Modelling the structure of Short Gastrulation and generation of a toolkit for studying its function in *Drosophila*

Sophie L. Frampton¹, Catherine Sutcliffe¹, Clair Baldock²*+ and Hilary L. Ashe¹*+

**ABSTRACT**

A BMP gradient is essential for patterning the dorsal-ventral axis of invertebrate and vertebrate embryos. The extracellular BMP binding protein Short Gastrulation (Sog) in *Drosophila* plays a key role in BMP gradient formation. In this study, we combine genome editing, structural and developmental approaches to study Sog function in *Drosophila*. We generate a sog knockout fly stock, which allows simple reintegration of altered versions of the sog coding sequence. As proof-of-principle, we test the requirement for two cysteine residues that were previously identified as targets for palmitoylation, which has been proposed to enhance Sog secretion. However, we show that the sog⁶⁷⁷,⁷⁸⁵ mutant is viable with only very mild phenotypes, indicating that these residues and their potential modification are not critical for Sog secretion in vivo. Additionally, we use experimental negative stain EM imaging and hydrodynamic data to validate the AlphaFold structure prediction for Sog. The model suggests a more compact shape than the vertebrate ortholog Chordin and conformational flexibility between the C-terminal von Willebrand C domains. We discuss how this altered compactness may contribute to mechanistic differences in Sog and Chordin function during BMP gradient formation.

This article has an associated First Person interview with the first author of the paper.

**KEY WORDS:** *Drosophila*, Short gastrulation, CRISPR-Cas9, AlphaFold, Model, Palmitoylation

**INTRODUCTION**

Bone morphogenetic proteins (BMPs) are a large class of highly conserved signalling molecules that belong to the TGF-beta superfamily. BMPs perform essential roles during animal development and adult tissue homeostasis, the significance of which is reflected in the variety of human diseases attributed to aberrant BMP activity (Bandyopadhyay et al., 2013; Wang et al., 2014). BMPs bind to their receptors resulting in the phosphorylation of a receptor regulated Smad, which then forms a complex with the common mediator Smad that accumulates in the nucleus to regulate target gene transcription (Schmierer and Hill, 2007). A key developmental role for BMPs is the patterning of the dorsal-ventral (DV) axis in early vertebrate and invertebrate embryos. BMP gradient formation is mediated by a conserved network of regulators, including two BMP binding proteins, Sog/Chordin and Twisted Gastrulation (Tsg), as well as a protease, Tolloid (Tld) (Madamanchi et al., 2021; Matsuda et al., 2016; Shilo et al., 2013).

The most potent BMP signalling molecule in the early *Drosophila* embryo is a heterodimer of the BMP ligands Decapentaplegic (Dpp) and Screw (Scw) (Shimmi et al., 2005b), which have uniform expression in the dorsal ectoderm (St Johnston and Gelbart, 1987; Arora et al., 1994; Shimmi et al., 2005b). *tsg* and *tld* are also expressed in the dorsal ectoderm, while *sog* is expressed ventro-laterally in the neuroectoderm (Francois et al., 1994; Marquès et al., 1997; Mason et al., 1994). During embryogenesis, reciprocal gradients of Dpp/Scw and Sog are established across the dorsal ectoderm (Ferguson and Anderson, 1992; Ashe and Levine, 1999; Srinivasan et al., 2002; Shimmi et al., 2005b; Wang and Ferguson, 2005). A narrow stripe of peak BMP signalling occurs along the dorsal midline and is flanked by lower signalling levels (Dorfman and Shilo, 2001; Rushlow et al., 2001; Sutherland et al., 2003; Shimmi et al., 2005b) thereby subdividing the dorsal ectoderm into amnioserosa and dorsal epidermis, respectively (Raffery and Sutherland, 2003).

A favoured model of BMP gradient formation requires the shuttling of BMP ligands dorsally in a multi-protein complex (Holley et al., 1996; Marquès et al., 1997; Eldar et al., 2002; Shimmi et al., 2005b; Umulis et al., 2006, 2009; Sawala et al., 2012). The model proposes that secreted Dpp/Scw binds to the extracellular matrix protein collagen IV (Col IV), which acts as a scaffold to promote formation of a Dpp/Scw-Sog-Tsg complex (Sawala et al., 2012; Wang et al., 2008). In this inhibitory complex, the Dpp/Scw ligand is unable to interact with its receptors but can diffuse dorsally (Ross et al., 2001; Eldar et al., 2002; Shimmi et al., 2005b; Sawala et al., 2012). Cleavage of Sog within this complex by Tld liberates Dpp/Scw, allowing the ligand to re-bind Col IV. In dorso-lateral regions, close to the sog expression domain, the Dpp/Scw-Sog-Tsg complex is reassembled, resulting in inhibition of signalling and diffusion of the complex towards the dorsal midline. At the dorsal midline and in the absence of Sog, however, Dpp/Scw is free to interact with receptors, resulting in a graded BMP signal across the dorsal ectoderm (Sawala et al., 2012; Wang et al., 2008; Winstanley et al., 2015).

Sog function is also important in *Drosophila* pupal wing vein patterning, including formation of the posterior crossvein (PCV), which depends on signalling by Dpp and Glass bottomed boat (Gbb) ligands (Serpe et al., 2005; Wharton et al., 1999; Yu et al., 1996). As in embryogenesis, Sog functions with a Tsg-like protein, Crossveinless (Cv), and a Tolloid-related (Tlr) metalloprotease to both locally inhibit BMP signalling and enhance it at a distance from the source in the pupal wing (Ralston and Blair, 2005; Serpe
et al., 2005; Vilmos et al., 2005; Shimmi et al., 2005a). In this model, Dpp/Gbb is transported from the longitudinal veins in a Dpp/Gbb-Sog-Cv complex to the presumptive PCV, where it is released from the inhibitory complex by Tlr-mediated Sog cleavage, enabling ligand-receptor interactions (Shimmi et al., 2005a; Serpe et al., 2005).

Sog and its vertebrate ortholog Chordin each contain four cysteine rich von Willebrand type C (vWC) domains which mediate protein interactions. These domains are 60–80 residues in length and have been identified in approximately 500 extracellular matrix proteins (Garcia Abreu et al., 2002; O’Leary et al., 2004; Zhang et al., 2007). Sog/Chordin vWC1 is separated from vWC2/3/4 domains by a ‘stem’ region comprising four Chordin specific (CHRD) domains (Francois et al., 1994). Structures have been solved for the human Procollagen IIA and the zebrafish Crossveinless-2 vWC1 domains (O’Leary et al., 2004; Xu et al., 2017; Zhang et al., 2008); however, there is currently no experimental structure for these domains in Sog, or the Sog/Chordin specific 4x CHRD ‘stem’ region.

Sog secretion is critical to its function, and a previous study has reported a potential role for palmitoylation during this process (Kang and Bier, 2010). Palmitoylation is a lipid modification that can influence protein interactions, membrane association, and trafficking between sub-cellular compartments (Bannan et al., 2008; Kang and Bier, 2010; Linder and Deschenes, 2003). The Drosophila palmitoyl-transferase Huntingtin-interacting protein 14 (dHIP14) was identified as an interacting partner of Sog, and Sog was shown to be palmitoylated in tissue culture cells (Giot et al., 2003; Kang and Bier, 2010). Mis-expression of dHIP14 in embryos and wings reduced BMP activity, similar to Sog overexpression phenotypes. In addition, mutation of cysteines 27 and 28 of Sog prevented the dHIP14-mediated increase in Sog secretion in tissue culture, suggesting that these two residues are the primary palmitoylation targets (Kang and Bier, 2010).

In this study, we generate a sog knockout (KO) ‘reintegration-ready’ fly stock that we use to test the effect of mutating the palmitoylation sites in vivo. Our data show that these residues are not critical for Sog function. In addition, we combine EM imaging

![Fig. 1. A CRISPR-Cas9 generated sog KO mutant.](image-url)
and AlphaFold prediction to model Sog structure, which reveals a curved compact shape. This Sog structure, along with the sog KO fly stock that we describe, will facilitate a complete molecular dissection of this critical extracellular BMP regulator.

RESULTS

Generation of a sog KO with CRISPR

To facilitate analysis of Sog in vivo, we used CRISPR genome editing to generate a sog KO line in which the translation start codon of the endogenous sog locus on the X chromosome was replaced with a phiC31 recombination landing site (Baena-Lopez et al., 2013) (Fig. 1A). Specifically, CRISPR-Cas9 mediated homology directed repair (HDR) was used to delete 800 bp including the ATG and signal sequence (Fig. 1Ai) and replace these with an attP recombination site. The resulting Drosophila stock facilitates simple insertion of modified sog sequences, such as point mutants (Fig. 2A), for expression under the endogenous sog promoter. In addition, regulatory elements located within sog introns, for instance the sog primary enhancer, remain intact (Markstein et al., 2002). The white gene was included in the HDR template and used as a marker to identify successful CRISPR events (Fig. 1Aii), before removal by Cre-Lox recombination (Fig. 1Aiii).

The resulting sogattP stock is maintained with an X-chromosomal balancer, and the absence of non-balancer males in the stock is consistent with a loss of Sog function. In addition, insertion of the attP recombination sequence at the sog locus was confirmed by sequencing. We used single molecule FISH (smFISH) to quantitate the amount of transcription from the sogattP locus. The smFISH probes detect both sog transcription and mRNAs in early nuclear cycle (nc) 14 wild-type and sogattP embryos (Fig. S1). To estimate sog mRNA number/cell, we assigned sog mRNAs to the closest nucleus as these embryos are only starting to cellularise. This

Fig. 2. Insertion of specific sog coding sequences into the endogenous sog locus. (A) Specific sog coding sequences fused to a C-terminal mNG tag were inserted into the genome by phiC31 recombination at the inserted attP landing site of the sogattP line. Wild-type Sog and a Cys27,28Ser mutant were reintegrated. (B) Embryos (nc14) of the indicated genotypes showing fluorescent RNA in situ hybridisation staining with sog and mNG probes (magenta and green, respectively). Scale bars: 10 μm. Expanded views of the areas outlined in the merged images are shown, scale bars: 5 μm. Colocalisation of sog (magenta) and mNG (green) transcription sites is indicated by white spots in the merged images. (C) sog smFISH (magenta) and mNG immunostaining (green) of the indicated nc14 embryos. A merged image with DAPI staining (blue) is shown, with single images of the sog and mNG channels for clarity. Scale bars: 50 μm.
analysis reveals that there is a ~2.5-fold reduction in the peak numbers of sog mRNAs/cell in the sog<sup>attP</sup> embryos (Fig. S1). The presence of sog mRNAs is consistent with the deletion removing sequences downstream of the sog promoter, although the reduced mRNA number suggests that there is an effect on transcription and/or mRNA stability, potentially due to nonsense-mediated decay.

A scan for cryptic start codons at the modified sog locus identified one large and several smaller open reading frames (ORFs) (Fig. S2). Although the large ORF encodes sog sequences starting within the first vWC domain, the signal sequence is absent. No cryptic signal sequence in this truncated Sog ORF was predicted using various software tools, e.g. Phobius webserver (Käll et al., 2004; 2007), that were able to predict endogenous Sog’s signal sequence/transmembrane domain (data not shown). As Sog is a secreted protein (Francois et al., 1994), no Sog function is predicted following the deletion and attP insertion made in the sog locus.

To confirm loss of Sog function in sog<sup>attP</sup> embryos, RNA fluorescence in situ hybridisation (FISH) was used to visualise expression of the peak BMP target gene Race (Fig. 1B). We also probed for lacZ mRNAs as fzt-lacZ is present on the balancer chromosome. The absence of lacZ expression indicates that the embryos are sog<sup>attP</sup> males. In these sog<sup>attP</sup> embryos Race expression is expanded in the anterior and lost in the presumptive amnioserosa (Fig. 1B), consistent with that described in sog mutant embryos (Ashe and Levine, 1999). These data, and the lethality of sog<sup>attP</sup> males, support successful removal of the sog translation start site and its replacement with an attP recombination sequence to generate a sog<sup>attP</sup> KO allele.

**Reintegration of transgenes at the sog locus**

The presence of an attP landing site at the sog locus facilitates targeted insertion of specific sog coding sequences into the genome (Fig. 2A). A previous study proposed that palmitoylation at two cysteines, at positions 27 and 28, is important for Sog secretion and stability of a membrane bound form of Sog (Kang and Bier, 2010). To test how the disruption of palmitoylation affects Sog function in vivo we used the sog<sup>attP</sup> line we generated to integrate a sog cDNA in which Cys27 and 28 are mutated to Ser. Wild-type and palmitoylation mutant versions of the sog cDNA, to which a C-terminal mNeonGreen (mNG) tag was added (referred to as sog-mNG and sog<sup>C27,28S-mNG</sup>, respectively), were integrated into the endogenous locus (Fig. 2A). In total, ~12.6 kb of DNA was inserted at the sog locus, including the sog CDS, white<sup>+</sup> marker, and LoxP sites (Fig. S3). Although endogenous sog sequences remain downstream of the integration site, cryptic initiation within the reintegration sequences and readthrough is not predicted to result in a Sog ORF longer than the truncated one described above. Therefore, if transcription of the remaining endogenous sog locus occurs due to cryptic initiation, the mRNA is only predicted to encode a truncated Sog ORF lacking a signal sequence (Fig. S2). This truncated Sog lacks activity based on the phenotype of the sog<sup>attP</sup> embryos and lethality of the sog<sup>attP</sup> males, as described above.

Both male and female flies carrying only the reintegrated sog-mNG and sog<sup>C27,28S-mNG</sup> sequences are viable (see later). Transcription of the integrated sog sequences in lateral stripes in the embryos was confirmed by FISH using sog and mNG probes (Fig. 2B). In this experiment, the control embryos carry an unmodified sog locus, so a signal is only detected with the sog probe (Fig. 2B). However, both the sog and mNG probes detect co-localised signals in the sog-mNG and sog<sup>C27,28S-mNG</sup> embryos, as expected for transcription of the reintegrated sequences. We next used smFISH to test for any differences in sog expression between sog-mNG and sog<sup>C27,28S,mNG</sup> early stage 6 embryos compared to wild-type (Fig. S4A). We found no significant difference in the number of sog expressing cells between embryos of these genotypes (Fig. S4B). We were unable to quantitate absolute sog mRNA numbers at this stage due to their clustering. However, quantitation based on fluorescence intensity in equivalent areas of the expression domain suggests that there are no significant differences in sog expression levels in the reintegration embryos (Fig. S4C). The same result was obtained by analysing the fluorescence intensity along a line through the whole expression domain (data not shown). Finally, mNG immunostaining in combination with sog smFISH showed accumulation of Sog-mNG and Sog<sup>C27,28S-mNG</sup> protein (Fig. 2C).

**Cysteines 27 and 28, putative palmitoylation targets, are not essential for Sog function**

As a fly stock homozygous for sog<sup>C27,28S-mNG</sup> was successfully established and maintained, cysteines 27 and 28 are not essential for Sog function. Due to difficulties associated with detecting palmitoylation in vivo we were unable to directly compare palmitoylation levels of wild-type and the mutant Sog. However, as the sog<sup>C27,28S</sup> mutant was less able to inhibit BMP activity in a tissue culture assay (Kang and Bier, 2010), we investigated whether these mutations reduce viability in vivo. To test this, the survival of embryos to pupal and adult stages was quantified (Fig. S5A). sog-mNG or sog<sup>C27,28S-mNG</sup> embryos were raised at 25°C and the number of pupae and eclosed adults counted. The proportion of pupae and adults show some lethality at each of these stages for both the sog-mNG or sog<sup>C27,28S-mNG</sup> lines. This could be due to the presence of the mNG tag or differences in the reintegration locus compared to wild-type (see Discussion). Despite a trend towards lower survival rates for the sog<sup>C27,28S-mNG</sup> allele, there is no significant difference between the number of sog<sup>C27,28S-mNG</sup> and sog-mNG embryos that developed into pupae and successfully eclosed as adults.

To further test the functionality of the sog-mNG and sog<sup>C27,28S-mNG</sup> sequences, the extent to which these alleles can rescue the sog<sup>6</sup> loss-of-function allele or the sog<sup>attP</sup> KO allele was assayed. sog-mNG and sog<sup>C27,28S-mNG</sup> males were crossed to sog<sup>6</sup>/FM7c or sog<sup>attP</sup>/FM7c females (Fig. S5B), and the numbers of female offspring with either the sog-mNG or sog<sup>C27,28S-mNG</sup> allele versus the FM7c balancer were counted. No significant difference in the ability of the sog-mNG or sog<sup>C27,28S</sup>-mNG alleles to rescue either sog mutant allele relative to wild-type was observed (Fig. S5Bii). Although the different viability assays appear to have different sensitivities (Fig. S5A,B), together the data are consistent with the sog<sup>C27,28S</sup> mutations having only a very minor effect, if any, on Sog function.

**BMP signalling readouts in sog-mNG and sog<sup>C27,28S-mNG</sup> embryos**

Although sog-mNG and sog<sup>C27,28S-mNG</sup> flies are viable, we investigated whether there are minor effects on Dpp gradient formation and interpretation. Sog functions in the early Drosophila embryo to concentrate BMP ligands dorsally, resulting in a stripe of the activated pMad transducer at the dorsal midline (Montanari et al., 2022). Therefore, pMad distribution was visualised in early stage 6 embryos by immunostaining and the width of the pMad stripe was measured at 50% embryo length.
The pMad stripes in sog-mNG and sogC27,28S-mNG embryos are generally broader than those in wild-type control embryos; however, these differences are not significant (Fig. 3B). Peak BMP/pMad signalling specifies amnioserosa cell fate. Therefore, to test whether subtle differences in pMad stripe width affected amnioserosa specification, embryos were stained for the amnioserosa cell marker Hindsight (Hnt, Fig. 3C). Both sog-mNG and sogC27,28S-mNG embryos have a small but significant reduction in the number of amnioserosa cells compared to wild-type embryos. However, the sog-mNG and sogC27,28S-mNG embryos have a similar reduction in the number of amnioserosa cells, suggesting that the C27,28S mutations do not affect embryonic BMP-dependent cell fate decisions. Together, these data suggest that BMP signal reception is marginally affected in both sog-mNG and sogC27,28S-mNG embryos, however this level of disruption is tolerated during development.

Quantitative analysis of BMP target gene expression

As the number of amnioserosa cells was reduced in sog-mNG and sogC27,28S-mNG embryos, we used smFISH and quantitative analysis to assess effects on transcription of BMP target genes. smFISH was performed for the BMP target genes Race and u-shaped (ush), which respond to peak and intermediate levels of BMP signalling, respectively (Fig. 4A,B) (Ashe et al., 2000). Both sog-mNG and sogC27,28S-mNG embryos show similar ush expression patterns to wild-type (unedited) embryos, indicating that there is a BMP gradient, consistent with Sog function (Ashe et al., 2000). However, while the number of mature ush mRNAs in sog-mNG embryos is similar to that in wild-type embryos, sogC27,28S-mNG embryos have, on average, around half the number of ush mRNAs (Fig. 4C; Fig. S6). Race expression levels in sog-mNG embryos are, in general, lower than in controls, but Race is restricted to the dorsal midline as in control embryos (Fig. 4D; Fig. S6). The mean number of Race mRNAs across the three biological repeat embryos is slightly lower in sog-mNG relative to control embryos, whereas there is an even greater reduction in Race expression in sogC27,28S-mNG embryos (Fig. 4D). In addition, the levels of Race expression observed in sogC27,28S-mNG embryos show more variation: while one sogC27,28S-mNG embryo has a weak stripe of Race expression along the dorsal midline (Fig. 4Di), it is almost absent in the others (Fig. S6). As Race expression in sogC27,28S-mNG embryos is weaker than in sog-mNG embryos, this suggests that sogC27,28S-mNG may be less able to promote peak BMP signalling. Overall, this highly sensitive assay of BMP target gene transcription identifies subtle deficiencies in the responses, particularly with sogC27,28S-mNG, even though these do not have major effects on viability.

The sogC27,28S mutant shows weakly penetrant PCV patterning defects

Sog also regulates BMP signalling during pupal wing vein patterning, including PCV patterning. Therefore, we used this as
an alternative developmental context to test whether the requirement for palmitoylation of Sog may be context dependent. The wings of adult female flies, raised at either 18°C or 25°C, were examined for defects in PCV specification and patterning (Fig. 5). A low proportion of sog$^{C27,S-mNG}$ wings displayed a mutant PCV phenotype: a small extension to the distal side of the PCV (Fig. 5A). Ectopic PCV development was observed in a slightly higher proportion of flies that developed at 18°C compared to 25°C, suggesting that the phenotype is exacerbated by mild cold temperature stress (Fig. 5B,C). Disruption to the PCV only in sog$^{C27,S-mNG}$ wings at 18°C suggests that mutation of cysteines 27 and 28 has mildly impacted Sog function or levels, resulting in reduced BMP signal refinement during PCV patterning (Antson et al., 2022) (see Discussion).
Sog has a curved shape

The data presented above demonstrate the utility of our \textit{sog}\textsuperscript{mNG} line for testing and elucidating the effect of specific \textit{sog} mutations. One limitation for targeted mutagenesis of \textit{sog}, however, is the absence of structural information. Therefore, to investigate Sog structure, we purified Sog (with C-terminal His and V5 tags) from the conditioned media of a stable, Sog-expressing, HEK293 EBNA cell line by affinity purification followed by two rounds of size exclusion chromatography (SEC) (Fig. 6A; Fig. S7). Sog purity was assessed by SDS-PAGE and western blot analysis where, after an initial round of SEC, a prominent doublet band was typically observed (Fig. S7C). This doublet band is likely to represent full-length Sog and Sog lacking the N terminal vWC1 domain due to Tld cleavage, as the cleavage product is detected by a His antibody (Fig. S7C) and expression of Chordin in the same cell line results in co-purification of a Tolloid cleavage product lacking the vWC1 domain (Troilo et al., 2014). A second round of SEC was included but could not completely separate the lower molecular weight species (Fig. S7C). Negative stain transmission electron microscopy (TEM) was used to investigate the three-dimensional (3D) structure of the purified Sog protein. During single particle analysis, two-dimensional (2D) classification aligned Sog particle images and the purified Sog (with C-terminal His and V5 tags) from the conditioned media of a stable, Sog-expressing, HEK293 EBNA cell line by affinity purification followed by two rounds of size exclusion chromatography (SEC) (Fig. 6A; Fig. S7). Sog purity was assessed by SDS-PAGE and western blot analysis where, after an initial round of SEC, a prominent doublet band was typically observed (Fig. S7C). This doublet band is likely to represent full-length Sog and Sog lacking the N terminal vWC1 domain due to Tld cleavage, as the cleavage product is detected by a His antibody (Fig. S7C) and expression of Chordin in the same cell line results in co-purification of a Tolloid cleavage product lacking the vWC1 domain (Troilo et al., 2014). A second round of SEC was included but could not completely separate the lower molecular weight species (Fig. S7C). Negative stain transmission electron microscopy (TEM) was used to investigate the three-dimensional (3D) structure of the purified Sog protein. During single particle analysis, two-dimensional (2D) classification aligned Sog particle images and produced 2D class averages that were used to generate and refine a Sog 3D model (Fig. 6C,D). The final 3D reconstruction, with an estimated resolution of 22.8 Å (Fig. 6D,E), reveals that Sog has an asymmetric, curved shape and dimensions of 13.6 nm×9.9 nm×9.2 nm (Fig. 6D). Due to the similarity in size, the Sog cleavage product cannot be separated from the full-length protein during image analysis, therefore the cleavage product will also have a contribution to the 3D model.

Comparing AlphaFold predictions to experimental data

Next, we probed the AlphaFold protein structure database, a recently developed resource based on the machine learning prediction of protein structures to atomic resolution (Jumper et al., 2021; Varadi et al., 2022), to analyse the predicted Sog structure and investigate how it fits within the Sog 3D reconstruction described above. For reference, the domain organisation of Sog is shown in Fig. 7A. Given the presence of the hydrophobic N-terminally located Sog transmembrane domain/signal peptide, it is likely that much of the Sog N-terminus is cleaved prior to secretion into the perivitelline space. The Sog N-terminus has therefore been removed from the AlphaFold model, at a site between R79 and H80 previously identified as a putative cleavage target for separation of a hydrophobic N-terminal signal sequence/transmembrane domain from mature extracellular Sog (Shimmi and O’Connor, 2003). AlphaFold generates a pLDDT score, which is a useful metric from which to infer confidence in the local structural information. Therefore, to investigate Sog structure, we included for testing and elucidating the effect of specific \textit{sog} mutations. One limitation for targeted mutagenesis of \textit{sog}, however, is the absence of structural information. Therefore, to investigate Sog structure, we purified Sog (with C-terminal His and V5 tags) from the conditioned media of a stable, Sog-expressing, HEK293 EBNA cell line by affinity purification followed by two rounds of size exclusion chromatography (SEC) (Fig. 6A; Fig. S7). Sog purity was assessed by SDS-PAGE and western blot analysis where, after an initial round of SEC, a prominent doublet band was typically observed (Fig. S7C). This doublet band is likely to represent full-length Sog and Sog lacking the N terminal vWC1 domain due to Tld cleavage, as the cleavage product is detected by a His antibody (Fig. S7C) and expression of Chordin in the same cell line results in co-purification of a Tolloid cleavage product lacking the vWC1 domain (Troilo et al., 2014). A second round of SEC was included but could not completely separate the lower molecular weight species (Fig. S7C). Negative stain transmission electron microscopy (TEM) was used to investigate the three-dimensional (3D) structure of the purified Sog protein. During single particle analysis, two-dimensional (2D) classification aligned Sog particle images and produced 2D class averages that were used to generate and refine a Sog 3D model (Fig. 6C,D). The final 3D reconstruction, with an estimated resolution of 22.8 Å (Fig. 6D,E), reveals that Sog has an asymmetric, curved shape and dimensions of 13.6 nm×9.9 nm×9.2 nm (Fig. 6D). Due to the similarity in size, the Sog cleavage product cannot be separated from the full-length protein during image analysis, therefore the cleavage product will also have a contribution to the 3D model.

Comparing AlphaFold predictions to experimental data

Next, we probed the AlphaFold protein structure database, a recently developed resource based on the machine learning prediction of protein structures to atomic resolution (Jumper et al., 2021; Varadi et al., 2022), to analyse the predicted Sog structure and investigate how it fits within the Sog 3D reconstruction described above. For reference, the domain organisation of Sog is shown in Fig. 7A. Given the presence of the hydrophobic N-terminally located Sog transmembrane domain/signal peptide, it is likely that much of the Sog N-terminus is cleaved prior to secretion into the perivitelline space. The Sog N-terminus has therefore been removed from the AlphaFold model, at a site between R79 and H80 previously identified as a putative cleavage target for separation of a hydrophobic N-terminal signal sequence/transmembrane domain from mature extracellular Sog (Shimmi and O’Connor, 2003). AlphaFold generates a pLDDT score, which is a useful metric from which to infer confidence in the local predicted protein structure. AlphaFold predicts the folds of each Sog vWC domain, with the majority of residues predicted with a high pLDDT score of 70–90 (Fig. 7B). Furthermore, AlphaFold predicts a novel fold for each of the four CHRD specific regions. (Fig. 7A) with high confidence (pLDDT >90 and 70–90) (Fig. 7B). Residues linking vWC domains to other vWC or CHRD domains, however, are predicted with low pLDDT scores (50–70 and <50), indicating potentially disordered and/or flexible regions.

A predicted aligned error (PAE) score is also calculated for each residue pair by AlphaFold (Fig. 7C) (Jumper et al., 2021; Varadi et al., 2022). This score is a measure of the confidence with which the positions of amino acid pairs are predicted, thereby indicating the confidence of relative domain positions. The relative positions of each CHRD domain, vWC1 and the 4x CHRD region, and vWC2 and vWC3 are predicted with high confidence. Other inter-domain distances, for instance between vWC3 and vWC4, are predicted with lower confidence, suggesting some flexibility in the full-length protein. The Tld cleavage sites in both Sog and its ortholog Chordin
are located within these flexible interdomain regions (Fig. 7A; Fig. S8) (see Discussion).

The asymmetric nature of the AlphaFold model, with the bulky 4x CHRD region on one side (Fig. 7B), resembles the Sog 3D negative stain EM model (Fig. 6D). Indeed, overlay of the AlphaFold model with the Sog 3D EM volume, using ChimeraX (Pettersen et al., 2021), demonstrates how the 4x CHRD region might sit within the larger arm of the Sog EM model, as well as the arrangement of the other domains within the density (Fig. 7D). For further insight into the hydrodynamic properties of Sog, purified Sog protein was subject to sedimentation velocity analytical ultracentrifugation (AUC) (Fig. 7E). cf(S) analysis calculated a sedimentation coefficient of 5.23 (S20,W), and a frictional ratio of 1.75, reflecting the relatively large size of Sog and indicating an elongated shape (Fig. 7F). To further test the level of agreement between the Sog AlphaFold structure and purified Sog, the sedimentation coefficient and frictional ratio of the N-terminally cleaved Sog AlphaFold model was predicted with US-SOMO (Brookes and Rocco, 2018; Rai et al., 2005; Brookes, et al., 2010a, b). A sedimentation coefficient of 5.14 (S20,W) and frictional ratio of 1.59 were predicted for the Sog AlphaFold model (Fig. 7F), lending further confidence to the accuracy of this predicted atomic structure. Together, similarities in shape between the AlphaFold Sog prediction and the Sog EM model, and between experimentally and theoretically derived hydrodynamic parameter values, suggest an overall domain organisation that provides a framework for future studies.

**DISCUSSION**

In this study, we have generated a sog KO line that allows simple reintegration of altered sog cDNAs. While the Sog 3D model that we have described will allow targeted mutagenesis in future studies, here we showed the utility of our sogflox KO stock by using it to investigate the effect of mutating two residues implicated in Sog palmitoylation. Palmitoylation at cysteines 27 and 28 has previously been suggested to play an important role in membrane targeting and Sog secretion (Kang and Bier, 2010). However, our data show that mutation of these two residues to serine in vivo resulted in viable flies, consistent with these residues not being essential for Sog function.

Our data reveal that the wild-type sog-mNG reintegration embryos accumulate less Race mRNAs and have a minor reduction in the number of amnioserosa cells compared to wild-type embryos. As Race expression and amnioserosa fate are dependent on peak signalling, we speculate that there is a subtle defect in BMP gradient formation arising from slightly reduced shuttling of BMP heterodimers to the dorsal midline in the sog-mNG embryos. Although no significant difference in the width of the pMad stripe in the reintegration embryos was observed, the reduced Race mRNA numbers suggest that there is a minor defect in pMad levels. Co-staining of wild-type and sog-mNG or sogC27,28S-mNG embryos with the pMad antibody, along with sophisticated quantitation of the staining intensities (Gavin-Smyth et al., 2013; Umulis et al., 2010) will allow this to be addressed in future studies.
There are various potential explanations for the mildly reduced peak outputs in the sog-mNG reintegration embryos. Fusion of the C-terminal mNeonGreen tag to Sog may have slightly reduced its function. It is also possible that we have perturbed the timing of expression as we reintegrated the sog cDNA. As the cDNA is much shorter than the sequences present in the endogenous locus that
contains long introns, there is potentially more rapid accumulation of sog mRNAs, which may be particularly important in the short nc13 (Sandler et al., 2018). However, we did not detect any global effects on sog expression levels relative to wild-type embryos at the end of nc14, which may be expected if there is earlier accumulation of full length sog mRNAs. Introduction of the sog cDNA in the reintegration embryos also prevents expression of a truncated sog transcript that includes intron-derived sequence. The encoded short Sog protein has been shown to suppress early BMP signalling and prevent ectopic target gene expression during nc13 (Sandler et al., 2018). However, as we observe a reduction in Race mRNAs at nc14, this phenotype does not seem compatible with a loss of Short Sog expression. Reintegration also introduces additional sequences including an attR scar in the 5′UTR, which could affect translational regulation of the sog mRNA. In addition, the endogenous sog locus sequences (starting at the end of the CRISPR deletion) are present downstream of the reintegration sequences. While we find no evidence that the remnants of the endogenous locus can encode a secreted functional Sog protein, reintegration plasmid sequences can be removed in future studies by Cre recombination.

Using the sog-mNG flies as controls, we identified two minor phenotypes associated with sogC27,28S-mNG embryos and wings. There is a reduction in both Race and ush expression in the sogC27,28S-mNG embryos, suggesting that there are slightly lower levels of the pMad activator. We speculate that this is due to a further minor reduction in BMP shuttling by the mutant Sog, due to a slight decrease in either its levels or activity, resulting in a lower concentration of pMad at the dorsal midline. Although sogC27,28S-mNG embryos have ∼4-fold reduction in Race mRNA numbers relative to the sog-mNG embryos, there is no difference in the number of amnioserosa cells specified. As Race expression is exquisitely sensitive to reductions in peak Dpp/Swc signalling (Ashe and Levine, 1999; Rusch and Levine, 1997), it is possible that other peak BMP targets show a more modest reduction in mRNA numbers. Consistent with this, the intermediate target gene ush is only reduced ∼2-fold in the sogC27,28S-mNG embryos. We also observed a weakly penetrant ectopic PCV phenotype in sogC27,28S-mNG wings, as a minor proportion have a small extension to the distal side of the PCV. This phenotype is consistent with reduced pMad, potentially due to lower activity of the sogC27,28S mutant in promoting BMP transport to the PCV (Antson et al., 2022). While both the embryonic target gene and wing PCV phenotypes suggest slightly reduced pMad levels, overall these are very mild phenotypes and the flies with sogC27,28S are viable.

While our data suggest that C27 and C28 in Sog are not critical, previously it was shown that overexpression of the dHIP14 palmitoyl transferase in the anterior of the early embryo or wing reduced pMad and disrupted wing vein patterning, respectively, similar to the phenotypes associated with sog overexpression (Kang and Bier, 2010). It is possible that dHIP14 overexpression has pleiotropic effects. Recent evidence suggests that the inhibitory Smad Daughters Against Dpp (Dad) is palmitoylated, which is important for its function (Li et al., 2017). While dad is not expressed in the early embryo, perhaps effects on Dad palmitoylation contribute to the wing phenotypes observed on dHIP14 overexpression.

As the low levels of Sog in vivo prevented us from directly measuring its level of palmitoylation for the wild-type and C27,28S mutant, we cannot rule out another palmitoylation target site in Sog. A cys residue within the predicted TMD/SP is another putative target (Kang and Bier, 2010). However, conditioned media collected from cells expressing the sogC27,28S mutant was less able to inhibit Dpp signalling compared to wild-type Sog in a tissue culture assay, consistent with reduced Sog secretion (Kang and Bier, 2010). This result suggests that the C27,28S mutation would be sufficient to reveal some defect in BMP signalling regulation in vivo, if Sog palmitoylation at these residues is important. Together, our data suggest that Sog secretion is much less dependent on C27,28 and palmitoylation in vivo than in tissue culture cells.

Our 3D model of Sog constructed from negative stain EM data reveals a curved shape similar to the ‘horseshoe’ shape of its vertebrate ortholog Chordin. Due to the known BMP-Chordin vWC domain binding affinities, BMP dimers have been predicted to cooperatively bind Sog vWC1 and vWC4, and Chordin vWC1 and vWC3 domains (Larrain et al., 2000; Troilo et al., 2014; Zhang et al., 2007). In addition, Sog vWC1 and vWC4 have also been shown to interact with Col IV (Sawala et al., 2012). Sog and Chordin have therefore been predicted to adopt a curved conformation that would position the N- and C-termini in close proximity (Larrain et al., 2000; Sawala et al., 2012; Troilo et al., 2014). For Chordin, this ‘horseshoe-like’ shape is supported by a 3D reconstruction generated by single particle analysis of negative stain EM data (Troilo et al., 2014). The curved shape of the Sog EM and AlphaFold structures is therefore consistent with models of cooperative BMP binding, and simultaneous vWC1- and vWC4-Col IV interactions.

In comparison to Chordin, the model of Sog presented here shows Sog to be slightly more compact, with dimensions of 13.6×9.9×9.2 nm versus 15×13×8 nm of Chordin (Troilo et al., 2014). While Tld can only cleave Sog when it is bound to BMP, Chordin alone is processed by Tld (Marqués et al., 1997; Peluso et al., 2011; Piccolo et al., 1997). It has been suggested that a BMP induced conformational change in Sog is required for Sog cleavage by Tld (Marqués et al., 1997; Peluso et al., 2011). The PAE scores for AlphaFold Sog suggest that it is less flexible than Chordin, which together with the more compact shape of Sog could contribute to the requirement for this conformational change. The lower confidence in the relative positions of C-terminal vWC domains of the AlphaFold model is consistent with a level of Sog flexibility, potentially facilitating a ligand dependent conformational change, and permitting Tld access to target residues. The predicted greater flexibility of Chordin interdomain regions, where Tld cleavage sites are located, could therefore reflect the absence of required co-factors for Tld-mediated Chordin processing. The dependency of Sog on BMP ligand binding for Tld cleavage appears to underpin the ‘shuttling’ function of Sog during Drosophila embryogenesis (Peluso et al., 2011). In contrast, a ‘source-sink’ model of BMP gradient formation is most likely to operate during vertebrate embryo dorsal-ventral patterning (Pompreiné et al., 2017; Tuazon et al., 2020; Zinski et al., 2017). Future studies probing how the structures of Sog and Chordin differ will illuminate how these proteins use the same interacting proteins but different mechanisms to generate a BMP gradient.

**MATERIALS AND METHODS**

**Sog purification**

Lipofectamine 3000 reagent (Thermo Fisher Scientific) and Xfect (Takara Bio) transfection reagents were used to transfect HEK293 EBNA cells (Ballock lab stock, not recently authenticated) with pCep-Pu-Sog-V5-His to establish a stable cell line. pCep-Pu-Sog-V5-His cells were maintained in growth media (10% FBS, DMEM:F-12 Hams, 1% penicillin-streptomycin (P/S), 1% Gla) containing 1 µg/ml puromycin. For protein expression, cells were cultured in HYPERflasks (Corning) with expression media (DMEM: F-12 Hams, 1% P/S, 1% Gla, 50 mM L-Arginine). Collected conditioned media was stored at −20°C.
Recombinant Sog protein was isolated from collected conditioned media using affinity chromatography. 1 ml HisTrap Excel columns (Cytiva) were used to pull down Sog via the C-terminal His-tag. For this, 2 mM imidazole was added to filtered conditioned media, and a loading buffer was used (10 mM Tris, 800 mM NaCl, pH 7.4) to equilibrate the His-trap column. Filtered conditioned media was pumped over the column at 4°C. The column was washed with 20 column volumes (CV) loading buffer with 10 mM imidazole. Protein elution was eluted from the column with elution buffer (10 mM Tris, 800 mM NaCl, 500 mM imidazole, pH 7.4).

Affinity chromatography fractions were subject to SEC for further purification with a Superdex 200 Increase 10/30 gel filtration column (Cytiva). The column was washed with 1.5 CV filtered and degassed Milli-Q water (Millipore), followed by 1.5 CVs SEC buffer (20 mM Tris, 800 mM NaCl, pH 7.4) to equilibrate. Affinity chromatography fractions were passed across the column at 0.5 mln/min. Protein elution was monitored using UV absorbance (280 nm). Eluted 0.5 mln fractions were collected and screened by SDS-PAGE under reducing conditions with Instant Bluestain (Abcam), and by western blot with anti-His [clone AD1.11.10] (1:1000, R&D Systems, cat. #MAB050, RRID: AB_357353) primary antibody and IRDye 800CW Donkey anti-Mouse (1:10,000, LI-COR Biosciences, cat. #926-32212, RRID: AB_621847/LI-COR) secondary antibody. Purified Sog was stored at −80°C, before undergoing a second round of SEC (as above) to improve sample purity for production of negative stain EM grids. Collected fractions were screened by SDS-PAGE under reducing conditions with Coloidal Coomassie stain, and by western blot (as above). The Sog sample used for AUC underwent one round of SEC with the following SEC buffer: 800 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4.

Negative stain electron microscopy

A 5 nm (approximately) layer of carbon was deposited onto a mica sheet using a Cresington coating system 308R. The carbon layer was floated onto the carbon side of the TEM grids which were left to dry overnight at room temperature. Carbon coated grids were glow discharged at 25 mA for 1 min. After SEC, Sog (~13 µg/ml) was adhered to carbon coated EM grids for 1 min and stained with 2% uranyl acetate. The stain was wicked away with filter paper before drying. Negatively stained protein molecules were imaged in low dose mode on an FEI Tecnai 12 Transmission Electron Microscope at 100 kV. Images were captured with a Gatan Orius camera at 30,000× magnification. Data collection parameters were according to (Hoppe and Ashe, 2021), using y′ w67−23 (BDSC Stock 6599), w1118, F;Bac[y+;mDint2] GEP[3;E.3xP3]=vas-Cas9::K00027 (BDSC Stock 51324); y′ w67−23; MKRS, P[y+[y+;y+]=hxFlp];86ETM6B, P[w+;w+;Crewl]/DH2, Tb1 (BDSC Stock 1501); y′ sog112/5FMC7c, sn+ (BDSC Stock 2497); brk5869/7FMC7c-fiz-lacZ (Jazwinska et al., 1999), y′ w67−23 flies were used as controls throughout.

CRISPR-Cas9 genome editing and phiC31 reintegration

The sog112/5FMC7c Drosophila line was generated by CRISPR-Cas9 with HDR in a two-step CRISPR approach (Baena-Lopez et al., 2013; Hoppe and Ashe, 2021). For a detailed outline of the strategy see (Hoppe and Ashe, 2021). PAM sites (NGG) located either side of the sog start codon were identified using the CRISPR OptimalTarget Finder tool on the flyCRISPR website (Gratz et al., 2014). Two guide RNA sequences were designed 3 nucleotide upstream of the selected PAM sites to target these sites for Cas9 nuclease digestion and the creation of double stranded breaks. For oligonucleotide sequences encoding sense and antisense strands for guide sequence see Table S1. Complementary guide oligonucleotides were annealed and inserted into the pU6-Bbs1-gRNA plasmid (RRID: Addgene_45946; Gratz et al., 2013) as previously described (Hoppe and Ashe, 2021). Homology arm (HA) sequences were obtained from Drosophila genomic DNA (BL51324) by PCR. Homology arms were inserted into the pTVCherry donor plasmid (Drosophila Genomics Research Center, DGRP_1338) (Baena-Lopez et al., 2013) at SpeI and KpnI restriction sites. Donor plasmids and gRNA plasmids were injected into Cas9 expressing embryos (BL51324) at the University of Cambridge Fly Injection Facility. Flies that developed from injected embryos were crossed according to (Hoppe and Ashe, 2021), using y′ w67−23 and brk5869/7FMC7c-fiz-lacZ (Jazwinska et al., 1999) stocks. The white+ marker was removed by crossing sog112/5FMC7c females with males that carried FM7c on the X and Cre-recombinase on the third chromosome. The sog112/5FMC7c CRISPR mutation results in a deletion of 800 bp (ChrX:15:625,466, -15,626,265, dm6), including the endogenous sog start codon and signal sequence, that is replaced by 103 nucleotides containing attP and LoxP sites.

Reintegration plasmids were generated from the RIVWhite plasmid (gift from the Vincent lab). A partial sog 5′UTR (p5′UTR) sequence (source: pBS-Sog-CDS; Ashe and Levine, 1999), followed by the sog CDS, mNeonGreen and the sog 3′UTR (source: pBS-Sog-CDS; Ashe and Levine, 1999), was inserted between the RIVWhite attB and LoxP sites. To summarise the construction of these plasmids, the partial 5′UTR and the sog CDS were ligated together in a pAC5.1/V5-His (Thermo Fisher Scientific, V411020) vector, as were the mNeonGreen and sog 3′UTR sequences, using In-fusion cloning (Takara Biosciences). A linker sequence was added downstream of the sog CDS. The p5′UTR-sog CDS-linker, and mNeonGreen-3′UTR sequences were inserted into RIVWhite using In-Fusion cloning. To create the sog112/27,28S mutant, Cys 27 and 28 were replaced by two Ser residues with In-Fusion cloning. Reintegration vectors were injected with a phiC31 encoding plasmid into embryos of the sog112/5FMC7c CRISPR stock. Female flies that developed from the injected embryos were cross to y′ w67−23 males and w+ offspring were crossed to brk5869/7FMC7c-fiz-lacZ to balance. Successful phiC31 mediated recombination was confirmed by sequencing genomic DNA. Flies in which successful phiC31 recombination events had occurred were backcrossed to each other to make homozygous stocks. The w+ marker gene was not removed from the sog-mNg and sog112/27,28S/mNg stock generated here. This could be done by Cre-recombinase if necessary for future work. See Table S1 for primers and oligonucleotide sequences. SnapGene viewer software (from Insightful Science; available at snapgene.com) was used to scan DNA sequences for ORFs.

Viability and lethality assays

For viability assays, 50 embryos were placed on an apple juice agar plate which was transferred into a food bottle. Embryos were incubated at 25°C in the bottle. The number of pupae and adults were counted. For the sog rescue assay, virgins of y′ sog112/5FMC7c, sn+ (BDSC Stock 2497) or sog112/5FMC7c females were cross to y′ w67−23 males and w+ offspring were crossed to brk5869/7FMC7c-fiz-lacZ to balance. Successful phiC31 mediated recombination was confirmed by sequencing genomic DNA. Flies in which successful phiC31 recombination events had occurred were backcrossed to each other to make homozygous stocks. The w+ marker gene was not removed from the sog-mNg and sog112/27,28S/mNg stock generated here. This could be done by Cre-recombinase if necessary for future work. See Table S1 for primers and oligonucleotide sequences. SnapGene viewer software (from Insightful Science; available at snapgene.com) was used to scan DNA sequences for ORFs.
were crossed to male yw, sog-mNG; or sog\textsuperscript{C27.285}; flies. The number of FM7c and non-FM7c female offspring were counted to assess the degree of rescue from each genotype in the presence of the mutant sog allele.

**Wing dissection**

Adult flies were incubated at 18°C or 25°C for 24 h, before transfer to a fresh vial. Flies were then allowed to lay eggs before being discarded. Embryos were allowed to develop to adulthood at the designated temperature condition. Wings were then removed from adult females, placed on a slide and washed briefly in isopropanol. Wings were mounted in DPX mounting media (Fisher D:5319/05) under a No.1 coverslip. Samples were imaged with a light microscope (Zeiss AxioSkop) with a 5X objective (Zeiss CP-Achromat 5X/0.12). Images were acquired with Agilent OpenLab 2.2.2 software. For analysis, statistical tests were performed on raw count data in RStudio.

**In situ hybridisation and immunofluorescence**

Embryos (2–4 h) were collected and stained by RNA in situ hybridisation with sog-digoxygenin-UTP, Race-Biotin-UTP, lacZ-digoxygenin-UTP or mNeon-Green-biotin-UTP probes as described (Hoppe et al., 2020; Kosman et al., 2004). An mNeonGreen-biotin-UTP probe was synthesised as previously described (Kosman et al., 2004) with primers listed in Table S1. Antibodies used were mouse anti-biotin (1:250, Roche, cat. #1297597), sheep anti-digoxygenin Fab fragments antibody, AP conjugated (1:200, Roche, cat. #11093274910 RRID:AB14497), donkey anti-mouse IgG secondary antibody, Alexa Fluor 647 (1:500, Thermo Fisher Scientific, cat. #A-31571, RRID:AB162542), and donkey anti-sheep IgG secondary antibody, Alexa Fluor 488 (1:500, Thermo Fisher Scientific, cat. #A-11015, RRID: AB_2534082). For pMad immunostaining, anti-Smad3 (phospho S423+S425 [E1823Y]) (1:500, Abcam, cat. Fab52903, RRID: AB_882596) primary antibody and Donkey anti-rabbit IgG secondary antibody, Alexa Fluor 647 (1:500, Thermo Fisher Scientific, cat. #A-31573, RRID: AB_2536183) were used. To stain embryo nuclei, samples were incubated with DAPI (1:1000, NEB 4083). Samples were mounted in ProLong™ Diamond Antifade Mountant (Thermo Fisher Scientific, P36961).

**smiFISH (Stellaris), smiFISH and immunofluorescence**

For smiFISH, 2–4 h or 1–3 h embryos were processed as described (Hoppe et al., 2020) with anti-sog Stellaris (Fig. 2C, Fig. S1), or sog single molecule inexpensive FISH (smiFISH) (Fig. S4) (Tsanov et al., 2016), ush Stellaris, lacZ Stellaris, and Race smiFISH probes (Tsanov et al., 2016). ush probe sequences are available from (Hoppe et al., 2020), while sog Stellaris, sog smiFISH, lacZ Stellaris, and Race smiFISH probe sequences are provided in Table S2. Race and sog smiFISH probes were annealed to a 570-conjugated Y-FLAP and Z-FLAP, respectively (Tsanov et al., 2016). For immunostaining against mNeonGreen and Spectrin, mouse anti-mNeonGreen [32F6] (1:500, ChromoTek, cat. #32F6-100, RRID: AB_2827566), mouse anti-Spectrin against mNeonGreen, were imaged in a Leica TCS SP8 AOBS confocal microscope with a HCX PL APO 63×/1.40 oil objective with 1×zoom. The confocal settings used were as follows, pinhole 1 airy unit, scan speed 400 Hz bidirectional, format 2048×512 pixels, at 8 bit and Z step size 0.3 μm. Images were collected using hybrid detectors using the white light laser with 548 nm (20%), 405 nm (14%), and 4X line averaging. Raw images were deconvolved with Huygens Remote Manager software v3.7.1 (SVI). Images shown are maximum intensity projections.

**smFISH**

Embryos for smFISH were fixed with 4% PFA in PBS, pH 7.4 for 1 h at RT, washed with PBS, pH 7.4 and then incubated in PBST with 5% goat serum for 1 h at RT. After blocking with 5% goat serum with 1 h at RT, embryos were incubated with primary antibodies overnight at 4°C. Primary antibodies used were anti-Spectrin antibody, were imaged using the white light laser with 467 nm (20%), 548 nm (20%), 405 nm (6%). To age embryos using the anti-Spectrin antibody stain, at the centre of the embryo stacks of 5–10 images (Z step size=0.3 μm) were collected with hybrid detectors using the white light laser with 548 nm (20%), 488 nm (12%) 405 nm (6%). To collect images for analysis embryos were imaged with a HCX PL APO 63×/1.40 oil objective and 0.75× zoom. The confocal settings used were as follows, pinhole 1 airy unit, scan speed 400 Hz bidirectional, format 2048×512 pixels, at 8 bit and Z step size 0.3 μm. Images were collected using hybrid detectors using the white light laser with 548 nm (20%), 405 nm (6%) with 3X line accumulation. Raw images were deconvolved with Huygens Remote Manager software v3.7.1 (SVI). Images shown are maximum intensity projections.

**Imaging stained embryos**

Fixed embryos stained with Race and lacZ RNA in situ hybridisation probes were imaged with a Leica TCS SP5 AOBs inverted microscope using a HCX PL APO lambda blue 20.0×0.70 IMM UV oil objective. The following confocal settings were used: pinhole 1 airy unit, scan speed 400 Hz unidirectional, 512×512 pixel format, Z step size of 1.5 μm at 8 bit and 1.3× zoom. Images shown are maximum intensity projections. Images were deconvolved with Huygens Professional software (SVI, Scientific Volume Imaging, RRID:SCR_014237).

**smFISH image analysis**

To age sog\textsuperscript{opp} and y\textsuperscript{+} w\textsuperscript{67C3} control embryos stained with anti-sog and anti-lacZ smiFISH and anti-Spectrin antibody, maximum intensity projections of images were made, and the length of the cell membrane ingestion (Spectrin antibody stain) was measured (Calvo et al., 2021). Embryos with cell membranes of 3.5–5.5 μm were used for analysis. Only male embryos were analysed. y\textsuperscript{+} w\textsuperscript{67C3} males were identified by the number of sog transcription sites, and sog\textsuperscript{opp} males were identified by the absence of lacZ smFISH stain.

Quantitative analysis of ush and Race, and sog smFISH images was performed in Imaris 9.2 (Bitplane, Oxford Instruments). For efficiency, analysis of ush/Race images was performed for only the central 1048×3144 area of the 3144×3144 images. For sog smFISH data, the entire 2048×512
images were analysed. For quantification of *ush*, *Race*, or *sog* mRNA number, individual mRNAs were detected with the ‘spots’ function. Spots of diameter 0.3 μm (X/Y direction) and 0.8 μm (Z direction) were used. Nuclear locations were determined using the ‘surfaces’ function to identify nuclei based on DAPI staining. To determine the number of spots per cell, spots were assigned to surfaces using the spotMe_V2.py python script (Vinter et al., 2021) (script available at https://github.com/TMinchington/sass). Output from this analysis was processed in Rstudio to remove duplicated nuclei values and ‘NAs’. Data for the number of mRNAs for each of the three embryos analysed were pooled and divided into 5 μm bins (approximate cell size) to permit calculation of the mean number of mRNAs at a given distance from the dorsal/expression domain midline. To account for imaging of embryos that were not perfectly lateral, data embryos stained with anti-*sog* smFISH, were cropped —60 and +80 μm from the expression domain midline, as these were the boundaries shared by all embryos. Binned data were plotted in Rstudio, while data for individual embryos were plotted in GraphPad Prism 9 (RRID:SCR_002798).

For analysis of *sog* expression in *sog-mNG* and *sogC27,28S-mNG* and control embryos, *sog* smFISH images were analysed in Fiji ImageJ (Schindelin et al., 2012). To quantify the height of the *sog* expression domain, maximum intensity projections of the imaging stacks were made, and the maximum vertical number of *sog* positive cells was counted in a 25 μm region of interest (ROI) located 50 μm to the posterior of the cephalic furrow. To quantify *sog* expression level within *sog* positive cells in sum of slices projected images, a 40×40 μm ROI was drawn 50 μm away from the cephalic furrow and bordering the ventral edge of the *sog* expression domain. The mean grey value of the ROI was calculated. To correct for background fluorescence, the mean grey value along a 20 μm line situated outside the *sog* expression domain was calculated. This value was subtracted from the mean grey value within the ROI. Statistical analysis was performed in GraphPad Prism 9 (RRID:SCR_002798).

Amnioserosa counts

Fixed embryos were stained with mouse anti-Hnt 1G9 (1:40, DSHB Cat# 1G9, RRID: AB_528278) and Anti-Mouse IgG (H+L) AP Conjugate S372B 1:500 (Promega) by standard techniques (Kosman et al., 2004). Stage 11 embryos were imaged on a Leica DM6000 microscope with a 20x objective using brightfield. Total amnioserosa cells were counted on ImageJ using the Cell Counter plugin. 50 embryos across three biological repeats were analysed per genotype. Counts were plotted in GraphPad Prism 9 (RRID:SCR_002798).

Acknowledgements

We thank Drs Michael Lockhart-Cairns and Alan Godwin for helpful discussions and support with EM data collection and analysis, Dr Thomas Jowitt (University of Manchester Biomolecular Analysis Facility) for AUC data collection and analysis, the Bloomberg Drosophila Stock Centre for flies, JP Vincent lab for plasmids, and the Cambridge Fly Facility for microinjections. We are grateful to the Baron lab for wing imaging training and Osamu Shimmri for helpful discussions. We also thank staff of The University of Manchester Biomaging Facility, and Fly Facility Manchester for their support and staff of the FBMH EM Core Facility (RRID:SCR_021147) for their assistance.

Competing interests

The authors declare no competing or financial interests.

Author contributions

Conceptualization: S.L.F., C.B., H.L.A.; Investigation: S.L.F., C.S.; Writing - original draft: S.L.F., C.B., H.L.A.; Writing - review & editing: S.L.F., C.S., C.B., H.L.A.

Funding

This research was supported by a Biotechnology and Biological Sciences Research Council project grant (BB/V008099/1) to H.A. and C.B. and a BBBSRC DTP PhD studentship (BB/M011208/1) to S.F. Open Access funding provided by The University of Manchester. Deposited in PMC for immediate release.
Hoppe, C. and Ashie, H. L. (2021). CRISPR-Cas9 strategies to insert MS2 stem-loops into endogenous loci in Drosophila embryos. STAR Protocols 2, 103308. doi:10.1016/j.xpro.2021.103308

Hoppe, C., Bowles, J. R., Milinchuk, T. G., Sutcliffe, C., Upadhyay, P., Ratrav, M. and Ashie, H. L. (2020). Modulation of the promoter activation rate dictates the transcriptional response to graded BMP signaling levels in the Drosophila embryo. Dev. Cell 54, 727-741.e7. doi:10.1016/j.devcel.2020.07.007

Jazwinska, A., Rushlow, C. and Roth, S. (1999). The role of brinker in mediating the graded response to Dpp in early Drosophila embryos. Development 126, 3323-3334. doi:10.1242/dev.126.15.3323

Jumper, J., Evans, R., Prutzal, A., Green, T., Figurnov, M., Ronneberger, O., Tynnyravuokko, K., Bates, R., Židek, A., Potapenko, A. et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589. doi:10.1038/s41586-021-03819-2

Käll, L., Krogh, A. and Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. J. Mol. Biol. 338, 1027-1036. doi:10.1016/j.jmb.2004.03.016

Käll, L., Krogh, A. and Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—The Phobius web server. Bioinformatics 23, 4137-4144. doi:10.1093/bioinformatics/btm455

Kang, K.-H. and Bier, E. (2010). dHIP14-dependent palmitoylation promotes secretion of the BMP antagonist Sog. Dev. Biol. 346, 1-10. doi:10.1016/j.ydbio.2010.06.024

Kosman, D., Mizutani, C. M., Lemons, D., Cox, W. G., Mcginnis, W. and Bier, E. (2004). Multiplex detection of RNA expression in Drosophila embryos. Science 305, 846. doi:10.1126/science.1099247

Larramendi, M. A., Roschiller, L., Bu, A., Gius, E., Piccolo, S. and De Robertis, E. M. (2000). BMP-binding modules in chordin: a model for signalling regulation in the extracellular space. Development (Cambridge, England) 127, 821-830. doi:10.1242/dev.127.4.821

Li, W., Li, Zou, L., Ji, S., Li, C., Liu, K., Zhang, G., Sun, Q., Xiao, F. and Chen, D. (2017). Membrane targeting of inhibitory Smads through palmitoylation controls the Bmp-Smad pathway. Proc. Natl Acad. Sci. USA 114, 13206-13211. doi:10.1073/pnas.1705401114

Linder, M. E. and Deschenes, R. J. (2003). New insights into the mechanisms of protein palmitoylation. Biochemistry 42, 4317-4320. doi:10.1021/bi0211989

Madamanchi, A., Mullins, M. C. and Umulis, D. M. (2021). Diversity and similarity of bone morphogenetic protein extracellular transport in the Drosophila embryo. Dev. Dyn. 250, 1122-1127. doi:10.1002/dvdy.25274

Markstein, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B. et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676-682. doi:10.1038/nmeth.1971

Schmierer, B. and Hill, C. S. (2007). TGFβ-SMAD signal transduction: molecular specificity and functional versatility. Nat. Rev. Mol. Cell Biol. 8, 970-982. doi:10.1038/nrm2297

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophys. J. 78, 1693-1699. doi:10.1016/S0006-8993(00)7713-0

Serpe, M., Ralston, A., Blair, S. S. and Connor, M. B. (2005). Matching catalytic activity to developmental function: Tolloid-related processes Sog in order to help specify the posterior crosvenin in the Drosophila wing. Development 132, 2645. doi:10.1242/dev.01838

Shilo, B.-Z., Haskel-Iltah, M., Ben-Zvi, D., Schejter, E. D. and Barkai, N. (2013). Creating gradients by morphogen shuttling. Trends Genet. 29, 339-347. doi:10.1016/j.tig.2013.01.001

Shimmi, O. and O‘Connor, M. B. (2003). Physical properties of Tld, Sog and Dpp protein interactions are predicted to help create a sharp boundary in Bmp signals during dorsal-ventral patterning of the Drosophila embryo. Dev. Dyn. 229, 484-497. doi:10.1002/dvdy.10747

St Johnston, R. D. and Gelbart, W. M. (1986). Solution structure and dynamics of a prototypical chordin-related protein, Crh. J. Mol. Biol. 190, 375-383. doi:10.1016/S0022-2836(86)80067-1

Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. and Schulte-Mickett, D. (2012). UCSF ChimeraX: structure visualization for macromolecular and nanoscale bioimaging. J. Struct. Biol. 181, 38-46. doi:10.1016/j.jsb.2012.05.009

Tirolo, H., Zük, A. V., Tunyasuvunakool, K., Bates, R., Raftery, L. A. and Sutherland, D. J. (2003). UCSF ChimeraX: Structure visualization for macromolecular and nanoscale bioimaging. J. Struct. Biol. 181, 38-46. doi:10.1016/j.jsb.2012.05.009

Tsunov, N., Samacoits, A., Chouaib, R., Traboulsi, A. M., Gostan, T., Weber, C., Zimmer, C., Zibara, K., Walter, T., Peter, M. et al. (2016). smiFISH and FISH-quant – a flexible single RNA detection approach with super-resolution capability. Nucleic Acids Res. 44, e165. doi:10.1093/nar/gkw784

Uchizono, T., Wang, X., Liu, J., Mullins, D. and Mullins, M. C. (2020). Proteolytic restriction of chordin undergoes BMP binding pattern formation. Cell Reports 32, 108039. doi:10.1016/j.celrep.2020.108039
Umulis, D. M., Serpe, M., O’Connor, M. B. and Othmer, H. G. (2006). Robust, bistable patterning of the dorsal surface of the Drosophila embryo. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11613-11618. doi:10.1073/pnas.0510398103

Umulis, D., O’Connor, M. B. and Blair, S. S. (2009). The extracellular regulation of bone morphogenetic protein signaling. *Development* 136, 3715-3728. doi:10.1242/dev.031534

Umulis, D. M., Shimmi, O., O’Connor, M. B. and Othmer, H. G. (2010). Organism-scale modeling of early Drosophila patterning via bone morphogenetic proteins. *Dev. Cell* 18, 260-274. doi:10.1016/j.devcel.2010.01.006

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A. et al. (2022). AlphaFold protein structure database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.* 50, D439-D444. doi:10.1093/nar/gkab1061

Vilmos, P., Sousa-Neves, R., Lukacsovich, T. and Lawrence Marsh, J. (2005). Crossveinless defines a new family of Twisted-gastrulation-like modulators of bone morphogenetic protein signalling. *EMBO Rep.* 6, 262-267. doi:10.1038/sj.embor.7400347

Vinter, D. J., Hoppe, C., Minchington, T. G., Sutcliffe, C. and Ashe, H. L. (2021). Dynamics of hunchback translation in real time and at single mRNA resolution in the Drosophila embryo. *Development* 148, dev.196121.

Wang, Y.-C. and Ferguson, E. L. (2005). Spatial bistability of Dpp-receptor interactions during Drosophila dorsal-ventral patterning. *Nature* 434, 229-234. doi:10.1038/nature03318

Wang, X. M., Harris, R. E., Bayston, L. J. and Ashe, H. L. (2008). Type IV collagens regulate BMP signalling in Drosophila. *Nature* 455, U72-U49. doi:10.1038/nature07214

Wang, R. N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S. et al. (2014). Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis.* 1, 87-105. doi:10.1016/j.gendis.2014.07.005

Wharton, K. A., Cook, J. M., Torres-Schumann, S., de Castro, K., Borod, E. and Phillips, D. A. (1999). Genetic analysis of the bone morphogenetic protein-related gene, gbb, identifies multiple requirements during Drosophila development. *Genetics* 152, 629-640. doi:10.1093/genetics/152.2.629

Winstanley, J., Sawala, A., Baldock, C. and Ashe, H. L. (2015). Synthetic enzyme-substrate tethering obviates the Tolloid-ECM interaction during Drosophila BMP gradient formation. *Elife* 4, e05508. doi:10.7554/eLife.05508

Xu, E.-R., Blythe, E. E., Fischer, G. and Hyvönen, M. (2017). Structural analyses of von Willebrand factor C domains of collagen 2A and CCN3 reveal an alternative mode of binding to bone morphogenetic protein-2. *J. Biol. Chem.* 292, 12516-12527. doi:10.1074/jbc.M117.788992

Yu, K., Sturtevant, M. A., Biehs, B., Francois, V., Padgett, R. W., Blackman, R. K. and Bier, E. (1996). The Drosophila decapentaplegic and short gastrulation genes function antagonistically during adult wing vein development. *Development* 122, 4033. doi:10.1242/dev.122.12.4033

Zhang, J.-L., Huang, Y., Qiu, L.-Y., Nickel, J. and Sebald, W. (2007). Von Willebrand factor type C domain-containing proteins regulate bone morphogenetic protein signaling through different recognition mechanisms. *J. Biol. Chem.* 282, 20002-20014. doi:10.1074/jbc.M700456200

Zhang, J., Qiu, L., Kotzsch, A., Weidauer, S., Patterson, L., Hammerschmidt, M., Sebald, W. and Mueller, T. D. (2008). Crystal structure analysis reveals how the chordin family member crossveinless 2 blocks BMP-2 receptor binding. *Dev. Cell* 14, 739-750. doi:10.1016/j.devcel.2008.02.017

Zinski, J., Bu, Y., Wang, X., Dou, W., Umulis, D. and Mullins, M. C. (2017). Systems biology derived source-sink mechanism of BMP gradient formation. *ELife* 6, e22199. doi:10.7554/eLife.22199

---

**RESEARCH ARTICLE**

Biology Open (2022) 11, bio059199. doi:10.1242/bio.059199