OA Microbiology
Open Access

Ternan NG. Small regulatory RNA molecules in bacteria. OA Microbiology 2013 Dec 01;1(1):1.

Competing interests: none declared. Conflict of interests: none declared.
All authors contributed to conception and design, manuscript preparation, read and approved the final manuscript.
All authors abide by the Association for Medical Ethics (AME) ethical rules of disclosure.

Small regulatory RNA molecules in bacteria

NG Ternan*

Northern Ireland Centre for Food and Health (NICHE), School of Biomedical Sciences, University of Ulster, Coleraine, N. Ireland.

*Corresponding author: Northern Ireland Centre for Food and Health (NICHE), School of Biomedical Sciences, University of Ulster, Coleraine, Co. Londonderry, N. Ireland, United Kingdom BT52 1SA.

Tel ++ 44 (0) 2870 123063

Fax ++ 44 (0) 2870 124965

Email: ng.ternan@ulster.ac.uk

Abstract

Small non coding RNA molecules are widespread in all kingdoms of life, where they serve to regulate and fine tune gene expression. They can act in *cis* or *trans*, depending upon their structural relationship with genes whose expression they influence, and function by interacting with target messenger RNA molecules to inhibit or accelerate translation. Thus, they can exert rapid control on cellular protein levels. Within bacteria, many sRNAs have been described in Gram negative model organisms but developments in our understanding of their role in Gram positive organisms has been slower. It is clear that sRNAs influence a wide range of cellular processes, including adaptation to environmental stresses, and virulence processes in pathogens.

Historically, identification of sRNAs has been challenging but recent developments in sequencing technology and computational analysis have led to over 45000 predicted sRNAs being catalogued in the last few years. However many of these *in silico* predictions are yet to be validated and the complexity, in terms of sRNA interactions with gene networks, means we are really only beginning to understand how wide-ranging their effects can be within bacteria. It is clear that sRNAs play a critical role in all aspects of bacterial physiology. Within the genus *Clostridium*, the role of sRNAs in the pathogens *C. perfringens*, *C. botulinum* and *C. difficile* is much less well understood, despite hundreds of sRNAs having been predicted within these organisms. These predictions represent a platform for novel discoveries regarding post transcriptional regulatory strategies mediated by these molecules in Clostridia.

Keywords: Regulatory RNA, Post transcriptional regulation, Stress, Virulence, Riboswitch, *Clostridium*, *difficile*

Introduction.

The genus *Clostridium* encompases a heterogeneous group of Gram-positive endospore forming obligately anaerobic microorganisms that are ubiquitous in soils and the intestines of higher organisms. Certain species are economically useful and relatively benign – for example, *C. acetobutylicum*, *C. beijerinckii* and *C. cellulovorans* are employed in the industrial production of biofuels[1]. However the genus is also infamous for the toxin-producing pathogens *C. difficile*, *C. perfringens* and *C. botulinum*, whose yearly socioeconomic impact is considerable[2,3]. *C. difficile* infection (CDI) causes infectious diarrhoea with associated abdominal pain, cramping and low grade fever up to 40.6°C[4]. *C. difficile* pathogenesis and many of the factors underlying CDI are well understood, but CDI can still be life-threatening if not treated promptly[2,4,5].

The availability of well over 30 *C. difficile* genome sequences[6-9] has afforded researchers excellent opportunities to better understand the evolution and lineages of these organisms. Generation of comparative functional genomics datasets has lagged somewhat and as a consequence, comparatively little is known about the adaptive ability of *C. difficile*. Thus in our laboratory we have taken a systems approach to understanding the response of *C. difficile* to clinically relevant heat stress, using comparative proteomics and transcriptomics[10,11]. While a classical heat shock response and class I chaperone induction was observed at 41°C, we also observed downregulation of the flagellum, *FliC* (CD0239) and several other recognised virulence factors, such as *cwp20* (CD1469), *cwp5* (CD2786) and *TcdA* (CD0663), strengthening the hypothesis that virulence of *C. difficile* is 'set' at 37°C. We also determined that the correlation between changes in protein abundances and their cognate transcripts was inconsistent. Several factors could explain this

observation including protein/mRNA stability, transcription efficiency or unrecognised post-transcriptional regulatory mechanisms[12]. Recently Chen et al.[13] demonstrated the presence of small, non coding, regulatory RNA molecules (small RNAs, sRNAs) in *C. acetobutylicum* and proposed a role for them in gene regulation in this microorganism. This review seeks to provide an overview of the key elements of sRNA biology and to summarise what is known of their role in Clostridia.

What are Small RNAs?

Eukaryotic sRNAs

Small non coding RNAs, including microRNAs (miRNAs) and small interfering RNAs (siRNAs) have been identified as regulators of a variety of cellular processes in plants and animals[14]. First described in *Caenorhabditis elegans*[15] several hundred miRNAs, generally ~21-22 nucleotides in length, have now been described. They are generated by cleavage of longer, precursor RNA transcripts that have formed a self complementary foldback loop by the RNAseIII-like enzyme Dicer, and function by base pairing with target mRNA, initiating its degradation. Small interfering RNAs (siRNAs, <30 nucleotides) are generated by Dicer mediated cleavage of double stranded RNA and play a role in RNA interference (RNAi) via the RNA-induced silencing complex (RISC), where they guide sequence-specific cleavage of RNAs. Thus these small RNAs are functionally interchangeable. miRNAs have also been predicted and experimentally verified in DNA viruses, with the herpesviridae containing the largest number of viral miRNAs: for such viruses that undergo persistent infection, the invisibility of miRNAs to the adaptive immune response is a useful trait[16,17]. Indeed, the fact that double stranded (ds) RNAs are quite stable *in vivo* and non immunogenic means that RNAi has great potential for therapeutic

use[18,19]. It is known that miRNAs encoded by both host and infecting viruses enable these protagonists to do battle with each other during infection[20] and as a result, miRNA profiles are becoming recognised as novel means of diagnosis[21].

Prokaryotic sRNAs

Small RNAs were initially identified in bacteria with the identification of 6s RNA in *E. coli*[22] but it is only relatively recently that their influence on bacterial cellular processes and their varied modes of action have become recognised[23]. In contrast to shorter eukaryotic or viral miRNAs, bacterial sRNAs (i.e. not tRNA, rRNA or 5S RNA) are typically between 50 and 500 nucleotides in length and as with many developments in microbial sciences *E. coli* was, and is, the model organism for study of sRNAs. Knowledge of sRNA biology in Gram positive organisms and in archaea has developed more slowly, due in part to a lack of efficient genetic tools[24,25]. Initial elucidation of an individual sRNA's function in *E. coli* came in 1984. Mizuno and colleagues[26] showed that an mRNA-interfering complementary RNA (micRNA) that was complementary to the 5' end region of the *ompF* gene mRNA, served to inhibit production of the ompF protein by interfering with translation. This new field of RNA biology, dubbed RNOmics[27] has since developed and expanded exponentially assisted in no small part by technical advances in DNA sequencing technologies and the development of computational algorithms for identification of sRNA sequences in genomic information (Figure 1).

Functionality of bacterial sRNAs.

Bacterial sRNAs regulate and fine tune gene expression in bacteria and it is thought that they enable a faster response to changing conditions at relatively low

metabolic cost. Functional RNA molecules require only limited transcription energy compared to other cellular regulatory mechanisms and in addition, less time is required for a sRNA to be produced and to impact upon target protein levels[28]. A wide range of environmental stimuli impact upon sRNA expression and it is not surprising that many sRNAs are associated with bacterial stress responses[29]. sRNAs can exert global effects on gene expression. In the oxidative stress response in *E. coli*, for example, the 109 nucleotide OxyS sRNA is transcribed divergently from, and regulated by, the *oxyR* gene encoding the redox-sensitive transcriptional regulator which is the actual sensor of the oxidative shock. Upon expression of oxyS sRNA, translation of rpoS is inhibited with rapid and global effects upon cellular physiology[29,30]. In *E. coli*, *FlhDC* – the master regulator of flagellar biosynthesis – is regulated by multiple protein transcription factors that respond to different environmental stimuli including cell envelope stress and salt concentration. However the recent work of De Lay and Gottesmann[31] has shown that complexity, and thus regulatory power, is increased because the 5' untranslated region (5' UTR) of the *flhDC* mRNA is also subject to negative regulation by six different sRNA molecules (ArcZ, OmrA, OmrB, OxyS, SdsR and GadY) and positive regulation by one (McaS). Thus, the *flhDC* mRNA serves as a hub that allows integration of signals derived from environmental salt and oxygen concentrations, oxidative insult and the general stress response, into the decision to make flagella. The question of whether the flagellum is a primary *C. difficile* virulence factor is open to debate[11], but a flagellar filament requires some 2% of a bacterial cell's total energy consumption under optimal growth conditions, in order to synthesise the necessary ~20,000 subunits of FliC protein: it is clear why such precise regulation of flagellar biosynthesis might be necessary. It has been suggested that up to ~300 sRNAs will be present in the

average bacterial genome, a number equivalent to the complement of transcription factors[32]. As exemplified above, however, these sRNAs have many times the potential regulatory capacity of protein transcription factors and thus they are clearly of critical importance in bacterial physiology.

How do sRNA molecules exert their biological effect?

In the Gram-positive bacterial pathogens in which sRNAs have been characterised to date, their biological functions have been linked to adaptation or virulence. For example, in *C. perfringens*, the VR-RNA sRNA regulates collagenase and alpha toxin gene transcription[33]. Like Gram-negatives, Gram positive bacteria have many sRNA-mediated regulatory mechanisms that allow response to environmental and intercellular signals via a number of different mechanisms[24]. Bacterial sRNAs are generally found in the intergenic regions of the genome and they fall into two main categories depending upon their genomic context in relationship to the target gene. Those that are transcribed independently from the target gene are encoded in *trans*, while those that are co-transcribed, usually from within the 5' UTR of the target transcript, are encoded in *cis*[29] (Figure 2). *Cis* encoded sRNAs can also be transcribed from the antisense strand at the same genetic locus as the target and these antisense RNAs (aRNAs) will therefore exhibit perfect complementarity with their target, allowing interactions that impact positively or negatively upon gene expression[32,34]. *Cis* and *trans* sRNAs can be further categorized into two subgroups based upon their mode of action. Certain sRNAs pair with mRNA targets to affect their stability or translation while others act as molecular decoys that bind to protein targets and affect their activity[35-39] (Figure 3). RNA thermosensors (Figure 4a) have been demonstrated to play pivotal regulatory

roles in not only the heat stress response, but also in the coordination of expression of virulence genes in number of human pathogens[40,41] while Riboswitches, a further class of *cis* acting RNA element, control expression of downstream genes via metabolite-induced alteration of sRNA secondary structures (Figure 4b). Riboswitches can function in a variety of ways but in brief, different metabolites can allow them either to induce or repress transcription or translation, as recently reviewed by Serganov and Nudler[42]. The bacterial sRNAs that have been characterised in Gram positive microorganisms are expressed mainly in a growth phase-dependent manner and while it may be hypothesised that, like in *E. coli*, they are part of complex regulatory processes our current knowledge of factors affecting sRNA expression in Gram positive bacteria is lacking[24]. Thus, while sRNAs have been characterised in *Bacillus subtilus*, *Listeria monocytogenes*, *Staphylococcus aureus*, *Streptococcus pyogenes*, *Clostridium acetylobutylicum* and *Clostridium perfringens*, very little is known about their role in *C. difficile*.

The role of the Hfq RNA chaperone protein.

While sRNA modes of action are fairly similar between Gram positive and Gram negative bacteria, one aspect of sRNA biology that is less well conserved is the role of the Hfq RNA chaperone. Hfq is highly conserved in prokaryotes and belongs to the Sm family of proteins that are known to interact with RNA in both eukaryotes and prokaryotes[43]. Hfq has been shown to interact with a considerable number of *trans* encoded sRNA molecules in Gram negative microbes, where it plays a key role in stabilising sRNA molecules or facilitating interaction with mRNA targets[24,44]. Thus, Hfq plays a key role in one of the most complex post transcriptional networks known[45]. In low GC Gram-positive bacteria, however, the

function of Hfq is still unclear[43] although in *L. monocytogenes*, Hfq is required for function of several sRNAs (LhrA–C)[46]. However, other *L. monocytogenes* sRNAs do not require Hfq for target interaction[47] and in *S. aureus*, Hfq does not seem to be required for sRNA-mRNA interactions at all[48]. There is also the consideration that not all bacterial genomes contain an Hfq homologue, raising the possibility that other proteins may be able to substitute for Hfq in certain organisms[49].

Identification and validation of sRNAs in bacteria.

Initial identification of sRNAs in bacteria is challenging, not least because until recently there was no general approach that provided a comprehensive solution to their prediction[21]. Furthermore, sRNA target prediction is awkward because many sRNA:mRNA hybridisations occur over relatively short regions of imperfect complimentarity[50]. The initial work on sRNAs some 40 years ago used gel electrophoresis to fractionate radiolabelled total bacterial RNA, followed by elution of low molecular mass RNA molecules from the gels and subsequent analysis[51]. In the 30 years since their first discovery, only around a dozen sRNAs were identified and characterised in *E. coli* but since then, developments in genomics and computational biology have allowed the field of sRNA biology to expand massively. In the past decade or so, sRNA gene finders based upon well-characterised sequences and algorithms to predict the minimum free energy of structured RNAs have been applied to newly catalogued bacterial genomes[52,53]. In addition, comparative genomics approaches that allow researchers to make sRNA predictions based upon the presence of rho independent terminators and promoters and other features in the intergenic regions have also been used to predict sRNAs[13,50,54].

A workflow for sRNA characterisation, therefore, might proceed from *in silico* identification of sRNAs to demonstration of their expression by qRT-PCR or Northern Blotting and the subsequent identification of direct and indirect targets of individual sRNA molecules using *in silico* prediction algorithms followed by wet lab methods to validate the interactions. For example, in the work of Chen et al.,[13] the only report to date on genome wide characterisation of sRNAs in clostridia, *in silico* methods were used to predict sRNAs in 21 clostridial species. The authors then used qRT-PCR to validate 30 sRNAs of 113 predicted in *C. acetobutylicum*, and 21 from *C. botulinum*, thus showing that qRT-PCR is a useful first screening step. Highly expressed sRNAs (by qRT-PCR) were then analysed using Northern blotting to validate transcript sizes against those predicted by the *in silico* analysis. A number of additional experimental approaches can also be used including tiling oligonucleotide microarrays, cDNA cloning and high throughput RNAseq[54-56]. In addition, the identification of sRNA:Hfq associations can provide further evidence that transcripts are sRNAs[45].

Databases for sRNA research.

Concomitantly with these predictive methods and experimental validations, the development of user friendly, browser based databases and software tools to allow information retrieval and analysis has proceeded apace. As with other post-genomic fields, for example mass spectrometry-driven proteomics,[57] these developments have been crucial to the expansion of sRNA biology as a field of research. Within even the last few years, the number of sRNAs identified in a wide range of bacteria, including in Gram positives, has increased at an incredible rate. A natural consequence of this success is an increasing urgency for identification of their cellular targets and functional roles, a facet of the research which has lagged considerably behind

identification studies[58]. A number of groups have presented a variety of tools for the purposes of sRNA identification[59]. One of the longest standing is the Rfam database (http://rfam.sanger.ac.uk), a collection of non-coding RNA families represented by multiple sequence alignments and secondary structure predictions that was first developed a decade ago[60,61]. The work of Livny et al. introduced the powerful SIPHT tool (sRNA identification *p*rotocol using high-throughput technologies), which incorporates a number of programs and adjustable search parameters to identify sRNAs and other features in an automated fashion[50]. SIPHT identifies conserved sequences along with rho independent terminators and promoters in intergenic regions and incorporates BLAST, genomic synteny and transcription factor binding site analyses into a workflow that yields an output that can be opened in Excel. This work has allowed prediction of candidate sRNA encoding loci from over 900 bacterial genomes and plasmids within the NCBI database, thus expanding the number of predictions from several hundred candidate sRNAs to over 45,000. However all databases will have perceived drawbacks, regardless of how they are implemented. They might not allow further analysis, or they may be restricted to a limited number of bacterial species, or be reliant upon published data. Two recent publications have sought to redress this deficiency: sRNAdb, developed by Pischimarov and colleagues[50] is a user-friendly searchable database allowing comprehensive comparative analysis of sRNAs from Gram positive microorganisms. In addition, further features of interest may be incorporated by the end user into a local customised database. The work of Li et al. describes BSRD – a repository for bacterial small regulatory RNA[62] which is said to contain more experimentally validated sRNAs than any other database and enables researchers to identify and characterise sRNAs in large scale transcriptome sequencing projects. Thus,

researchers interested in a particular bacterial group now have at their disposal a comprehensive range of predictions, databases and *in silico* analysis tools to underpin their investigations.

Identification of sRNA targets – dissection of roles and functions.

Having validated the existence of a population of sRNAs, there remains the issue of what individual sRNA molecules actually do. It is clear that only a relatively small proportion of the sRNAs predicted to date have had their targets experimentally verified, although targets can initially be inferred computationally. Many sRNAs are antisense regulators and bioinformatics searches for complementarity can assist with target identification – although in reality, the base pairing between sRNAs and their targets is often imperfect, making this task difficult[58]. One such tool, sTarPicker, is based upon a mathematical model of hybridisation between sRNA and mRNA and is said to predict sRNA targets with higher efficiency than competing programmes[63]. sRNATarBase, developed by the same group, seeks to provide a resource of sRNA targets that have been experimentally verified, thus providing support for predictive models and subsequent *in silico* and functional analyses. The authors systematically and manually collected sRNA:target interaction data from published papers in order to develop their database of sRNA targets[64]. However, where targets are as yet only inferred, is still necessary to validate these sRNA:target predictions and to this end, several interesting approaches can be used. In addition, the determination of what constitutes a primary target (direct interaction with the sRNA) and what is a secondary target, such as a transcription factor, is also of considerable importance[58].

Analysis of the sRNA and proposed target mRNA expression under different conditions is one approach to target identification. As reported by Chen et al., a

conserved novel sRNA (CAC610) in *C. acetobutylicum* and a downstream gene (CAC0528) both responded to the antibiotic clindamycin. As the distance between the sRNA and the gene was conserved across a number of clostridial strains at ~185bp (although neither exist within in *C. difficile*), the authors concluded that there was a functional relationship between the two, although the exact mechanism by which the sRNA might modulate gene expression (or *vice versa*) was not determined[13]. Another method for determination of sRNA targets has been described as a 'biochemical fishing expedition'. The use of sRNA molecules as the bait in order to capture a mRNA target is an approach that can be further refined by incorporating a recombinant affinity tagged Hfq protein. As many sRNAs interact with Hfq, its subsequent purification, complete with sRNA and the sRNA target, can allow sRNA target identification. In this instance, creation of cDNA clones, and their hybridisation to whole genome microarrays could be employed[65]. Functional genomics analyses, for instance with mutants constructed in validated sRNA encoding regions of the genome, allows the subsequent determination of the effect of these deletions on both host cell physiology and on the expression of predicted targets[66,67]. With mutants in hand, tiling oligonucleotide microarrays, or RNAseq analysis, would provide a genome-wide picture of their effect. Furthermore, it should be possible to experimentally express a high level of a given sRNA in a host cell, and compare global cellular responses with those of either the wild type or a deletion mutant.

Conclusions and future perspectives

RNOmics is still a rapidly expanding field and it is clear that advances in our understanding will be driven by the use of high throughput post genomic technologies

such as transcriptome sequencing. Focus will also be required to determine the functions of individual bacterial sRNAs which is a not inconsiderable task given the potential for widespread interactions of sRNAs with multiple targets and within gene networks. There is still much to be done to experimentally validate sRNA predictions in clostridia, where it appears that the number of sRNAs is related to the physiology of the organism. Greater numbers of sRNAs have been predicted in the genomes of pathogenic clostridia – for example *C. difficile* 630 is predicted to contain 264 sRNAs, none of which have been experimentally verified as yet[13]. Clostridial sRNAs appear to be phylogenetically restricted to these organisms and are not conserved in, for example, Bacilli, thus it will be of interest to determine precisely under what conditions these sRNAs are expressed, and whether strain to strain variations exist – between different *C. difficile* ribotypes, for example.

The work so far on *C. acetobutylicum* suggests that certain sRNAs may play a role in antibiotic resistance and this observation provides new avenues for research into antibiotic tolerance mechanisms, drug targets and diagnostic methods. At present, there is no data on the role of the Hfq homologues that exist in the genomes of *Clostridium* spp, although with functional genomics tools such as ClosTron[68] it should be possible to construct gene knockouts and determine the role of Hfq. Our understanding of small RNAs in Clostridia is at present incomplete, presenting the research community with an opportunity to define the roles of these molecules within these anaerobic microorganisms.

References

1. Ezeji TC, Qureshi, N, Blasckek, HP. Industrially relevant fermentations. In: Dürre, P, editor. Handbook on Clostridia. Boca Raton, Florida: CRC Press Boca Raton, Florida; 2005.p797-812.
2. Rupnik M, Wilcox MH, Gerding DN. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. Nature Rev Microbiol. 2009 July;7(7):526-536.
3. Peck MW, Stringer SC, Carter AT. *Clostridium botulinum* in the post-genomic era. Food Microbiol. 2011 Apr; 28(2):183-191.
4. Bartlett JG. Historical perspectives on studies of *Clostridium difficile* and *C. difficile* infection. Clin Infect Dis. 2008 Jan;48(Suppl 1):s4-s11.
5. Pépin J, Saheb N, Coulombe MA, Alary ME, Corriveau MP, Authier S, Leblanc M, Rivard G, Bettez M, Primeau V, Nguyen M, Jacob CE, Lanthier L. Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*-associated diarrhea: A cohort study during an epidemic in Quebec. Clin Infect Dis. 2005 Nov;41:1254-1260.
6. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeño-Tárraga AM, Wang H, Holden MT, Wright A, Churcher C, Quail MA, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, Dowd L, Fraser A, Feltwell T, Hance Z, Holroyd S, Jagels K, Moule S, Mungall K, Price C, Rabbinowitsch E, Sharp S, Simmonds M, Stevens K, Unwin L, Whithead S, Dupuy B, Dougan G, Barrell B, Parkhill J. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. Nat Genet. 2006 Jul;38(7):779-786.

7. Stabler RA, He M, Dawson L, Martin M, Valiente E, Corton C, Lawley TD, Sebaihia M, Quail MA, Rose G, Gerding DN, Gibert M, Popoff MR, Parkhill J, Dougan G, Wren BW. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. BMC Genome Biol. 2009 Sep;10:R102.

8. He M, Sebaihia M, Lawley TD, Stabler RA, Dawson LF, Martin MJ, Holt KE, Seth-Smith HM, Quail MA, Rance R, Brooks K, Churcher C, Harris D, Bentley SD, Burrows C, Clark L, Corton C, Murray V, Rose G, Thurston S, van Tonder A, Walker D, Wren BW, Dougan G, Parkhill J. Evolutionary dynamics of *Clostridium difficile* over short and long time scales. Proc Natl Acad Sci USA. 2010 Apr;107:7527-7532.

9. Forgetta V, Oughton MT, Marquis P, Brukner I, Blanchette R, Haub K, Magrini V, Mardis ER, Gerding DN, Loo VG, Miller MA, Mulvey MR, Rupnik M, Dascal A, Dewar K. Fourteen-genome comparison identifies DNA markers for severe-disease-associated strains of *Clostridium difficile*. J Clin Microbiol. 2011 Jun;49: 2230-2238.

10. Jain S, Graham C, Graham RLJ, McMullan G, Ternan, NG. A quantitative proteomic analysis of the heat stress response in *Clostridium difficile* strain 630. J Proteome Res. 2011 Sep;10:3880-3890.

11. Ternan NG, Jain S, Sirivastava M, McMullan G. Comparative transcriptional analysis of clinically relevant heat stress response in *Clostridium difficile* strain 630. PLoS ONE. 2012 Jul;7(7)e36671.

12. Olivares-Hernández R, Bordel S, Nielsen J. Codon usage variability determines the correlation between proteome and transcriptome fold changes. BMC Syst Biol. 2011 Feb;5:33.

13. Chen Y, Indurthi DC, Jones SW, Papoutsakis ET. Small RNAs in the genus *Clostridium*. MBio. 2011 Jan-Feb;2(1):e00340-10.

14. Carrington JC, Ambros, V. Role of MicroRNAs in Plant and Animal Development. Science. 2003 Jul;303, 336-338.

15. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature. 1998 Feb;391:806-811.

16. Pfeffer, S, Zavolan, M , Grasser, FA , Chien, MC ; Russo, JJ, Ju, JY , John, B ; Enright, AJ ; Marks, D ; Sander, C ; Tuschl, T. Identification of virus-encoded microRNAs. Science. 2004 Apr;304:5671:734-736

17. Kincaid RP, Sullivan CS. Virus-Encoded microRNAs: An Overview and a Look to the Future. PLoS Pathog. 2012 Dec;8(12):e1003018.

18. Esquela-Kerscher A, Slack, FJ. Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer. 2006 Apr;6(4):259-269.

19. Siomi MC. Short interfering RNA-mediated gene silencing; towards successful application in human patients. Adv Drug Deliv Rev. 2009, Jul;61(9):668-671.

20. Houzet L, Jeang KT. MicroRNAs and human retroviruses. Biochim Biophys Acta. 2011 Nov-Dec;1809(11-12):686-693.

21. Mraheil MA, Billion A, Kuenne C, Pischimarov J, Kreikemeyer B, Engelmann S, Hartke A, Giard JC, Rupnik M, Vorwerk S, Beier M, Retey J, Hartsch T, Jacob A, Cemič F, Hemberger J, Chakraborty T, Hain T. Comparative genome-wide analysis of small RNAs of major Gram-positive pathogens: from identification to application. Microb Biotechnol. 2010 Nov;3(6):658-676.

22. Brownlee GG. Sequence of 6S RNA of *E .coli*. Nature New Biology. 1971 Feb;229(5): 147-149.

23. Wassarman KM, Zhang AX, Storz G. Small RNAs in *Escherichia coli.* Trends Microbiol. 1999 Jan;7(1): 37-45.

24. Romby P, Charpentier, E. An overview of RNAs with regulatory functions in Gram-positive bacteria. Cell Mol Life Sci. 2010 Jan;67:217-237.

25. Prasse D, Ehlers C, Backofen R, Schmitz RA. Regulatory RNAs in archaea: first target identification in Methanoarchaea. Biochem Soc Trans. 2013 Feb;41(1):344-349.

26. Mizuno T, Chou MY, Inouye M. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). Proc Natl Acad Sci USA. 1984 Apr;81(7):1966-1970.

27. Vogel J, Bartels V, Tang TH, Churakov G, Slagter-Jager JG, Huttenhofer A, Wagner EGH. RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucl Acids Res. 2003 Nov;31(22):6435-6443.

28. Beisel CL, Storz G. Base pairing small RNAs and their roles in global regulatory networks. FEMS Microbiol Rev. 2010 Sep;34(5):866-82..

29. Repoila F, Majdalani N, Gottesman S. Small non-coding RNAs, co-ordinators of adaptation processes in *Escherichia coli*: the RpoS paradigm. Mol Microbiol. 2003 May; 48(4):855-861.

30. Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. Cell. 1997 Jul;90(1):43-53.

31. De Lay N, Gottesman S. A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. Mol Microbiol. 2012 Nov;86(3):524-538.

32. Irnov I, Sharma CM, Vogel J, Winkler WC. Identification of regulatory RNAs in *Bacillus subtilis*. Nucl Acids Res. 2010 Jun;38(19):6637-6651.

33. Shimizu T, Yaguchi H, Ohtani K, Banu S, Hayashi H. Clostridial VirR/VirS regulon involves a regulatory RNA molecule for expression of toxins. Mol Microbiol. 2002 Jan;43(1):257-265.

34. Salim NN, Feig AL. An Upstream Hfq Binding Site in the *fhlA* mRNA Leader Region Facilitates the OxyS-*fhlA* Interaction. PLoS ONE. 2010 Sep;5(9). pii: e13028.

35. Babitzke P, Romeo T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. Curr Opin Microbiol. 2007 Apr;10(2):156-163.

36. Gottesman S. Stealth regulation: biological circuits with small RNA switches. Genes Dev. 2002 Nov;16(22):2829-2842.

37. Toledo-Arana A, Repoila F, Cossart P. Small noncoding RNAs controlling pathogenesis. Curr Opin Microbiol. 2007 Apr;10(2):182-188.

38. Podkaminski, D, Vogel, J. Small RNAs promote mRNA stability to activate the synthesis of virulence factors. Mol Microbiol. 2010 Dec;78(6):1327-1331.

39. Desnoyers G, Bouchard MP, Masse E. New insights into small RNAdependent translational regulation in prokaryotes. Trends Genet. 2013 Feb;29(2):92-98.

40. Toledo-Arana A, Repoila F, Cossart P. Small noncoding RNAs controlling pathogenesis. Curr Opin Microbiol. 2007 Apr;10(2):182-188.

41. Johansson J, Mandin P, Renzoni A, Chiaruttini C, Springer M, Cossart P. An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. Cell. 2002 Sep;110(5):551-561.

42. Narberhaus F. Translational control of bacterial heat shock and virulence genes by temperature-sensing mRNAs. RNA Biol. 2010 Jan/Feb;7(1):84-89.

43. Serganov A, Nudler E. A Decade of Riboswitches. Cell. 2013 Jan;152(1-2):17-24.

44. Nielsen JS, Lei LK, Ebersbach T, Olsen AS, Klitgaard JK, Valentin-Hansen P, Kallipolitis BH. Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in *Listeria monocytogenes*. Nucl Acids Res. 2010 Jan;38(3):907-919.

45. Sobrero P, Valverde, C. The bacterial protein Hfq: much more than a mere RNA-binding factor. Crit Rev Microbiol. 2012 Feb;38(4):276-299.

46. Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J. An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. EMBO J. 2012 Oct;31(20):4005-4019.

47. Christiansen JK, Nielsen JS, Ebersbach T, Valentin-Hansen P, Sogaard-Andersen L, Kallipolitis BH. Identification of small Hfq-binding RNAs in *Listeria monocytogenes*. RNA. 2006 Jul;12(7):1383-1396.

48. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA, Soubigou G, Regnault B, Coppee JY, Lecuit M, Johansson J, Cossart P. The *Listeria* transcriptional landscape from saprophytism to virulence. Nature. 2009 Jun;459(7249):950-956.

49. Bohn C, Rigoulay C, Bouloc P. No detectable effect of RNA-binding protein Hfq absence in *Staphylococcus aureus*. BMC Microbiol. 2007 Feb;7,10.

50. Pandey SP, Minesinger BK, Kumar J, Walker GC. A highly conserved protein of unknown function in *Sinorhizobium meliloti* affects sRNA regulation similar to Hfq. Nucl Acids Res. 2011 Feb;39(11):4691-4708.

51. Livny J, Teonadi H, Livny M, Waldor MK. High-throughput, kingdom-wide prediction and annotation of bacterial non-coding RNAs. PLoS One. 2008 Sep;3(9):e3197.

52. Griffin BE. Separation of 32P-labelled ribonucleic acid components. The use of polyethylenimine-cellulose (TLC) as a second dimension in separating oligoribonucleotides of '4.5S' and 5S from *E. coli*. FEBS Lett. 1971 Jun;15(3):165-168.

53. Rivas E, Klein RJ, Jones TA, Eddy SR. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. Curr Biol. 2001 Sep;11(17):1369-1373.

54. Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EGH, Margalit H, Altuvia S. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. Curr Biol. 2001 Jun;11(12):941-950.

55. Wassarman KM, Repoila F, Rosenow C, Storz G, Gottesman S. Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev. 2001 Jul;15(13):1637-1651.

56. Pichon C, du Merle L, Caliot ME, Trieu-Cuot P, Le Bouguénec, C. An *in silico* model for identification of small RNAs in whole bacterial genomes: characterization of antisense RNAs in pathogenic *Escherichia coli* and

Streptococcus agalactiae strains. Nucl Acids Res. 2012 Dec;40(7):2846-2861.

57. Qu Y, Bi L, Ji X, Deng Z, Zhang H, Yan Y, Wang M, Li A, Huang X, Yang R, Han Y. Identification by cDNA cloning of abundant sRNAs in a human-avirulent *Yersinia pestis* strain grown under five different growth conditions. Future Microbiol. 2012 Apr;7(4):535-547.
58. Graham C, McMullan G, Graham RLJ. Proteomics in the microbial sciences. Bioeng Bugs. 2011 Jan/Feb;2(1):17-30.
59. Vogel J, Wagner EGH. Target identification of small noncoding RNAs in bacteria. Curr Opin Microbiol. 2007 Jun;10(3) 262-270.
60. Pischimarov J, Kuenne C, Billion A, Hemberger J, Cemič F, Chakraborty T, Hain T. sRNAdb: A small non-coding RNA database for Gram-positive bacteria. BMC genomics. 2012 Aug;13:384.
61. Griffiths-Jones S, Bateman A,. Marshall M, Khanna A, Eddy, SR. Rfam: an RNA family database. Nucl Acids Res. 2003 Jan;31(1):439-441.
62. Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. Rfam 11.0: 10 years of RNA families. Nucl Acids Res. 2013 Jan;41(Database issue):D226-232.
63. Li L, Huang D, Cheung MK, Nong W, Huang Q, Kwan HS. BSRD: a repository for bacterial small regulatory RNA. Nucleic Acids Res. 2013 Jan;41(Database issue):D233-238.
64. Ying X, Cao Y, Wu J, Liu Q, Cha L, Li W. sTarPicker: A Method for Efficient Prediction of Bacterial sRNA Targets Based on a Two-Step Model for Hybridization. PLoS ONE. 2011 Jul;6(7):e22705.

65. Cao Y, Wu J, Liu Q, Zhao Y, Ying X, Cha L, Wang L, Li W . sRNATarBase: A comprehensive database of bacterial sRNA targets verified by experiments. RNA. 2010 Nov;16(11):2051-2057.

66. Antal M, Bordeau V, Douchin V, Felden B. A small bacterial RNA regulates a putative ABC transporter. J Biol Chem. 2005 Mar;280(9):7901-7908.

67. Barra-Bily L, Fontenelle C, Jan G, Flechard M, Trautwetter A, Pandey SP, Walker GC, Blanco C. Proteomics alterations explain phenoptypic changes in *Sinorhizobium meliloti* lacking the RNA chaperone Hfq. J Bacteriol. 2010 Mar;192(6):1719-1729.

68. Kim JN, Kwon YM. Genetic and phenotypic characterization of the RyhB regulon in *Salmonella* Typhimurium. Microbiol Res. 2013 Jan;168(1):41-49.

69. Heap JT, Kuehne SA, Ehsaan M, Cartman ST, Cooksley CM, Scott JC, Minton NP. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. J Microbiol Methods. 2010 Jan;80(1):49-55.

Figure 1: The rise in RNomics over the last decade.

Year on year (a) and Cumulative total (b) publications from the Web of Knowledge database that contain both “sRNA*” and “bacteria*” in the title. It is clear that we are in exponential phase of this new and exciting research area and also that there is a significant body of literature.

Figure 2: Generalised genomic context of *cis* and *trans* acting small RNAs.

Cis acting sRNAs are generally found in the 5’ untranslated region of the mRNA (5’UTR), although less commonly, they may be encoded in the 3’UTR. Riboswitches and RNA thermometers fall into this class of sRNAs.

Trans acting sRNAs are encoded in intergenic regions of the genome (characterised by the presence of rho independent terminators and promoters in their sequence) and are transcribed independently of the target. They usually act by base pairing (often assisted by the Hfq RNA chaperone protein) with the target mRNA, influencing the output from that mRNA.

Figure 3: Small RNA molecules can act to modulate gene expression in a variety of ways.

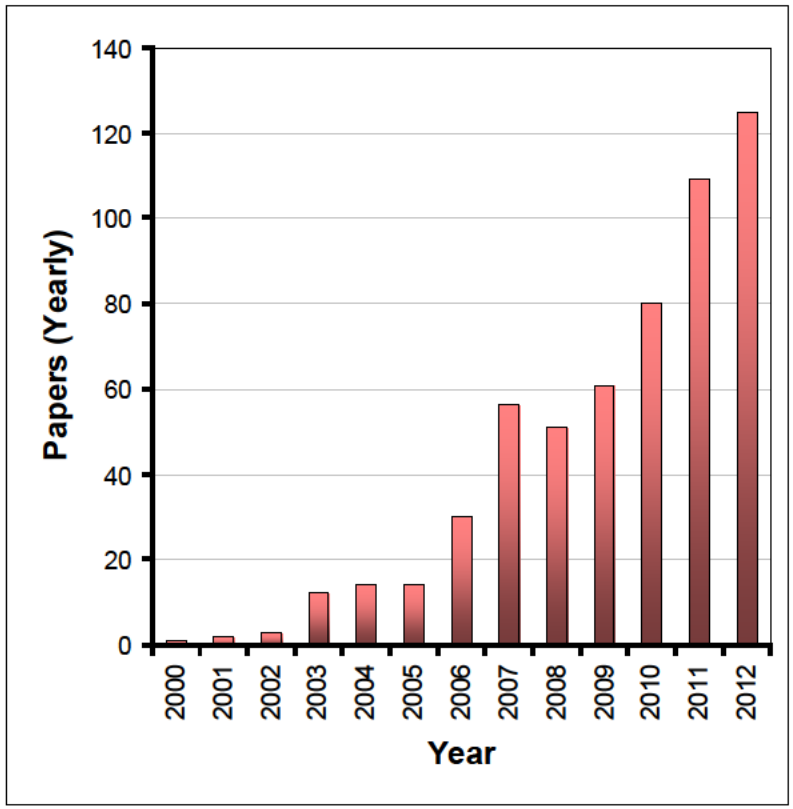
Base pairing of the sRNA with a target mRNA sequence can lead to (a) termination of transcription, (b) degradation of the mRNA, (c) occlusion of the ribosome binding site (RBS) and decreased translation or (d) changes in the secondary structure of mRNA such that the RBS is more accessible by the 30S ribosome and translation is increased. In an alternative mechanism, the (*trans* encoded) sRNA acts as a molecular decoy – here, binding of an inhibitor protein to the mRNA prevents

translation but if the inhibitor is sequestered by binding to the decoy sRNA, repression is lifted.

Figure 4. RNA thermometers and Riboswitches are examples of *cis* encoded small RNA molecules.

(A) At low temperature, the 30S ribosome is prevented from accessing the shine dalgarno (SD) sequence and the start codon (AUG) due to the complex secondary structure of the mRNA. Upon increasing temperature the secondary structure gradually melts and the ribosome can access the SD and AUG. This is thus a faster, direct, temperature sensing mechanism which is known to regulate heat shock gene expression and virulence in bacteria. Sequence conservation in the 5' aptamer domain enables database searches for identification of these thermosensing elements.

(B) Generalised mechanism for expressional control via metabolite binding to *cis* acting riboswitches. The riboswitch consists of a sensor aptamer domain which can bind the metabolite (for example, anions, metal ions, co factors, purines and amino acids are all known to direct switching) and an expression platform. Riboswitches sense different concentrations of a single metabolite and upon highly discriminatory binding of the metabolite to the aptamer domain, the secondary structure of the element changes to allow changes in transcription, translation, splicing and mRNA stability.


Papers (Yearly)
140
120
100
80
60
40
20
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
Year

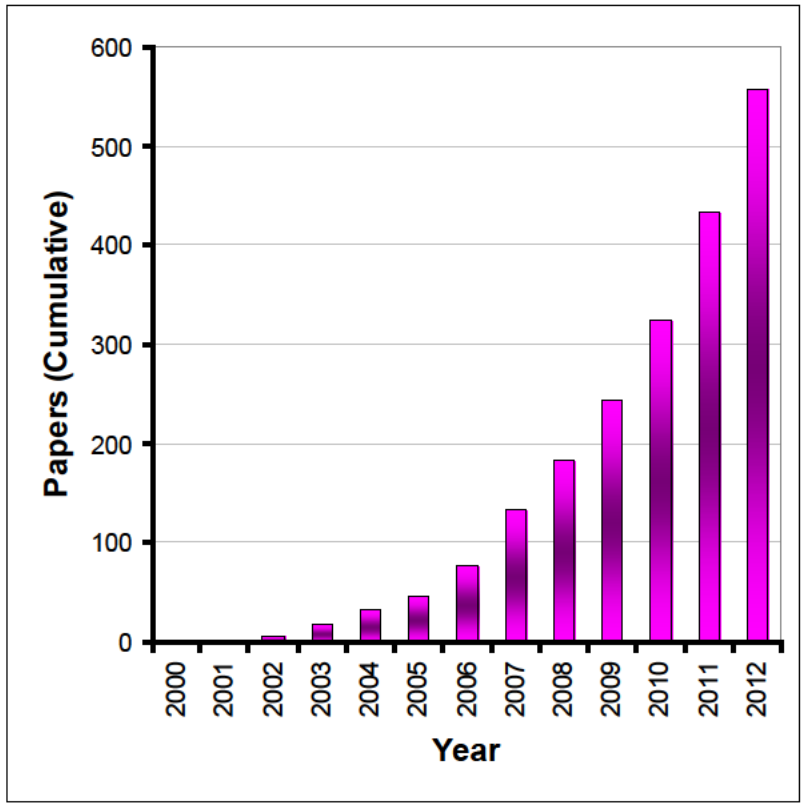
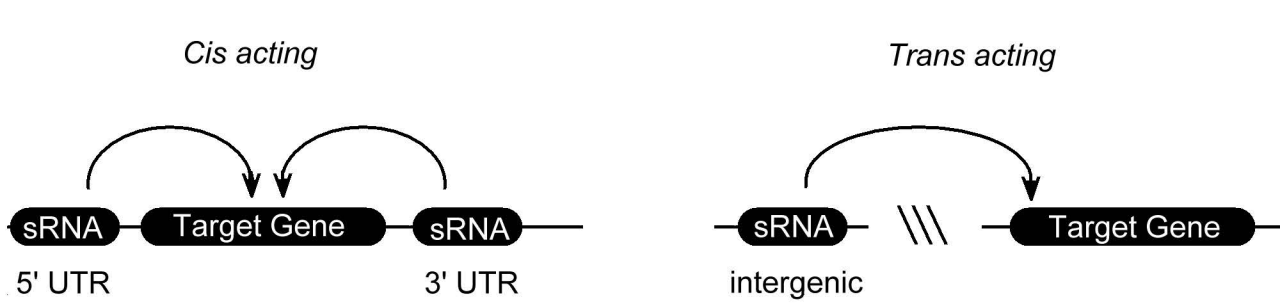
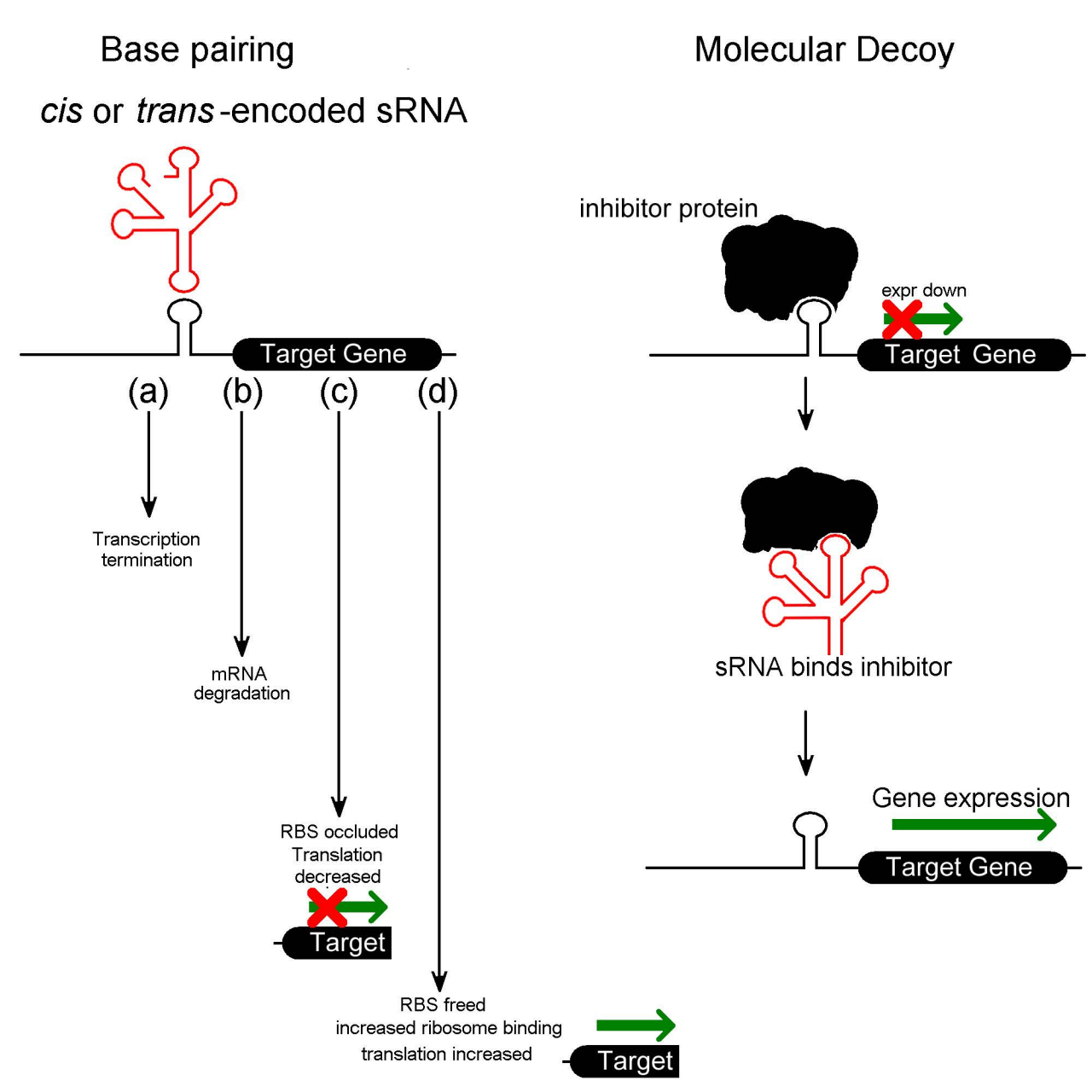
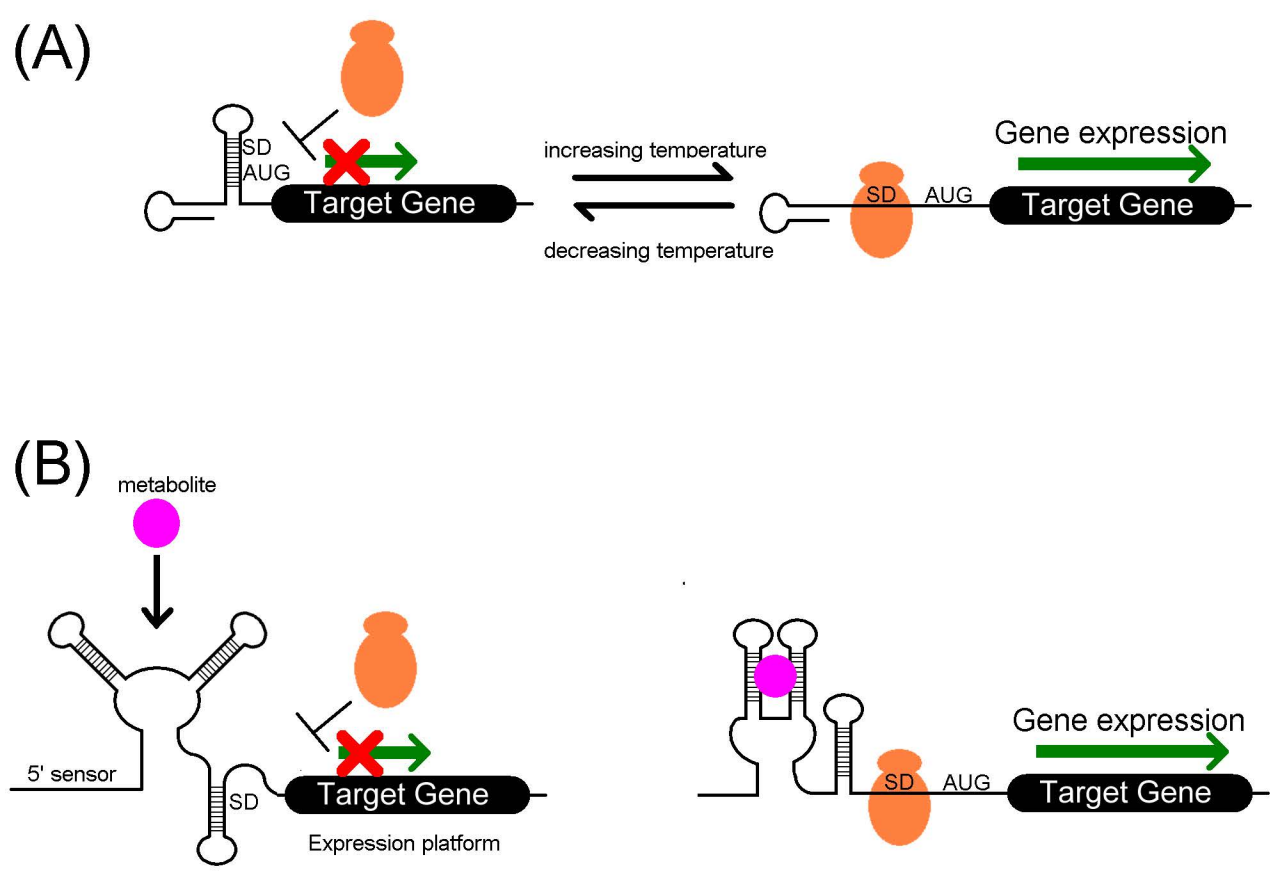

Papers (Cumulative)
600
500
400
300
200
100
0
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
Year

Figure 1


Cis acting
sRNA
Target Gene
sRNA
5' UTR
3' UTR
Trans acting
sRNA
Target Gene
intergenic


Base pairing
cis or trans-encoded sRNA
Target Gene
(a)
(b)
(c)
(d)
Transcription termination
mRNA degradation
RBS occluded
Translation decreased
Target
RBS freed
increased ribosome binding
translation increased
Target
Molecular Decoy
inhibitor protein
expr down
Target Gene
sRNA binds inhibitor
Gene expression
Target Gene


(A)
SD
AUG
Target Gene
increasing temperature
decreasing temperature
Gene expression
SD
AUG
Target Gene
(B)
metabolite
5' sensor
SD
Target Gene
Expression platform
Gene expression
SD
AUG
Target Gene