Investigating the Mechanism of Follistatin Function with Bioinformatics

Heather L. Stickney
Computational Molecular Biology
Prof. Doug Brutlag, Lee Kozar
June 6, 2002

Background and Significance

During embryonic development, a small number of conserved families of signaling molecules are used over and over again for intercellular communication, functioning to regulate a diverse array of developmental processes. Bone morphogenetic proteins (BMPs) and their relatives, over 20 of which have been identified in vertebrates (e.g., zebrafish BMP2) and invertebrates (e.g., Dpp), comprise one such family [1]. The functional BMP unit is a secreted dimer. These dimers bind to Type II receptor serine/threonine kinases, which then recruit Type I receptor serine/threonine kinases. This heterotetrameric complex then signals downstream through an intracellular signal transduction pathway comprised of various Smads, eventually bringing about the transcriptional regulation of target genes (for review, see [1]).

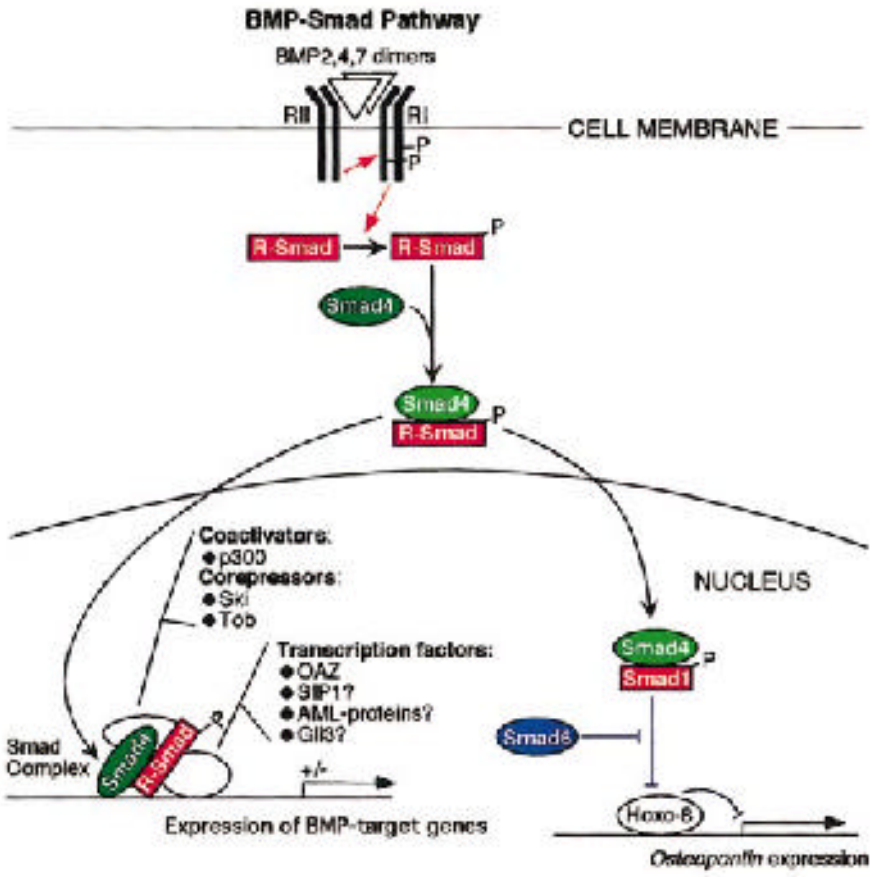

Figure 1. A BMP dimer binds to a Type II BMP receptor, which then recruits a Type I receptor to form a heterotetrameric complex. The Type II receptor phosphorylates the Type I receptor, which then transduces the BMP signal through various Smads. From [2].

BMP signaling has been implicated in such fundamental processes as dorsal-ventral patterning, primary neural induction, somite patterning, eye development, limb patterning, and organogenesis, to name but a few (for review, see [1, 3]). Given the multifunctional nature of

these proteins, it is clearly important for their activity to be tightly regulated both spatially and temporally. Although some of this regulation is under transcriptional control, a number of post-translational regulators of BMP activity, such as Chordin and Noggin, have been identified (for review, see [4]. Recent evidence suggests that Follistatin, a secreted glycoprotein of 32-39kDA, may be another such post-translational regulator.

Follistatin has long been known to have a biological activity opposite that of the BMPs. Xenopus animal cap experiments performed in the early nineties identified Follistatin as a putative neural inducer [5], while BMP-4 exerts a potent anti-neuralizing effect in the same assay [6]. More recently, overexpression experiments in Xenopus and zebrafish embryos have demonstrated that misexpression of *follistatin* RNA results in an expansion of dorsal fates at the expense of ventral ones [7], whereas injection of RNAs encoding various BMP family members causes severe ventralization of the embryos [8, 9].

However, evidence that Follistatin functions as a BMP antagonist has only come within the last five years. Coinjection of *bmp4* or *bmp7* and *follistatin* mRNAs into zebrafish embryos completely rescued both the ventralizing effects of *bmp* injection and the dorsalizing effects of *follistatin,* suggesting that a functional antagonism exists between the two secreted proteins when ectopically expressed [7, 10]. To investigate whether this functional antagonism occurs between endogenous proteins, Bauer and colleagues examined the spatiotemporal pattern of *follistatin, noggin, bmp2b,* and *bmp4* RNA expression in zebrafish [11]. They found that although *follistatin* and the known *bmp* antagonist *noggin* are expressed in distinct domains, the expression patterns of the two genes are similar in one intriguing way. The domains of *noggin* and *follistatin* expression often border, but do not overlap, domains of *bmp* expression. Thus, endogenous *follistatin* is expressed at the right time and place to act as a BMP antagonist.

Yet the mechanism through which this functional antagonism is mediated is unclear. Injection of *follistatin* mRNA is insufficient to rescue embryos ventralized by the overexpression of an mRNA encoding a constitutively active form of one of the BMP receptors, suggesting that Follistatin antagonizes BMP signaling upstream of BMP receptor activation. Two well-characterized BMP antagonists, Chordin and Noggin, bind directly to the BMPs, interfering with their ability to bind to the type I BMP receptors and hence inhibiting BMP activity in a competitive manner [12, 13]. I*n vitro* binding assays have demonstrated that Follistatin can also bind directly to BMPs [7, 10, 14], though no experiments have confirmed that this interaction occurs *in vivo*. However, the binding of Follistatin to a BMP dimer does not appear to prevent the dimer from binding to its receptor, as a trimeric complex of Follistatin, BMP4, and a soluble form of a BMP receptor can form *in vitro* [10]. Follistatin must inhibit BMPs by some mechanism other than interference with the formation of the BMP-BMPR complex.

With the recent growth of interest in both applied and basic research, particularly in the fields of genomics and proteomics, has come growth in the field of bioinformatics. There are currently a plethora of tools available to help the molecular biologist in his/her research. I have utilized several of these tools in an attempt to gain insight into the mechanism by which Follistatin antagonizes the BMP signaling pathway.

Bioinformatic Methods and Results[1]

I. Identification of Putative Motifs

[1] see Appendix A for URLs for searches used

I first utilized a number of motif searching programs, including *e*Motif-search, *e*Matrix-search, *e*Blocks-search, Blocks+, SMART, Pfam, InterProScan, ScanProsite, and Profilescan, to search the zebrafish Follistatin protein sequence for motifs that might be indicative of its function (see Appendix A for URLs). Although these programs search overlapping datasets of motifs, I tried all of them, because after testing a few, I found that the hits I got from a particular database varied based on the search method I used. For example, one of the statistically significant hits I got using InterProScan was from the SMART database, but when I used a SMART search, that motif was not statistically significant. However, for simplicity, I will discuss only the results of the *e*Motif, Blocks+, and InterProScan, which include almost all of the domains found by the other searches.

An *e*Motif search with the zebrafish Follistatin sequence retrieved nine entries, but only 3 distinct motifs. There were six hits for slightly different variations of the Kazal-type serine protease inhibitor family motif, all in the same region (~aa134-176), with statistically significant expectation values ranging from 1.66e-07 to 1.83e-03. *e*motif scan with several of those motifs returned only known serine protease inhibitors, agrins, and other follistatins. Though the presence of this kazal-type serine protease inhibitor motif is suggestive of a serine protease activity for Follistatin, it is quite possible that this is a structural rather than functional motif in Follistatin, as agrins are not known to have serine protease inhibitor activity. The search also found a motif for an osteonectin domain (aa98-121) and one for a trypsin inhibitor-like cysteine rich domain (aa228-248), both with statistically significant expectation values. The osteonectin domain is a domain of unknown function that is found in extracellular proteins; the trypsin-inhibitor-like domain is found in trypsin-inhibitors, but also in many extracellular proteins. While the osteonectin and trypsin-inhibitor domains are likely to be biologically significant, as Follistatin is an extracellular protein, they give little insight into the function of Follistatin. However, it is important to note that some Kazal-type serine proteinase inhibitors are trypsin inhibitors [15].

A Blocks+ search with the same sequence also returned statistically significant hits on Kazal-type serine protease inhibitor and osteonectin motifs. Interestingly, this search found three, not one, kazal-type serine protease inhibitor domains (aa127-167, E=1.5e-14; aa202-242, E=1.7e-09; aa279-319, E=1.2e-09) and three osteonectin domains (aa96-128, E=7.1e-06; aa169-201, E=0.45; aa246-278, E=0.0089), although in the latter case only the first and possibly the last could be considered statistically significant and all three domains matched in only 1/5 blocks. It's intriguing that in each of the three cases, the putative osteonectin domains seem to closely precede the putative kazal-type serine protease inhibitor domains, but what that means is unclear given the unknown function of the osteonectin domain. The Blocks+ search also identified a cysteine-knot domain that matched 3/3 blocks (aa98-122, 165-183, and 253-305) with a statistically significant combined E value of 0.00012. Note that the three blocks of the cysteine-knot domain overlap with the three putative osteonectin domains and that both domains are cysteine-rich. To me this suggests that only one of these is in fact biologically significant and the other is simply a spurious hit, but which is which can not be deduced from this data. Finally, the Blocks+ search retrieved a keratin, high sulfur B2 protein domain, but this domain was not statistically significant and, given what is thus far known about Follistatin, is unlikely to be biologically significant.

The InterproScan of zebrafish Follistatin identified three putative kazal-type serine protease inhibitor domains (in similar regions as the Blocks+ scan), three Follistatin-like, N-

terminal domains, and a Factor I membrane attack complex domain. The Follistatin-like domains, which are found in Follistatins, Agrins, and prespore vesicle proteins, are obviously biologically significant. These Follistatin-like domains fall in the same location as the cysteine-knot blocks and putative osteonectin domains found in previous searches. Given that one of the distinguishing features of these three domains is a richness of cysteines, perhaps it is not surprising that the same sequence fits all three motifs. The Factor I membrane attack domain is found in complement component proteins, complement component factor I, and Agrins. As there is no function associated with this domain, it is difficult to evaluate whether it is likely to be a biologically significant domain in Follistatin with only this data.

II. Evaluation of motif search results

The objective underlying the above motif searches was to identify potential mechanisms for Follistatin function. The results above suggest two hypotheses. First, given that all but one (ProfileScan) of the motif searches I ran identified at least one statistically significant Kazal-type serine protease inhibitor domain, Follistatin could be a Kazal-type serine protease inhibitor. The second is subtler. I noticed that a number of the domains retrieved by the Follistatin sequence are also found in Agrins, including the Kazal-type serine protease inhibitor, Follistatin-like N-terminal, and Factor I membrane attack complex domains. This indicated that Follistatin might in fact function in a manner similar to that of the Agrins, especially since the Factor I membrane attack complex is not found in any of the known serine protease inhibitors.

III. The hunt for potential mechanisms

After identifying two possible functions for Follistatin, it was then important to see if either or both of those functions could lead to inhibition of BMPs. In other words, could the putative functions form part of an elegant mechanism for the inhibition of BMP activity by Follistatin? In order to determine whether a serine protease inhibitory or Agrin-like function for Follistatin could explain BMP inhibition, it was necessary to perform literature searches to discover what was known about Kazal-type serine protease inhibitors, Agrins, and BMP function.

I searched the PubMed database with MESH search for “Serine proteinase inhibitors” and “bone morphogenetic proteins”. The papers retrieved with this search revealed that BMPs are translated as proproteins that must be cleaved for activity [16]. Interestingly enough, the cleavage of the prodomain from the mature peptide is carried out by the proprotein convertase family of serine proteinases [17, 18], which suggested an elegant mechanism for BMP inhibition by Follistatin. Follistatin could function to inhibit the proprotein convertases, thus preventing the formation of mature BMPs and activation of the BMP-BMP receptor complex (Figure 2). For this to be possible, several things have to be true. First, because the domains of *follistatin* and *bmp* expression do not overlap, cleavage of the prodomain would have to occur extracellularly. Additional literature searches through Pubmed revealed that little is known about whether the prodomain cleavage of BMPs occurs intra- or extracellularly [1]. It is entirely possible that it occurs extracellularly as both the uncleaved form of BMPs and active proprotein convertases have been shown to be present in the extracellular space [19]. Second, noncleaved BMPs must have a dominant negative activity if the inhibition of proteolytic cleavage is to inhibit BMP signaling. This is the case. When coexpressed, a cleavage mutant can inhibits the

processing of wildtype proteins [20]. My literature searches also revealed two facts that slightly weaken the case for a Follistatin serine proteinase inhibitor activity: (1) the only known protein inhibitor of the proprotein convertases is a serpin rather than a Kazal-type serine proteinase inhibitor [21] and (2) other Kazal-type serine proteinase inhibitors (the ovomucoids) have been shown to be ineffective inhibitors of proprotein convertases [15]. However, even given that information, it is still possible that Follistatin functions as a proprotein convertase inhibitor.

Figure 2. Potential mechanism for Follistatin action. (A) In the absence of Follistatin, proprotein convertases (blue) cleave the prodomains from the BMP proproteins (red). The mature protein can then activate the BMP receptors (green) and transduce a signal. (B) In the presence of Follistatin (purple), cleavage by the proprotein convertases is inhibited. Although the dimer can still bind to the BMP receptors, the receptors are not activated and no signal is transducted.

To determine whether a function similar to that of the Agrins could explain Follistatin activity, I first did a PubMed search using the Mesh terms “bone morphogenetic proteins” and “Agrin”, which pulled up nothing. I followed that search with another for “Agrin” limited to publication type “review” to discover exactly how Agrins functioned, thinking that it might give insight into how Follistatin might function. This search revealed that Agrins are basement membrane proteins that cause the aggregation of acetylcholine esterase receptors at the surface of neuromuscular fibers [22]. The N-terminal half of the protein mediates binding to the other membrane components and thus anchors Agrin in the extracellular matrix, the C-terminal half is responsible for interaction with the receptor and aggregation of the acetylcholine esterase receptors. This information suggested that perhaps Follistatin could function by causing the aggregation of BMPs. Though binding of Follistatin to a single dimer does not prevent receptor binding [10], aggregation of BMPs might as a result of steric hindrance (Figure 3).

Figure 3. Model for Follistatin action. (A) In the absence of Follistatin, BMP dimers (red) activate the BMP receptors (green) and transduce the signal. (B) In the presence of Follistatin (purple), BMP dimers aggregate and are no longer able to bind to the receptors.

IV. Conservation of Motifs

The first step in evaluating the above hypotheses was to determine which of the domains identified in the zebrafish Follistatin sequence were conserved across various species. Assuming a conserved function for Follistatin (which appears to be the case for Xenopus and zebrafish, at least) across species, domains necessary for Follistatin function should be conserved.

To find all of the known homologues to zebrafish Follistatin, I ran a Psi-Blast search on the NCBI server starting with the zebrafish Follistatin sequence. Given that I wanted to be sure only to have Follistatin homologues, for the second iteration I selected only the 20 proteins that were actually called Follistatin, rejecting numerous Follistatin-like proteins, Agrins, laminins, and serine proteinase inhibitors. However, the fact that Agrins and serine proteinase inhibitors were picked up by the blast search suggests that Follistatin shares a fair bit of sequence similarity to these proteins, which in turn supports both of the above hypotheses for Follistatin function. The second, third, and fourth iterations pulled up additional proteins, but no Follistatins so my search set remained the same. By the fifth iteration the search had converged. Of the 20 Follistatin sequences that I retrieved with this search, nine were redundant. In the end, I obtained the Follistatin sequences for 11 species: zebrafish, frog, horse, pig, sheep, cow, rat, mouse, chick, human, and fly. To confirm that these were indeed homologs, I made a UPGMA evolutionary tree with the 11 sequences using SeqWeb GrowTree software (Figure 4). This tree concurred with the commonly accepted evolutionary phylogeny for these species, so the assumption that they are homologs is justified.

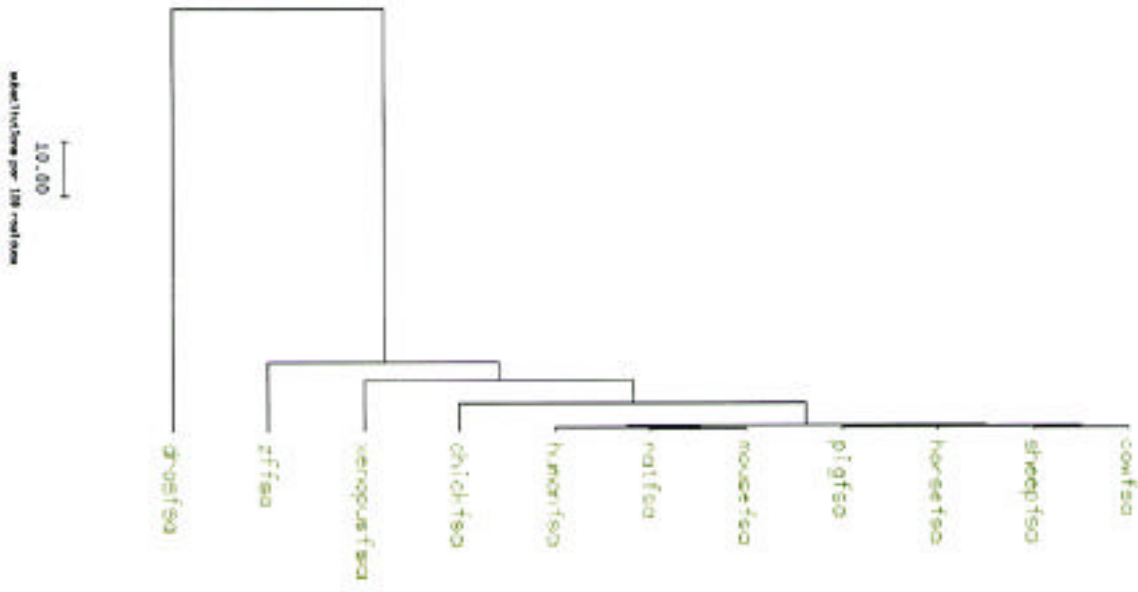

Figure 4. GrowTree phylogram of Follistatin sequences from (left to right) *Drosophila melanogaster*, zebrafish, xenopus, chick, human, rat, mouse, pig, horse, sheep, and cow.

I then performed a multiple alignment with ClustalW, the results of which can be found in Appendix B. This alignment demonstrated that the *Drosophila* protein differed greatly from the vertebrate proteins, which varied little amongst themselves in comparison. Because it was the most distant, the *Drosophila* sequence was perhaps the most important of the eleven sequences in determining conservation of the motifs. The other ten sequences were so similar that it would have been difficult to know whether motifs were conserved because of functional necessity or simply because there was not enough evolutionary distance between the proteins.

To determine what motifs were conserved between the 11 sequences, I performed a motif search with Lama on the Fred Hutchinson Cancer Research Center Blocks WWW server. This program searches for sequence similarities between conserved regions of protein families—it compares a multiple alignment input by the user to the Blocks or Prints databases, which are comprised of other multiple alignments. Comparison of the Follistatin multiple alignment with the Blocks+ databases identified the three Kazal-like serine protease inhibitor domains (aa137-166; 221-261; 302-340; note these are offset slightly from the numbers given above due to gaps in the multiple alignment) observed in the Block+ search of the zebrafish Follistatin sequence as well as two of the osteonectin domains (aa104-134; 179-203) with expectation values of 0e+00. A cysteine-knot domain (aa177-193) was also identified, although with a fairly high expectation value of 1.9e-01. Unfortunately, as the motif for the Factor I membrane attack complex in the zebrafish Follistatin was not identified by the Blocks server and the LAMA search was run on the same server, it was unclear whether that domain was conserved or not. For this reason, I then searched several of the sequences for motifs with InterProScan, which had originally detected a Factor I membrane attack complex in the zebrafish Follistatin sequence. That domain was not identified in either the human or the *Drosophila* Follistatin sequence.

In short, the above results suggested that both the osteonectin and kazal-type serine protease inhibitor domains were conserved across species and thus are likely to be essential in some manner for Follistatin function. My confidence in these results would have been greater had I at least one more protein as divergent as the *Drosophila* Follistatin to decrease the chances that some of these domains were conserved by chance rather than necessity. However, the conservation of the Kazal domain supports the hypothesis that Follistatin functions as a serine proteinase inhibitor, preventing cleavage of the BMP prodomain. Conservation of the osteonectin and kazal-type serine protease domains also supports the possibility that Follistatin functions like Agrins, as both are found in Agrins. Conversely, the Factor I membrane attack complex domain identified in the zebrafish Follistatin sequence is likely to be insignificant biologically, as it was not conserved in Follistatin homologs. This nonconservation weakens the hypothesis that Follistatins function in a similar mechanism to Agrins, although it is possible that that particular domain is not required for the aggregation of target molecules but for some other purpose that, though not essential for Agrin function, is required for Follistatin function.

IV. Comparison of Follistatin and a known Kazal-type serine protease inhibitor

To further investigate the possibility that Follistatin functions as a serine protease inhibitor, I compared both the sequence and the structure of Follistatin to a known kazal-type serine protease inhibitor, human isk5. First, I attempted to align the sequence of human Follistatin to human isk5 using the ALIONment software. I originally used a BLOSUM62 matrix with the default gap opening and gap extension penalties, which aligned a small portion of the two sequences with a 45% identity within the aligned region. I then repeated the alignment using a BLOSUM45 matrix, but this did not affect the alignment. Only 35 amino acids were aligned between the two sequences, with a 45% identity. Interestingly, the stretch of aligned sequence did fall within one of the Kazal domains in the Follistatin sequence. However, the sequences are clearly not similar enough to draw any conclusions about the function of Follistatin as a serine proteinase inhibitor.

I then compared an InterProScan for motifs in isk5 to the motifs found in Follistatin. As expected, the Kazal domain was present in the isk5 sequence (13 distinct domains). There was also a cytochromeC heme-binding site identified in the isk5 sequence. This is not likely to be biologically significant, as isk5 is not of the cytochrome C family. Thus, it appears that Follistatin does not lack any domains necessary for function as a Kazal-type serine proteinase inhibitor.

To investigate whether the structure of Follistatin is similar to that of kazal-type serine proteinase inhibitors, I used GorIV to predict secondary structural elements in both isk5 and Follistatin. The resulting predictions were similar in that only alpha helices, extended strands, and random coils were predicted for both structures, which is consistent with the two having a similar function though the percentages of the proteins that formed the various types of structures were slightly different. For both, random coil was the most common (68.3% for Follistatin and 53.38% for isk5). However, alpha helices were predicted to comprise 34.40% of the isk5 structure but only 7.69% of Follistatin, and extended strands 12.22% of isk5 and 23.96% of Follistatin.

I also used Fugue software with the ClustalW multiple sequence alignment performed previously to search the pdb database for structures predicted to be present in Follistatin. The Fugue search reported the presence of a domain of structure similar to that of Kazal-type serine

protease inhibitors with a confidence of 99%. All other predictions had a confidence of less than 50% and are unreliable. The results from this search are also consistent with the possibility that Follistatin functions as a kazal-type serine proteinase inhibitor.

V. Comparison of Follistatin and a proprotein convertase inhibitor

To explore the possibility that Follistatin functions to inhibit the proprotein convertases responsible for BMP prodomain cleavage, I attempted to align the sequence of the one protein known to inhibit the proprotein convertases, alpha1 anti-trypsin Portland (ref), to the human Follistatin sequence to look for sequence similarities. I used a BLOSUM35 matrix with the default gap opening and gap extension penalties, which aligned a small portion (~16aa) of the two sequences with a 37% identity within the aligned region, suggesting that there is no significant sequence similarity between Alpha1 antitrypsin-PDX and Follistatin.

I then performed an InterproScan search of motifs in Alpha1 antitrypsin-PDX in order to compare motifs with Follistatin. The only motif found in Alpha1 antitrypsin-PDX was that of the serpin family. Serpins are also serine proteinase inhibitors, but do not belong to the Kazal family.

I also used the LIBRA application program to evaluate the compatibility of the Follistatin sequence with the known structure of Alpha1 antitrypsin PDX via threading. The sequence was not compatible, with a raw score of –0.16 as compared to the raw scores of the top 100 compatible sequences, which had scores ranging between –6.88 and –3.89.

This lack of similarity implies that Follistatin does not function in a manner similar to that of Alpha1 antitrypsin-PDX. However, Follistatin still could function to inhibit proprotein convertases by a different mechanism than Alpha1 antitrypsin-PDX, which originally acts as a substrate for the proprotein convertases but cannot be released after cleavage.

VI. Comparison of Follistatin and Agrin

To further explore the possibility that Follistatin functions like Agrin, I performed a sequence alignment of the chick Agrin and chick Follistatin sequences with the ALIONment program. I originally used a BLOSUM62 matrix with the default gap opening and gap extension penalties, which aligned a portion of the two sequences with a 29% identity within the aligned region. I then repeated the alignment using a BLOSUM30 matrix with other default settings. This aligned almost the complete Follistatin sequence with the N-terminal half of the Agrin sequence, suggesting that the N-terminal half of Agrin shares significant sequence similarity with Follistatin.

I also performed an InterProScan to search for motifs in the chick Agrin sequence. If Follistatin functions in a similar way to Agrins, it should contain many of the same motifs as Follistatin. While Agrin had Follistatin domains and Kazal-type serine protease inhibitor domains like Follistatin, it had more (9 Kazal and 5 Follistatin) than did Follistatin. In addition, there were a number of other domains present in the Agrin sequence that were not found in Follistatin, including laminin, EGF, SEA, and Factor I membrane attack complex domains. Furthermore, the Kazal and Follistatin domains fell within the N-terminal half of the protein; no domains within the C-terminal half of Agrin were found in Follistatin.

The fact that Follistatin aligns with the N-terminal half of Agrin and that all of the domains found in Follistatin that are present in Agrin are located within the N-terminal half of

the protein makes it highly unlikely that Follistatin functions like Agrin. As mentioned above, the N-terminal half of Agrin is only known to be required for anchoring in the extracellular matrix. It is the C-terminal half of Agrin that mediates aggregation, and this part of the Agrin protein does not seem to resemble Follistatin at all.

Summary and Conclusion

In summary, searching the zebrafish Follistatin sequence for motifs identified two potential functions for Follistatin—a function as a Kazal-type serine proteinase inhibitor or a function similar to that of Agrin. PubMed searches of the literature enabled me to develop comprehensive hypotheses for the mechanism of Follistatin inhibition of BMP activity based upon the putative Follistatin functions. I evaluated these hypotheses with other bioinformatic methods, the results of which were consistent with Follistatin functioning as a Kazal-type serine proteinase rather than in a manner similar to Agrin. There is still much experimental work to be done to test this hypothesis, as bioinformatics alone cannot definitively prove the hypothesis that Follistatins function by inhibiting the serine proteases necessary for maturation of the BMPs. However, use of these tools was invaluable for identifying and evaluating a potential mechanism for Follistatin action that makes a number of predictions that can be tested in the laboratory.

References

1. Hogan, B.L.M., *Bone morphogenic proteins: multifunctional regulators of vertebrate development.* Genes Dev, 1996. **10**: p. 1580-94.
2. von Bubnoff, A. and K.W. Cho, *Intracellular BMP signaling regulation in vertebrates: pathway or network?* Dev Biol, 2001. **239**(1): p. 1-14.
3. Hild, M., et al., *The roles of BMPs, BMP antagonists, and the BMP signaling transducers Smad1 and Smad5 during dorsoventral patterning of the zebrafish embryo.* Ernst Schering Res Found Workshop, 2000(29): p. 81-106.
4. Thomsen, G.H., *Antagonism within and around the organizer: BMP inhibitors in vertebrate body patterning.* Trends Genet, 1997. **13**(6): p. 209-11.
5. Hemmati-Brivanlou, A., O.G. Kelly, and D.A. Melton, *Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neuralizing activity.* Cell, 1994. **77**(2): p. 283-95.
6. Sasai, Y., et al., *Regulation of neural induction by the Chd and Bmp-4 antagonistic patterning signals in Xenopus.* Nature, 1995. **376**(6538): p. 333-6.
7. Fainsod, A., et al., *The dorsalizing and neural inducing gene follistatin is an antagonist of BMP-4.* Mech Dev, 1997. **63**(1): p. 39-50.
8. Dale, L., et al., *Bone morphogenetic protein 4: a ventralizing factor in early Xenopus development.* Development, 1992. **115**(2): p. 573-85.
9. Jones, C.M., et al., *DVR-4 (bone morphogenetic protein-4) as a posterior-ventralizing factor in Xenopus mesoderm induction.* Development, 1992. **115**(2): p. 639-47.
10. Iemura, S., et al., *Direct binding of follistatin to a complex of bone-morphogenetic protein and its receptor inhibits ventral and epidermal cell fates in early Xenopus embryo.* Proc Natl Acad Sci U S A, 1998. **95**(16): p. 9337-42.
11. Bauer, H., et al., *Follistatin and noggin are excluded from the zebrafish organizer.* Dev Biol, 1998. **204**(2): p. 488-507.
12. Piccolo, S., et al., *Cleavage of Chordin by Xolloid metalloprotease suggests a role for proteolytic processing in the regulation of Spemann organizer activity.* Cell, 1997. **91**(3): p. 407-16.
13. Zimmerman, L.B., J.M. De Jesus-Escobar, and R.M. Harland, *The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4.* Cell, 1996. **86**(4): p. 599-606.
14. Amthor, H., et al., *Follistatin regulates bone morphogenetic protein-7 (BMP-7) activity to stimulate embryonic muscle growth.* Dev Biol, 2002. **243**(1): p. 115-27.
15. Lu, S.M., et al., *Predicting the reactivity of proteins from their sequence alone: Kazal family of protein inhibitors of serine proteinases.* Proc Natl Acad Sci U S A, 2001. **98**(4): p. 1410-1415.
16. Cui, Y., et al., *The activity and signaling range of mature BMP-4 is regulated by sequential cleavage at two sites within the prodomain of the precursor.* Genes Dev, 2001. **15**(21): p. 2797-802.
17. Cui, Y., et al., *BMP-4 is proteolytically activated by furin and/or PC6 during vertebrate embryonic development.* Embo J, 1998. **17**(16): p. 4735-43.
18. Constam, D.B. and E.J. Robertson, *Regulation of bone morphogenetic protein activity by pro domains and proprotein convertases.* J Cell Biol, 1999. **144**(1): p. 139-49.
19. Molloy, S.S., et al., *Bi-cycling the furin pathway: from TGN localization to pathogen activation and embryogenesis.* Trends Cell Biol, 1999. **9**(1): p. 28-35.
20. Hawley, S.H., et al., *Disruption of BMP signals in embryonic Xenopus ectoderm leads to direct neural induction.* Genes Dev, 1995. **9**(23): p. 2923-35.
21. Anderson, E.D., et al., *Inhibition of HIV-1 gp160-dependent membrane fusion by a furin-directed alpha 1-antitrypsin variant.* J Biol Chem, 1993. **268**(33): p. 24887-91.
22. Hoch, W., *Formation of the neuromuscular junction. Agrin and its unusual receptors.* Eur J Biochem, 1999. **265**(1): p. 1-10.

Appendix A: URLs for web searches

*e*Motif-search	http://motif.stanford.edu/emotif-search/
*e*Matrix-search	http://motif.stanford.edu/ematrix-search/
*e*Blocks-search	http://eblocks.stanford.edu/eblocks/seqsearch.html)
Blocks+	http://blocks.fhcrc.org/blocks/blocks_search.html
SMART	http://smart.embl-heidelberg.de/
Pfam	http://www.sanger.ac.uk/Software/Pfam/search.shtml
InterProScan	http://www.ebi.ac.uk/interpro/scan.html
ScanProsite	http://www.expasy.ch/tools/scanprosite/
Profilescan	http://hits.isb-sib.ch/cgi-bin/PFSCAN
PubMed	http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?CMD=&DB=PubMed
Psi-Blast	http://www.ncbi.nlm.nih.gov/blast/Blast.cgi
GrowTree	http://pmgm2.stanford.edu/gcg-bin/seqweb.cgi
ClustalW	http://decypher.stanford.edu
LAMA	http://blocks.fhcrc.org/blocks-bin/LAMA_search.sh
ALIONment	http://motif.Stanford.EDU/alion/
GorIV	http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html
Fugue	http://www-cryst.bioc.cam.ac.uk/~fugue/prfsearch.html
LIBRA	http://www.ddbj.nig.ac.jp/E-mail/libra/LIBRA_I.html

Appendix B: Multiple alignment of Follistatin sequences

```
gi|15214003zebrafish              --------------------------------------------------
gi|1706914cow                     --------------------------------------------------
gi|1346041mouse                   --------------------------------------------------
gi|399513sheep                    --------------------------------------------------
gi|120549rat                      --------------------------------------------------
gi|120547human                    --------------------------------------------------
gi|17981441drosophilamelanogas    MSKTSAATATQLARATSRECRTILGLFIVSAWSALLELANCLRQDENRYV
gi|1079405chicken                 --------------------------------------------------
gi|627265xenopus                  --------------------------------------------------
gi|3062845horse                   --------------------------------------------------
gi|120548pig                      --------------------------------------------------


gi|15214003zebrafish              --------------------------------------------------
gi|1706914cow                     --------------------------------------------------
gi|1346041mouse                   --------------------------------------------------
gi|399513sheep                    --------------------------------------------------
gi|120549rat                      --------------------------------------------------
gi|120547human                    --------------------------------------------------
gi|17981441drosophilamelanogas    GHGRDVNASSDASDEVSMAREQQQHQRQHSKHSASSSPAKSMRCTRAKAA
gi|1079405chicken                 --------------------------------------------------
gi|627265xenopus                  --------------------------------------------------
gi|3062845horse                   --------------------------------------------------
gi|120548pig                      --------------------------------------------------


gi|15214003zebrafish              --------------------------------------------------
gi|1706914cow                     --------------------------------------------------
gi|1346041mouse                   --------------------------------------------------
gi|399513sheep                    --------------------------------------------------
gi|120549rat                      --------------------------------------------------
gi|120547human                    --------------------------------------------------
gi|17981441drosophilamelanogas    SASTGTAAATATLGAAAAEGATSGHQLGGQPLMALLIGLLLLNFRLTAAG
gi|1079405chicken                 --------------------------------------------------
gi|627265xenopus                  --------------------------------------------------
gi|3062845horse                   --------------------------------------------------
gi|120548pig                      --------------------------------------------------


gi|15214003zebrafish              --------------------------------------------------
gi|1706914cow                     --------------------------------------------------
gi|1346041mouse                   --------------------------------------------------
gi|399513sheep                    --------------------------------------------------
gi|120549rat                      --------------------------------------------------
gi|120547human                    --------------------------------------------------
gi|17981441drosophilamelanogas    TCWQTHLGSGKCGQVFSTDISRSECCGSSQSFSYTDRELSSVEYFFATAI
gi|1079405chicken                 --------------------------------------------------
gi|627265xenopus                  --------------------------------------------------
gi|3062845horse                   --------------------------------------------------
gi|120548pig                      --------------------------------------------------


gi|15214003zebrafish              --------------------------------------------------
gi|1706914cow                     --------------------------------------------------
gi|1346041mouse                   --------------------------------------------------
gi|399513sheep                    --------------------------------------------------
gi|120549rat                      --------------------------------------------------
gi|120547human                    --------------------------------------------------
gi|17981441drosophilamelanogas    GGGVECSPCMESCKGFKCGPNKKCVKRKGRPKCVCAPECGAALRRRTHQQ
gi|1079405chicken                 --------------------------------------------------
gi|627265xenopus                  --------------------------------------------------
gi|3062845horse                   --------------------------------------------------
```

```
gi|120548pig                        --------------------------------------------------


gi|15214003zebrafish                --------------------------------------------------
gi|1706914cow                       --------------------------------------------------
gi|1346041mouse                     --------------------------------------------------
gi|399513sheep                      --------------------------------------------------
gi|120549rat                        --------------------------------------------------
gi|120547human                      --------------------------------------------------
gi|17981441drosophilamelanogas      ELELEQEPELEEEQQEMKPPRESRSLSSKNTNFNLGLDGGQRQRADNQRK
gi|1079405chicken                   --------------------------------------------------
gi|627265xenopus                    --------------------------------------------------
gi|3062845horse                     --------------------------------------------------
gi|120548pig                        --------------------------------------------------


gi|15214003zebrafish                --------------------------------------------------
gi|1706914cow                       --------------------------------------------------
gi|1346041mouse                     --------------------------------------------------
gi|399513sheep                      --------------------------------------------------
gi|120549rat                        --------------------------------------------------
gi|120547human                      --------------------------------------------------
gi|17981441drosophilamelanogas      LQPRESKHRRLLIIDSSSRSSSNLPDQPTSGRRGRVEMTGYERRRHRNNH
gi|1079405chicken                   --------------------------------------------------
gi|627265xenopus                    --------------------------------------------------
gi|3062845horse                     --------------------------------------------------
gi|120548pig                        --------------------------------------------------


gi|15214003zebrafish                ---------------MLRMLKRQQLHPGMILLLFWLCYLIEDQKVQAGNC
gi|1706914cow                       -----------------MARPRHQPGGLCLLLLLLCQFMEDRSAQAGNC
gi|1346041mouse                     -----------------MVCARHQPGGLCLLLLLLCQFMEDRSAQAGNC
gi|399513sheep                      ------------------------PGGVCLLLLLLCQFMEDRSAQAGNC
gi|120549rat                        -----------------MVCARHQPGGLCLLLLLLCQFMEDRSAQAGNC
gi|120547human                      -----------------MVRARHQPGGLCLLLLLLCQFMEDRSAQAGNC
gi|17981441drosophilamelanogas      GKHLTRSMTTGANDSFPADKTPAQSPAADSILAQSRLANLANANEPANDI
gi|1079405chicken                   -----------------MLNQRIHPGML-VLLMFLYHFMEDHTAQAGNC
gi|627265xenopus                    -----------------MLNERIQPGMIFLLTVSLCHFMEYRAVQAGNC
gi|3062845horse                     -----------------MVRPRHQPGGLCLLLLLLCQFMEDRSAQAGNC
gi|120548pig                        -----------------MVRPKHQPGGLCLLLLLLCQFMEDRSAQAGNC
                                                      :*       :      *.:

gi|15214003zebrafish                WLQQGKNGRCQVLYMPGMSREECCRSGRLGTSWTEE------DVPNSTLF
gi|1706914cow                       WLRQAKNGRCQVLYKTELSKEECCSTGRLSTSWTEE------DVNDNTLF
gi|1346041mouse                     WLRQAKNGRCQVLYKTELSKEECCSTGRLSTSWTEE------DVNDNTLF
gi|399513sheep                      WLRQAKNGRCQVLYKTELSKEECCSTGRLSTSWTEE------DVNDNTLF
gi|120549rat                        WLRQAKNGRCQVLYKTELSKEECCSTGRLSTSWTEE------DVNDNTLF
gi|120547human                      WLRQAKNGRCQVLYKTELSKEECCSTGRLSTSWTEE------DVNDNTLF
gi|17981441drosophilamelanogas      NRQSSTRVRHQHARRIEHAPNPDNGLRKRQHKQQQQHQHQHQDRMNTEAQ
gi|1079405chicken                   WLRQARNGRCQVLYKTDLSKEECCKSGRLTTSWTEE------DVNDNTLF
gi|627265xenopus                    WLQQSKNGRCQVLYRTELSKEECCKTGRLGTSWTEE------DVPNSTLF
gi|3062845horse                     WLRQAKNGRCQVLYKTELSKEECCSTGRLSTSWTEE------DVNDNTLF
gi|120548pig                        WLRQAKNGRCQVLYKTELSKEECCSTGRLSTSWTEE------DVNDNTLF
                                      :.. . * *       : :      :   .  ::      *  :.

gi|15214003zebrafish                RWMIFNGGAPNCIPCKETCDNVDCGPGKRCKMNRRSKPRCVCAPDCSNVT
gi|1706914cow                       KWMIFNGGAPNCIPCKETCENVDCGPGKKCRMNKKNKPRCVCAPDCSNIT
gi|1346041mouse                     KWMIFNGGAPNCIPCKETCENVDCGPGKKCRMNKKNKPRCVCAPDCSNIT
gi|399513sheep                      KWMIFNGGAPNCIPCKETCENVDCGPGKKCRMNKKNKPRCVCAPDCSNIT
gi|120549rat                        KWMIFNGGAPNCIPCKETCENVDCGPGKKCRMNKKNKPRCVCAPDCSNIT
gi|120547human                      KWMIFNGGAPNCIPCKETCENVDCGPGKKCRMNKKNKPRCVCAPDCSNIT
gi|17981441drosophilamelanogas      SANTTVSGSGSGATNHRRISQLQHGSETAASSDLANTHDLAHLGGIYAPL
gi|1079405chicken                   KWMIFNGGAPNCIPCKETCENVDCGPGKKCKMNKKNKPRCVCAPDCSNIT
gi|627265xenopus                    KWMIFHGGAPHCIPCKETCENVDCGPGKKCKMNKKNKPRCVCAPDCSNIT
```

```
gi|3062845horse                    KWMIFNGGAPNCIPCKETCDNVDCGPGKKCRMNKKNKPRCVCAPDCSNIT
gi|120548pig                       KWMIFNGGAPNCIPCKETCENVDCGPGKKCRMNKKNKPRCVCAPDCSNIT
                                         .*:    . :.  .::: *. . .  :  ..   .   .

gi|15214003zebrafish               WK---GPVCGSDGKTYRDECALLKSKCK-GHPDLEVQYQGKCKKTCRDVL
gi|1706914cow                      WK---GPVCGLDGKTYRNECALLKARCK-EQPELEVQYQGKCKKTCRDVF
gi|1346041mouse                    WK---GPVCGLDGKTYRNECALLKARCK-EQPELEVQYQGKCKKTCRDVF
gi|399513sheep                     WK---GPVCGLDGKTYRNECALLKARCK-EQPELEVQYQGKCKKTCRDVF
gi|120549rat                       WK---GPVCGLDGKTYRNECALLKARCK-EQPELEVQYQGKCKKTCRDVF
gi|120547human                     WK---GPVCGLDGKTYRNECALLKARCK-EQPELEVQYQGRCKKTCRDVF
gi|17981441drosophilamelanogas     PPQHSNPVCGTDGRTYNTECQLRKRACRTNNAQLEVAYRGHCKNSCSGVH
gi|1079405chicken                  WK---GPVCGLDGKTYRNECALLKARCK-EQPELEVQYQGKCKKTCRDVL
gi|627265xenopus                   WK---GSVCGIDGKTYKDECALLKAKCK-GVPELDVQYQGKCKKTCRDVL
gi|3062845horse                    WK---GPVCGLDGKTYRNECALLKARCK-EQPELEVQYQGKCKKTCRDVN
gi|120548pig                       WK---GPVCGLDGKTYRNECALLKARCK-EQPELEVQYQGKCKKTCRDVF
                                        ..*** **:**. ** * *  *:   .:*:* *:*:**::* .*

gi|15214003zebrafish               CPGSSTCVVDQTNNAYCVTCNRICP--------EVMSPDQYLCGNDGIIY
gi|1706914cow                      CPGSSTCVVDQTNNAYCVTCNRICP--------EPTSSEQYLCGNDGVTY
gi|1346041mouse                    CPGSSTCVVDQTNNAYCVTCNRICP--------EPSSSEQYLCGNDGVTY
gi|399513sheep                     CPGSSTCVVDQTNNAYCVTCNRICP--------EPTSSEQYLCGNDGVTY
gi|120549rat                       CPGSSTCVVDQTNNAYCVTCNRICP--------EPSSSEQSLCGNDGVTY
gi|120547human                     CPGSSTCVVDQTNNAYCVTCNRICP--------EPASSEQYLCGNDGVTY
gi|17981441drosophilamelanogas     CLNGLTCVEDQYLMPHCIACRIECPWDNLDVDSSGYDERQAVCGVDGKTY
gi|1079405chicken                  CPGSSTCVVDQTNNAYCVTCNRICP--------EPTSPEQYLCGNDGITY
gi|627265xenopus                   CPGSSSCVVDQTNNAYCVTCNRICP--------EPTSPDQYLCGNDGITY
gi|3062845horse                    CPGSSTCVVDQTNNAYCVTCNRICP--------EPTSSEQYLCGNDGVTY
gi|120548pig                       CPGSSTCVVDQTNNAYCVTCNRICP--------EPTSSEQYLCGNDGVTY
                                   * .. :** **   .:*::*.  **        .  .  * :** **  *

gi|15214003zebrafish               ASACHLRRATCLLGRSIGVAYEGKCIKAK-SCDDIHCSAGKKCLWDAKMS
gi|1706914cow                      PSACHLRKATCLLGRSIGLAYEGKCIKAK-SCDDIQCTGGKKCLWDFKVG
gi|1346041mouse                    SSACHLRKATCLLGRSIGLAYEGKCIKAK-SCEDIQCGGGKKCLWDSKVG
gi|399513sheep                     PSACHLRKATCLLGRSIGLAYEGKCIKAK-SCEDIQCTGGKKCLWDFKVG
gi|120549rat                       SSACHLRKATCLLGRSIGLAYEGKCIKAK-SCEDIQCGGGKKCLWDFKVG
gi|120547human                     SSACHLRKATCLLGRSIGLAYEGKCIKAK-SCEDIQCTGGKKCLWDFKVG
gi|17981441drosophilamelanogas     RSACDINRMICKIGRSIAVAYPGPCRAGRVSCADIKCGPKDNCLVDLQTR
gi|1079405chicken                  ASACHLRKATCLLGRSIGLAYEGKCIKAK-SCEDIQCSAGKKCLWDFKVG
gi|627265xenopus                   GSACHLRKATCLLGRSIGLAYEGKCIKAK-SCEDIQCSAGKKCLWDSRVG
gi|3062845horse                    SSACHLRKATCLLGRSIGLAYEGKCIKAK-SCEDIQCTGGKKCLWDFKVG
gi|120548pig                       SSACHLRKATCLLGRSIGLAYEGKCIKAK-SCEDIQCTGGKKCLWDFKVG
                                    ***.:.:  * :****.:** * *  .: ** **:*   .:** * :

gi|15214003zebrafish               RGRCAVCAESCP--ESRSEEAVCASDNTTYPSECAMKQAACSLGVLLEVK
gi|1706914cow                      RGRCSLCGELCP--ESKSEEPVCASDNATYASECAMKEAACSSGVLLEVK
gi|1346041mouse                    RGRCSLCDELCP--DSKSDEPVCASDNATYASECAMKEAACSSGVLLEVK
gi|399513sheep                     RGRCSLCGELCP--ESKSEEPVCASDNATYASECAMKEAACSSGVLLEVK
gi|120549rat                       RGRCSLCDELCP--DSKSDEPVCASDNATYASECAMKEAACSSGVLLEVK
gi|120547human                     RGRCSLCDELCP--DSKSDEPVCASDNATYASECAMKEAACSSGVLLEVK
gi|17981441drosophilamelanogas     QPRCVTCRYKCPRKQQRPVHKICGYNNQTYNSWCEMHKHSCESRYFIGVK
gi|1079405chicken                  RGRCALCDELCP--ESKSDEAVCASDNTTYPSECAMKEAACSMGVLLEVK
gi|627265xenopus                   RGRCALCDDLCG--ESKSDDTVCASDNTTYPSECAMKQAACSTGILLEVK
gi|3062845horse                    RGRCSLCDELCP--DSKSEEPVCASDNATYASECAMKEAACSSGVLLEVK
gi|120548pig                       RGRCSLCDELCP--ESKSEEPVCASDNATYASECAMKEAACSSGVLLEVK
                                   : **  *   *   :.:. . :*. :* ** * * *:: :*.   :: **

gi|15214003zebrafish               HSGSCNCK-------------------------
gi|1706914cow                      HSGSCNSISEDTEDEEEDEDQDYSFPISSILEW
gi|1346041mouse                    HSGSCNSISEETEEEEEEEDQDYSFPISSILEW
gi|399513sheep                     HSGSCNSISEDTEDEEEDEDQDYSFPISSILEW
gi|120549rat                       HSGSCNSISEETEEEEEEEDQDYSFPISSTLEW
gi|120547human                     HSGSCN------EEEEEDEDQDYSFPISSILEW
gi|17981441drosophilamelanogas     SQGSC---------------------------
gi|1079405chicken                  HSGSCNSINEDPEEEEEDEDQDYSFPISSILEW
```

```
gi|627265xenopus                HSGSCNCK------------------------
gi|3062845horse                 HSGSCNSISEDTEEEEEDEDQDYSFPISSILEW
gi|120548pig                    HSGSCNSISEDTEEEEEDEDQDYSFPISSILEW
                                 .***
```