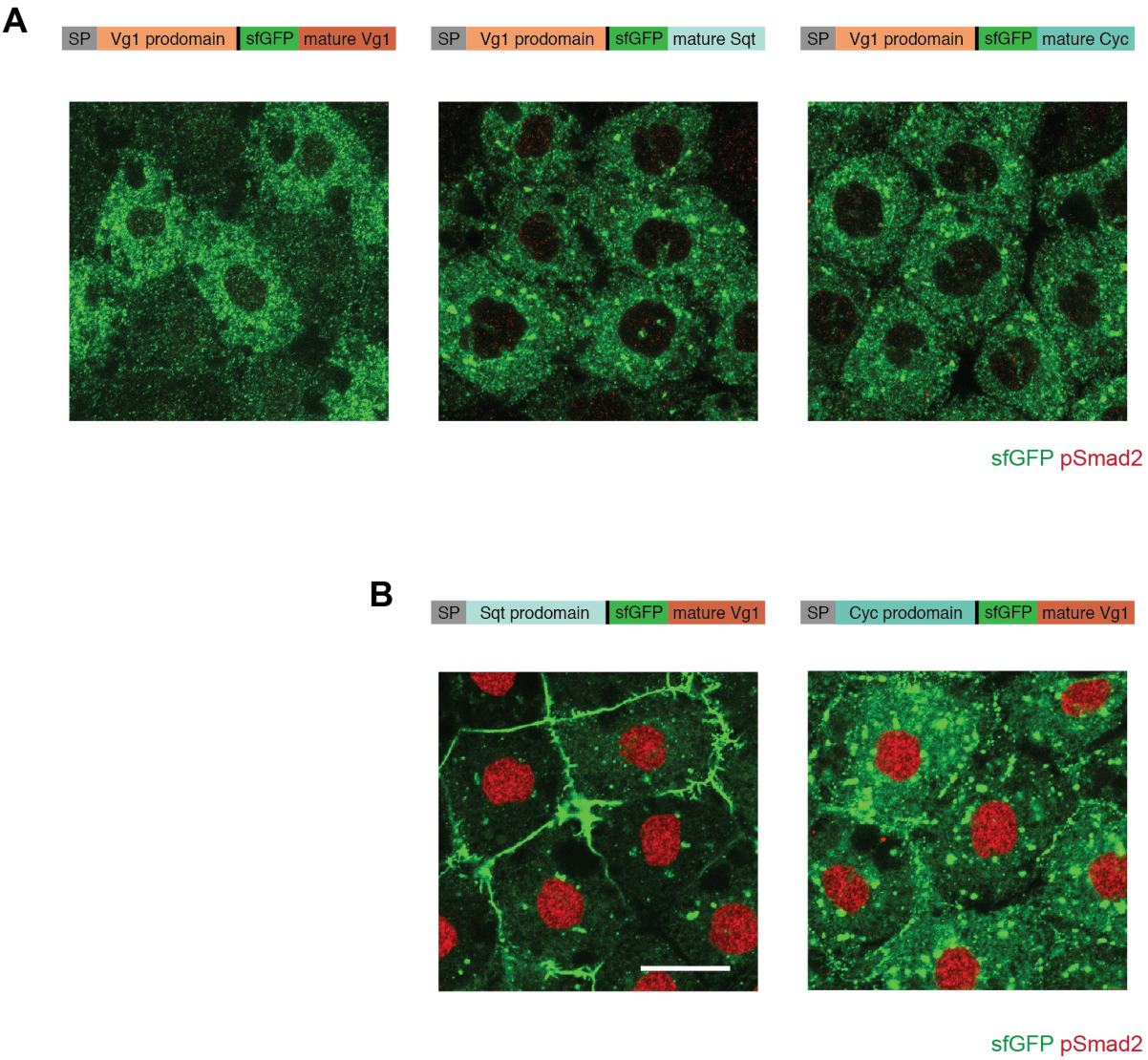


Figure 3—figure supplement 2: Vg1 mRNA Rescue of M*vg1* embryos

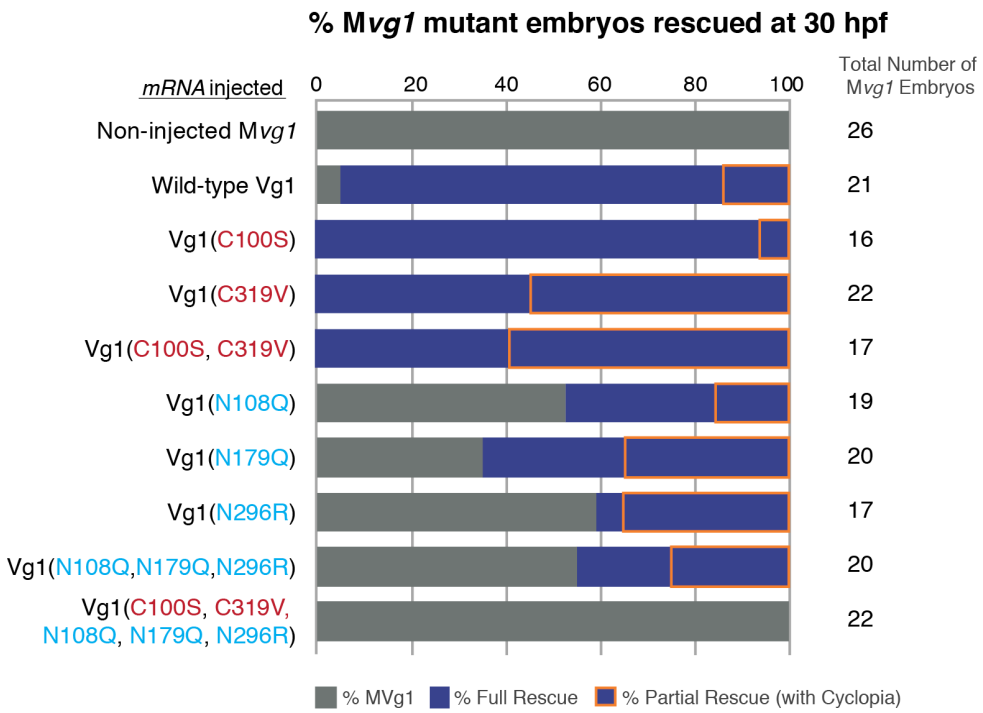


Figure 4: Binding motifs for BiP promote ER retention of the Vg1 mature domain

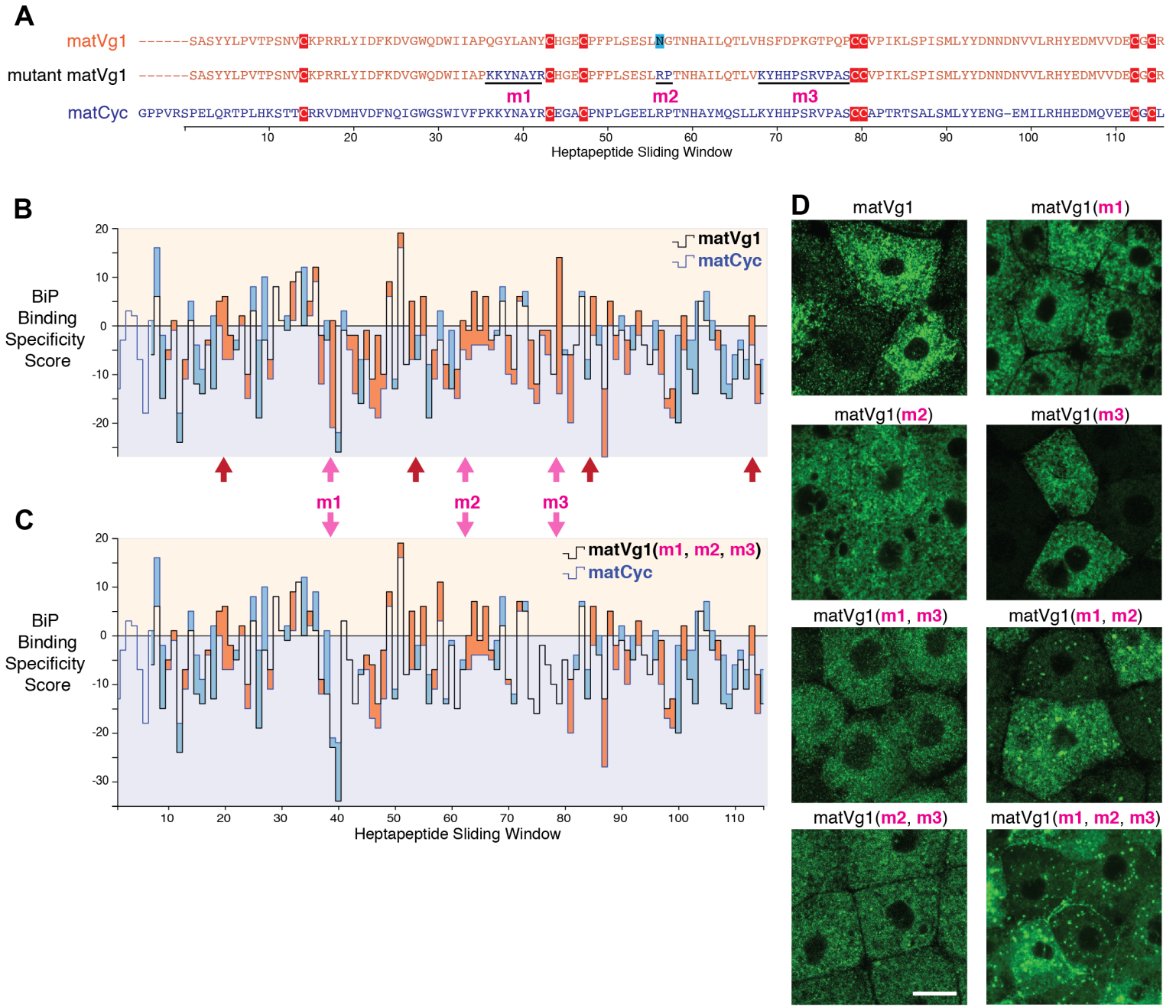


Figure 4—figure supplement 1: Localization and activity of Vg1 and Nodal mature domains

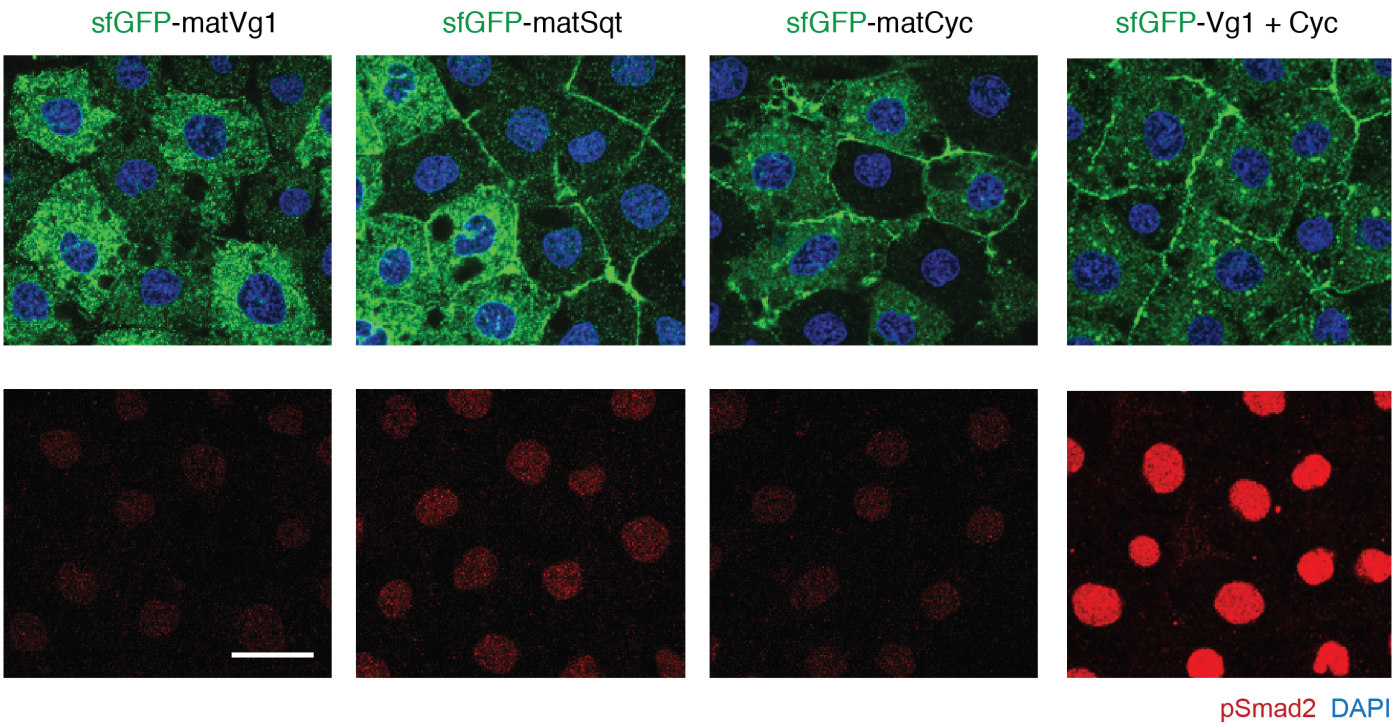


Figure 4—figure supplement 2: ER Retention of matCyc using BiP-binding features of matVg1

A

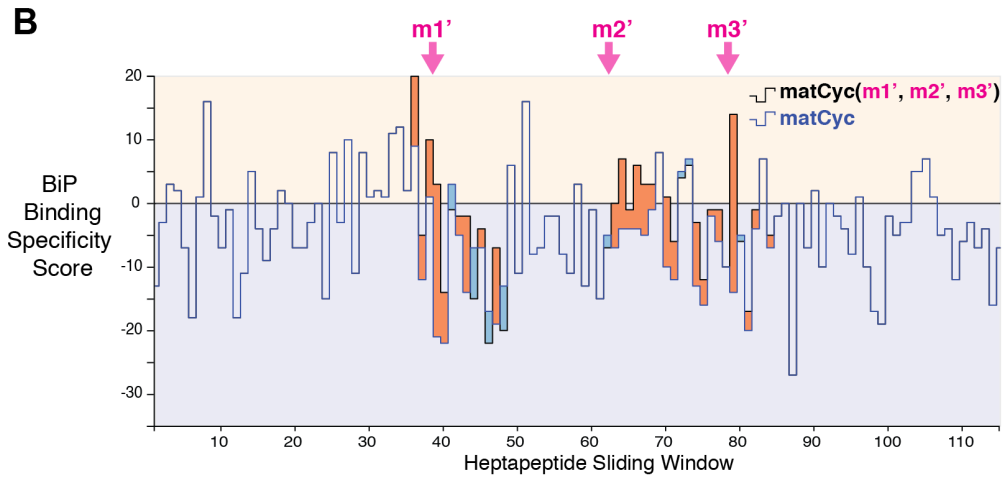
matVg1 ------SASYYLPVTPSNVCKPRRLYIDFKDVGWQDWIIAPQGYLANYCHGECPFPLSESLNGTNHAILQTLVHSFDPKGTPQPCCVPIKLSPISMLYYDNNDNVVLRHYEDMVVDECGCR

matCyc GPPVRSPELQRTPLHKSTTCRRVDMHVDFNQIGWGSWIVFPKKYNAYRCEGACPNPLGEELRPTNHAYMQSLLKYHHPSRVPASCCAPTRTSALSMLYYENG-EMILRHHEDMQVEECGCL

mutant matCyc GPPVRSPELQRTPLHKSTTCRRVDMHVDFNQIGWGSWIVFPQGYLANYCEGACPNPLGEELRPTNHAILQTLLHSFDPKGTPQPCCAPTRTSALSMLYYENG-EMILRHHEDMQVEECGCL

m1' m2' m3'

B



C

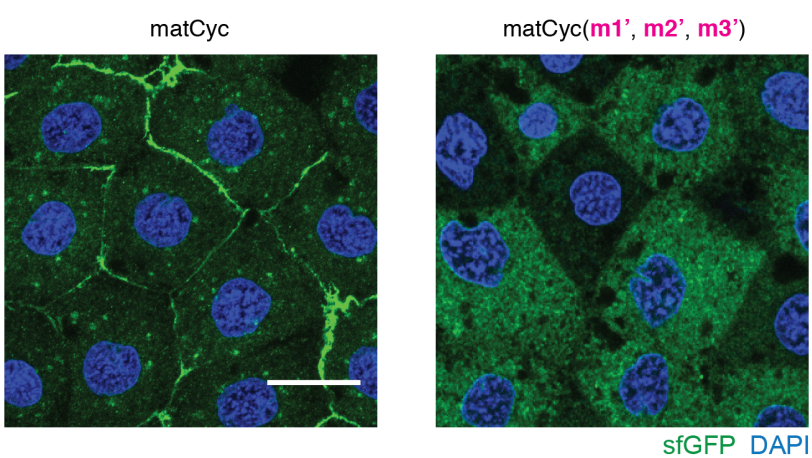
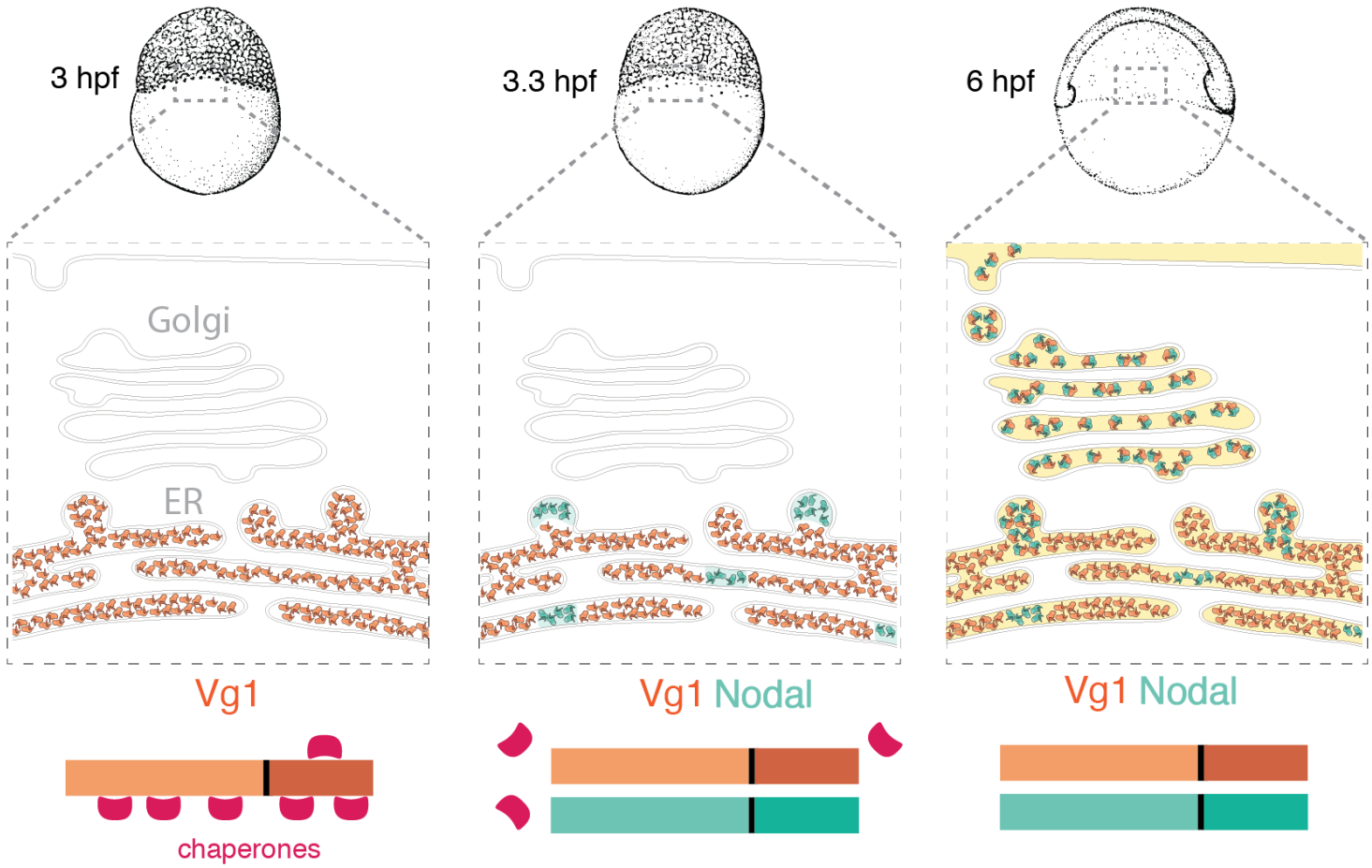


Figure 5: Model of Vg1-Nodal heterodimer formation and processing



Vg1 motifs promote retention in the ER

Nodal induces Vg1 release from retention

Prodomain processing is independent from dimer formation and secretion