

Research article

The SGS3 protein involved in PTGS finds a family

Alex Bateman

Address: The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA, UK

E-mail: agb@sanger.ac.uk

Published: 5 August 2002

Received: 11 July 2002

BMC Bioinformatics 2002, 3:21

Accepted: 5 August 2002

This article is available from: <http://www.biomedcentral.com/1471-2105/3/21>

© 2002 Bateman; licensee BioMed Central Ltd. This article is published in Open Access: verbatim copying and redistribution of this article are permitted in all media for any non-commercial purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Post transcriptional gene silencing (PTGS) is a recently discovered phenomenon that is an area of intense research interest. Components of the PTGS machinery are being discovered by genetic and bioinformatics approaches, but the picture is not yet complete.

Results: The gene for the PTGS impaired Arabidopsis mutant *sgs3* was recently cloned and was not found to have similarity to any other known protein. By a detailed analysis of the sequence of SGS3 we have defined three new protein domains: the XH domain, the XS domain and the zf-XS domain, that are shared with a large family of uncharacterised plant proteins. This work implicates these plant proteins in PTGS.

Conclusion: The enigmatic SGS3 protein has been found to contain two predicted domains in common with a family of plant proteins. The other members of this family have been predicted to be transcription factors, however this function seems unlikely based on this analysis. A bioinformatics approach has implicated a new family of plant proteins related to SGS3 as potential candidates for PTGS related functions.

Background

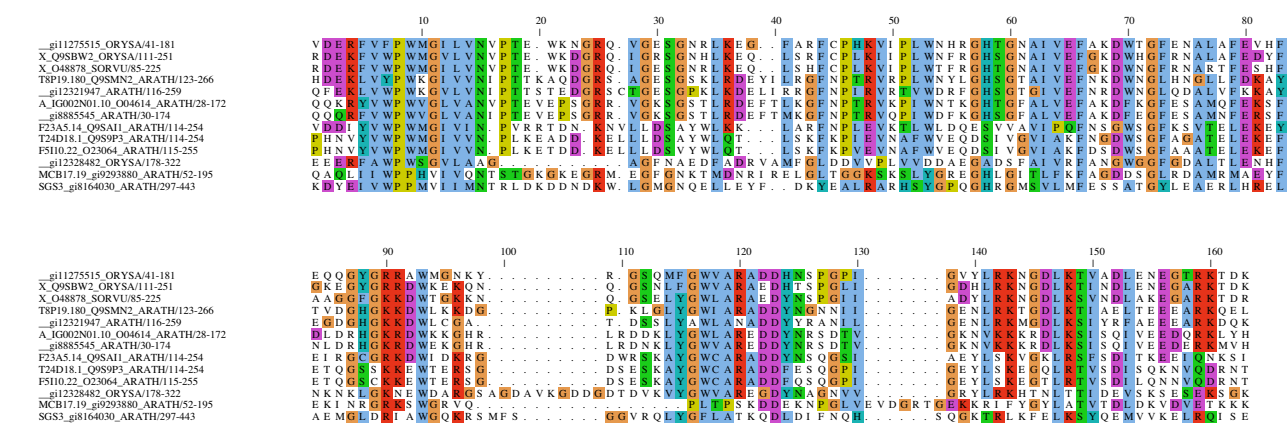
Post transcriptional gene silencing (PTGS) is a recently discovered phenomenon [1]. The components of PTGS are being cloned and experiment combined with sequence analysis is helping to elucidate its mechanisms. Study of PTGS is providing links between diverse biological processes such as defence against viruses, RNA metabolism [2,3] and development [4]. The gene for the PTGS impaired Arabidopsis mutant *sgs3* was recently cloned [5]. An initial analysis of the protein did not reveal any motifs, domains or similarity to any other protein. To help shed light on the function of SGS3 a more detailed analysis of the protein has been carried out.

Results

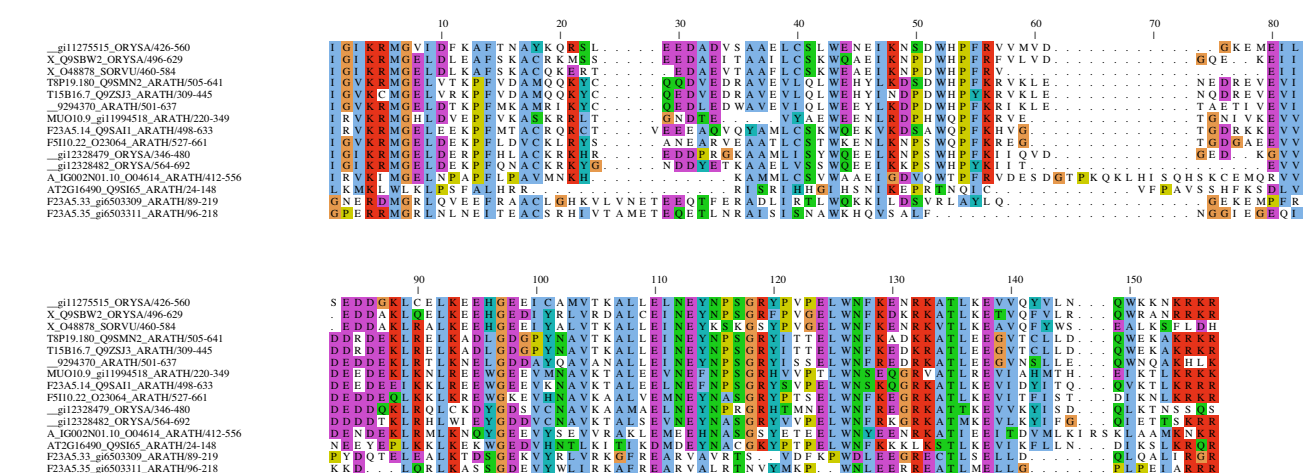
After initial PSI-BLAST searches with the sequence of SGS3, weak matches were found to a number of plant pro-

teins. Reciprocal matches can often verify the significance of weak matches. Using residues 85 to 225 of a weakly matching *Sorghum bicolor* protein (SWISSPROT accession O48878) as a PSI-BLAST [6] query at the NCBI site, using an inclusion E-value of 0.002, SGS3 was found in the second round with an E-value of 0.001. This search also found a number of other plant proteins including the rice gene X product (also known as gene X1) [7].

I have termed the main region of similarity the XS domain after 'rice gene X and SGS3'. This presumed domain is around 140 amino acid residues in length (see figure 1). The XS domain contains a completely conserved aspartate that could suggest an enzymatic active site. Prediction of the secondary structure using the Jnet server [8] suggests a mixed alpha and beta structure.

**Figure 1**

Multiple sequence alignments of the XS domain. Figures have been generated using the Jalview program written by Michele Clamp. The protein identifiers are given as name_accession number_species/start-end. The five letter species designations are those used in SWISS-PROT. Alignments are colored using the ClustalX scheme in Jalview (orange: glycine (G); gold: Proline (P); blue: small and hydrophobic amino-acids (A, V, L, I, M, F, W); green: hydroxyl and amine amino-acids (S, T, N, Q); magenta: negative-charged amino-acids (D, E); red: positive-charged amino-acids (R, K); dark-blue: histidine (H) and tyrosine (Y)).

**Figure 2**

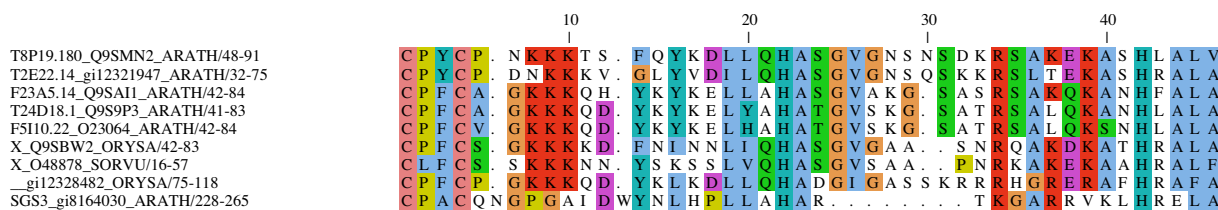
Multiple sequence alignments of the XH domain. See caption of figure 1 for details

After initial alignment of the protein sequences the DNA sequence for each was inspected for possible frameshift errors and incorrect splicing boundaries using tblastn and genewise.

The XS domain containing proteins are predicted by ncoils [9] to contain coiled-coils, which suggests that they will oligomerise. Most coiled-coil proteins form either a dimeric or a trimeric structure. It is possible that different

members of the XS domain family could oligomerise via their coiled-coils forming a variety of complexes.

Analysis of the C-terminal region after the central coiled-coil region of rice gene X identifies a second region of conservation termed the XH domain, for 'rice gene X Homology'. The XH domain is between 124 and 145 residues in length. All the members can be found with any XH domain sequences as a PSI-BLAST query. XH domains exist

**Figure 3**

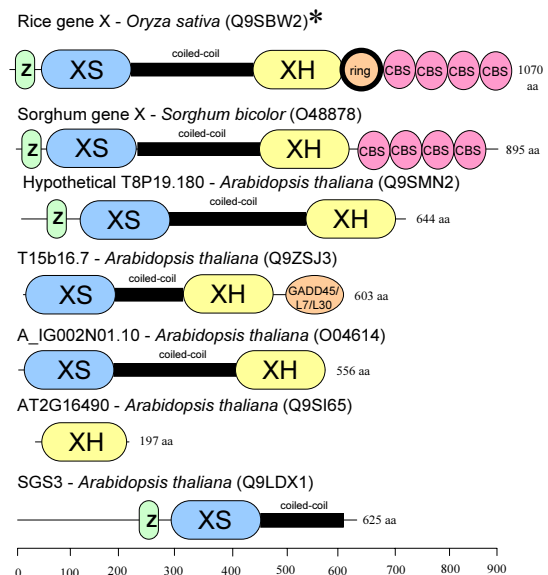
Multiple sequence alignments of zf-XS domain. An alignment of SGS3-type C2H2 zinc binding domain. See caption of figure 1 for details

in some proteins that do not contain an XS domain, for example AT2G16490 from *Arabidopsis thaliana*. Figure 2 shows an alignment of this presumed domain and figure 4 shows the complete domain organisation of these proteins. The XH domain was not found in the SGS3 protein. As the XS and XH domains are fused in most of these proteins, these two domains may interact. The XS domain of SGS3 may also interact with XH domains of other proteins. The XH domain contains one completely conserved glutamate that could potentially be part of an active site or other functionally important region.

A global alignment of the full length of the sequences with an XS domain using the T-Coffee alignment program [10], suggested that many of the proteins contained an N-terminal cysteine/histidine cluster. An alignment of the N-terminal cluster is shown in figure 3. This pattern of conservation suggests a zinc binding domain. Although SGS3 is included in the alignment and conserves the putative zinc ligands, there is no statistical support for its inclusion with standard methods. However given the conservation pattern and presence of the shared XS domain and coiled-coil it seems likely that this is an evolutionarily conserved domain. The rice gene X homologues conserve several lysines within the putative zinc binding domain that suggest it may be a nucleic acid binding domain. An RNA binding function would seem plausible if this larger family of proteins were involved in PTGS as the SGS3 protein appears to be.

Discussion

Can we infer the function of SGS3 based on its similarity to other XS domain proteins? Unfortunately the members of this family are functionally uncharacterised. However, rice gene X was predicted by Chen *et al.* to be a transcription factor based on two pieces of evidence: (i) the presence of the coiled-coil as found in other transcription factors such as GCN4 [11], (ii) the rice gene X contains a ring zinc finger [7]. However, many coiled-coils are found in non-transcription factor proteins weakening the first ar-

**Figure 4**

A schematic figure showing the architectures of proteins containing XS domains. More information about the ring finger domain, the CBS domain and the GADD45/L7/L30 domain can be found in the Pfam database [13] at [http://www.sanger.ac.uk/Software/Pfam] with accession numbers PF00097, PF00571 and PF01248 respectively. The asterisk denotes where the domain architecture shown is not from ring finger in rice gene X (Q9SBW2) sequence deposited in the protein database but is that found in the manuscript by Chen *et al.* [7]. The XS, XH and the zinc finger found in SGS3 have been submitted to Pfam and given accession numbers PF03468, PF03469, PF03470 respectively.

gument. Ring fingers are now thought to mediate the protein interactions of ubiquitin ligases rather than interact with DNA [12]. The Rice gene X is the only member of this family that contains a ring finger. This second piece of evidence no longer points to a transcription factor function,

but potentially to a role in ubiquitination. Therefore the evidence used to infer these protein are transcription factors is weak and the inference unlikely to be correct.

Conclusions

In summary this analysis suggests that SGS3 may have a nucleic acid binding function and that a large family of plant proteins containing the novel XS and XH domains may be uncharacterised components in PTGS.

Acknowledgements

AB is supported by the Wellcome Trust. I would like to thank William Mifsud, Richard Durbin for comments on the manuscript.

References

1. Finnegan EJ, Wang M, Waterhouse P: **Gene silencing: fleshing out the bones.** *Curr Biol* 2001, **11**:R99-R102
2. Hutvagner G, McLachlan J, Pasquinelli AE, Balint E, Tuschl T, Zamore PD: **A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA.** *Science* 2001, **293**:834-8
3. Grishok A, Pasquinelli AE, Conte D, Li N, Parrish S, Ha I, Baillie DL, Fire A, Ruvkun G, Mello CC: **Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing.** *Cell* 2001, **106**:23-34
4. Knight SW, Bass BL: **A Role for the RNase III Enzyme DCR-1 in RNA Interference and Germ Line Development in C. elegans.** *Science* 2001, **2**:2
5. Mourrain P, Beclin C, Elmayan T, Feuerbach F, Godon C, Morel JB, Jouette D, Lacombe AM, Nikic S, Picault N, et al: **Arabidopsis SGS2 and SGS3 genes are required for posttranscriptional gene silencing and natural virus resistance.** *Cell* 2000, **101**:533-542
6. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucl Acids Res* 1997, **25**:3389-3402
7. Chen M, Bennetzen JL: **Sequence composition and organization in the Sh2/Al-homologous region of rice.** *Plant Mol Biol* 1996, **32**:999-1001
8. Cuff JA, Barton GJ: **Application of multiple sequence alignment profiles to improve protein secondary structure prediction.** *Proteins* 2000, **40**:502-511
9. Lupas A, Van Dyke M, Stock J: **Predicting coiled coils from protein sequences.** *Science* 1991, **252**:1162-1164
10. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302**:205-217
11. O'Shea EK, Klemm JD, Kim PS, Alber T: **X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil.** *Science* 1991, **254**:539-544
12. Joazeiro CA, Weissman AM: **RING finger proteins: mediators of ubiquitin ligase activity.** *Cell* 2000, **102**:549-52
13. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer ELL: **The Pfam protein families database.** *Nucleic Acids Res* 2000, **28**:263-266

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright

Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>



BioMedcentral.com

editorial@biomedcentral.com