CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

INSTITUT PASTEUR

Small non-coding RNAs in bacteria

Massimo Vergassola

CNRS &

Pasteur Institute Research Unit

In Silico Genetics

A (partial and biased) selection of the numerous regulatory processes involving ncRNAs in the bacterial world

Two ubiquitous across bacteria: SsrS (6S) and SsrA (tmRNA)

• 6S RNA (SsrS in E. coli) was first detected as an abundant peak in *in vivo* gels by Hindley in **1967.** Its function was determined by Wassarman & Storz, (Cell, 2000): it confers selectivity to the RNAP and controls the equilibrium between the various polymerases.

•SsrA (tmRNA) is the transfer-messenger RNA, which rescues stalled ribosomes.

Jacob and Monod looked first for a RNA repressor and just afterwards started looking for a protein repressor C.R. Acad. Sci., 254, pp. 4214-4216, 1962; JBM 1961

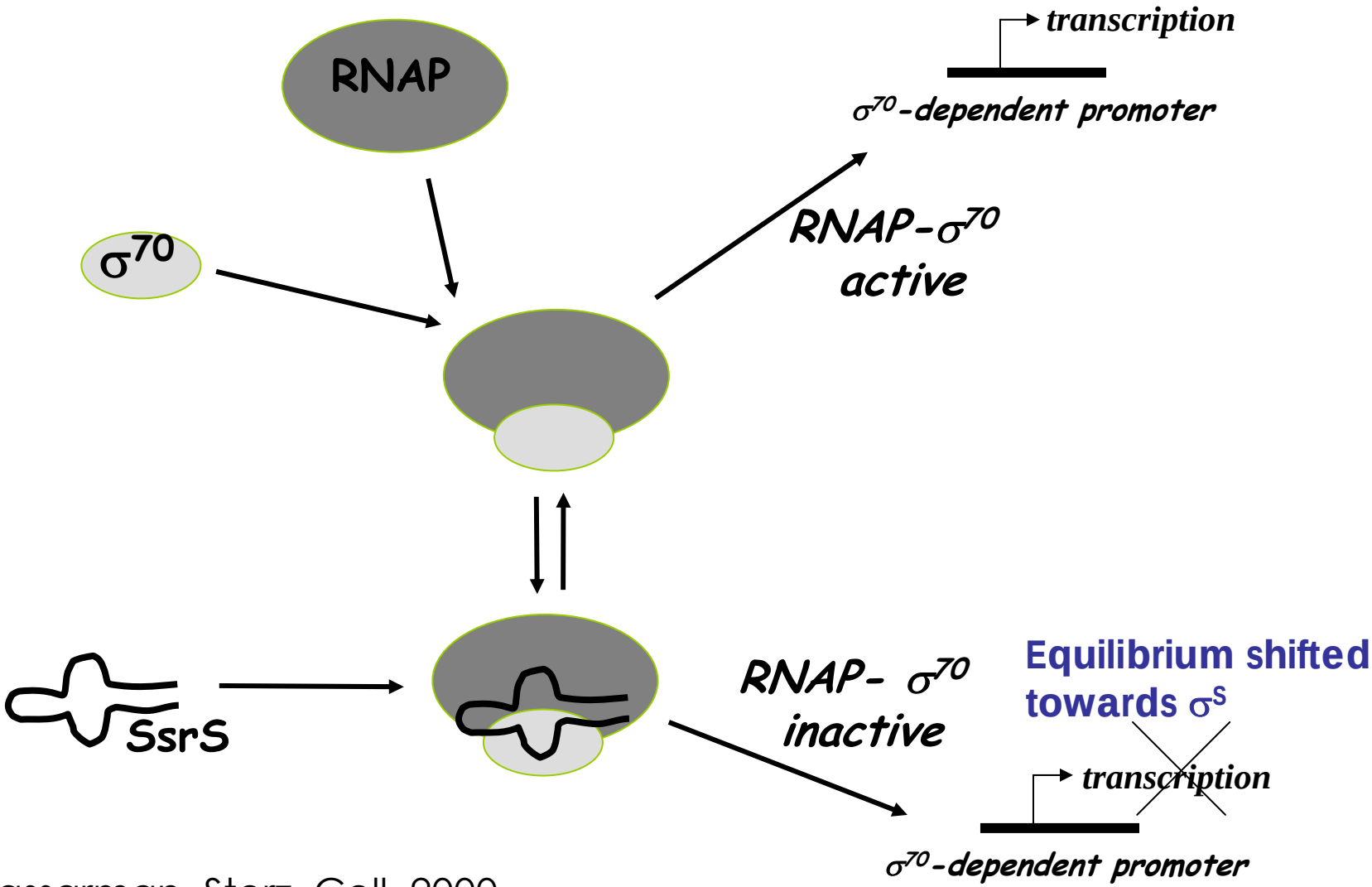
SsrS (6S) is encoded upstream of an ORF (ygfA) from a σ^{70} promoter

SsrS genomic locus in E. coli

Standard σ^{70} promoter

At the entry in stationary state there is an excess of about 10,000 copies of SsrS, to be compared with O(1000) of RNAP and σ^{70}.

SsrS (6S) confers selectivity to the RNAP



Wassarman, Storz, Cell, 2000

Wassarman, Storz, Cell, 2000

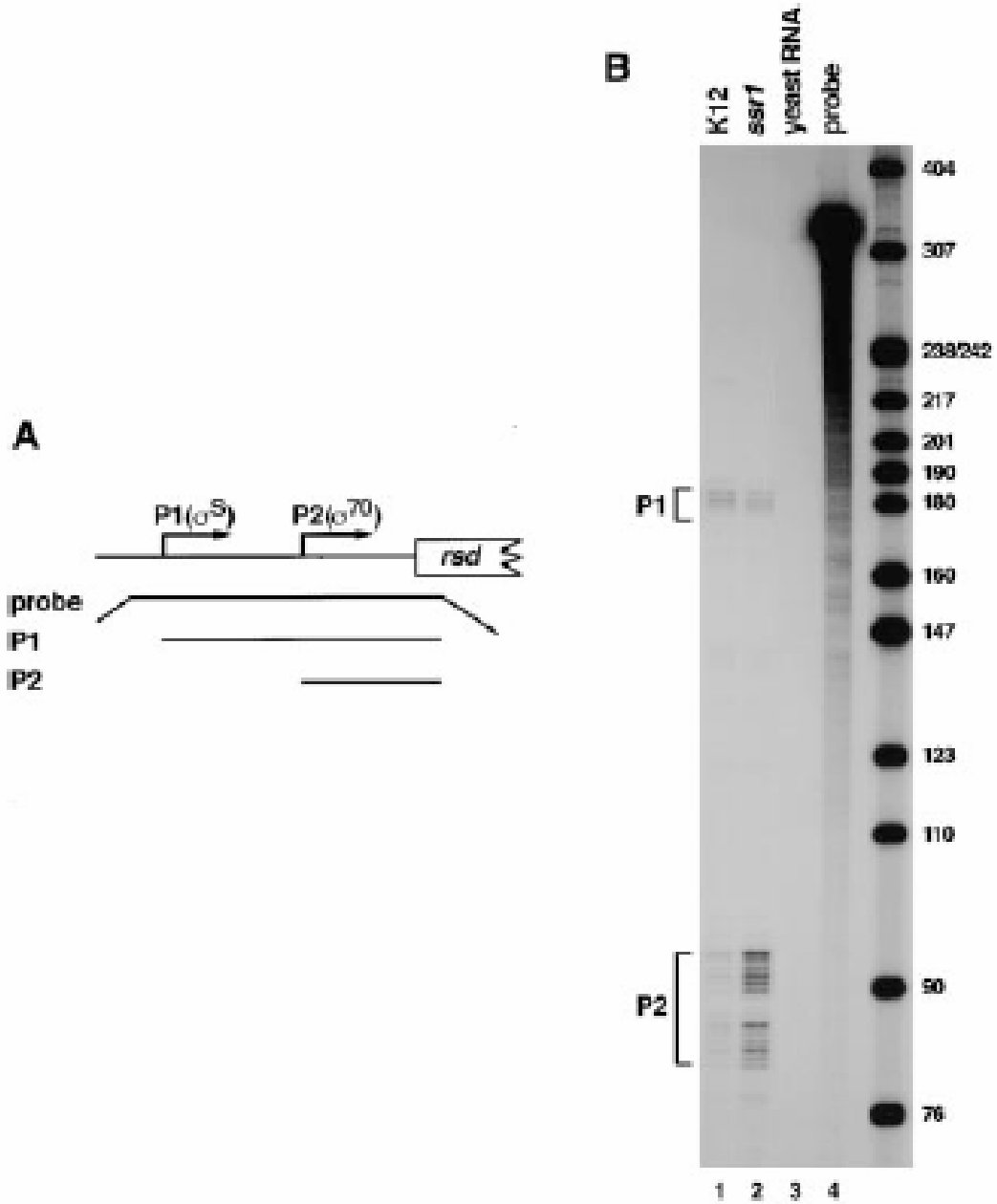


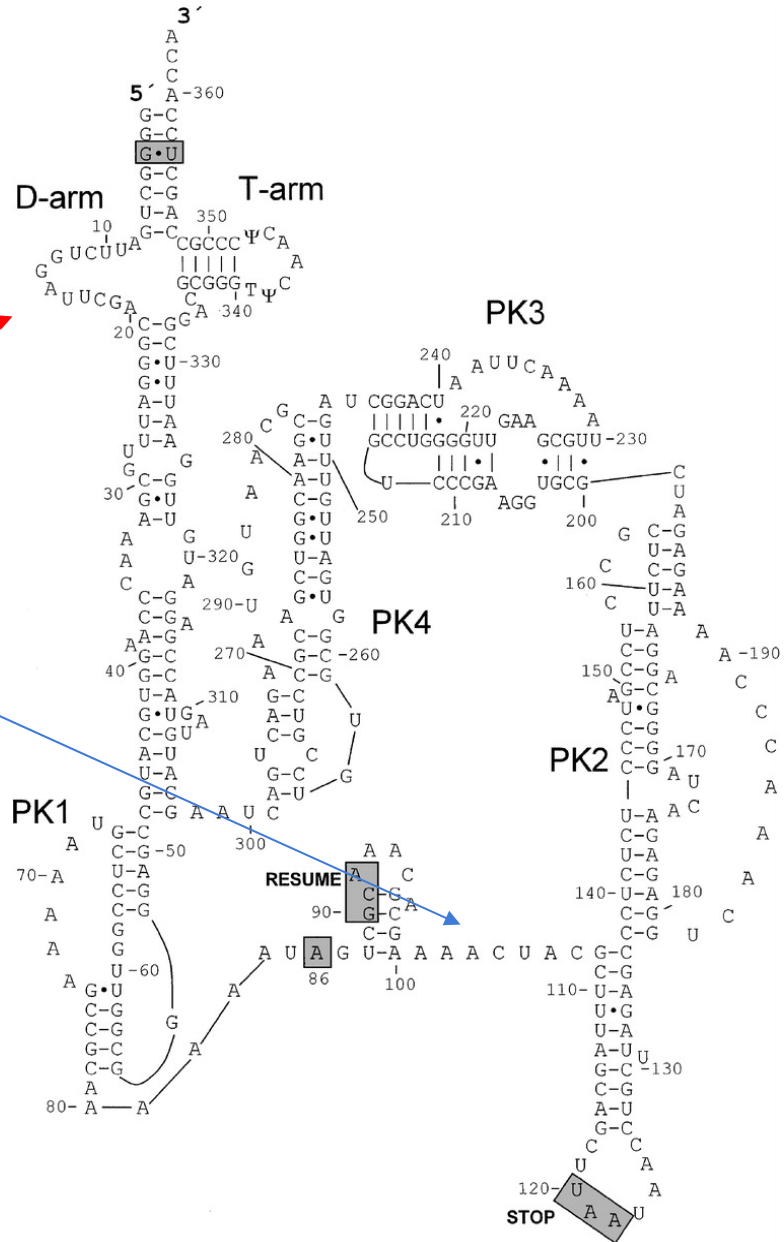
Figure 6. Utilization of a σ^{70} Promoter Is Altered in the Absence of 6S RNA

SsrA (tmRNA): a mimetic ncRNA

tmRNA:

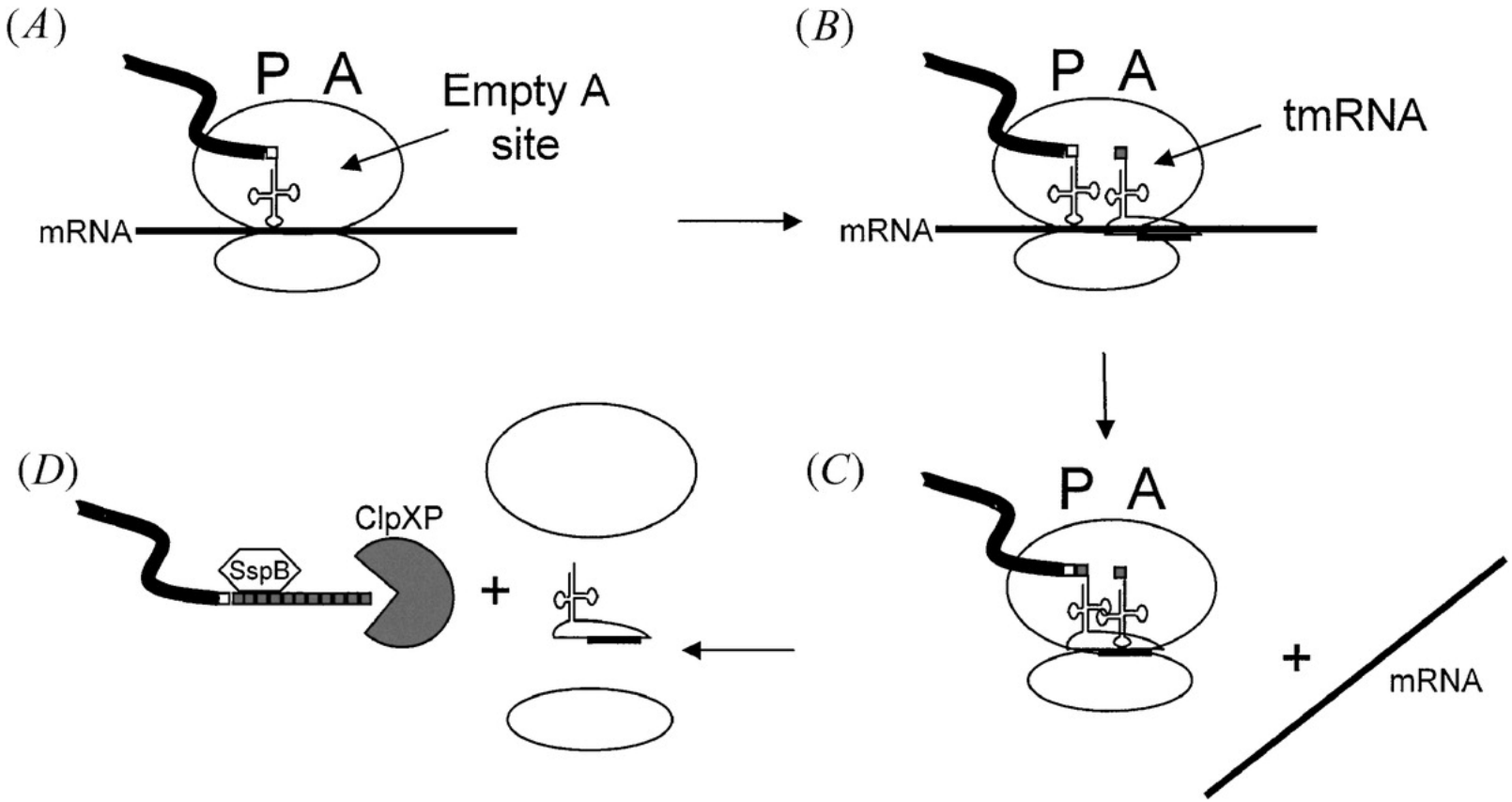
transfer-messenger

First detected in 1978 by Apirion and coll., J. Bact.



http://www.indiana.edu/~tmrna/

SsrA: rescuing ribosomes

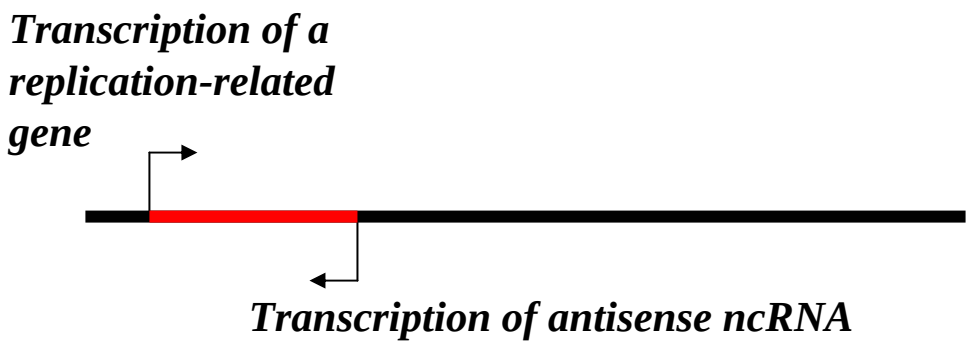


Cis-acting RNAs: antisense, riboswitches, thermosensors

Plasmid copy number control

The number of plasmids in bacterial cells is under regulation, preventing excessive reduction (risk of loss of the plasmid) and/or excessive replication (runaway is strongly deleterious).

For a number of plasmids this regulation features an essential ncRNA component of the type



Antisense: complementarity between the ncRNA and its interaction partner

First demonstrated by Itoh & Tomizawa, PNAS, 1980; Lacatena & Cesareni, Nature 1981 for control of ColE1 plasmid.

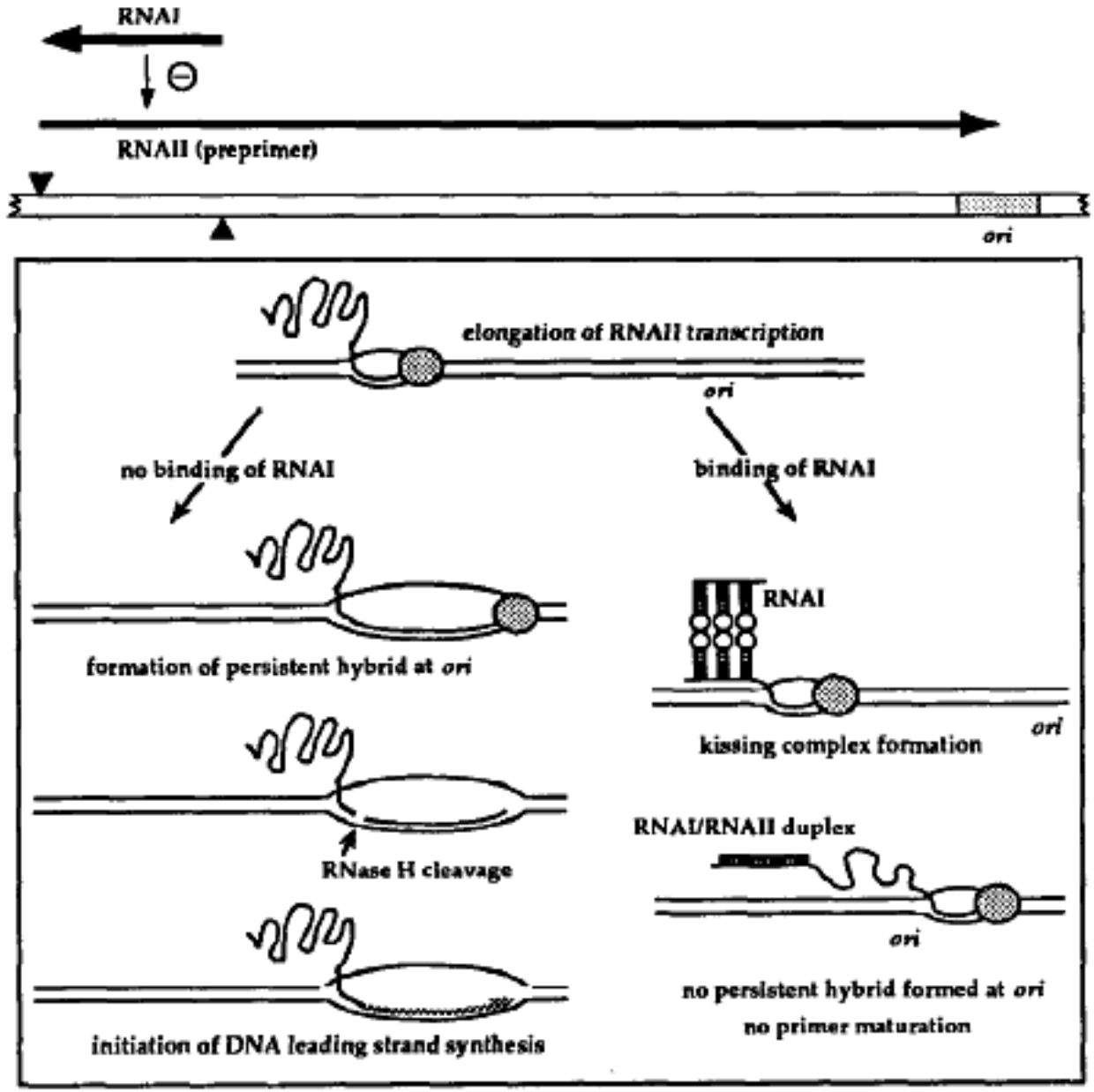
ColE1 plasmid antisense control

Reviews:

Eguchi, Itoh, Tomizawa, Ann. Rev. Biochem., 1991;

Wagner & Simons, Ann. Rev. Microb., 1994;

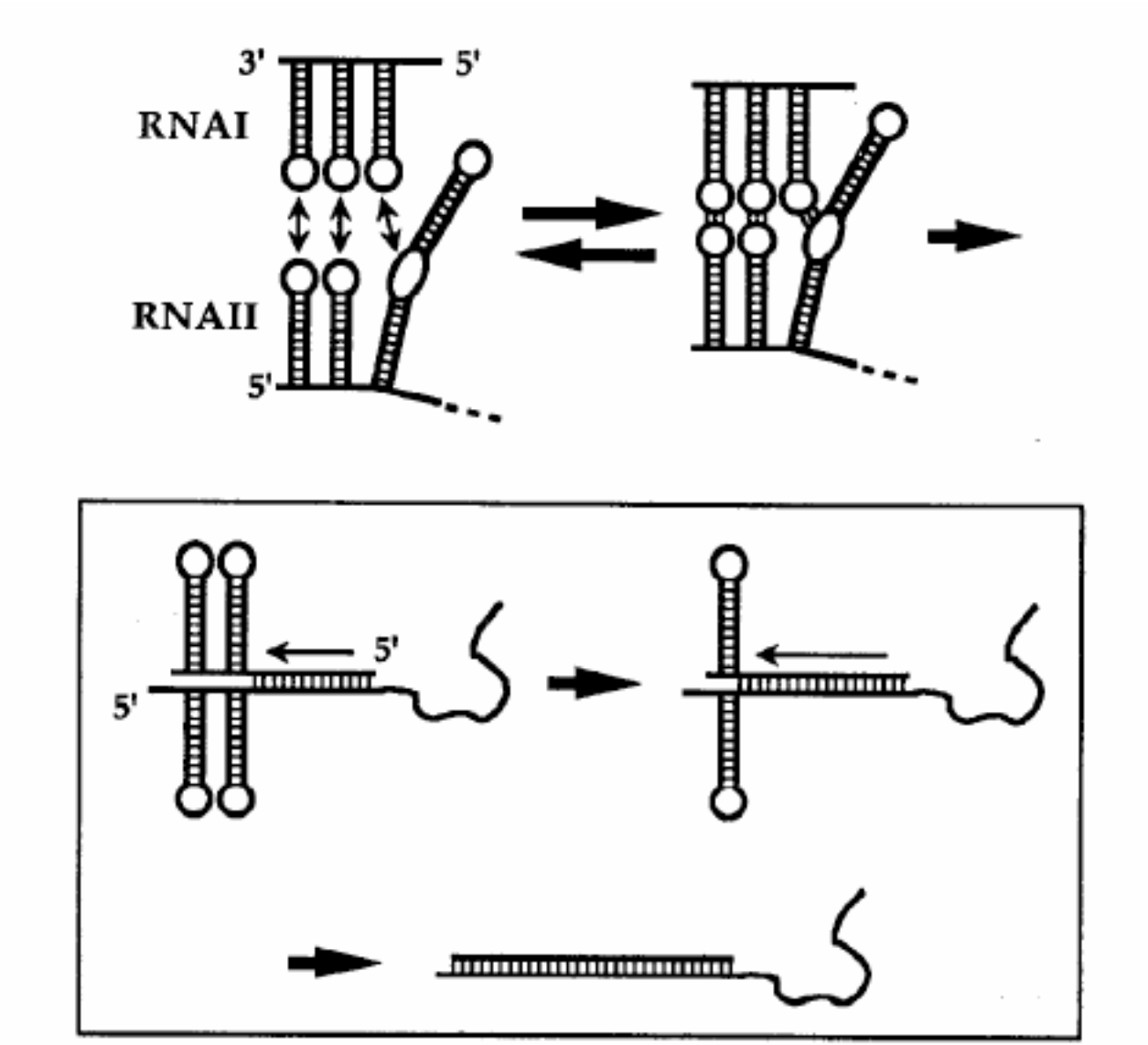
Wagner, Altuvia, Romby, Adv. In Genetics, 2002



RNAI-RNAII pairing dynamics

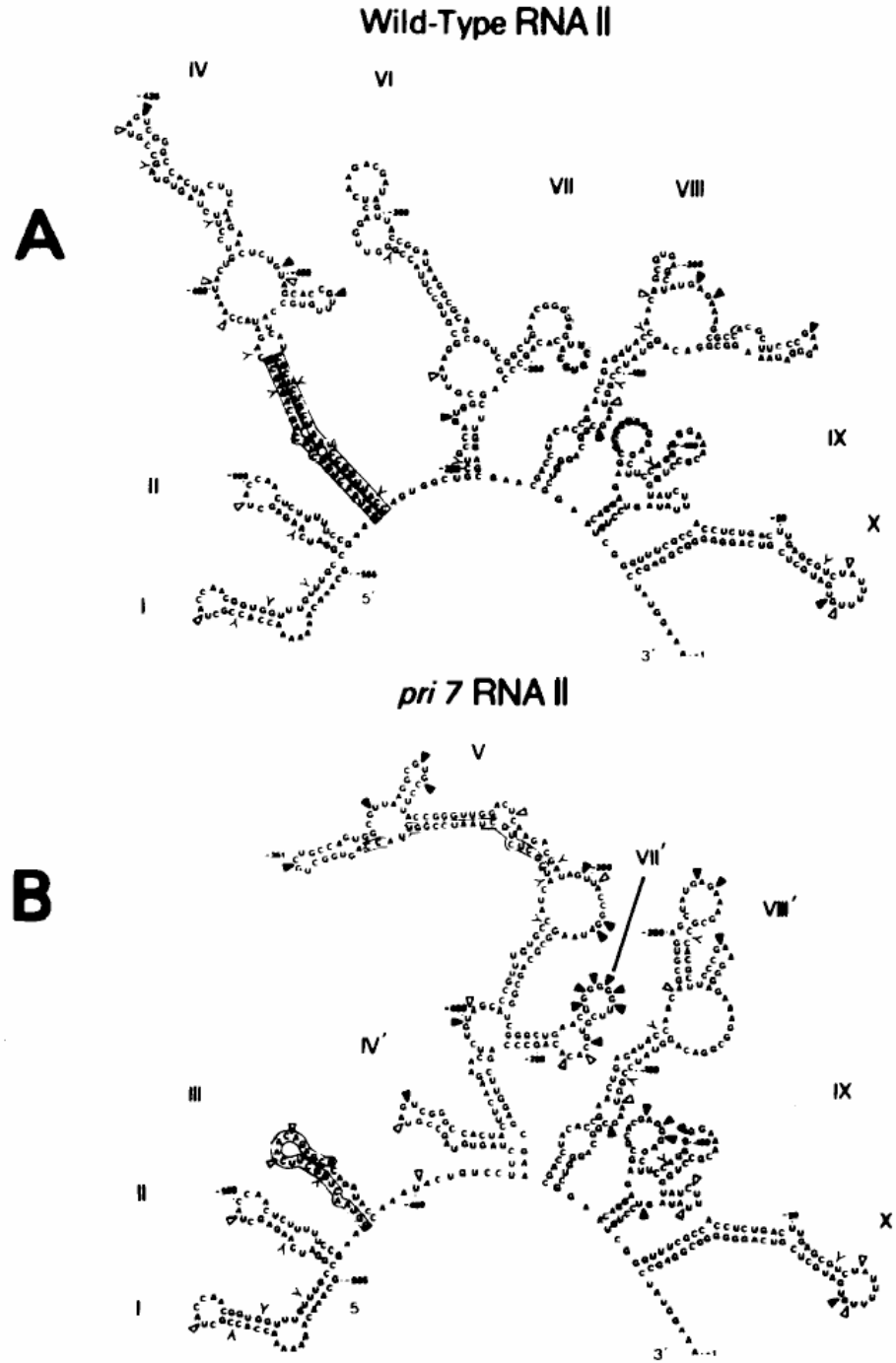
Cartoon from

Wagner & Simons, Ann. Rev. Microb., 1994



RNAI-RNAII folding structures

Mutant employed to test the model; obtained mutating **one** key base (G⌫A) to disrupt structure IV of the wild type.

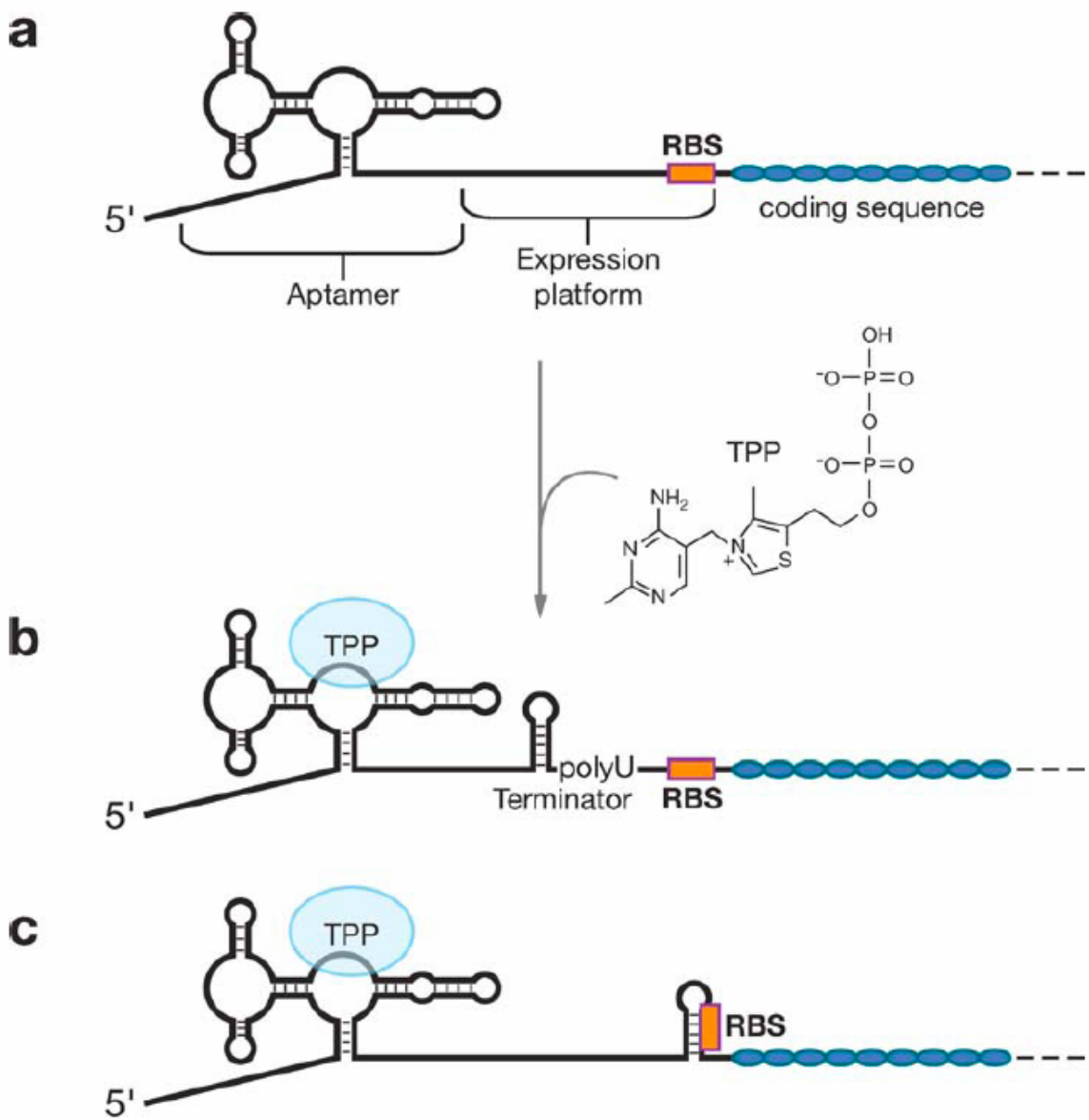


Schematic view of a riboswitch

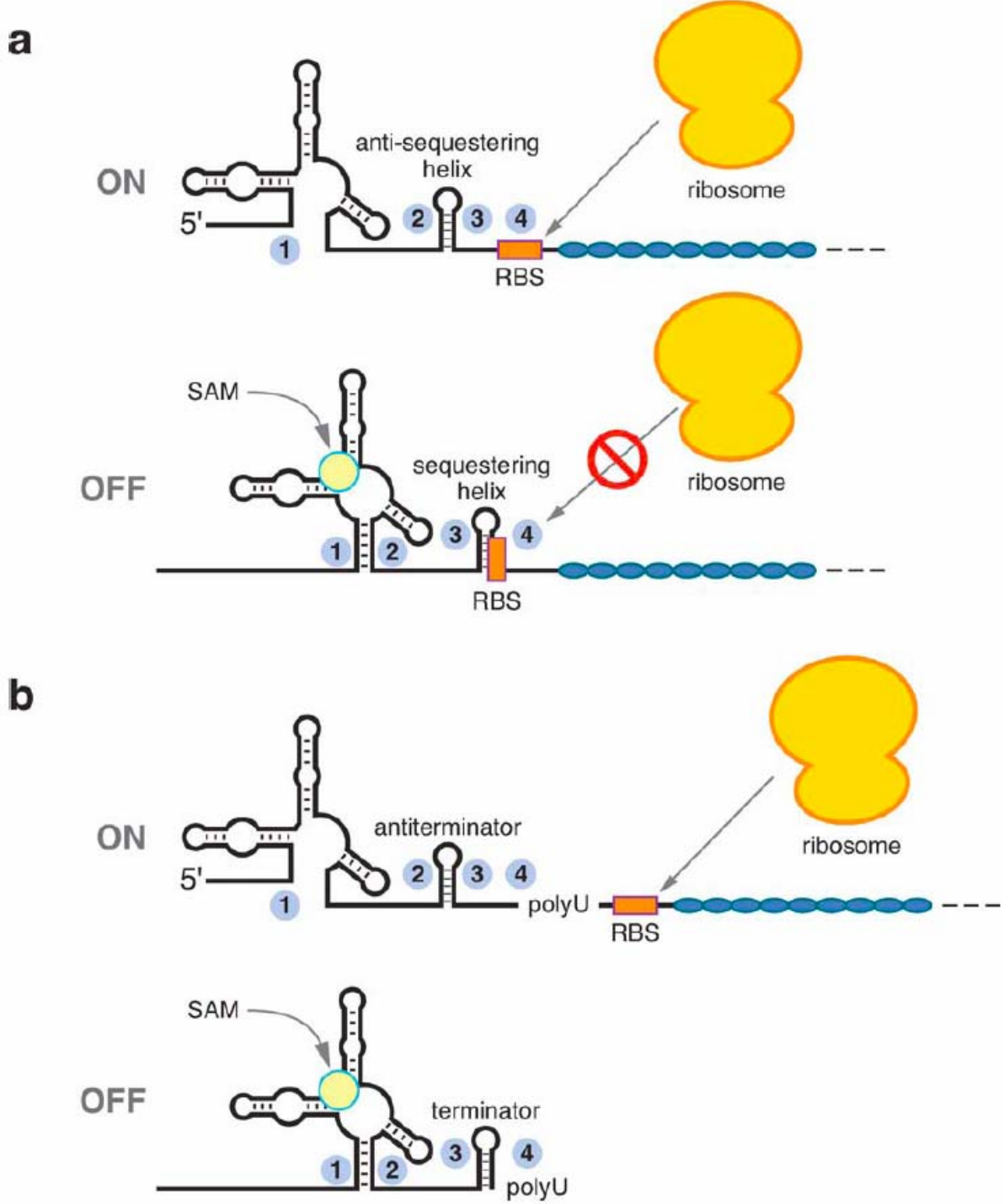
TPP is thyamine pyrophosphate, but there is a plethora of aptamers, specific to particular metabolites

Database RFAM to check if your favorite RNA is a riboswitch of some sort: www.sanger.ac.uk/Software/Rfam/

Winkler & Breaker, Ann. Rev. Microb. 2005; Coppins et al., Curr. Opin. Microb., 2007 for reviews

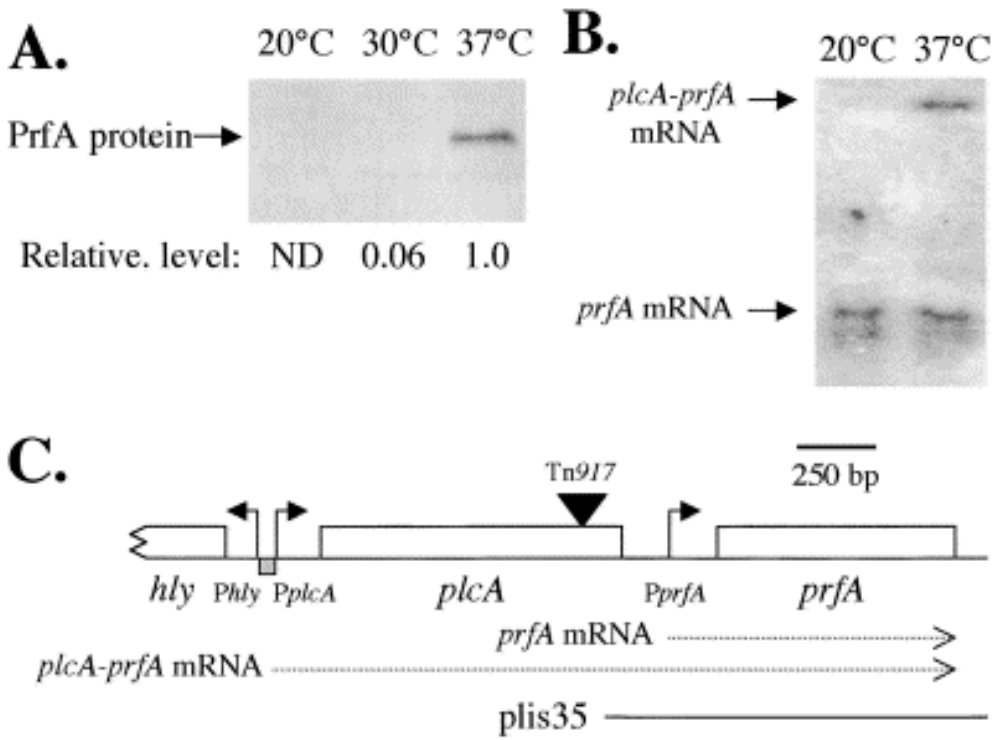


Cartoons of modes of action



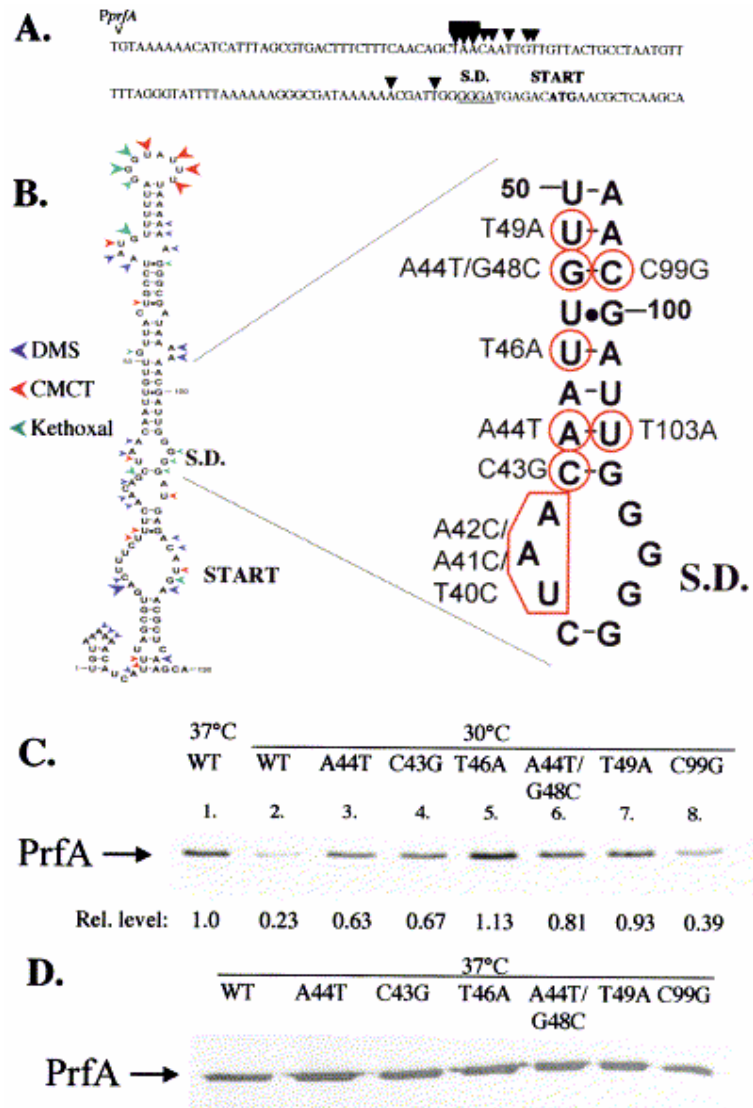
A RNA thermosensor

PrfA is the master regulator of the virulence of Listeria, the bacterium responsible for listeriosis.



Johansson et al., Cell 2002

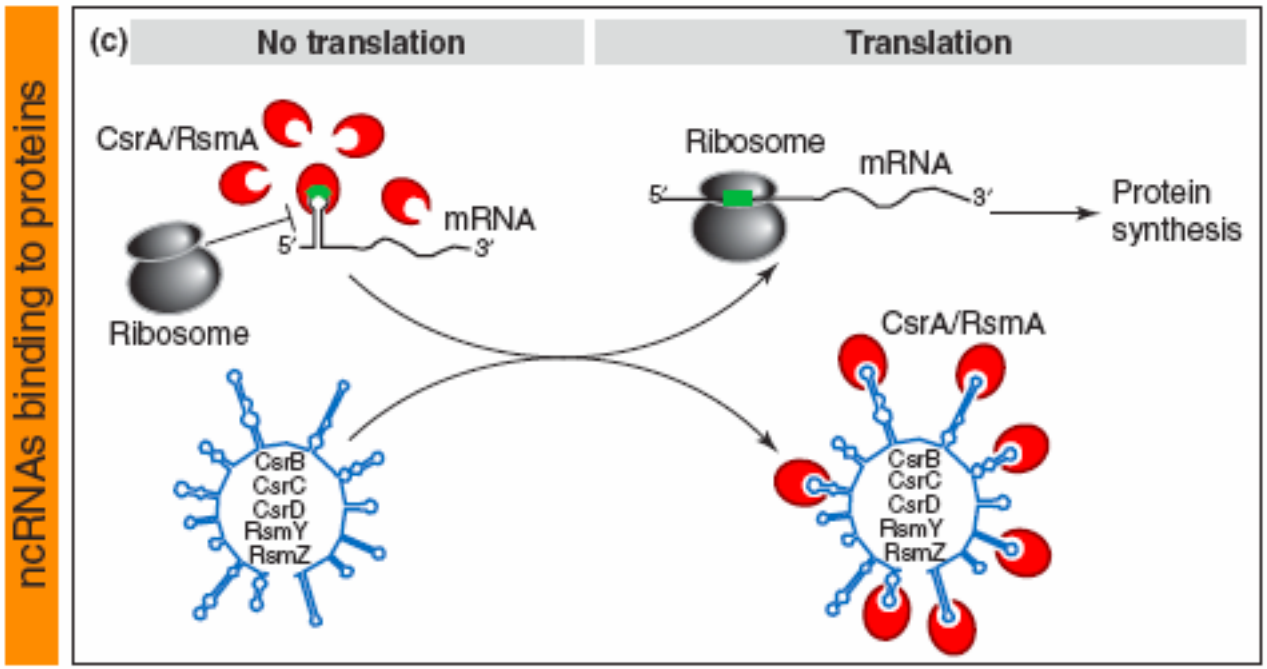
The effect is due to a stem, stable at low temperatures and sequestering SD, and getting unstable above ≈37°



Johansson et al., Cell 2002

Trans-acting ncRNAs: protein/mRNA targets

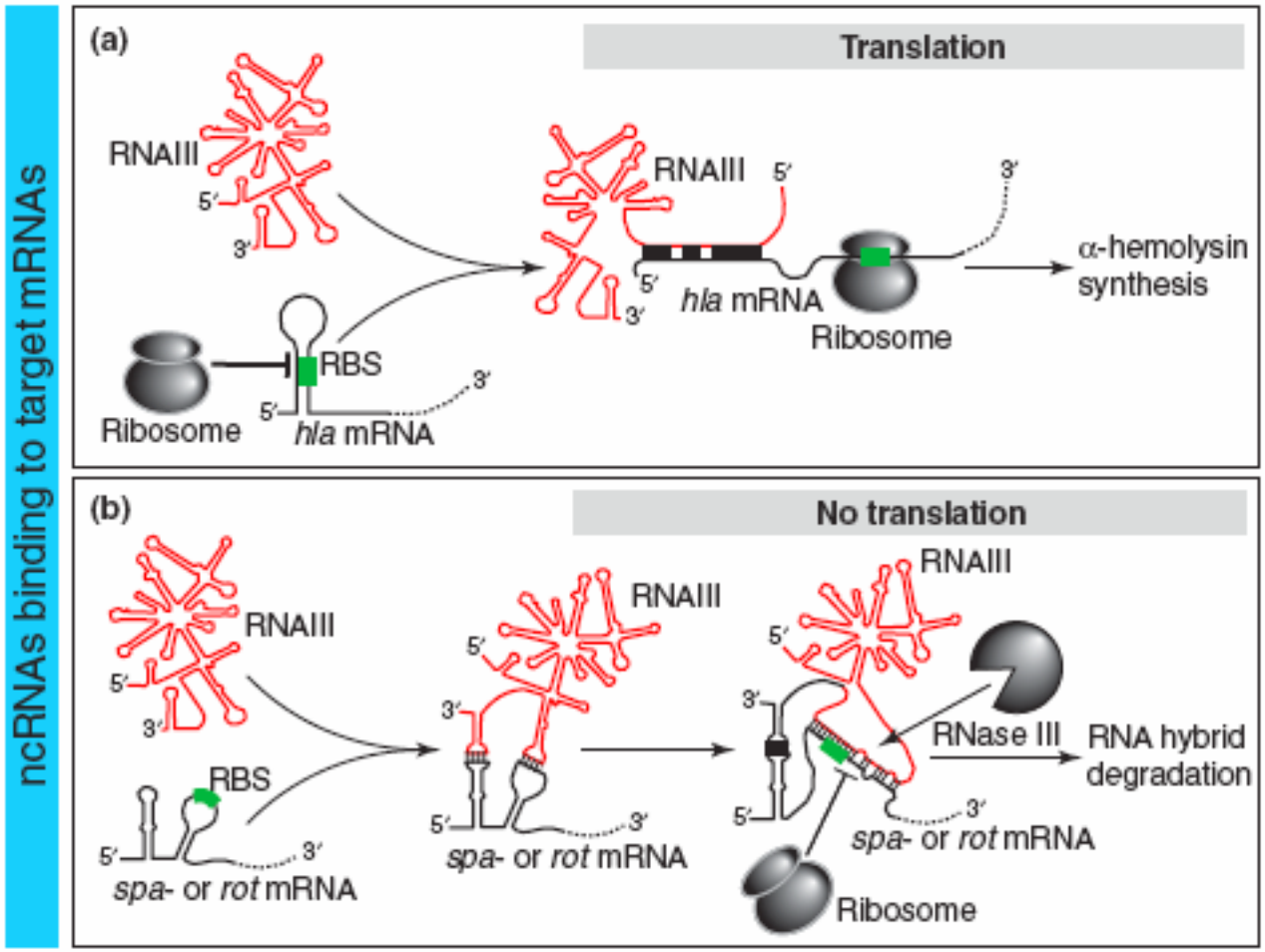
Protein-interacting ncRNAs

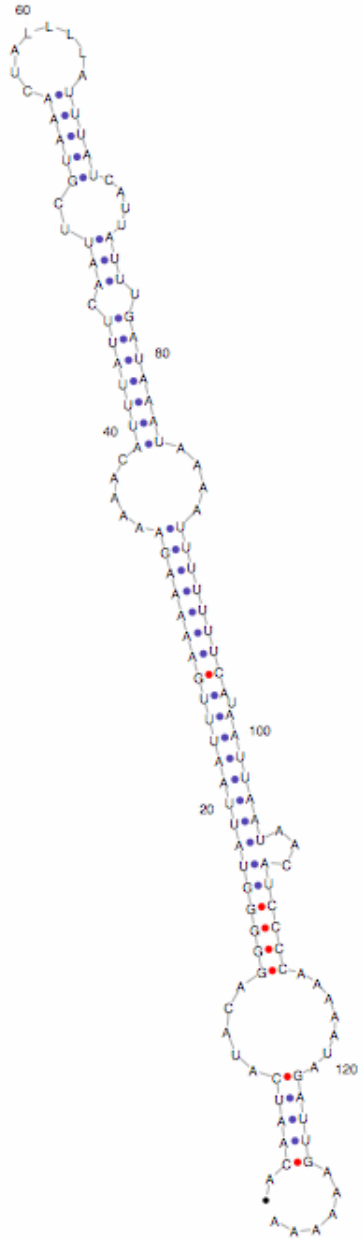


mRNA-interacting ncRNAs

RNAIII is a 514nt long transcript of *S.aureus.*

It encodes an ORF and a ncRNA that regulates a set of genes involved in virulence. The expression of RNAIII is regulated by the agrA/C system.

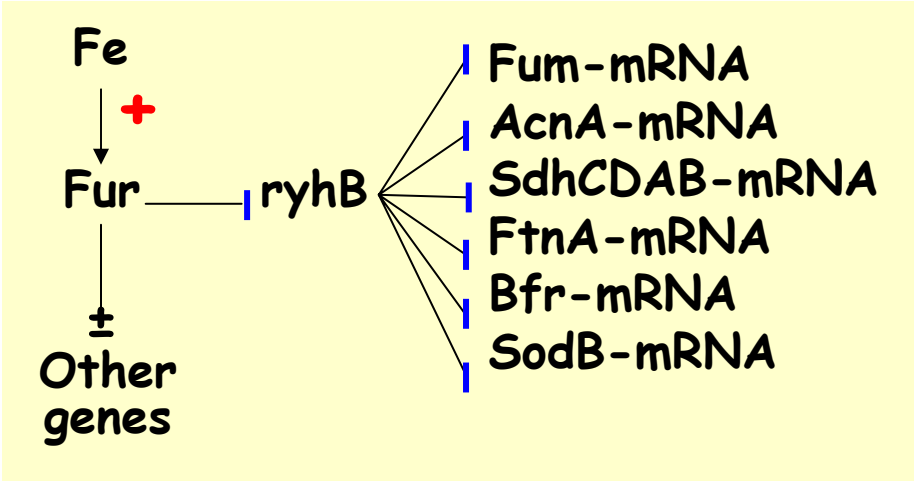




dG = -24.2 [initially -25.9] spa

RyhB regulates iron metabolism in E. coli

(Massé & Gottesman, 2002, 2003)



RyhB targets multiple mRNAs coding for iron-using and/or iron-storing proteins

RyhB expression is repressed by Fur (Fe^{2+} cofactor).

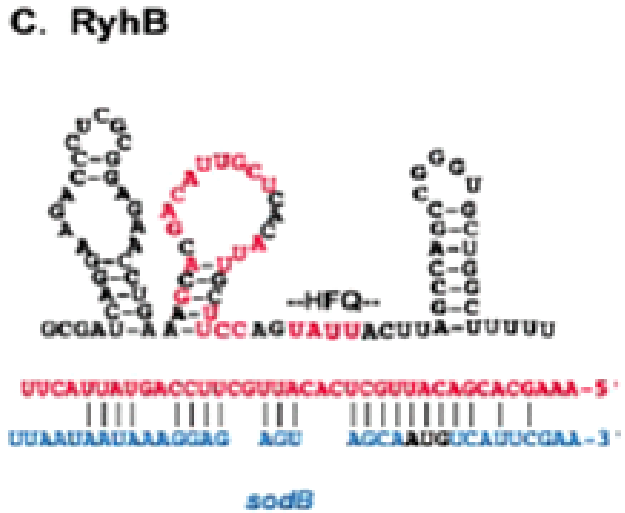
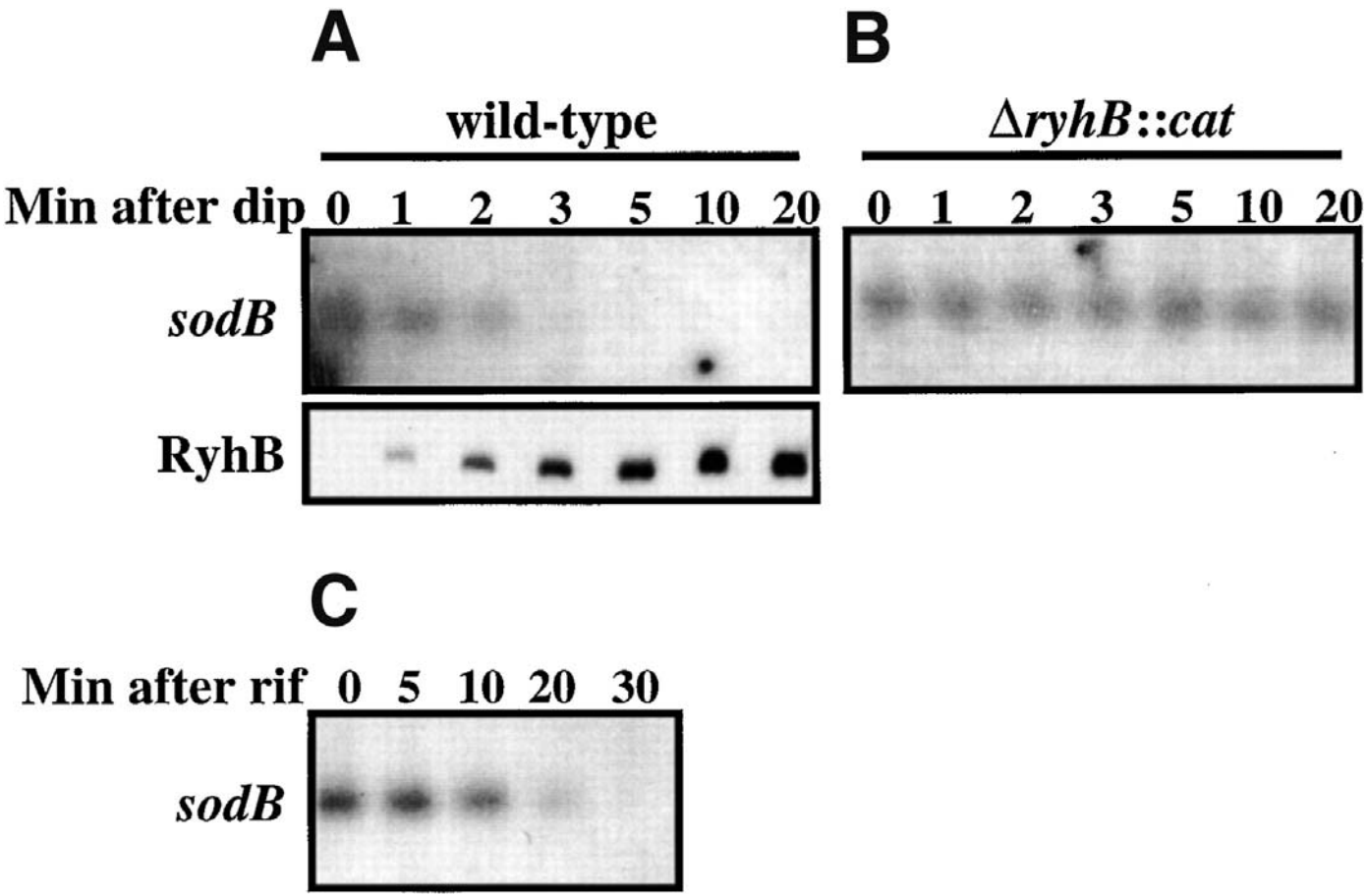


Figure 1. Degradation of full-length sodB mRNA



Eric Masse et al. Genes Dev. 2003; 17: 2374-2383

Spot42 regulates polarity in the *gal* operon

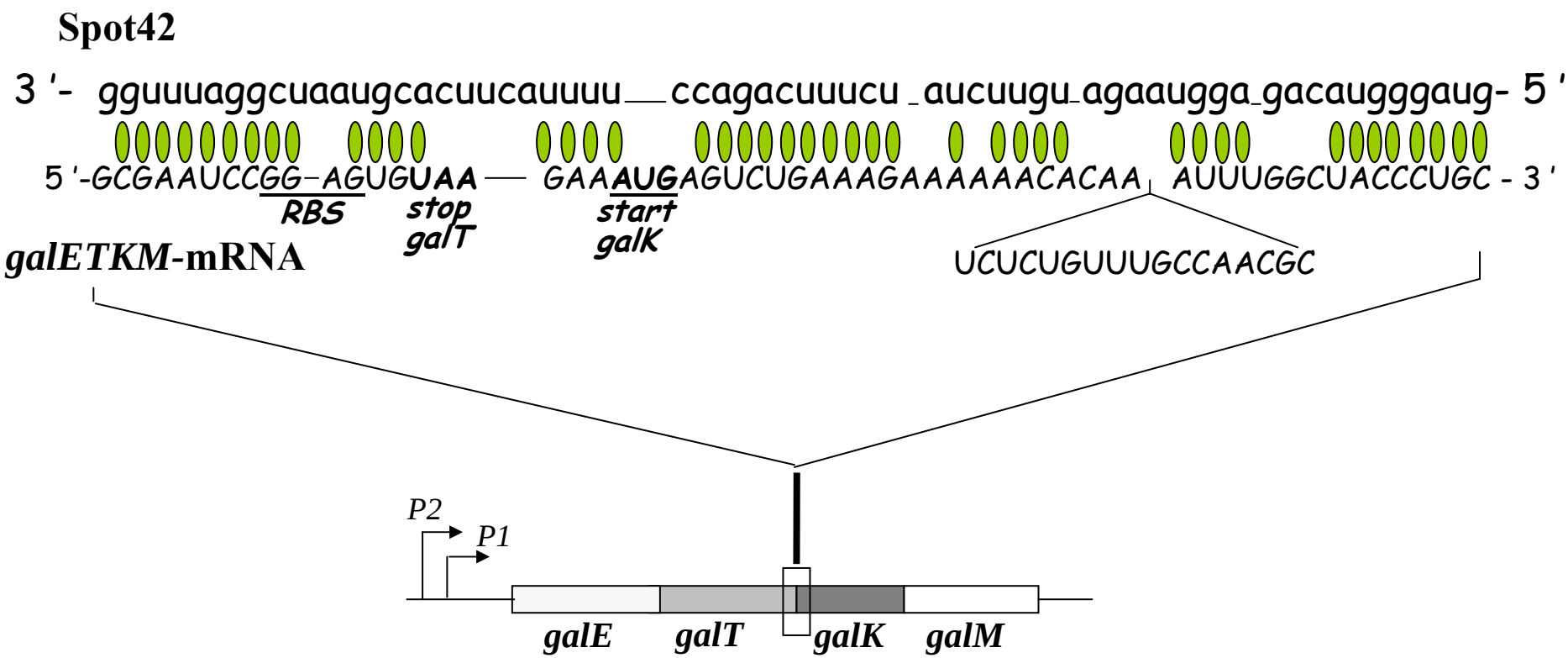
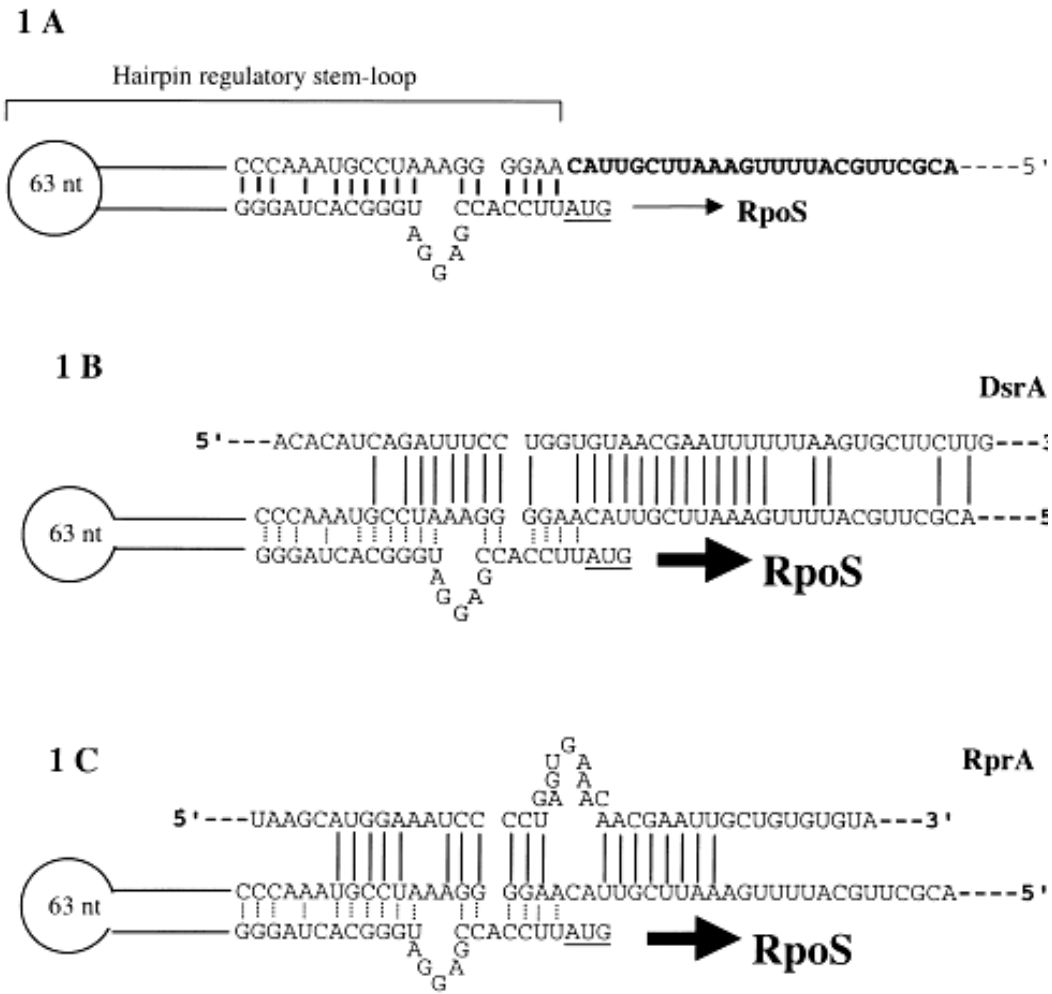
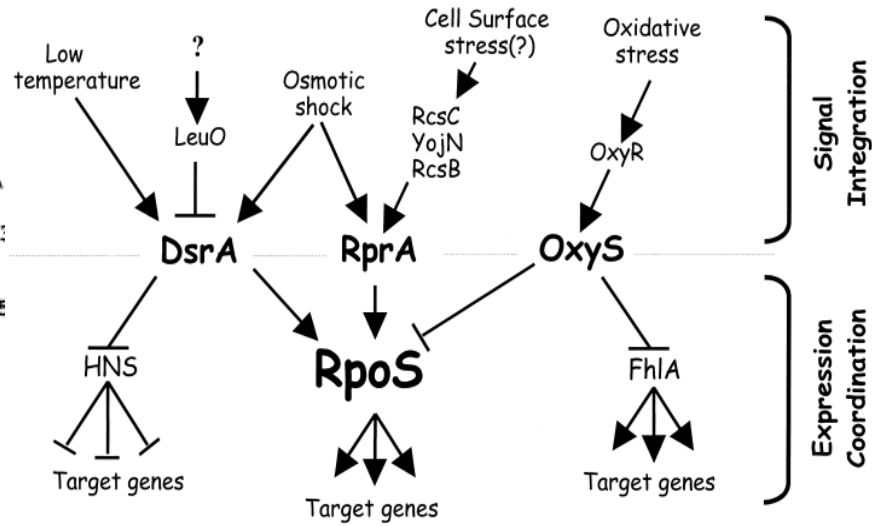


Figure 1. Base pairing between Spot42 and *galETKM*-mRNA. In the upper part, Spot42 (top sequence) pairs the translation initiation region of *galK* (bottom sequence). The translation start codon and the ribosome binding site (RBS) of *galK* are underlined. The translation stop codon of *galT* is indicated in bold. The lower part of the schematic shows the organization of the *galETKM* operon. The two promoters (P1 and P2) driving the expression of the *gal* operon are shown by the arrows. The rectangle on the operon indicates the location of the pairing Spot42- *galETKM*-mRNA. Adapted from [37].

Combinatorial regulation by ncRNAs





P Romby et al, 2006; G Storz et al 2005; S Gottesman, 2004; Toledo-Arana et al. 2007.
M. Guillier, S. Gottesman, and G. Storz Genes & Dev., 2006

How does one find them?

(There is no general approach and all have limits …)

- Functional genetic screens
- Microarrays of Igs
- Rnomics
- Computational screens
- Co-immunopreicipitation with Hfq

(Christiansen et al., RNA 2006; 3 in Listeria)

Vogel & Sharma, How to find small non-coding RNAs in bacteria, Biol. Chem. 2005, 386, 1219

Simple computational methods

- Comparative searches on IGs based on sequence conservation and folding (ncRNAs are often well-folding) (Wassarman et al.; Argaman et al., 2001).

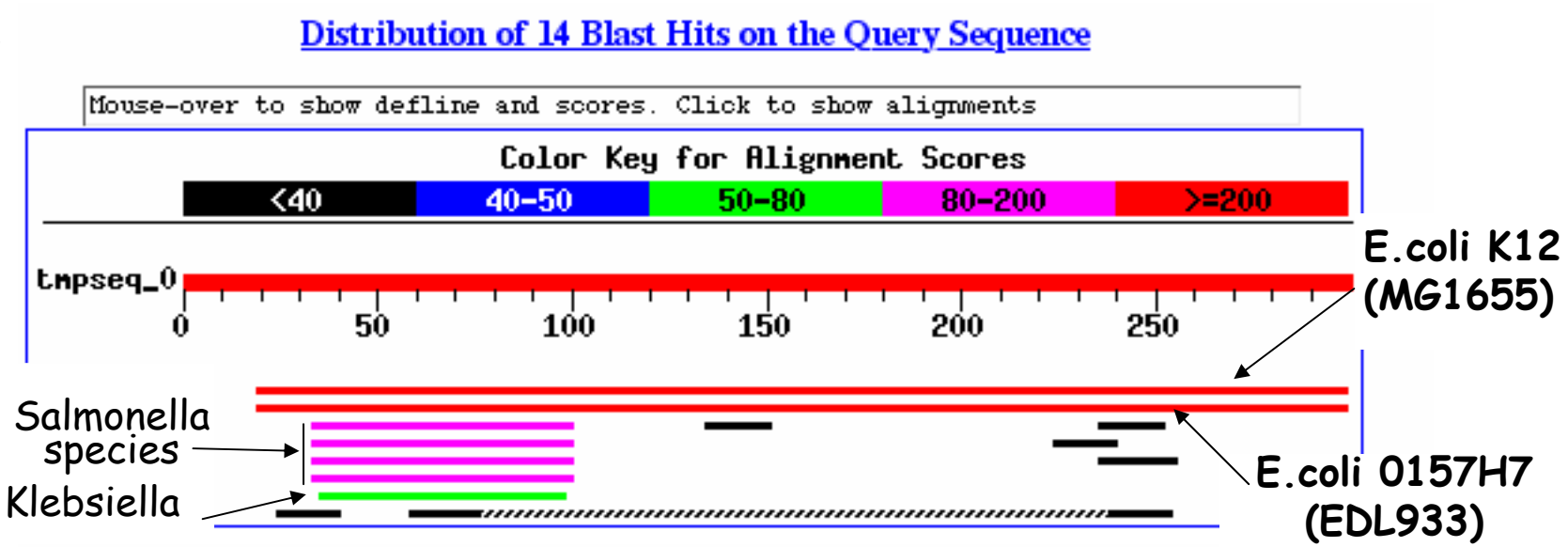
- “Weird” transcriptional structures in IGs such as orphan terminators and promoters.

- These search criteria were automatized in sRNAPredict(2) (Livny et al., Nucl. Acid Res., 2005, 2006). Predictions for P. aeruginosa gave 17/31.

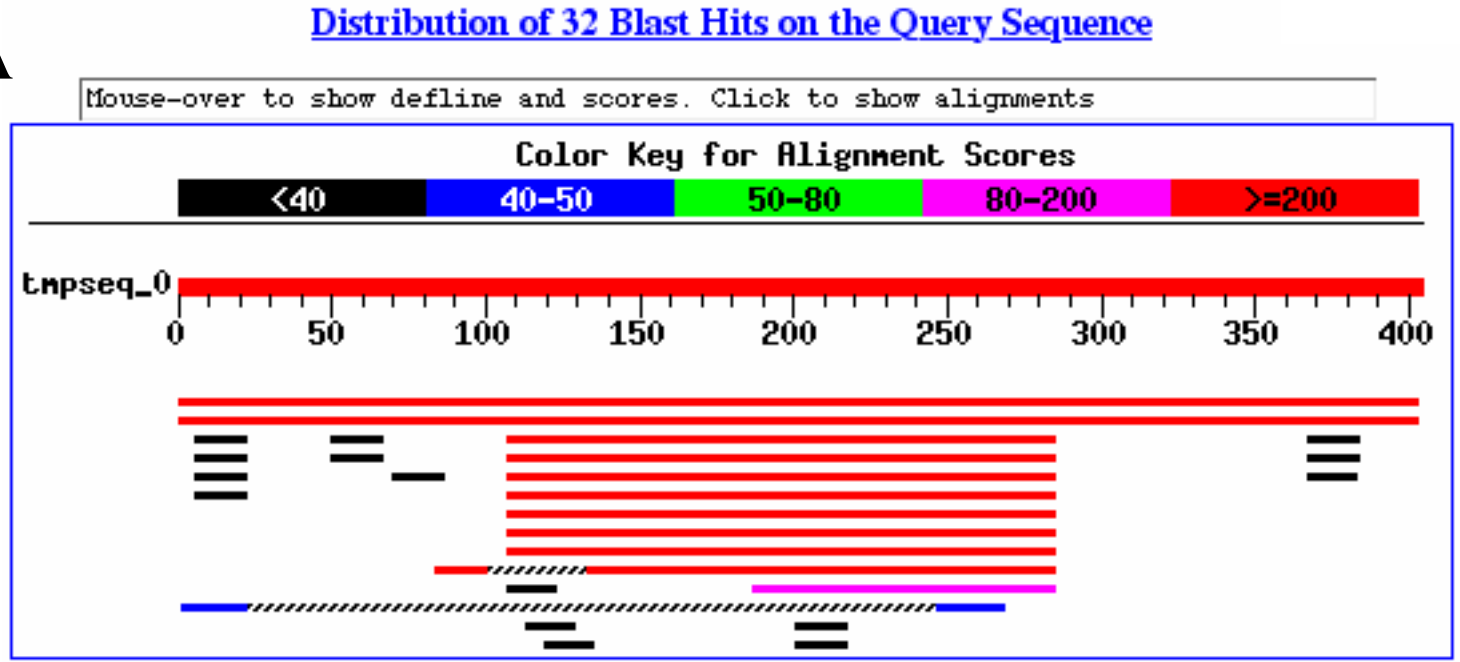
Candidates are then assayed either by individual Northern or using an array to spot all of them together (Pichon & Felden, PNAS 102, 14249, 2005: 12 new ncRNAs predicted for S. aureus).

Examples of IGs in *E. coli* carrying ncRNAs encoding genes

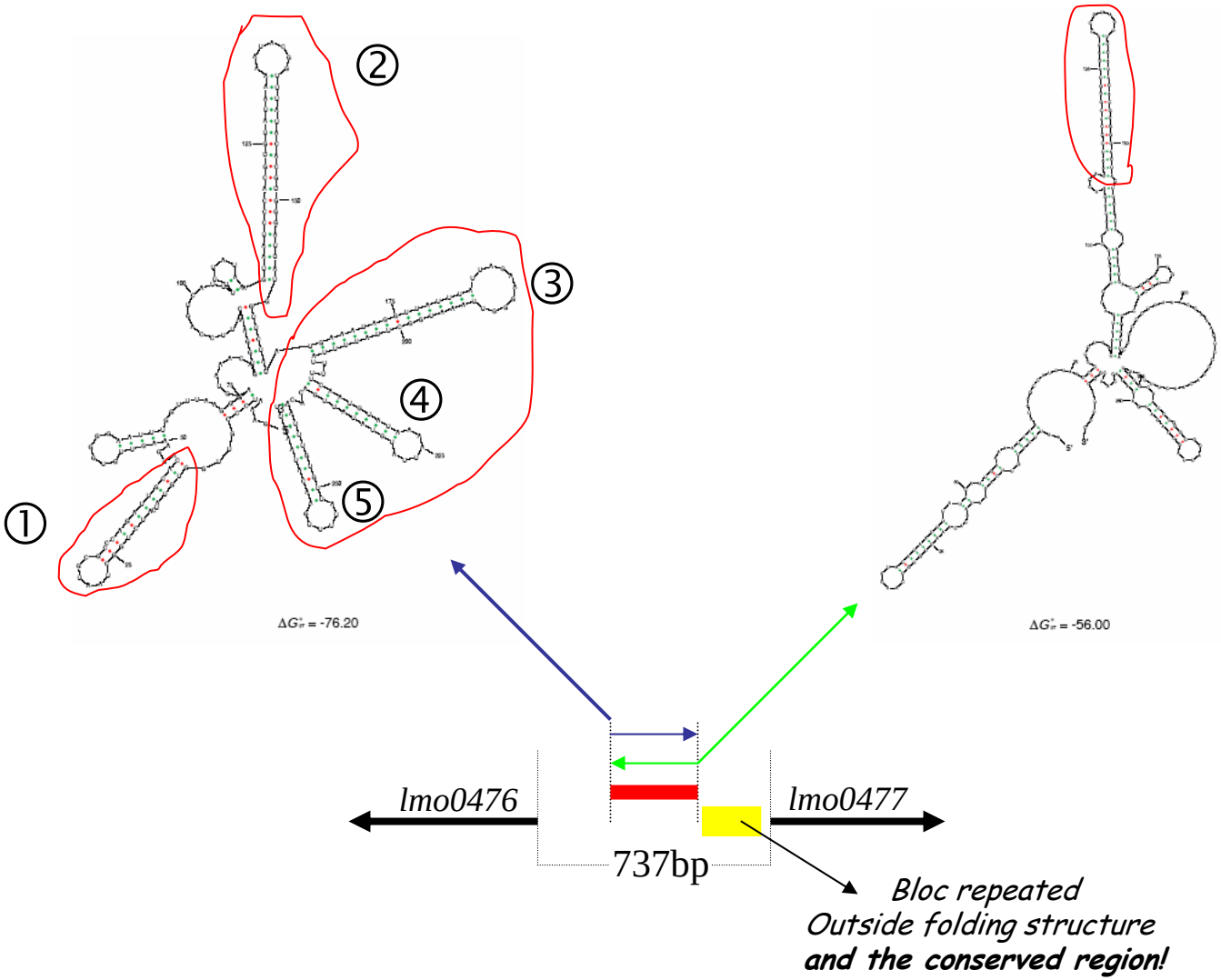
DsrA



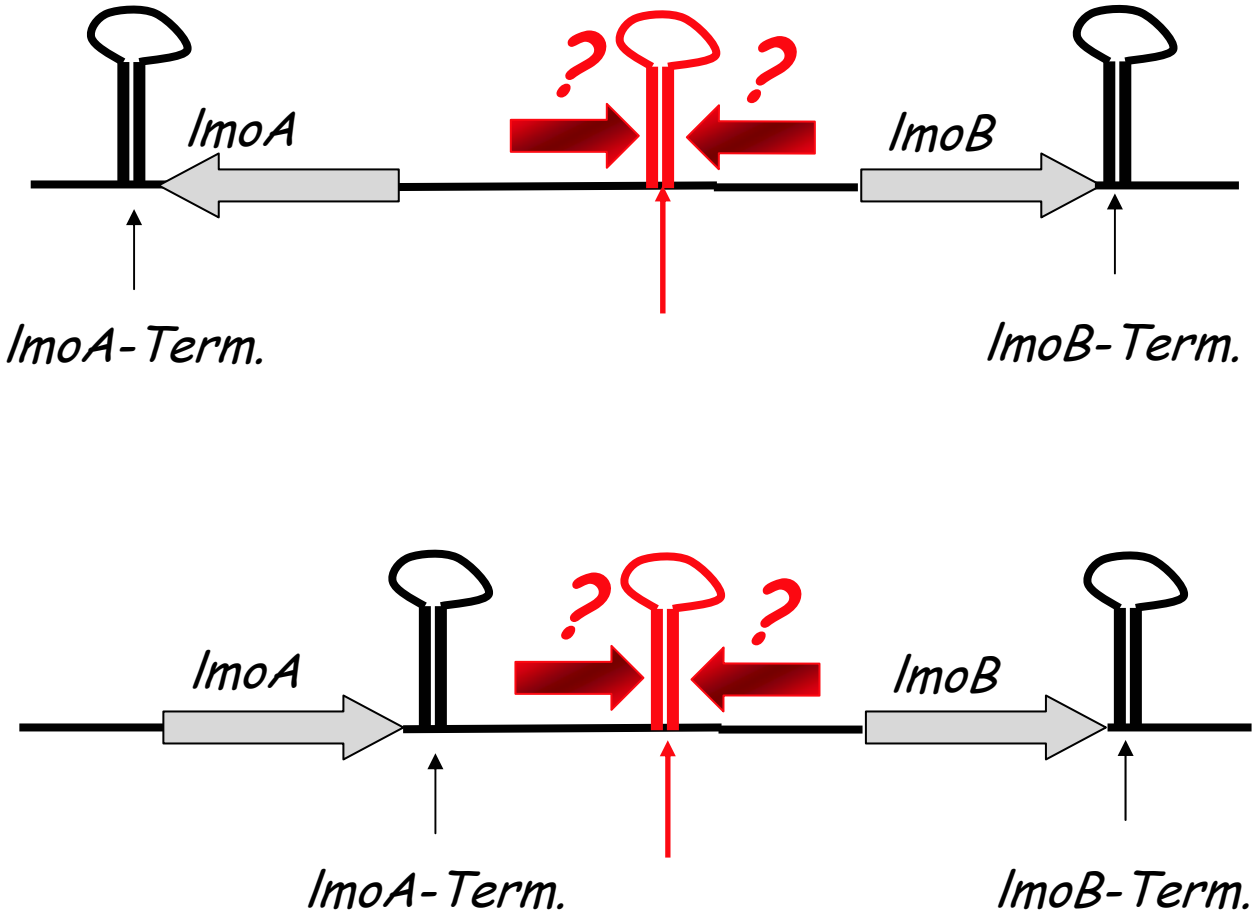
RprA



Folding analysis of lmo0476/0477 from 197 to 463

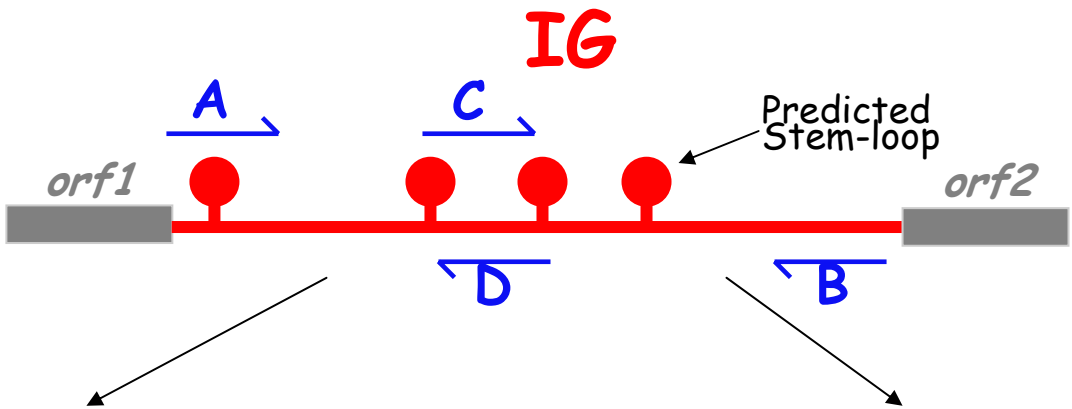


Identification by "orphan" terminators prediction



ρ-independent terminators are identified by standard pattern-searching methods.

Northern Blot analysis of IG candidates



PCR frgt. A/B
used as probe

↓

N.B. in low
stringency

↓

Detection of transcripts
within and from outside
the IG on both strands

Internal oligos C & D
used as probe

↓

N.B. in high
stringency

↓

Detection of transcripts
complementary to the oligo's
sequence

For comparative methods combining RNA structure prediction and alignment, see

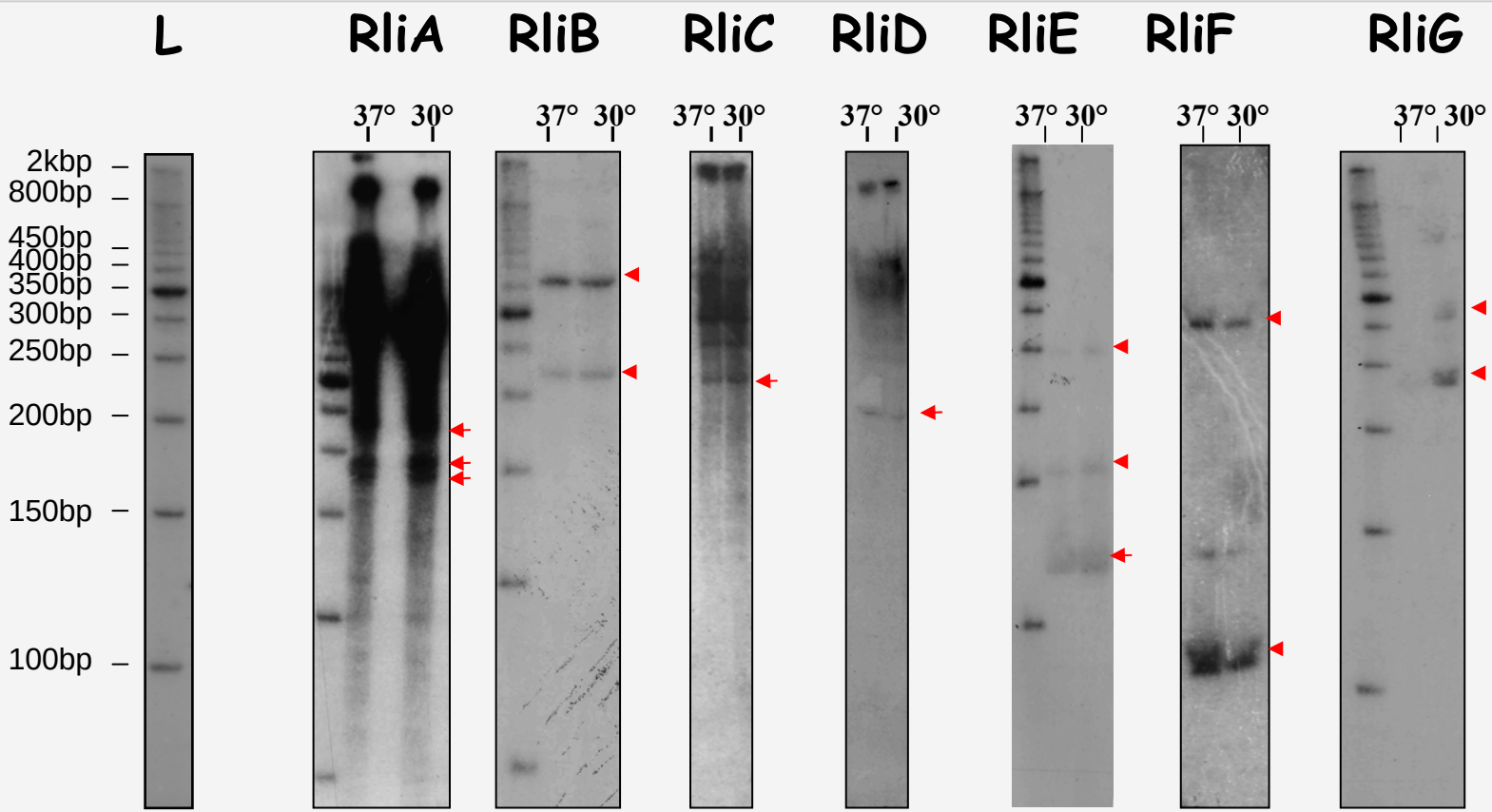
Uzilov AV, Keegan JM, Mathews DH *Detection of non-coding RNAs on the basis of predicted secondary structure formation free energy change* BMC Bioinformatics 2006, 7:173

Dowell R, Eddy SR *Efficient pairwise RNA structure prediction and alignment using sequence alignment constraints* BMC Bioinformatics 2006 7:400 and Rivas & Eddy, 2001

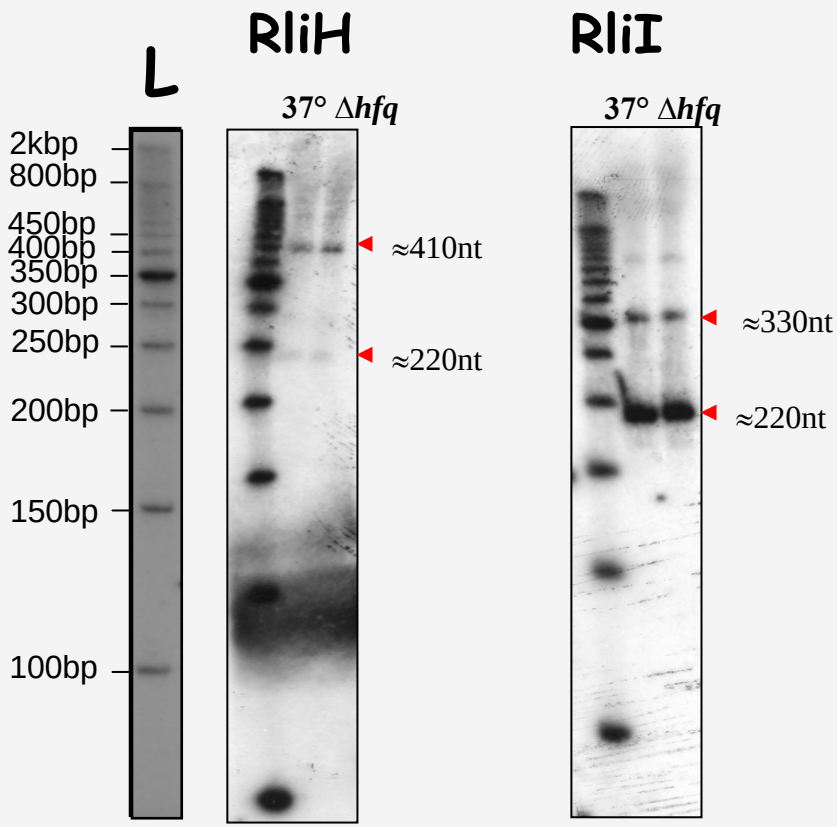
Washietl S, Hofacker IL, Stadler PF *Fast and reliable prediction of noncoding RNAs*. PNAS 2005, 102:2454-2459.

Benchmark comparisons for bacteria are made in *E. coli* and *Salmonella*, where many ncRNAs are already known.

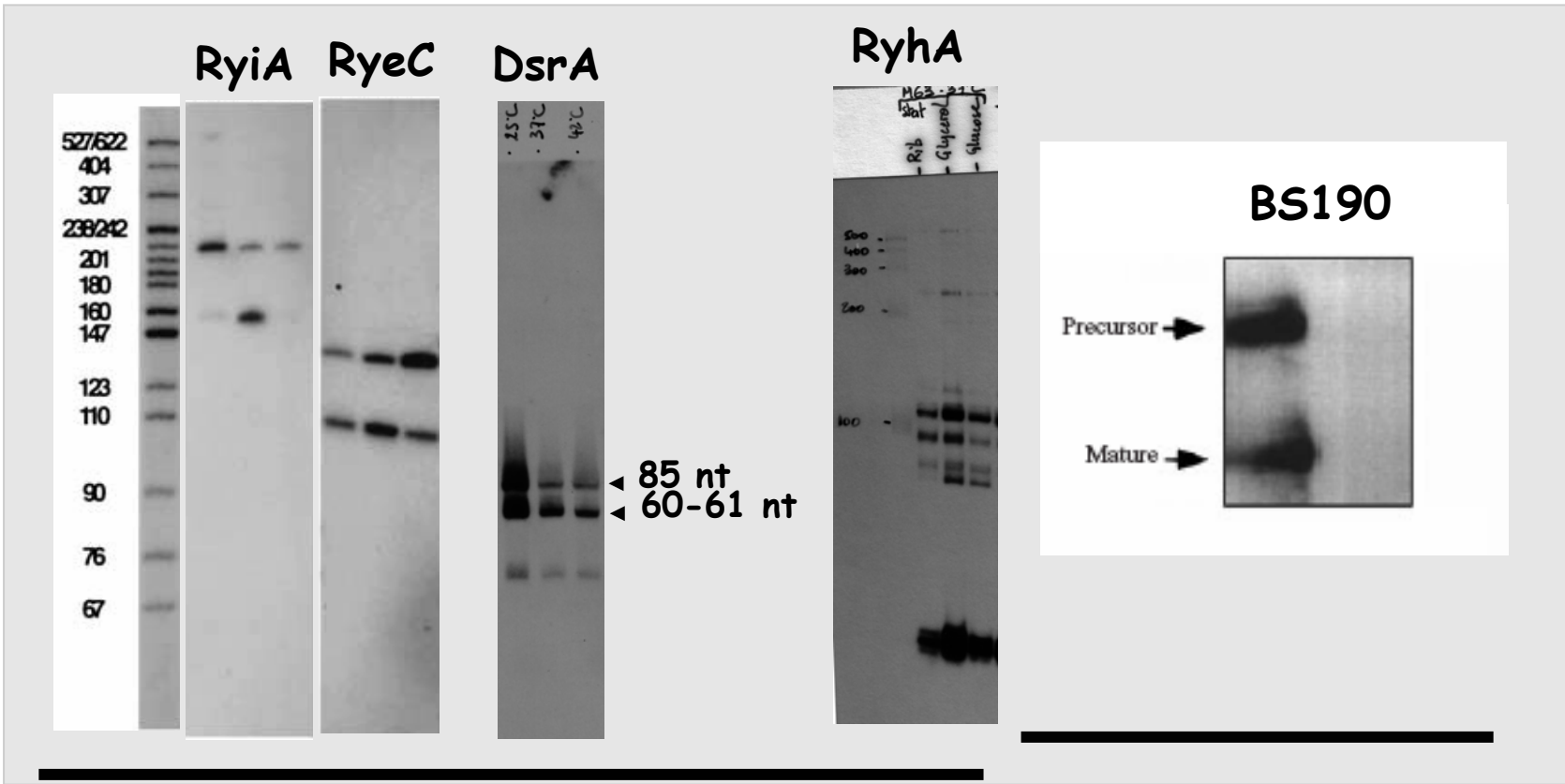
Northern Blot analysis for Listeria candidates



And two more...


L
RliH
RliI
37° Δhfq
37° Δhfq
2kbp
800bp
450bp
400bp
350bp
300bp
250bp
200bp
150bp
100bp
≈410nt
≈220nt
≈330nt
≈220nt

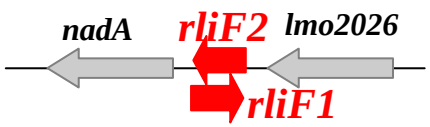
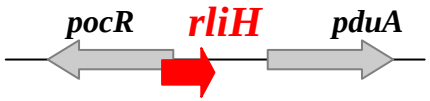
ncRNAs have frequently different forms



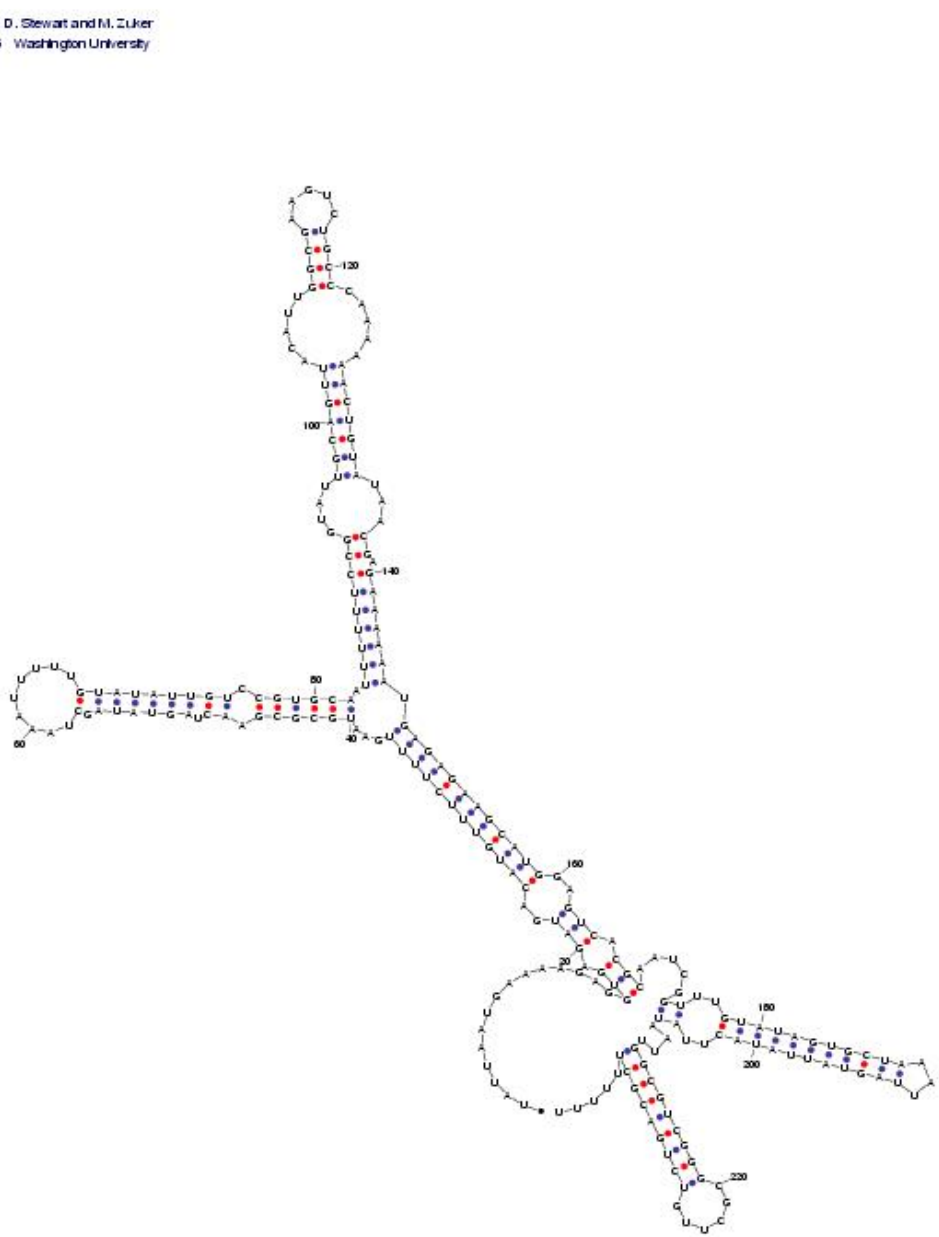
In E. coli K12
(Repoila & Gottesman, unp.)

In B. subtilis
(Suzuma et al. 2002)

New genes encoding ncRNAs (rli) in L. monocytogenes

ncRNA Locus	Presence in *L. innocua*	Function of flanking genes
	No	***lmo0476***: metal-dependent phosphohydrolase ***lmo0477***: putative secreted conserved in *l.monocyt.*
	No	***prs*** (like): Similar to phosphoribosyl pyrophosphate synthetase ***lmo0510*** : putative lipoprotein
	No	***lmo1117***: glyoxalase family protein ***lmo1118***: No function known.
	Yes	***rpsO***: ribosomal protein S15 ***pnpA***: polynucleotide phosphorylase
	Yes	***comC***: late competence protein ComC (B. subtilis) ***folC***: Folyl-polyglutamate synthetase
	No	***nadA***: quinolonate synthase ***lmo2026***: putative peptidoglycan bound protein (LPXTG motif)
	No	***lmo2302***: Unknown; absent of *L.inn.* ***lmo2303***: Gp66 of Bacteriophage A118; present in *L.inn.*
	Yes	***pocR:*** AraC family of regulators. Regulates *cob* and *pdu* ***pduA*:** Catabolism of 1,2-propanediol
	Yes	***lmo2760:*** ABC transporter ***lmo2761:*** beta-glucosidase

Folding of RIil

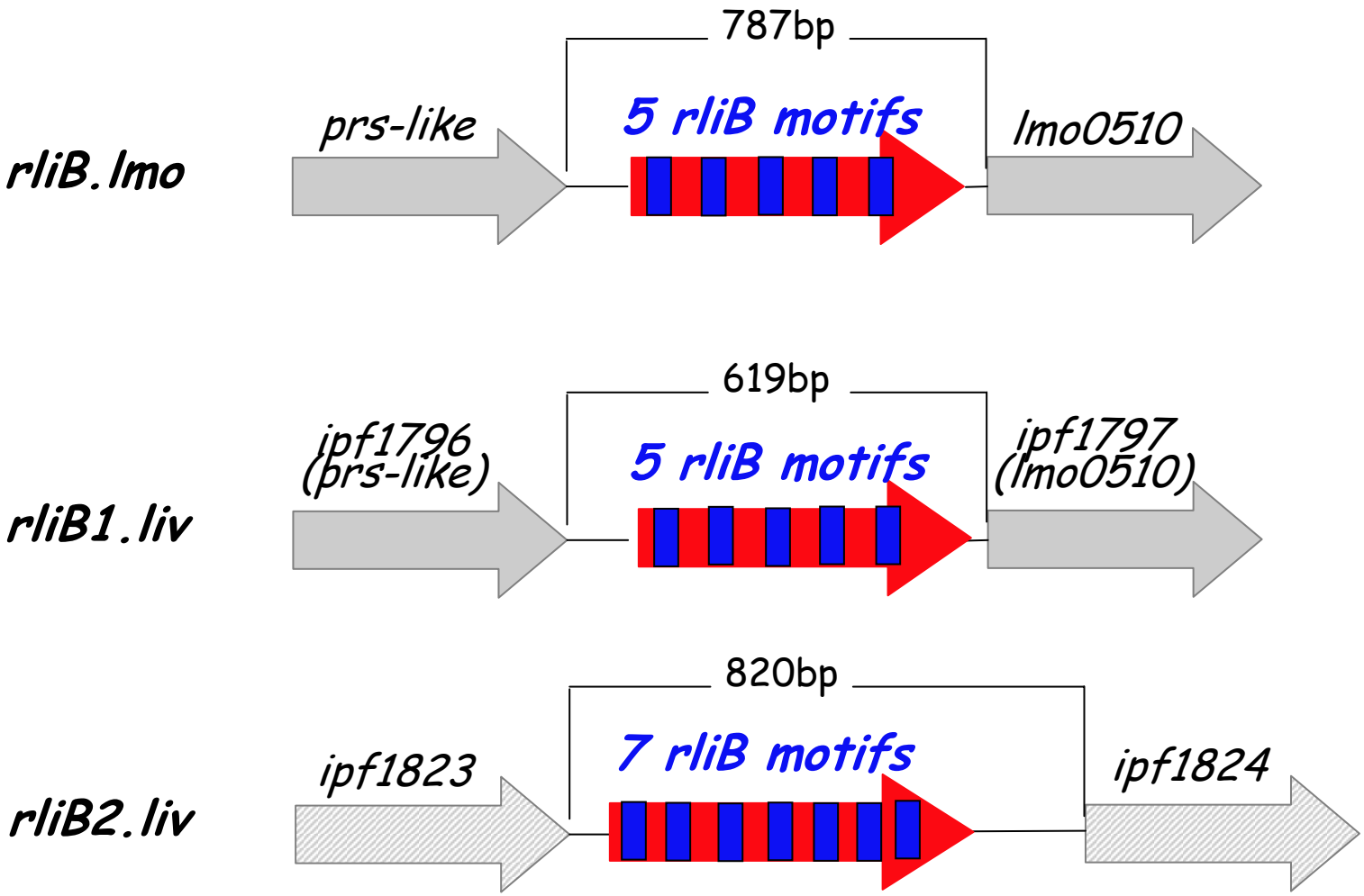


Sequence of rliB locus

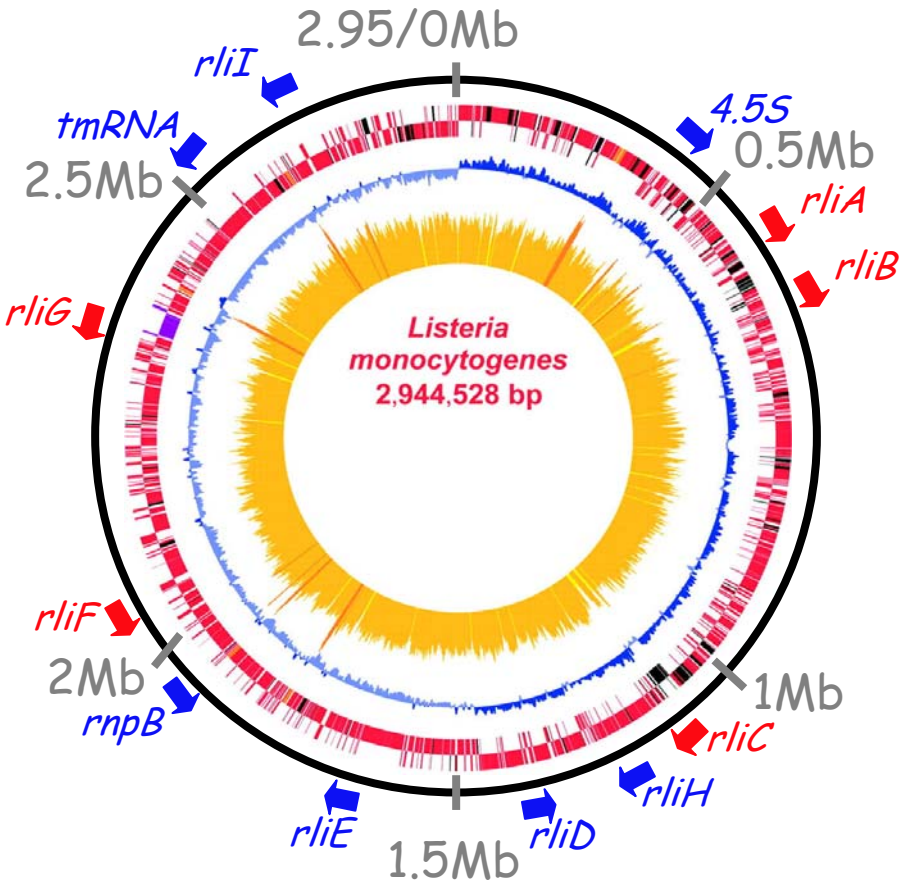
Mandin, P. et al. Nucl. Acids Res. 2007 35:962-974; doi:10.1093/nar/gkl1096

Nucleic Acids Research

RliB of *L.monocytogenes* is conserved and duplicated in *L ivanovii*

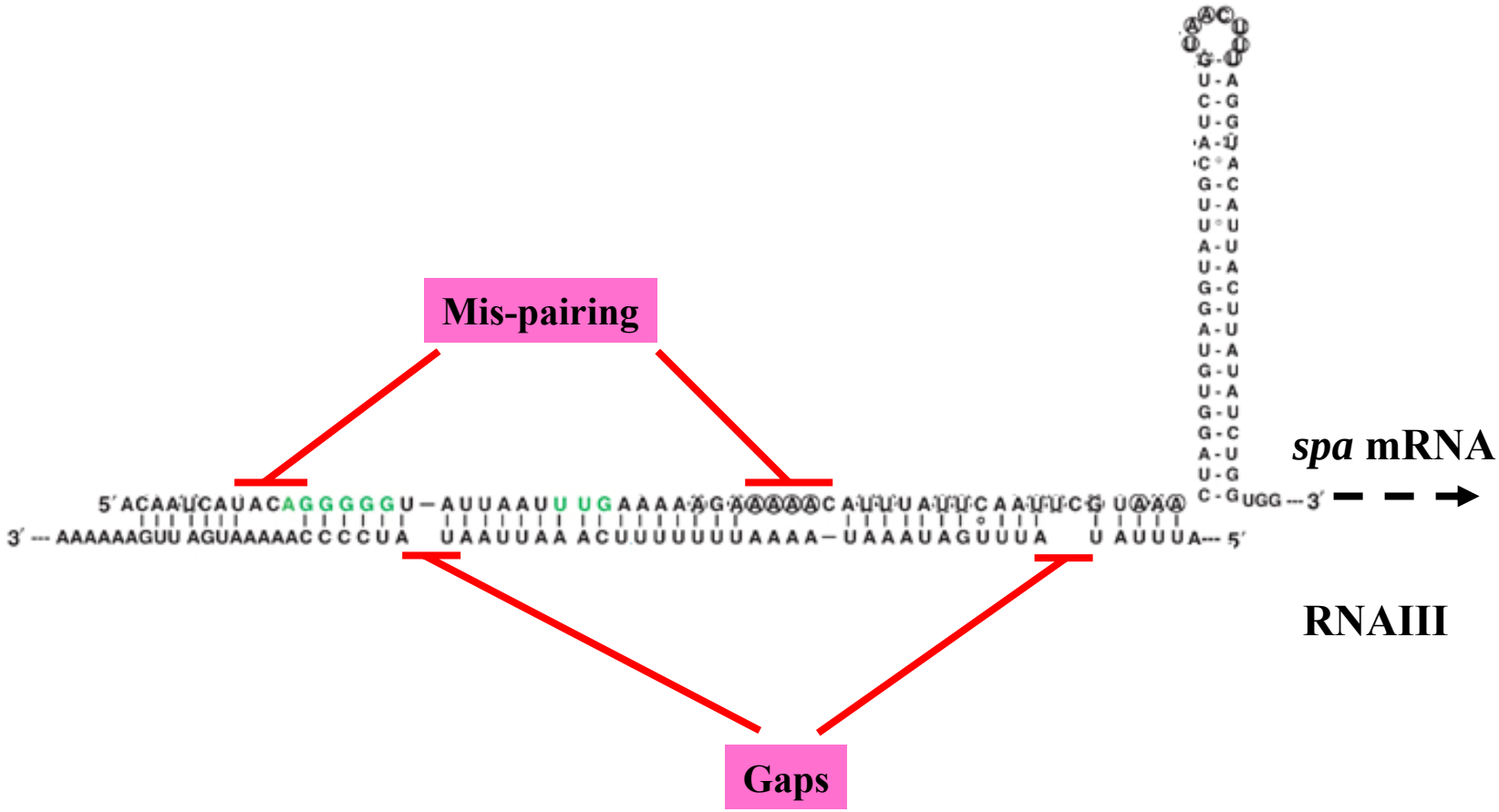


9 new genes encoding ncRNAs (rli) in L. monocytogenes EGD.e



In red *rli* gene not present in *L. innocua;*
In blue present in EGD.e and *L. innocua*

Basic problem with mRNA target prediction



Bacterial ncRNA/mRNA recognition does not occur *via* perfect pairing

Target prediction methods

(i) B. Tjaden et al. Nucleic Acids Research 2006 3:2791-2802
(ii) P Mandin et al. Nucleic Acids Research 2007 35:962-974

A Smith-Waterman alignment of the two sequences (ncRNA and putative target mRNA).

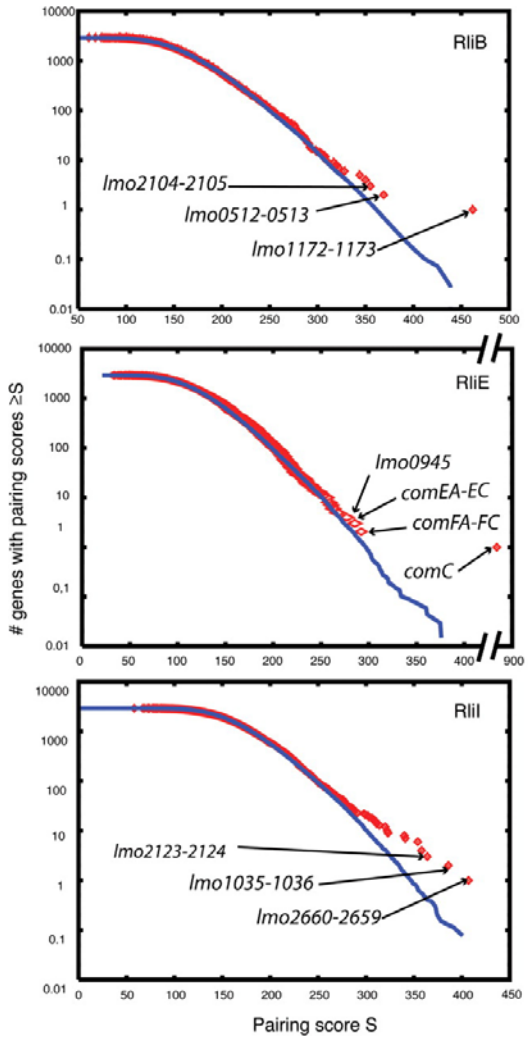
In the simplest version of (i) flat scores:

(A, C, G, U) x (A, C, G, U) = [(6, 6, 6, –5), (6, 6, –5, 6), (6, –5, 6, 1), (–5, 6, 1, 6)]. Mismatch opening penalty 12 and extension 3.

Problems with AT rich bacteria such as *Listeria* and no energetic weight of the pairing.

Stacking effects then included as in (ii), where bulge opening and extension, internal loop opening and extension are learnt from known examples of targeted mRNAs.

Prediction of mRNA targets for RliE, RliB and RliI



Genes	Function
lmo1172-1173	Two-component system
lmo0512-0513	Unknown function; transcriptional regulator
*lmo2104-2105**	Iron transport proteins A and B

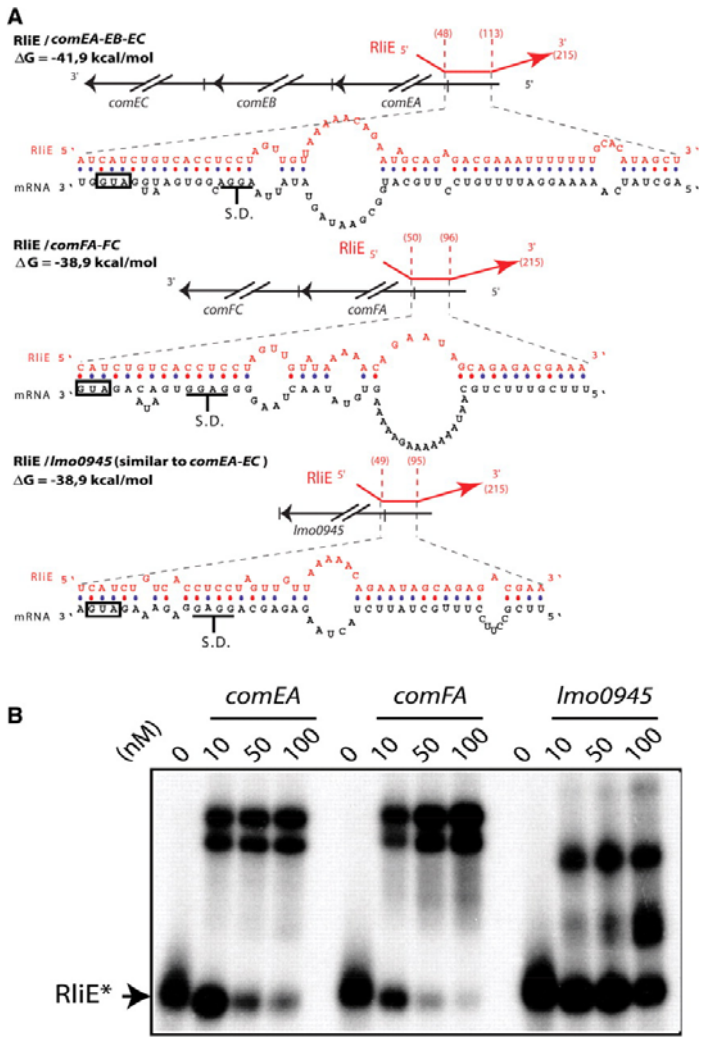
Genes	Function
comC	Late competence protein ComC
comFA-FC	Late competence protein required for DNA uptake
comEA-EC	Exogenous DNA binding protein
lmo0945	ComEC/ComEA like

Genes	Function
lmo2660-2659	Transketolase; Ribulose phosphate 3 epimerase
lmo1035-1036	Beta-glucoside enzyme II ABC Unknown function
lmo2123-2124	Maltodextrin ABC transporter

Mandin, P. et al. Nucl. Acids Res. 2007 35:962-974; doi:10.1093/nar/gkl1096

Nucleic Acids Research

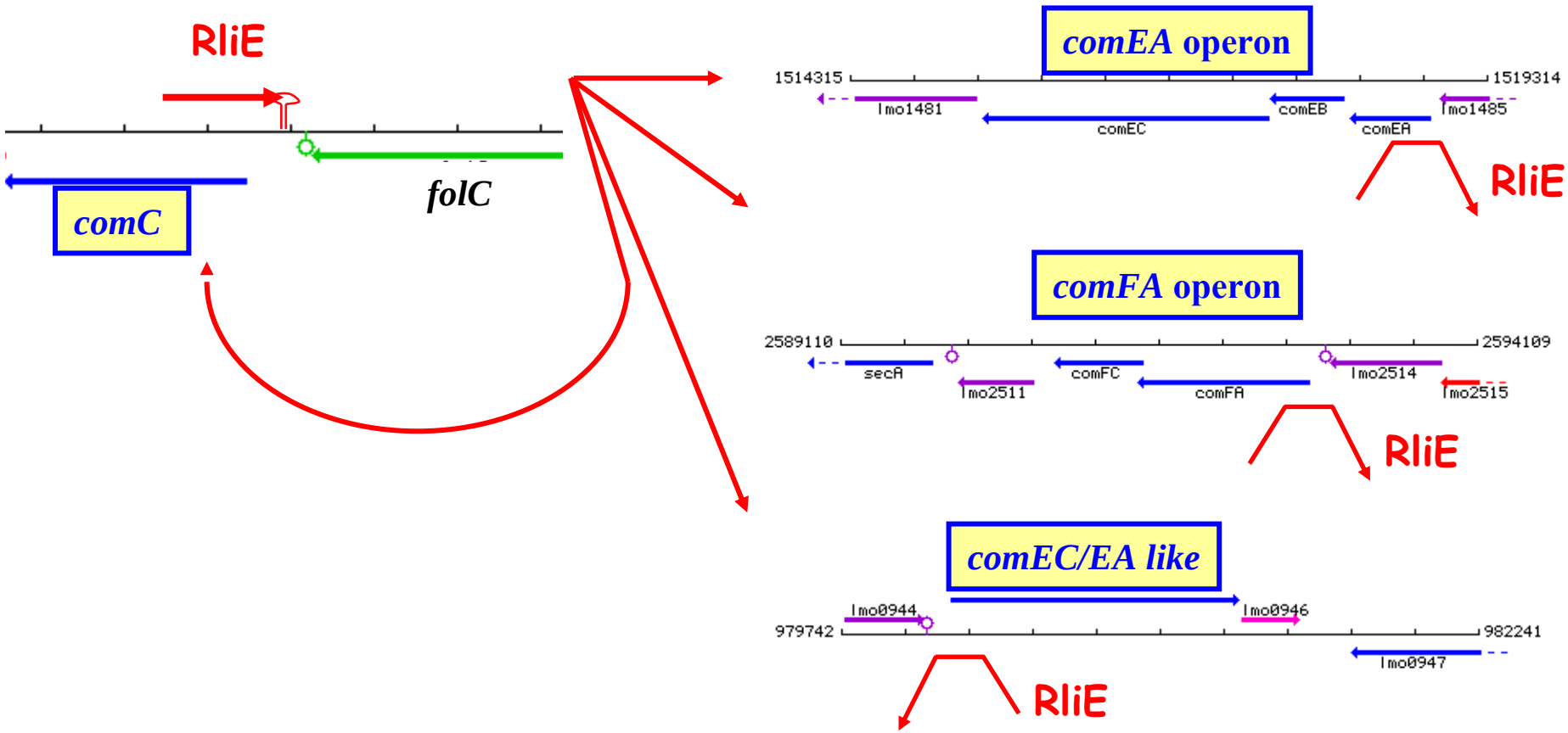
Hybrids between RliE and mRNA targets



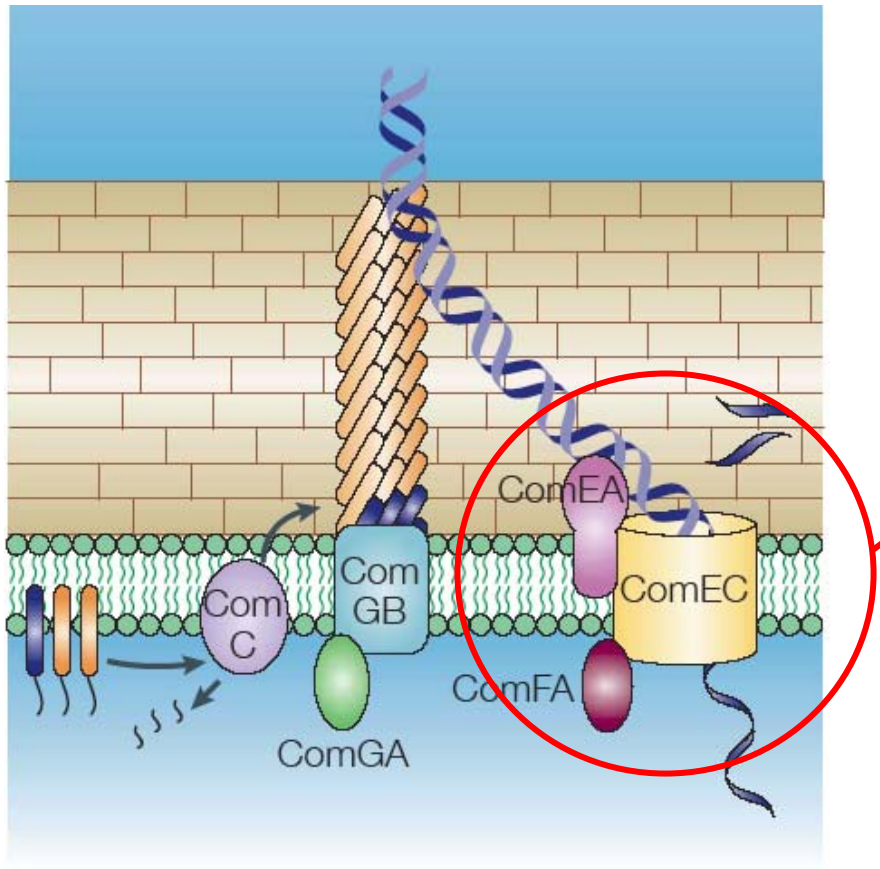
Mandin, P. et al. Nucl. Acids Res. 2007 35:962-974; doi:10.1093/nar/gkl1096

Nucleic Acids Research

Majority of RliE predicted targets are competence genes



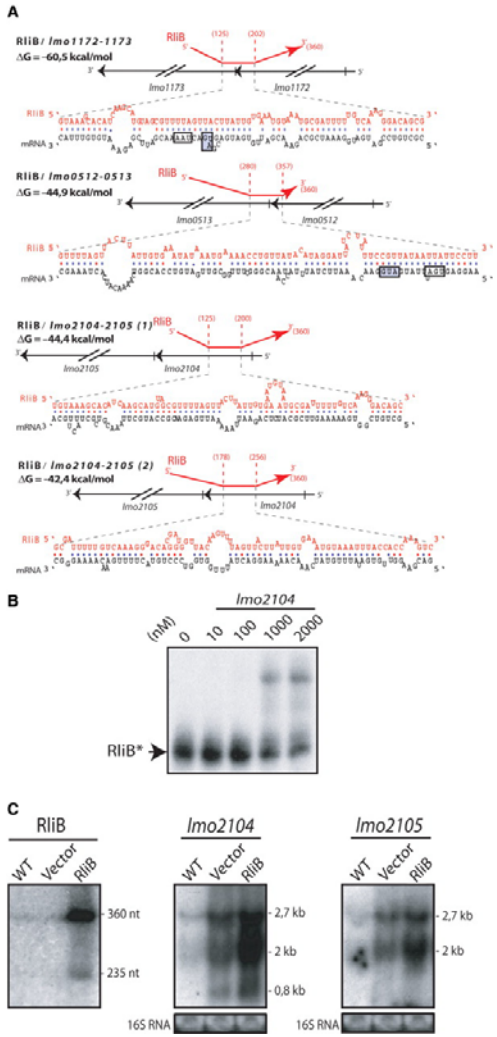
RliE predicted targets are part of the same complex



RliE predicted targets

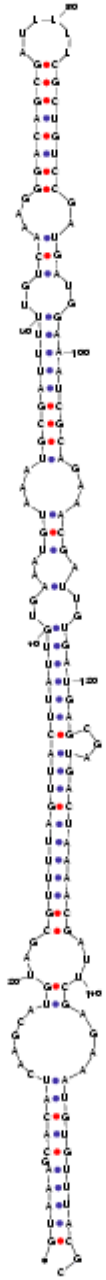
RliE might regulate competence of *Listeria (*but nobody knows as yet how to make it competent...) or the importing complex might have been recruited to import something else....

Hybrids between RliB and mRNA targets

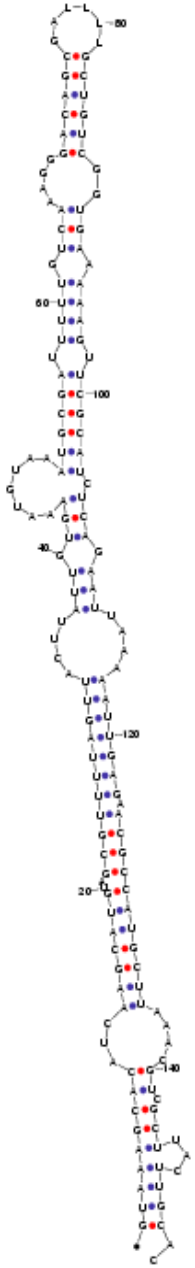


Mandin, P. et al. Nucl. Acids Res. 2007 35:962-974; doi:10.1093/nar/gkl109

Nucleic Acids Research

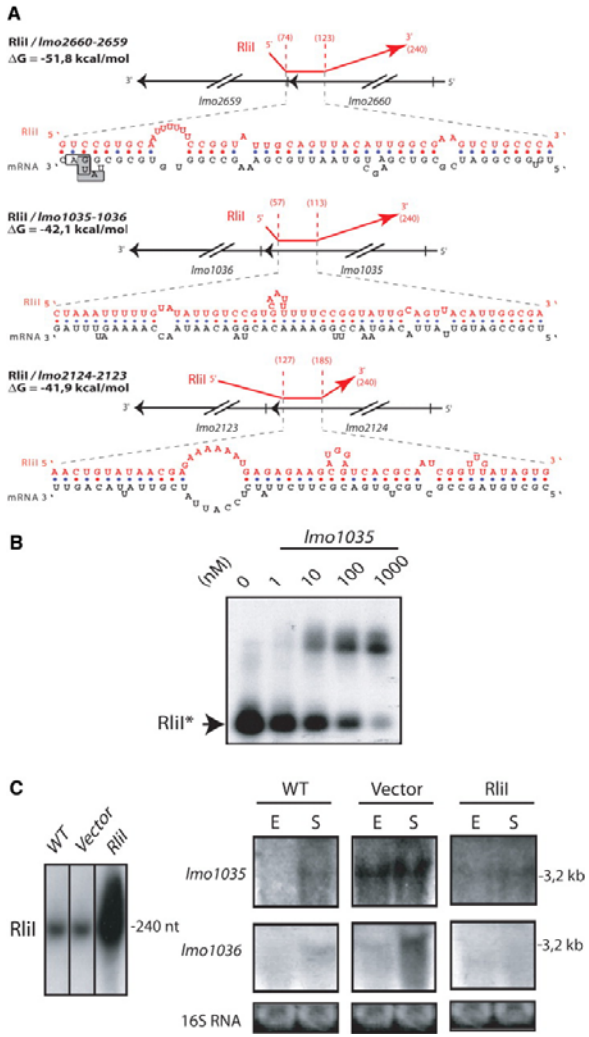


dG = -58.1 [initially -60.5] lmo1173



dG = -41 [initially -44.0] lmo2104

Hybrids between RliI and mRNA targets



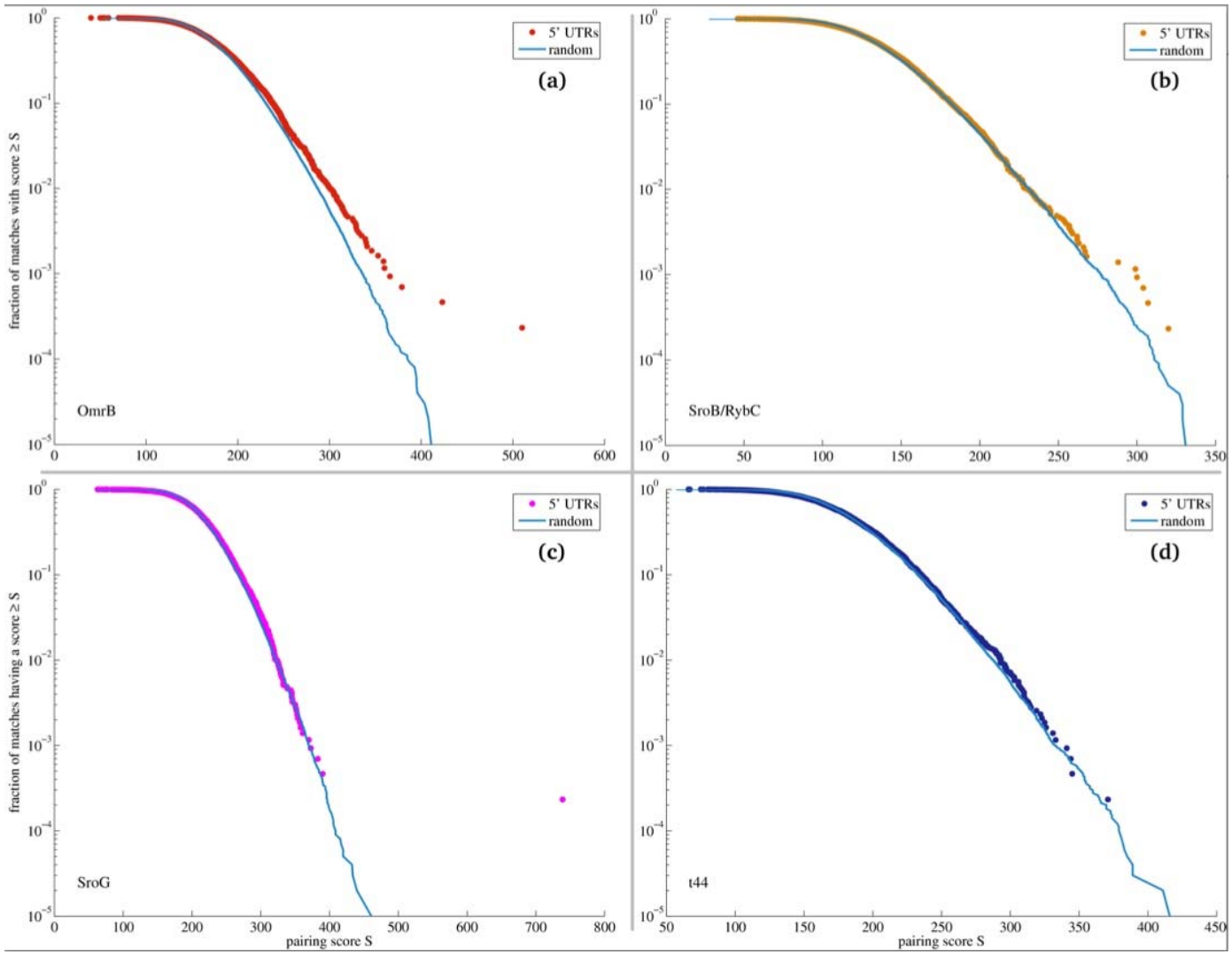
Mandin, P. et al. Nucl. Acids Res. 2007 35:962-974; doi:10.1093/nar/gkl1096

Nucleic Acids Research

E. coli prediction sketch

23

19



16

22

Acknowledgements

P. Cossart (Inst. Pasteur)

A. Danchin (Inst. Pasteur)

A. Fouquier d'Herouel (Inst. Pasteur)

T. Geissmann (IBMC, Strasbourg)

J-D Huang (Univ Hong-Kong)

P. Mandin (Inst. Pasteur, now at NIH)

F. Repoila (INRA, Paris)