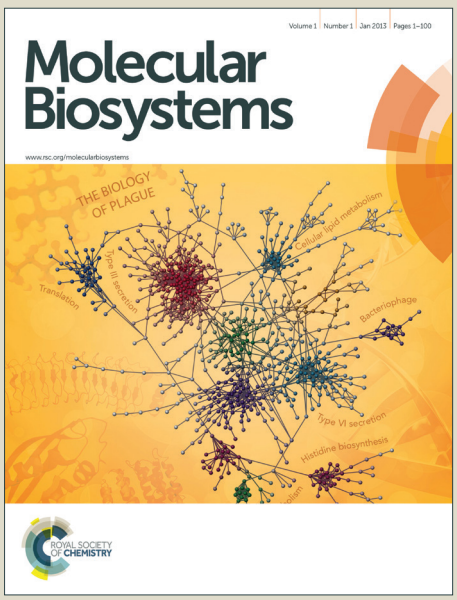
# Molecular BioSystems

Accepted Manuscript





ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

# Identification and molecular structure analysis of a new noncoding RNA, a sbRNA homologous, in the silkworm *Bombyx mori* genome

Francisco Ferreira Duarte Junior[1], Quirino Alves de Lima Neto[1], Fabiana dos Santos Rando[1], Douglas Vinícius Bassalobre de Freitas[1], José Renato Pattaro Júnior[1], Lorena Gomes Polizelli[1], Roxelle Ethienne Ferreira Munhoz[1], Flavio Augusto Vicente Seixas[2] and Maria Aparecida Fernandez[1*].

[1]Departamento de Biotecnologia, Genética e Biologia Celular, [2]Departamento de Bioquímica, Universidade Estadual de Maringá, Av. Colombo 5.790, 87020-900, Maringá, Paraná, Brasil

* Author to whom correspondence should be addressed:

Maria Aparecida Fernandez

Departamento de Biotecnologia, Genética e Biologia Celular

Universidade Estadual de Maringá

Av. Colombo, 5790, 87020-900

Maringá, Paraná, Brasil

Phone: 55 (44) 3011 5398;

E-mail address: mafernandez@uem.br; aparecidafernandez@gmail.com

**Abstract**

The small noncoding group of RNAs called stem-bulge RNAs (sbRNAs), first reported in *Caenorhabditis elegans*, is described as molecules homologous to the Y RNAs, a specific class of noncoding RNAs that is present in vertebrates. This homology indicates the possibility of the existence of sbRNAs in other invertebrate organisms. In this work, we used bioinformatic tools and conserved sequences of sbRNAs from *C. Elegans* and Y RNAs to search for homologous sbRNA sequences in the *Bombyx mori* genome. This analysis led to the discovery of one noncoding gene, which was translated into RNA segments and comparatively analysed with segments from human and hamster Y RNAs and *C. elegans* sbRNAs in molecular dynamic simulations. This gene represents the first evidence of a new sbRNA-like noncoding RNA, the *Bm*sbRNA gene, in this Lepidoptera genome.

Key words: sbRNA, *Bombyx mori*, noncoding RNA, *Bm*sbRNA

## Introduction

The stem-bulge RNAs (sbRNAs) were first described by constructing a cDNA library specific for noncoding RNAs (ncRNAs).[1] The characteristic secondary structure and the absence of sequences homologous to other RNAs yielded a place for sbRNAs to be a new family of ncRNAs. Though more than a dozen sbRNAs have been described, a solid characterisation of their functionality has not yet been described.[1,2,3] Some authors have suggested the possibility that sbRNA genes, which are found in *C.elegans,* are homologous to the Y RNAs, a specific class of ncRNAs that is present in vertebrates.[3,4]

Human Y RNAs (hY RNAs) were originally found as a particle of the Ro ribonucleoproteins (RNPs) that result from the disassociation of Y RNAs with the Ro 60 protein and can be detected by the auto antibodies present in the serum of patients with connective tissue disease.[5,6,7] The RNA polymerase III transcribes the Y RNAs from four genes in humans and two genes in rodents.[4,8] Y RNAs are overexpressed in a variety of human tumours[9] and can be detected circulating in the blood of humans, which suggests that they can be useful as biomolecular markers.[10] In recent years, noncoding RNAs have been described as regulators in diverse cellular processes.[11,12,13] The licensing of human DNA replication requires the participation of Y RNAs,[5,8,14] several proteins that act as factors in the initiation and DNA replication,[15,16] as was proposed before, could be influenced by factors related to chromatin structure.[17]

The absence of reports concerning the presence of sbRNAs in insects has led us in the attempt to identify new ncRNAs in *Bombyx mori*, because it is a model organism for studies in molecular and cellular biology.[18,19] This

Lepidoptera has a rich repertoire of genetic information on the mutations that affect its development, morphology and behaviour.[20] *B. mori* have also been used to produce recombinant protein,[21,22,23] and have been used as bioreactors,[24,25,26] and a source of silk biomaterials.[27] The pupa could also be used to remove heavy metals from contaminated waters.[28]

Silkworm breeding is important in the Brazilian industry of silk production. However, the cocoon productivity is affected by the Bombyx mori *Nucleopolyhedrovirus* infection.[29] The baculovirus infection and the involvement of noncoding RNAs rely on the importance of the DNA replication in these systems.[30] Therefore, the aim of this study is to use bioinformatics tools, molecular dynamics and molecular biology analysis to investigate the presence of noncoding RNAs that may belong to the group of sbRNAs and Y RNAs in the *B. mori* genome.

## Results and discussion

### Homology search and structure modelling

The search through BLASTn, using the conserved sequences (Fig. 1A), resulted in only one sequence (GTGGCTTATC) in the *B. mori* genome containing one of 16 expected nucleotide sequences, which is conserved in the sbRNA of *C. elegans,*[1] gene CeN76 and in the human hY5 and hamster chY5 genes (Fig. 1B and C). The sequence found in *B. mori* belongs to a 57-bp gene identified as *Bm*sbRNA, which presents both promoter and stopping sequences (Supplementary Figure 1S), according as described elsewhere.[31] The Multiple Sequence Alignment (Fig. 1D) has shown homology, primarily in the conserved domain.

To generate the 2D structures of the RNA segments, we first validated a methodological protocol for this purpose. The parameters and defaults of the MFOLD server[32] were set to generate an identical representation of the human Y1 RNA segment.[5,14] Once the protocol was validated, all of the other segments were generated using the same protocol. The 2D RNA structures generated from the segments of the four genes submitted were drawn using the Visualization Applet for RNA (VARNA),[33] (Fig. 2).

The 2D models shown in Fig. 2 present the following theoretical values of $\Delta G_{folding}$ in kcal·mol$^{-1}$: –26.10 (human hY5 RNA), –25.10 (hamster chY5 RNA), –8.40 (*B. mori*) and –30.80 (*C. elegans*). The highest value for the *B. mori* segment is due to its smaller size[32]. From these results, the structures were generated in the 3D format (.pdb files as additional data) required for the molecular dynamics simulations.

The finding of a promising sbRNA in the *B. mori* genome has led us to investigate the expression of this sbRNA in the total RNA samples from a pool of larva tissues. Using cDNA made with random primers, we carried out PCR using specific primers (Fig. 3A). After cloning and sequencing, we had a confirmation of the expected sequence (Fig. 3B), which indicates that this *Bm*sbRNA is expressed in *B. mori* larva tissues.

**Molecular dynamics**

Performing molecular dynamics simulations of RNA segments is not a very common procedure. One reason is due to the conformational diversity of these molecules, a fact that generates randomness in the results, there by complicating its interpretation. However, since the sbRNAs found in this work

presented a similar conformation and displayed a defined pattern (Fig. 2), it was possible to establish a protocol for their simulation and analysis.

The Fig. 4A shows the Root Mean Square Deviation (RMSD) value calculated for each of the four RNA segments along the MD trajectory. The figure shows that all RNA segments reached equilibrium after four nanoseconds of simulation. Fig. 4B shows that all segment of RNA exhibit radius of gyration values with significant fluctuations along the trajectory. Nevertheless, these fluctuations were concentrated around an average, which indicates that not all RNA segments do unfold along simulation. The continuing trend of hydrogen bonds number along simulation also indicates the stability of the segments (Supplementary Fig. 2S). It is noticeable, in Fig. 4B, that the behaviours of the RNA segment of *B.mori* and *C. elegans* were the most stable, a fact that we attribute to their smaller sizes.

The analysis of the Root Mean Square Fluctuation (RMSF) of the $C_1$' carbons for each nucleotide (Fig. 5) shows that the regions containing the functional sequence GUG, i.e., the minimal domain responsible for the functionality of Y RNAs when licensing DNA replication, were the most stable during the period in which the system was at equilibrium (last two nanoseconds). This prediction was recently confirmed by spectroscopic analysis of the thermal denaturation of Y RNA segments.[34]

The analysis of Fig. 5 shows a similar behaviour from all RNA segments. The highest fluctuation arose from the residues placed in the centre of each graph, a region that corresponds to the central loop observed in the 2D structures (Fig. 2). The other regions of high fluctuation coincide with the

unpaired regions of RNA segments, a fact that implies a higher degree of freedom of movement in these regions along the simulation.

The ncRNAs and microRNAs (miRNAs) have received increasing attention as regulatory factors of cell growth and development.[35-37] A catalogue of several ncRNAs has been described in eukaryotes with approximately 400 different types of RNAs, including sbRNAs and Y RNA.[7,38]

The sbRNAs were first described in the nematode *C.elegans* by analysing a cDNA library specific for ncRNAs.[1] The nine novel transcripts presented in the literature, represent a new family of ncRNAs because of three common characteristics: 1) the absence of homologous sequences to other non-coding RNAs; 2) the peculiar structure that contains a chain double helix with a bulge in a short stretch of conserved sequence (AACUU); and 3) a loop of single-stranded RNA with varying size[1] (Figure 2C). Later, another nine new molecules were characterised as sbRNAs in *C. elegans.*[2,3] However, a solid characterisation of their functionality has not been described for these ncRNAs. Initially discovered in mammals,[39] the homology of sequence and structure to the Y RNAs, which are involved in the replication of DNA,[3] has been described, suggesting that sbRNAs may also be associated with this process.[14,40,41] The literature proposes that Y RNAs may be associated with different proteins within the cell,[4,42,43] which indicates that a mechanism of action remains unclear. However, it is known that the expression of Y RNAs is more intense in pathological situations such as autoimmune diseases[44,45] and viral infections.[39,46]

The role of sbRNAs in *C. elegans* is still not clear, but because of their homology to vertebrate Y RNAs, it is possible that sbRNAs exhibit similar

function, acting in *B. mori* and other invertebrates at the same processes, such as RNA processing[47] and DNA replication.[5,9] In insects, the presence of Y RNAs is unlikely; leading us to hypothesise that an alternative class of noncoding RNAs replaces the function of Y RNAs in insects. The *B. mori* sbRNA (*Bm*sbRNA) sequence presented in this work showed structural homology to sbRNAs and is a possible candidate to receive such a classification.

In *C. elegans*, 18 sequences of sbRNA were found,[3] and the computer analysis performed in this study resulted in only one valid string in *B. mori*, which presents a potential promoter and transcribes a RNA molecule with a secondary structure similar to the pattern described for sbRNAs previously.[3] Although the sequence may represent a homolog of sbRNAs in *B. mori*, it is likely that a different class of noncoding RNAs has a functional role similar to sbRNAs in such organisms to compensate for the shortage of this type of RNA.

The structural similarity of sbRNAs with Y RNAs was explored in this work using molecular dynamics simulations (MD), where the analysis of the trajectories showed that the RNA segments analysed maintained their folding along the simulations, which indicates stable three-dimensional structures. The biological significance of this behaviour can be interpreted as functional segments of RNA that could have a specific biological function because the structure of a macromolecule is directly related to its function.[48]

The regions of the RNA segments that correspond to the minimal domain related to small binding RNA groups were used in BLASTn to search for similar genes in *B. mori*. Those regions, which are marked in red in Figs. 2 and 5, also correspond to regions of lower fluctuation along the molecular dynamics

simulations. Recent biophysical analyses from *in vitro* Y RNA segments support this finding,[33] which is attributed, in the most part; to its pairing with a complementary region of this segment (blue marks in Figs. 2 and 5), which reduces the freedom of movement of these regions. However, this result also means that these regions of the segments are quite stable and suggests that they should exert a biological function, e.g., recognition of their target within the cell.

The behaviour of the RNA segments from *B. mori* along the simulation closely resembles the behaviour of the Y RNA segments of hamster and human, suggesting a possible functional and structural correlation. To strengthen the hypothesis that *B. mori* sbRNA can be homologous to Y RNAs, it is necessary to conduct a functional analysis to evaluate whether this noncoding RNA can substitute for the hY RNAs in a human cell-free system.[8] For this purpose, the universal primer was constructed with the promoter for SP6 RNA polymerase, and the *in vitro* expression experiment is currently underway.

The 3D structures (.pdb files as additional data) were generated from the 2D model with lower free energy proposed by MFold server. The 2D *B.mori* sbRNA representation with the lowest free energy was –8.40 kcal.mol$^{-1}$ (Fig. 2C) with the UUAUC segment forming a stem structure. However, the *B. mori* sbRNA representation with the second lowest free energy generated by MFold was –8.30 kcal.mol$^{-1}$. In this alternative structure, the UUAUC segment is inserted in the central region forming a single hairpin loop (Supplementary Fig. 3S) as might be expected for sbRNAs.[49,50] Note that the difference between the two representations is only 0.10 kcal.mol$^{-1}$, i.e., within a margin of error less

than 5%. However, in this second representation, which appeared to be the most appropriate, it was observed inaccuracies in the 3' terminal. Therefore, for the choice of the final representation for *B. mori* sbRNA, we adopted the criteria validated in our methodology and we chose the structure shown in Figure 2C as a representative model for *B. mori* sbRNA, so as to ensure reproducibility of the results.

Furthermore, these are theoretical models and we believe that *in vivo*, the structure that actually represents the sbRNA of *Bombyx mori* must be something between the two models of lower free energy generated by MFOLD server. To clarify this issue once, structural studies of NMR will be carried out and the results published in a future paper. However, the most important result of this work does not lie in what would be the representative model of *B. mori* sbRNA, but rather in the fact that this study shows for the first time that the gene for this sbRNA is expressed in a pool of tissues from *Bombyx mori* larvae.

## Experimental

### Homology search and structure modelling

A stretch of conserved sequence found in all sbRNAs and Y RNAs (GTGNNTTATC, where N can range between any nucleotide bases) has been compared with the genome of *B. mori* in the NCBI database using BLASTn.[51] The sequences of the genes for *C. elegans* sbRNA (CeN76), human Y5 RNA (hY5 RNA), and Chinese hamster *(Cricetulus griseus)* Y5 RNA (chY5 RNA) were obtained from the NCBI Gene Bank. The obtained sequences were validated by comparing the sequences described for sbRNAs and Y RNAs

using ClustalW2[52] and WebLogo graphical representation.[53,54] The Vienna files were obtained from the mfold web server[32] using the gene sequences from human and hamster Y RNAs and *B. mori* and *C. elegans* sbRNAs. Those structures, which had a lower theoretical value of free energy, were selected to provide the 2D model structures and were drawn using VARNA.[33] The output file was then submitted to RNA Composer 3 server to generate the 3D structures. The nomenclature of the residues was then edited according to the topology for ribonucleotides, as described by the CHARMM c35b2-c36a24 force fields.[41]

**Molecular dynamics**

The software packages VMD[55] and NAMD2[56] were used to perform the molecular dynamics (MD) simulations. The equilibration MD was carried out in steps. Initially, the 3D structures were immersed in a water box simulating periodic boundary conditions and containing sufficient sodium counter ions to neutralise the system charges. Then, all systems were minimised by 20,000 steps of conjugate gradient. In the second step, the atoms of the nucleotides were fixed in space, and the water and salts were equilibrated for 60 ps. In the third step, all atoms of the system were minimised again by new 20,000 steps. In the fourth and last step, the whole system was equilibrated for 10 ns under conditions of constant pressure (1 atm) and temperature (300 K) (NPT ensemble).

The electrostatic interactions were calculated using the particle-mesh Ewald algorithm with a grid spacing of 1.0 Å, and the cut-off radius for van der Waals interactions was set to 10 Å. The integration time for the equations of

motion was 2 fs per step. The trajectory of the dynamics was generated in a file from the system settings extracted every 2 ps along the simulation time.

The simulations were analysed from the trajectory files in terms of RMSD and radius of gyration ($R_{gyr}$) with reference to all the atoms of nucleotides compared with the structure minimised in the third step. The last ten nucleotides of the 3' end segment from *C. elegans*, human and hamster and the last three nucleotides of the *B. mori* segment were excluded from these analyses due to their high degree of freedom of movement. The RMSF for the C1' atoms of each nucleotide were extracted from the last two nanoseconds of simulation.

The simulations were performed on four nodes of an SGI Altix ICE 8400 LX cluster running in parallel on two Intel 6-core 5680 3.33GHz processors (48 cores) and 36 GB RAM at Cenapad/Unicamp, Brazil.

**Molecular cloning and sequencing**

Total RNA was extracted from a pool of *B. mori* tissues and treated with DNase I (Invitrogen) right before the first-strand cDNA synthesis, which was carried out using the M-MLV Reverse Transcriptase (Invitrogen) and random primers according to the manufacturer's specifications. The primers used for molecular cloning were designed using the Fast PCR v.4.0.27 program and were synthesised by the Synthesis Biotechnology Company. The universal primer was constructed along with the SP6 promoter for *in vitro* expression experiments with *B. mori* sbRNA. The PCR product obtained (75 bp) was cloned into the TOPO® TA Cloning® Kit (Invitrogen). The transformation was performed in competent DH5α bacteria through heat shock[57]. The plasmid purification from selected clones was performed using the CTAB method[58], and

the presence of the insert was confirmed by PCR using aspecific pair of primers. The selected clones were amplified by PCR using the M13 primers pair, and the resulting product was purified with polyethylene glycol and sequenced using the DYEnamic ET Terminator (Amersham Biosciences) kit and MegaBACE 1000 (Molecular Dynamics) equipment.[59]

## Conclusions

Highly conserved DNA segments of the genes encoding hY5 RNAs from human and chY5 from Chinese hamster culture cells and the gene encoding sbRNA from *C. elegans* were used in the search for homologous genes with high identity in *B. mori* suggesting that the unnoted gene *Bm*sbRNA could be a possible Y RNA.

The 3D structural similarity and the similar behaviour observed in the molecular dynamics simulations of these four RNA segments strongly suggest that the *B. mori* sbRNA may act as a Y RNA. The functional characterisation of this ncRNA and the metabolic role of this possible *B. mori* Y RNA remain targets for further studies. However, the results presented here strongly suggest that this is the first evidence of a new noncoding gene in Lepidoptera.

## Acknowledgements

This work was supported by CNPq, CAPES/Fundação Araucária, FINEP/ COMCAP/UEM and Secretaria de Estado da Ciência, Tecnologia e Ensino Superior - FUNDO PARANÁ, Brasil. We also acknowledge CENAPAD (proj520), Sao Paulo, Brasil, for providing computational facilities.

## References


1. W. Deng, X. P. Zhu, G. Skogerbo, Y. Zhao, Z. Fu, Y.D. Wang, H.S. He, L. Cai, H. Sun, C. N. Liu, B. Li, B. Y. Bai, J. Wang, D. Jia, S. W. Sun, H. He, Y. Cui, Y. Wang, D. B. Bu, R. S. Chen, *Genome Res*, 2006, **16**, 20-29.
2. M. N. Aftab, H. He, G. Skogerbo and R. Chen, *Bmc Genomics,* 2008, **9**.
3. I. Boria, A. R. Gruber, A. Tanzer, S. H. Bernhart, R. Lorenz, M. M. Mueller, I. L. Hofacker, P. F. Stadler, *J Mol Evol,* 2010, **70**, 346-358.
4. J. Perreault, J. P. Perreault, G. Boire, *Mol Biol Evol*, 2007, **24**, 1678-89.
5. C. P. Christov, T. J. Gardiner, D. Szuts, T. Krude, *Mol Cell Biol,* 2006, **26**, 6993-7004.
6. T. Krude, *Subcellular Biochemistry*, 2010, **50**, 105-118.
7. M. Köhn, N. Pazaitis, S. Hüttelmaier, *Biomolecules*, 2013, **3**, 143-156.
8. T. Krude, C. P. Christov, O. Hyrien, K. Marheineke, *Journal of Cell Science*, 2009, **122**, 2836-2845.
9. C.P. Christov, E. Trivier, T. Krude, *Br J Cancer*, 2008, **98**, 981-8.
10. J. M. Dhahbi, *Ageing Res Rev*, 2014, ARR-501; No. of Pages 13.*Article in press*.
11. M. M. Mohammad, T. R. Donti, J. S. Yakisich, A. G. Smith, G. M. Kapler, *Embo Journal*, 2007, **26**, 5048-5060.
12. J. Norseen, A. Thomae, V. Sridharan, A. Aiyar, A. Schepers, P. M. Lieberman, *Embo Journal*, 2008, **27**, 3024-3035.
13. A. E. Hall, C. Turnbull, T. Dalmay, *BioMol Concepts*, 2013, **4**, 103–110.
14. T. J. Gardiner, C. P. Christov, A. R. Langley, T. Krude, *Rna*, 2009, **15**, 1375-85.
15. S. P. Bell, A. Dutta, *Annu Rev Biochem*, 2002, **71**, 333–374.

16. J. Méndez, B. Stillman, *Molecular And Cellular Biology,* 2000, **20**, 8602–8612.
17. Q. A. Lima Neto, F. S. Rando, D. V. B. Freitas, L. F. Rodrigues, F. R. Rosado, A. Fiorini, F. Gimenes, J. Tavares, M. A. Fernandez, *Biochemistry (Moscow)*, 2014, **79**, 3743.
18. J. Nagaraju, *CurrSciIndia*, 2000, **78**, 151-161.
19. R. Parthasarathy, K. P. Gopinathan, *Gene Expr Patterns*, 2005, **5**, 323-339.
20. Q. Xia, Et al., *Insect Biochem Mol Biol*, 2008, **38**, 1036–1045.
21. T. Tamura, C. Thibert, C. Royer, T. Kanda, E. Abraham, M. Kamba, N. Komoto, J. L. Thomas, B. Mauchamp, G. Chavancy, P. Shirk, M. Fraser, J. C. Prudhomme, P. Couble, T. Toshiki, T. Chantal, R. Corinne, K. Toshio, A. Eappen, K. Mari, K. Natuo, T. Jean-Luc, M. Bernard, C. Gerard, S. Paul, F. Malcolm, P. Jean-Claude, C. Pierre, *Nat Biotechnol*, 2000, **18**, 81-84.
22. J. L. Thomas, M. Da Rocha, A. Besse, B. Mauchamp, G. Chavancy, *Insect Biochem Mol Biol*, 2002, **32**, 247–253.
23. M. Tomita, H. Munetsuna, T. Sato, T. Adachi, R. Hino, M. Hayashi, K. Shimizu, N. Nakamura, T. Tamura, K. Yoshizato, *Nat Biotechnol*, 2003, **21**, 52–56.
24. J. Chen, X. Wu, Y. Z. Zhang, *J Biotechnol*, 2006, **123**, 236-247.
25. S. Hofmann, C. T. Foo, F. Rossetti, M. Textor, G. Vunjak-Novakovic, D. L. Kaplan, H. P. Merkle, L. Meinel, *J Control Release*, 2006, **111**, 219-227.
26. Y. Zhang, J. Chen, Z. Lv, Z. Nie, X. Zhang, X. Wu, *Eur J Pharm Sci*, 2006, **28**, 212-223.
27. G. H. Altman, F. Diaz, C. Jakuba, T. Calabro, R. L. Horan, J. Chen, H. Lu, and J. Richmond, D. L. Kaplan, *Biomaterials*, 2003, **24**, 401-416.

28. A. T. Paulino, F. A. Minasse, M. R. Guilherme, A. V. Reis, E. C. Muniz, J. Nazaki, *J Colloid Interface Sci*, 2006, **301**, 479-487.

29. L. F. C. Ribeiro, D.B. Zanatta, J.P. Bravo, R.M.C. Brancalhão, M.A. Fernandez, *Genet Mol Res* 2009, **8**, 144-153.

30. G. F. Rohrmann, Baculovirus Molecular Biology: Third Edition [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2013. Available from: http://www.ncbi.nlm.nih.gov/books/NBK114593/(accessed August 2014)**.**

31. L. Schramm, N. Hernandez, *Genes Dev*, 2002, **16**, 2593-2620.

32. M. Zuker, *Nucleic acids research*, 2003, **31**, 3406-3415.

33. K. Darty, A. Denise, Y. Ponty, *Bioinformatics*, 2009, **25**, 1974-1975.

34. I. Wang, M. P. Kowalski, A. R. Langley, R. Rodriguez, S. Balasubramanian, D. Sang-Te, D. Hsuand, T. Krude, 2014, *Biochemistry,* **53**, 5848-5863.

35. X. Yu, Q. Zhou, S. C. Li, Q. Lou, Y. Cai, W. C. Lin, H. Chen,Y. Yang, S. Hu, J. Yu, *PLoS one*, 2008, **3**, e2997.

36. Y. Zhang, X. Zhou, X. Ge, J. Jiang, M. Li, S. Jia, X. Yang X, Y. Kan, X. Miao, G. Zhao, *PLoS One*, 2009, **4**, e4677

37. J. Li, Y. Cai, L. Ye, S. Wang, J. Che, Z. You, J. Yu, B. Zhong, *BMC Genomics*, 2014, **15**, 410.

38. C. N. Liu, B. Y. Bai, G. Skogerbo, L. Cai, W. Deng, Y. Zhang, D. B. Bu, Y. Zhao and R. S. Chen, *Nucleic Acids Research*, 2005, **33**, D112-D115.

39. A. P. M. Verhagen, G. J. M Pruijn, *Bioessays*, 2011, **33**, 674-682.

40. M. Popenda, M. Szachniuk, M. Antczak, K. J. Purzycka, P. Lukasiak, N. Bartol, J. Blazewicz, R. W. Adamiak, *Nucleic acids research*, 2012, 40, e112.

41. A. D. Mackerell Jr, M. Feig, C. L. Brooks 3rd, *J Comput Chem*, 2004, **25**, 1400-1415.

42. A. T. Zhang, A. R. Langley, C. P. Christov, E. Kheir, T. Shafee, T. J. Gardiner, T. Krude, *J Cell Sci*, 2011, **124**, 2058-2069.

43. A. R. Langley, H. Chambers, C. P. Christov, T. Krude, *PLoS One*, 2010, **5**, e13673.

44. S. Sim, D. E. Weinberg, G. Fuchs, K. Choi, J. Chung, S. L. Wolin, *MolBiol Cell*, 2009, **20**, 1555-1564.

45. M. R. Lerner, J. A. Boyle, J. A. Hardin, J. A. Steitz, *Science*, 1981, **211**, 400-402.

46. E. L. Garcia, A. Onafuwa-Nuga, S. Sim, S. R. King, S. L. Wolin, A. Telesnitsky, *J Virol*, 2009, **83**, 12526-12534.

47. A. J. Stein, G. Fuchs, C. M. Fu, S. L. Wolin, K. M. Reinisch, *Cell*, 2005, **121**, 529-539.

48. D. L. Nelson, M. M. Cox, A. L. Lehninger, *in Lehninger principles of biochemistry*, ed. W.H. Freeman, Basingstoke, New York, 5th edn., 2008, pp. 1158.

49. T. Hermann, D. J. Patel, *Structure*, 2000, **8**, R47-54.

50. T. Li, H. He, Y. Wang, H. Zheng, G. Skogerbø, R. Chen, *BMC Mol Biol*, 2008, **9**, 71.

51. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J Mol Biol*, 1990, **215**, 403-410.

52. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Research*, 1994, **22**, 4673-4680.

53. G. E. Crooks, G. Hon, J. M. Chandonia, E. S. Brenner, *Genome Research*, 2004, **14**, 1188-1190.

54. T. D. Schneider, R. M. Stephens, *Nucleic Acids*, 1990, **18**, 6097-6100.

55. W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33-38, 27-28.

56. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, K. Schulten, *J Comput Chem*, 2005, **26**, 1781-1802.

57. J. Sambrook, D. W. Russell, in *The condensed protocols from Molecular cloning: a laboratory manual.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2006, pp 800.

58. G. Del Sal, G. Manfioletti, C. Schneider, *Biotechniques*, 1989, **7**, 514-520.

59. W. A. Silva, M. C. Costa, V. Valente, J. F. Sousa, M. L. Paco-Larson, E. M. Espreafico, S. S. Camargo, E. Monteiro, A. J. Holanda, M. A. Zago, A. J. Simpson, E. D. Neto, *Biotechniques*, 2001, **30**, 537-542.

**Figure legends**

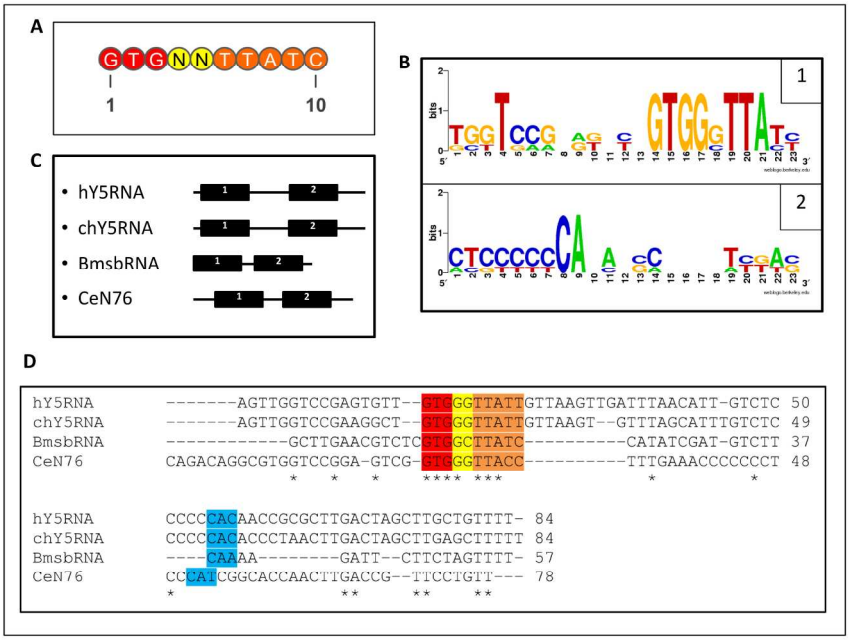
**Fig. 1.** Homology search of the new noncoding RNA in *Bombyx mori.* (A) Conserved sequence found in Y RNAs and sbRNAs, where "N" may be any nucleotide. Conserved motif for Y RNAs[14] in red, variable region in yellow, conserved sequence for sbRNAs[3] in orange. (B) Comparison and (C) relative position of the 5'-end conserved sequence (1) and 3'-end paired strand (2) between the sequences of Y5 RNA from human and hamster, *Bm*sbRNA and CeN76 sbRNA from *C. elegans* genes. (D) Alignment between Y5 RNA from human and hamster, *Bm*sbRNA and CeN76 sbRNA from *C. elegans* genes. The colours in the nucleotides are as described in (A), though the 3'-end paired nucleotides are in blue.

**Fig. 2.** RNA 2D structures. Secondary structures of Y5 RNA from human (A) and Chinese hamster (B), *B. mori* sbRNA (C) and *C. elegans* sbRNA (D), drawn by Varna[33]. Conserved motif for Y RNAs in red and blue, variable region in yellow, conserved sequence for sbRNAs in orange.

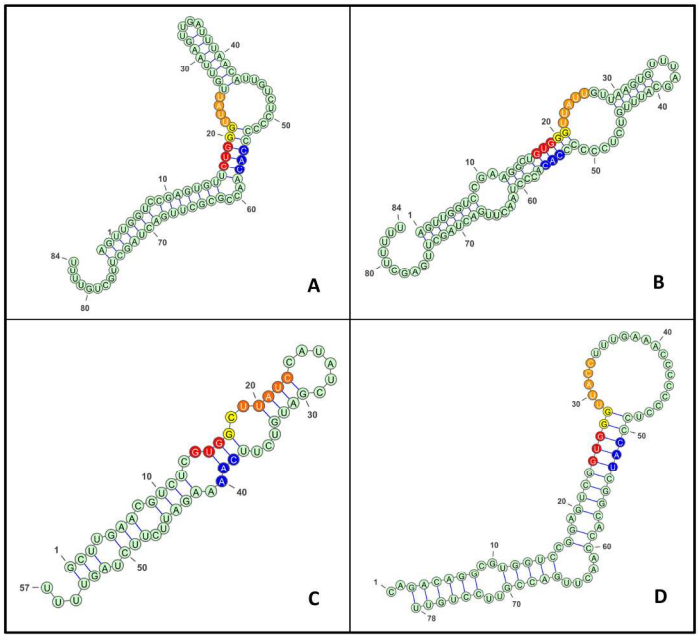
**Fig. 3.** *Bombyx mori* sbRNA expression analysis. (A) sbRNA expressed in a pool of larvae tissues, (M) 100bp DNA Ladder Invitrogen, (1) *B. mori* sbRNA. (B) *B. mori* sbRNA amplified sequence showed in (A). *Eco*RI restriction enzyme recognition site in light blue; SP6 promoter sequence in green; and *B. mori* sbRNA primers forward and reverse in yellow. The plasmid sequence is in Italic and the cloned sequence in bold.

**Fig. 4.** Behaviour of RNA segments along MD simulation. The colours indicate the RNA segments for human Y5 RNA, Chinese hamster Y5 RNA, *B. mori* sbRNA and *C. elegans* sbRNA in black, red, blue and green, respectively.

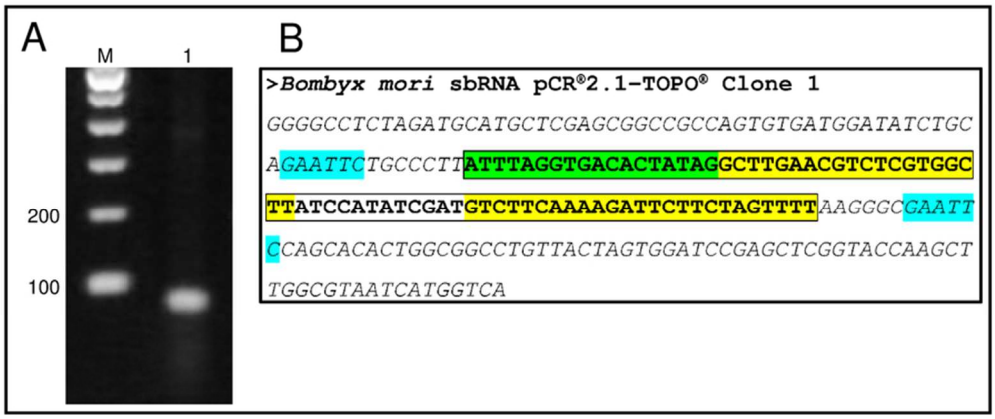
**Fig. 5.** RMSF from the movement of the $C_1$' atom evaluated for each nucleotide. Each graph represents the different segments of the species. (A) human Y5 RNA, (B) Chinese hamster Y5 RNA, (C) *B. mori* sbRNA and (D) *C. elegans* sbRNA.



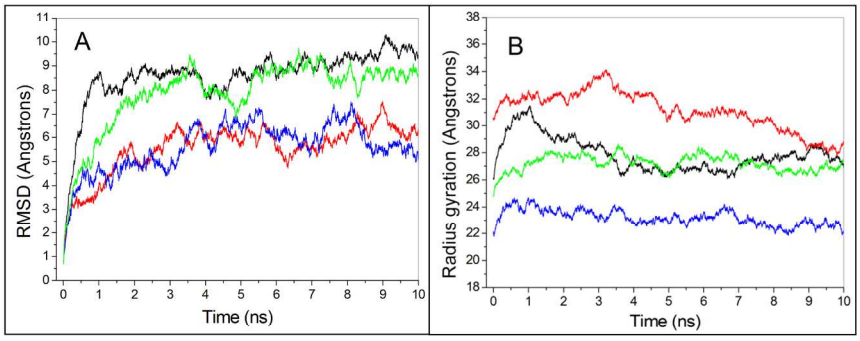

209x148mm (300 x 300 DPI)



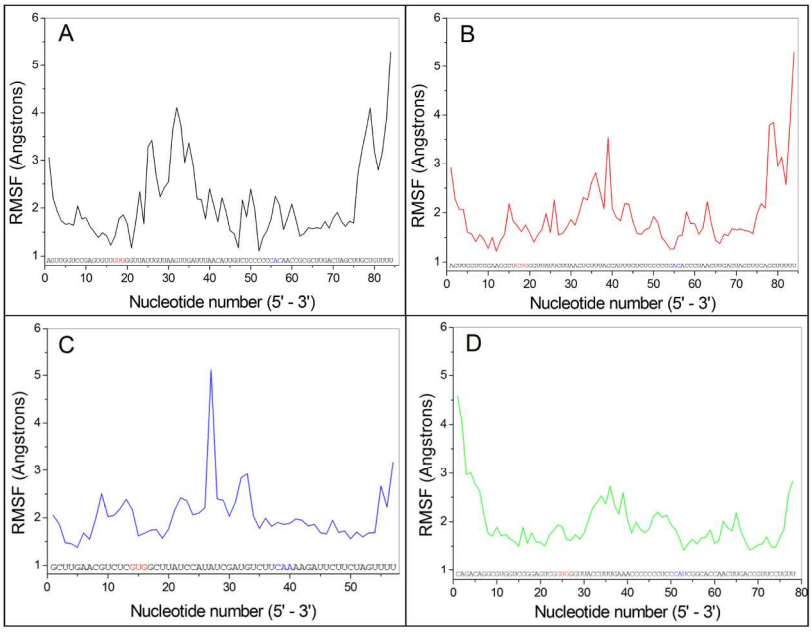

209x148mm (300 x 300 DPI)




76x32mm (300 x 300 DPI)




209x148mm (300 x 300 DPI)




209x148mm (300 x 300 DPI)