Predicting transcription terminator efficiency

Ronny Lorenz
ronny@tbi.univie.ac.at

Bioinf
University of Leipzig

Bled, Slovenia, April 12, 2014

De novo design of theophylline sensing riboswitches

Published online 28 December 2012 *Nucleic Acids Research, 2013, Vol. 41, No. 4* **2541–2551**
doi:10.1093/nar/gks1330

De novo design of a synthetic riboswitch that regulates transcription termination

Manja Wachsmuth[1], Sven Findeiß[2,3], Nadine Weissheimer[1], Peter F. Stadler[3,4,5,6,7,8] and Mario Mörl[1,*]

[1]Institute for Biochemistry, University of Leipzig, Brüderstr. 34, 04103 Leipzig, Germany, [2]Research group Bioinformatics and Computational Biology, Faculty of Computer Science, University of Vienna, Währingerstraße 29, A-1090 Vienna, Austria, [3]Institute for Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Vienna, Austria, [4]Bioinformatics Group, Department of Computer Science, University of Leipzig, Härtelstr. 16-18, 04107 Leipzig, Germany, [5]Interdisciplinary Center for Bioinformatics, Härtelstr. 16-18, 04107 Leipzig, Germany, [6]Max Planck Institute for Mathematics in the Science, Inselstr. 22, 04103 Leipzig, Germany, [7]Fraunhofer Institute for Cell Therapy and Immunology, Perlickstr. 1, 04103 Leipzig, Germany and [8]Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe NM 87501, USA

Received October 1, 2012; Revised November 26, 2012; Accepted November 27, 2012

ABSTRACT

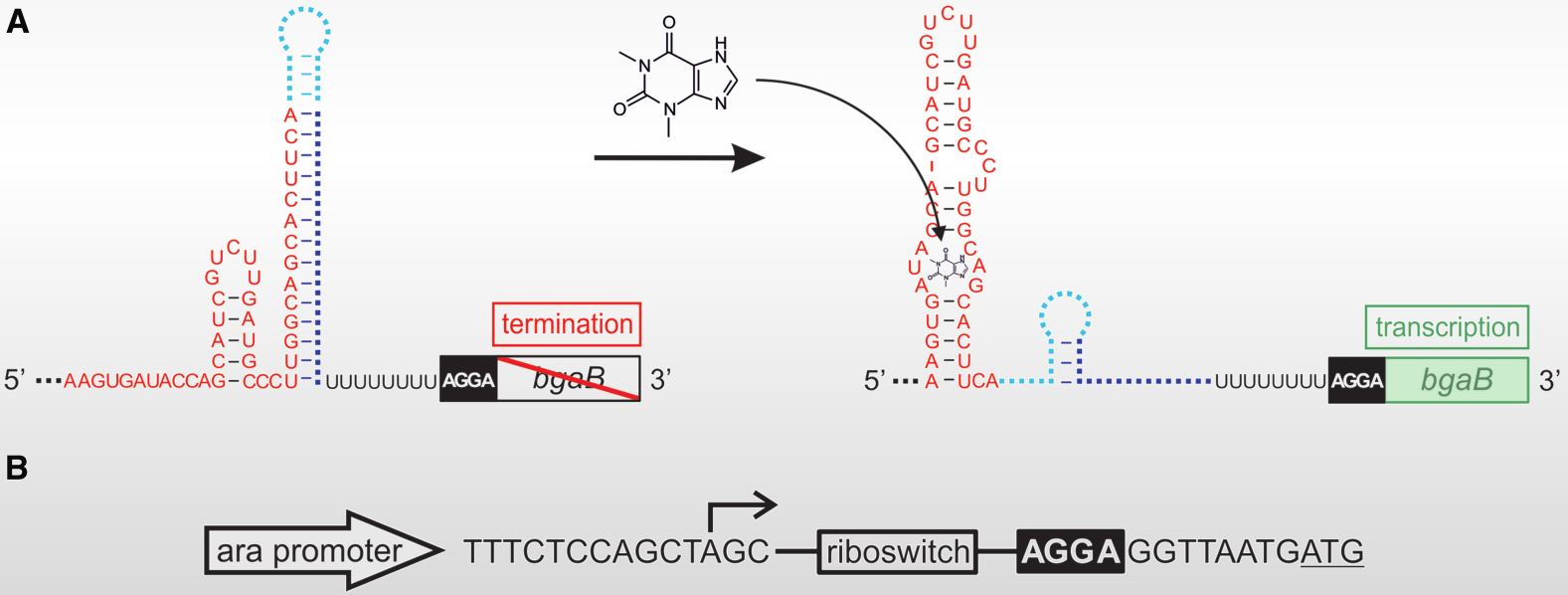
Riboswitches are regulatory RNA elements typically located in the 5′-untranslated region of certain mRNAs and control gene expression at the level of transcription or translation. These elements consist of a sensor and an adjacent actuator domain. The sensor usually is an aptamer that specifically interacts with a ligand. The actuator contains an intrinsic terminator or a ribosomal binding site for transcriptional or translational regulation, respectively. Ligand binding leads to structural rearrangements of the riboswitch and to presentation or masking of these regulatory elements. Based on this modular organization, riboswitches are an ideal target for constructing synthetic regulatory systems for gene expression. Although riboswitches for translational control have been designed successfully, attempts to construct synthetic elements regulating transcrip- [...] **is possible to engineer riboswitches not only for translational but also for transcriptional regulation.**

INTRODUCTION

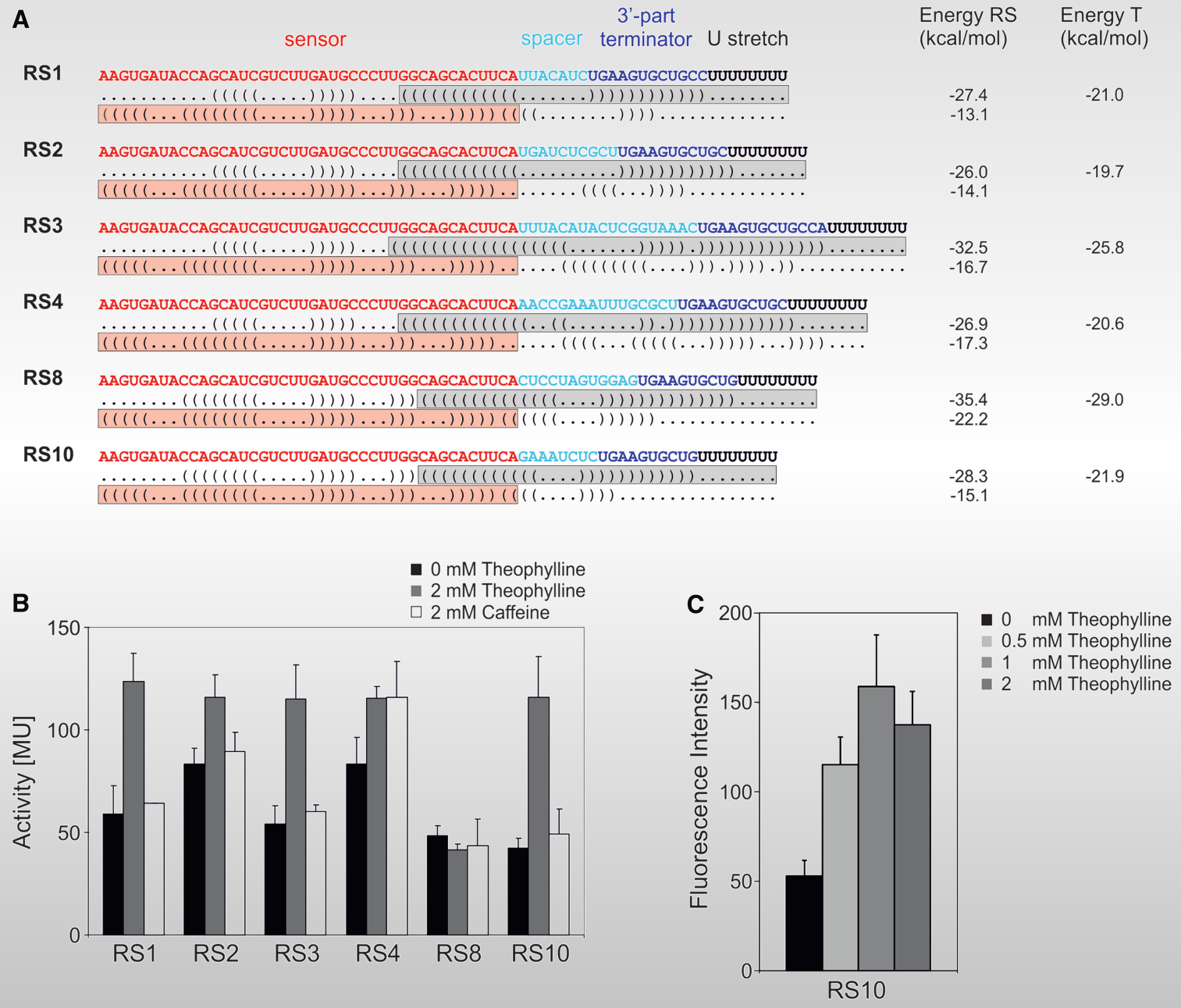
Synthetic biology has become a rapidly developing branch of life sciences. At its heart is the rational design and construction of biosynthetic and regulatory networks, pathways and their constituent components. As RNA plays a central role in a diverse set of regulatory pathways, efforts in synthetic biology have produced novel RNA components capable of regulating gene expression *in vivo* (1). Riboswitches are a particularly attractive type of gadget for synthetic biology applications (2,3). These naturally occurring RNA elements determine the expression state of a gene in response to an external signal (in most cases, a small molecule that is specifically recognized) through a conformational change, regulating either transcription or translation. The fundamental architecture of a riboswitch combines a sensor (aptamer

De novo design of theophylline sensing riboswitches

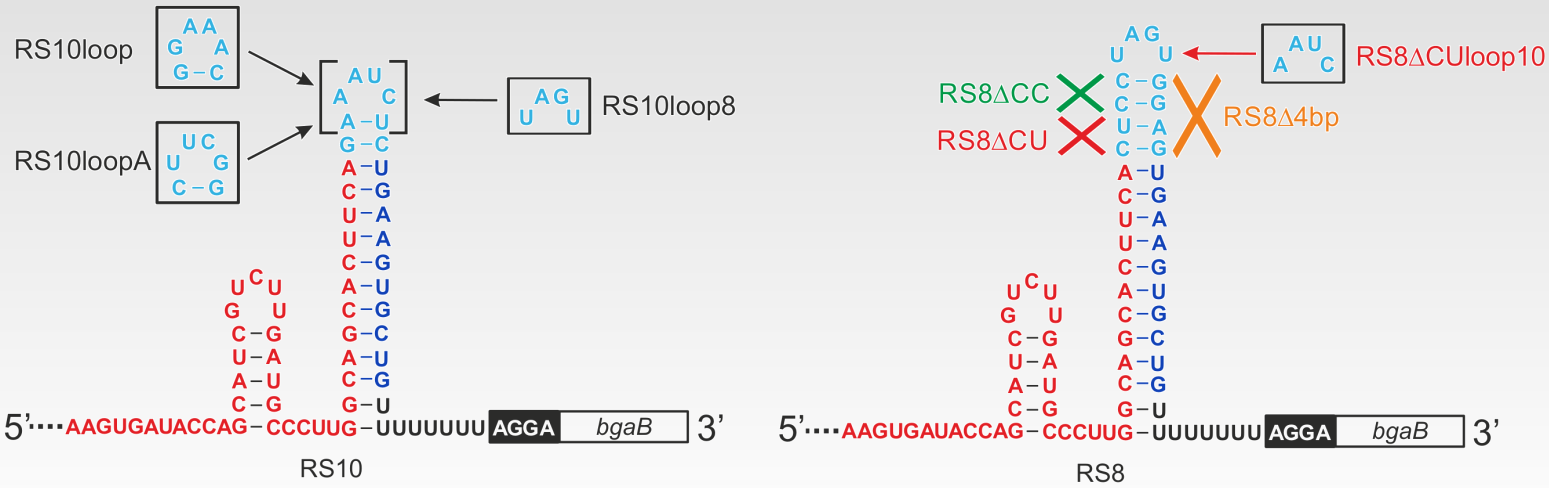
- in silico design, in vivo validation (E. coli)
- theophylline aptamer upstream of a terminator hairpin
- aptamer fold overlaps with terminator
- ON switch upon presence of theophylline
- iterative design with `RNAinverse` and `RNAfold`

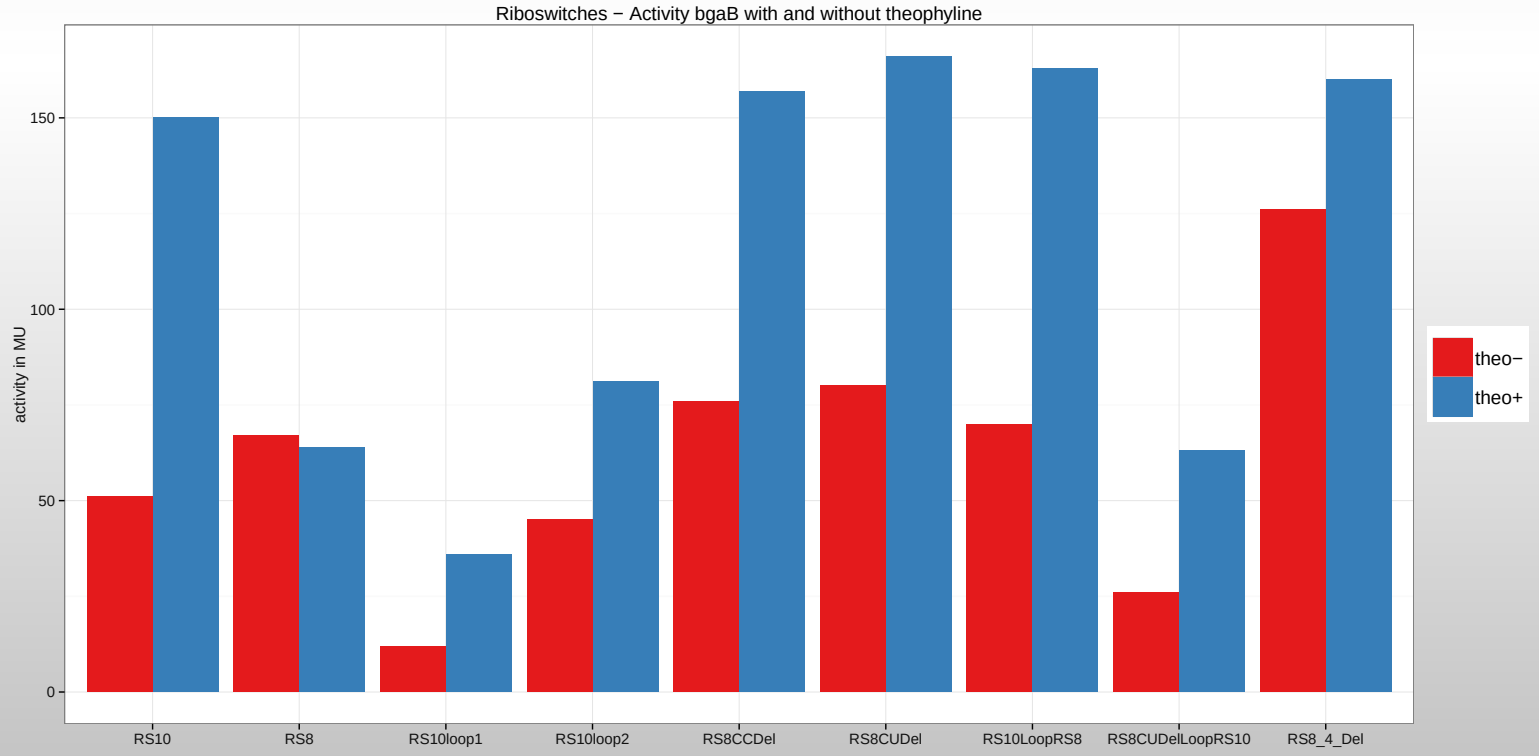


De novo design of theophylline sensing riboswitches

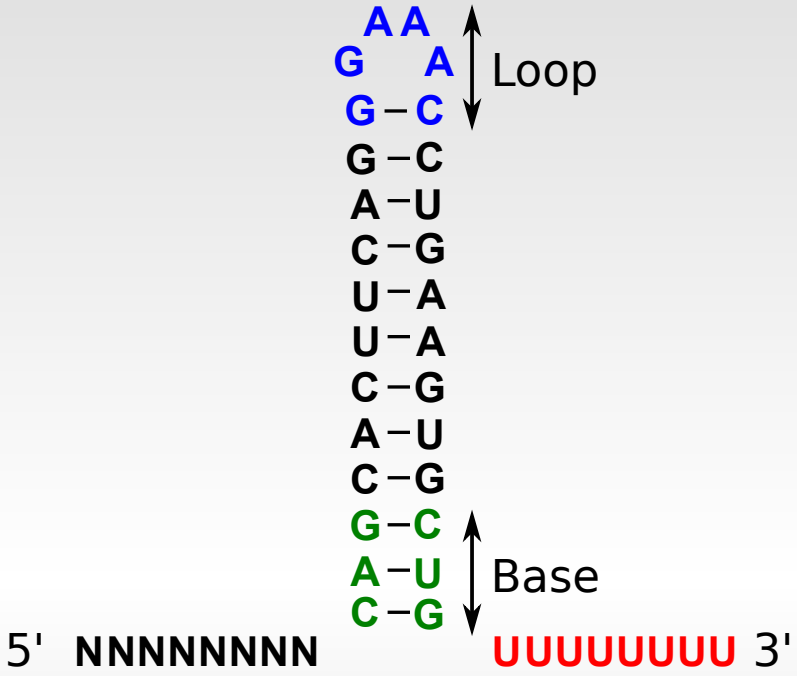


Mutational study





Bioinformatics analysis of the results

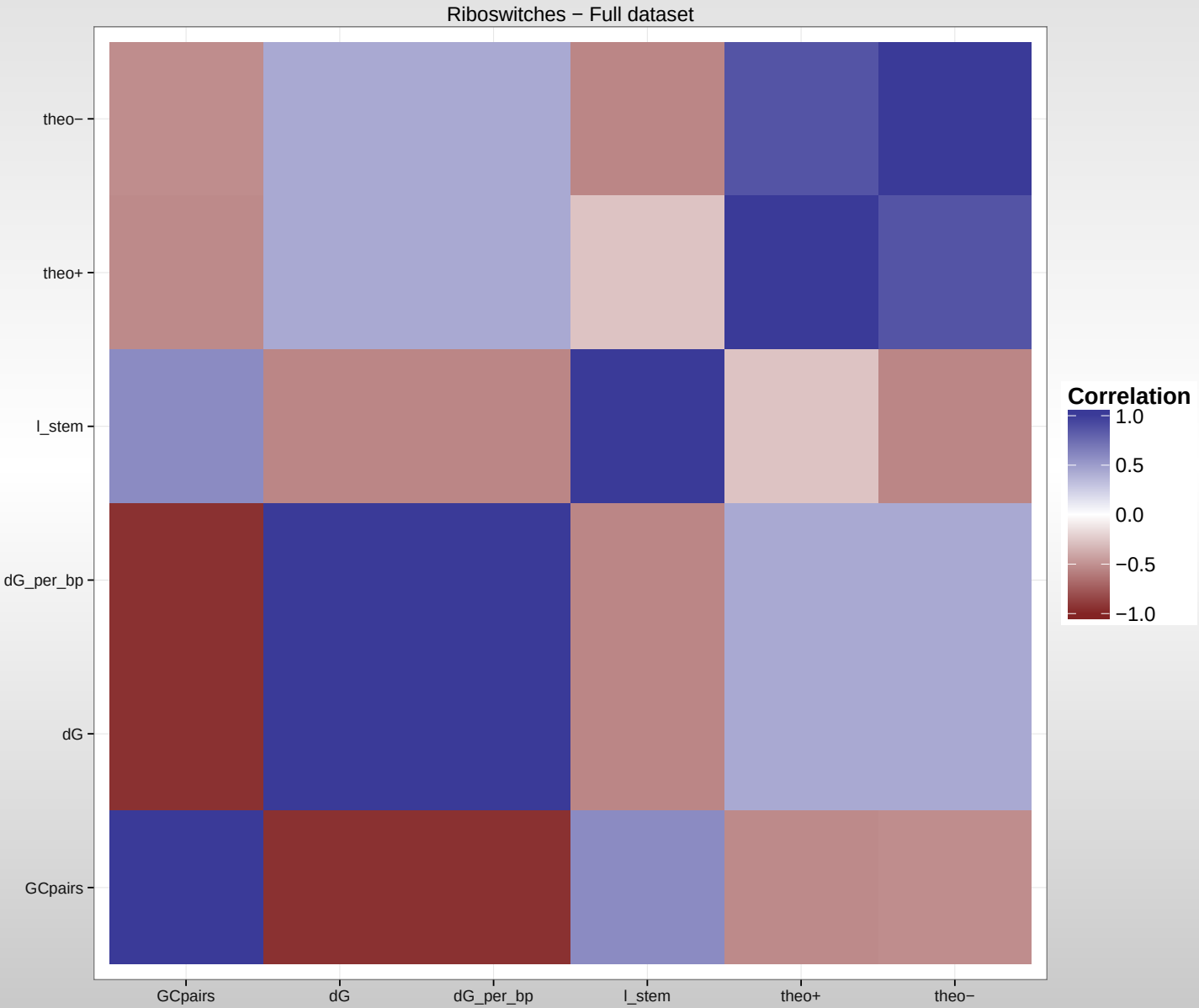


Obvious sequence / structure prediction measures:

- extraction and evaluation of 128 E.coli 3′ terminators[1]
- prediction of MFE and according structure of RS transcripts
- total and mean free energy per bp of the terminators
- GC content of terminator hairpin
- . . .

[1] Carafa et al. 1990

Correlation of obvious parameters



Terminator efficiency scores - Carafa et al. 1990[2]

Terminator descriptors:

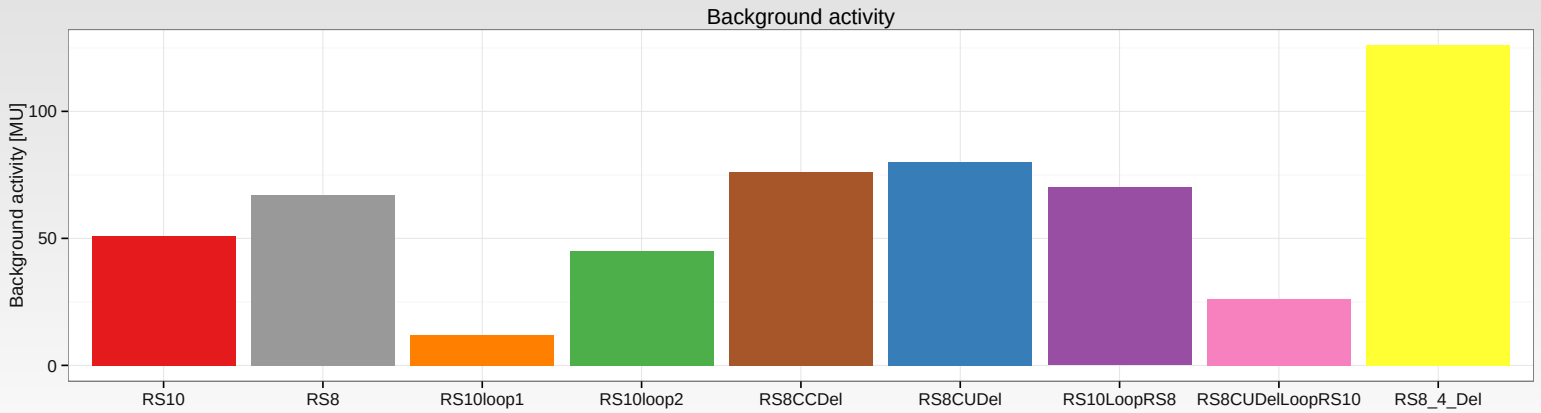
- n_{GC} . . . Number of GC pairs in the stem
- n_L . . . Number of nucelotides in the loop
- n_b . . . Distance of terminators 3′ end to the poly-U stretch
- L_H . . . Distance of terminators 5′ end to the poly-U stretch

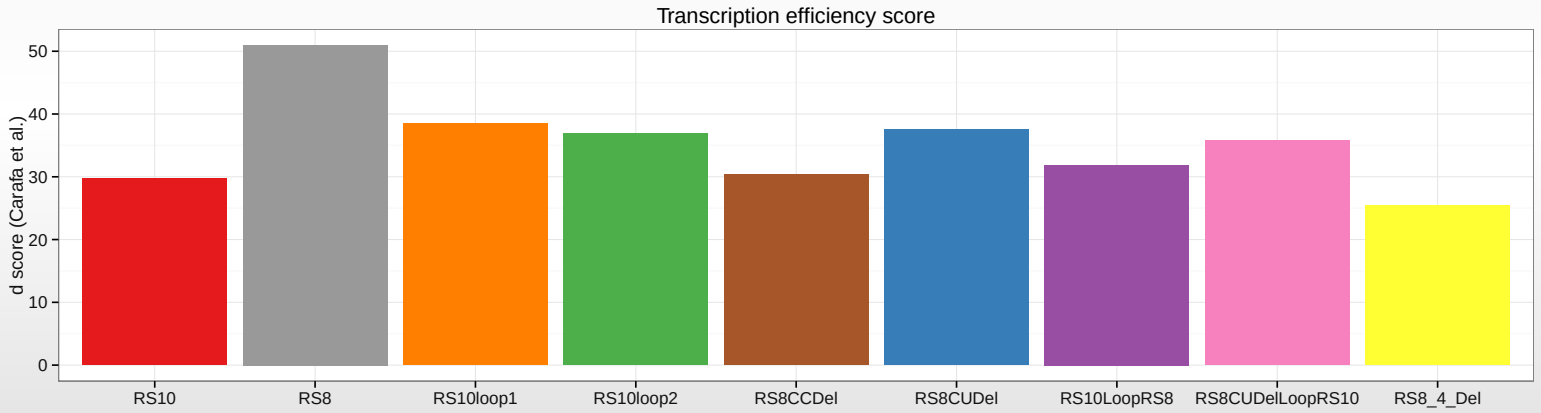
Their scoring:

$$
\begin{aligned}
d &= n_T \cdot 18.16 + Y \cdot 96.59 - 116.87 \\
n_T &= \sum_{1 \le i \le 15} x_i \cdot \delta_i \\
x_i &= x_{i-1} \cdot \begin{cases} 0.9 & \text{if } s[i] ==' T' \\ 0.6 & \text{otherwise.} \end{cases} \\
\delta_i &= \begin{cases} 1 & \text{if } s[i] ==' T' \\ 0 & \text{otherwise.} \end{cases} \\
Y &= \frac{-\Delta G}{L_H}
\end{aligned}
$$

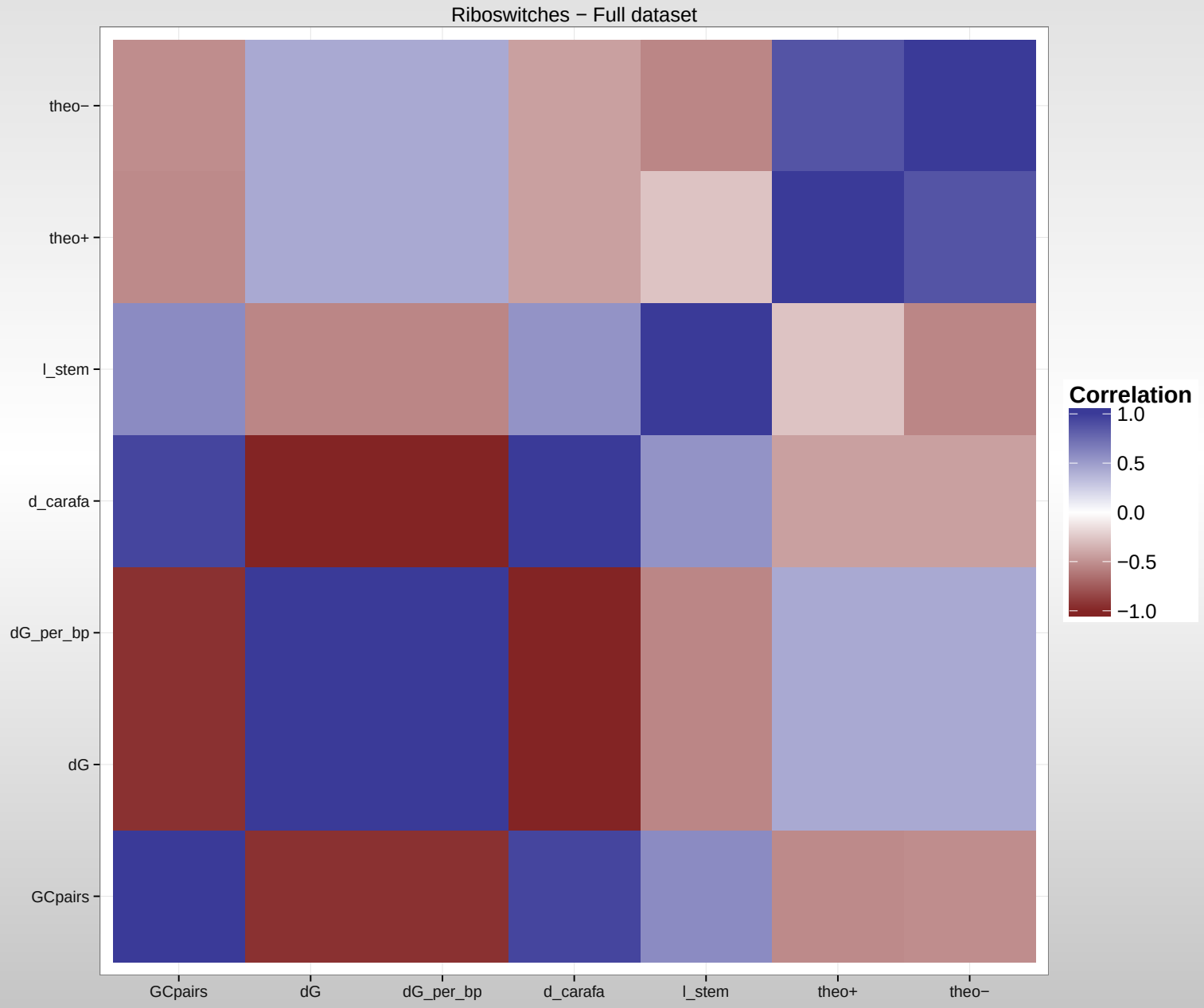
[2] "Prediction of Rho-independent Escherichia coli Transcription Terminators", J. Mol. Biol. (1990) 216, 835-858

Terminator efficiency scores - Carafa et al. 1990





Terminator efficiency scores - Carafa et al. 1990



Terminator efficiency scores - de Hoon et al. 2005[3]

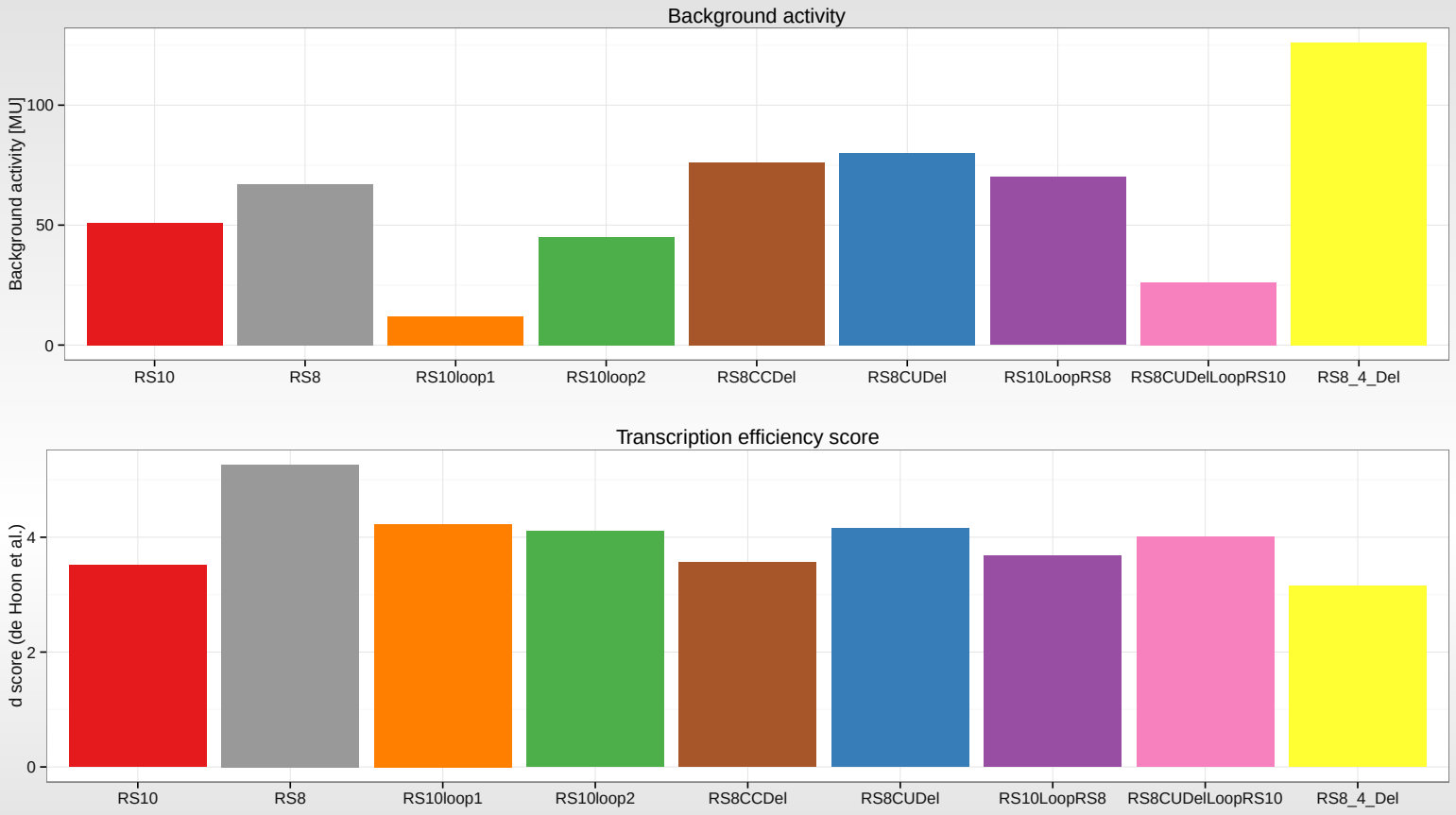
Same approach as Carafa et al.:

$$
\begin{aligned}
d &= n_T \cdot 2.67 + Y \cdot 7.9 - 14.91 \\
n_T &= \sum_{0 \le i \le 14} \exp(-\lambda i) \cdot \delta_i \\
\lambda_i &= 0.144
\end{aligned}
$$

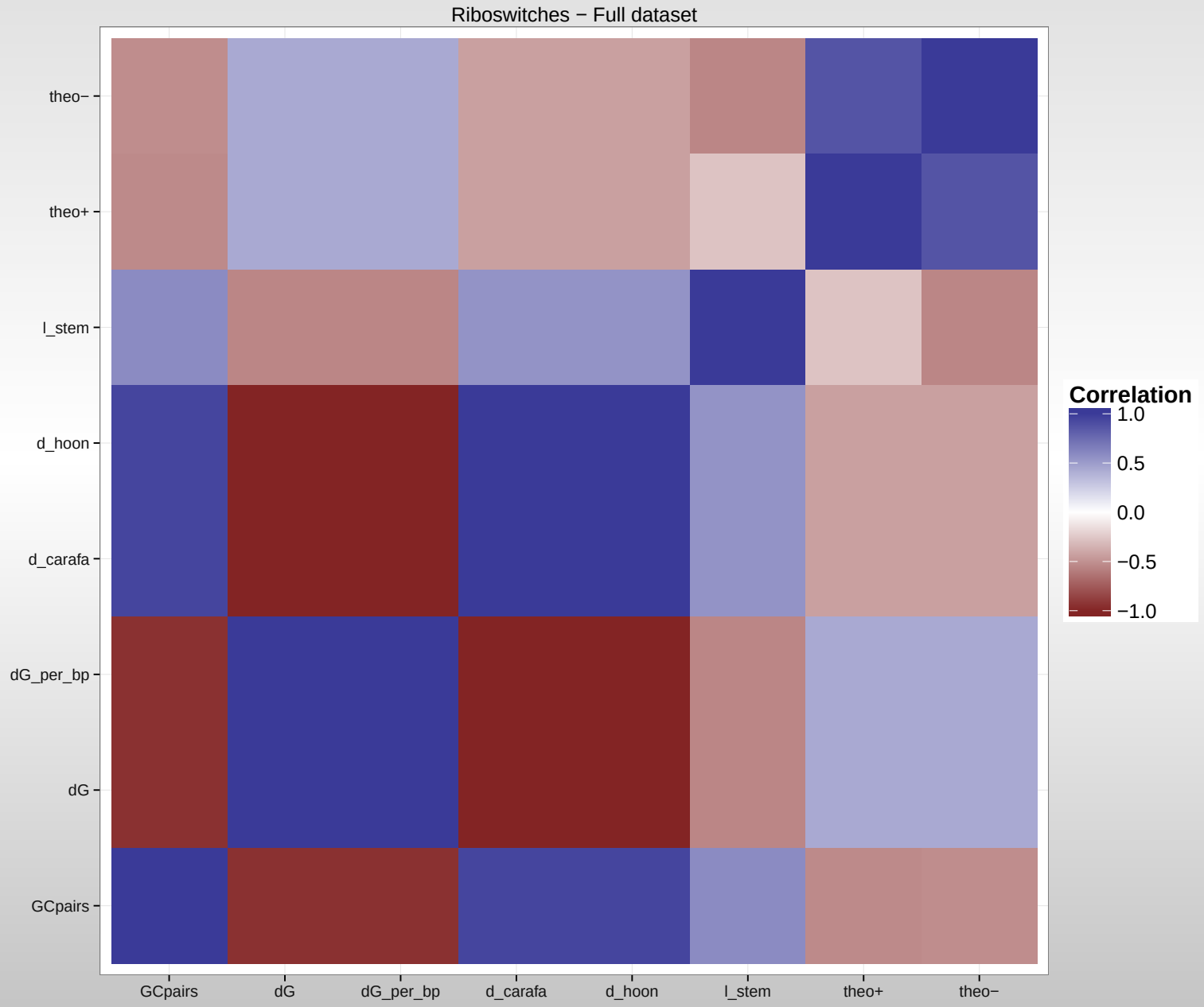
- 93.95% sensitivity, 94.36% specificity

[3] "Prediction of Transcriptional Terminators in Bacillus subtilis and Related Species", PLoS Comp Biol 1(3): e25

Terminator efficiency scores - de Hoon et al. 2005



Terminator efficiency scores - de Hoon et al. 2005



Terminator efficiency scores - Chen et al. 2013[4]

Thermodynamics only approach

$$T_S = 1 + \frac{1}{B_1 e^{\beta_1 \Delta G_L} + B_4 e^{\beta_4 (\Delta G_B + \Delta G_A - \Delta G_U)} \cdot (1 + B_1 e^{\beta_1 \Delta G_L})}$$

ΔG_U ... RNA : DNA hybrid of polyU stretch

ΔG_L ... Hairpin loop

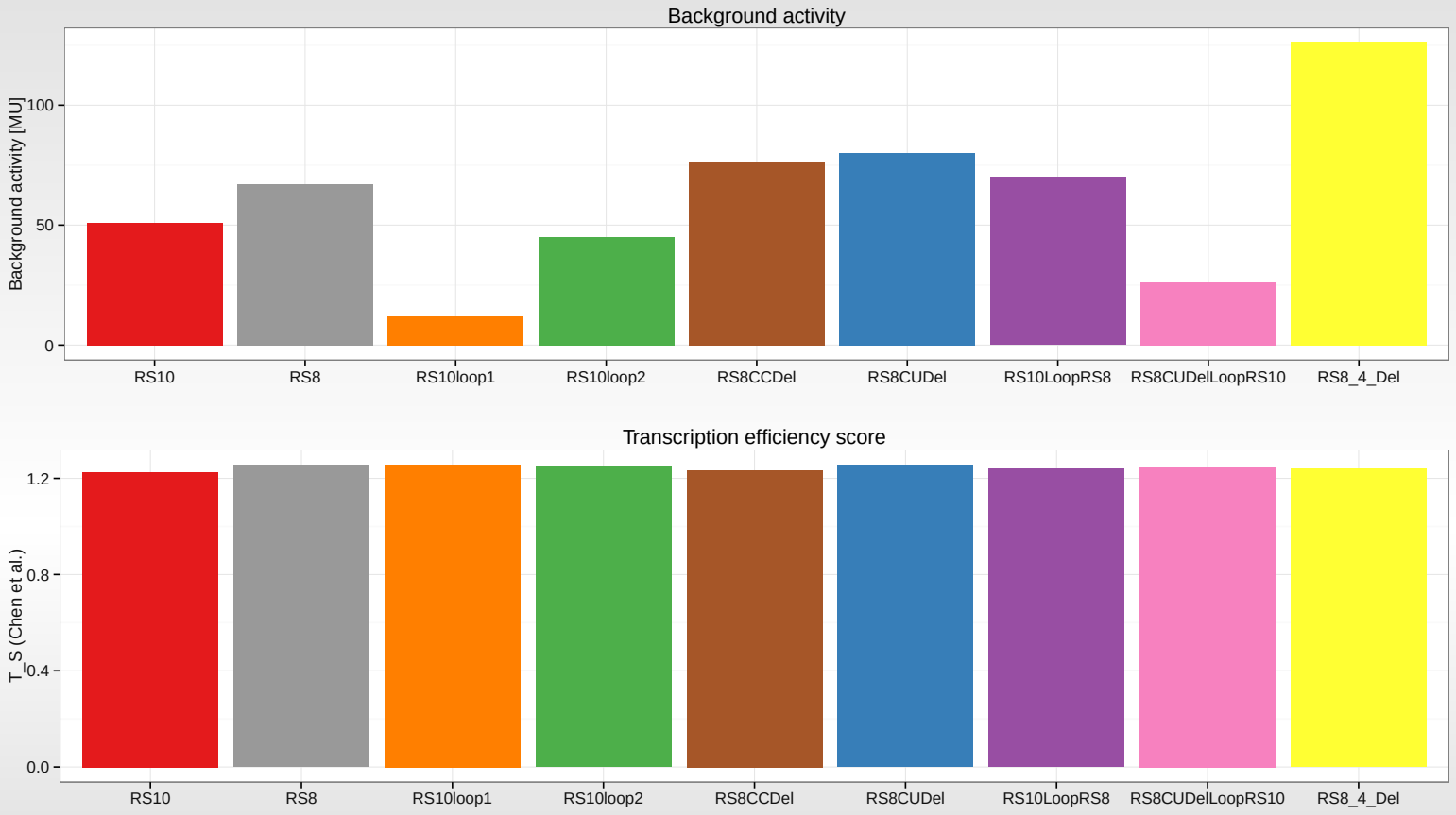
ΔG_B ... Stem base

ΔG_A ... RNA : RNA duplex of polyU stretch with $5'$ polyA

$B_1 = 0.005, B_4 = 6.0, \beta_1 = 0.6, \beta_4 = 0.45$

[4] "Characterization of 582 natural and synthetic terminators and quantification of their design constraints", Nature Methods (2013) Vol. 10 No. 7, 659-666

Terminator efficiency scores - Chen et al. 2013



Bioinformatics analysis of the results - revisited

Cotranscriptional structure prediction and more:

- prediction of MFE and structure with `Cofold`[5]
- prediction of cotranscriptional folding with `kinwalker`[6]
- evaluation of free energies for 3bp-,4bp- and 5bp-seed of the terminators as a measure of how fast the terminator will form

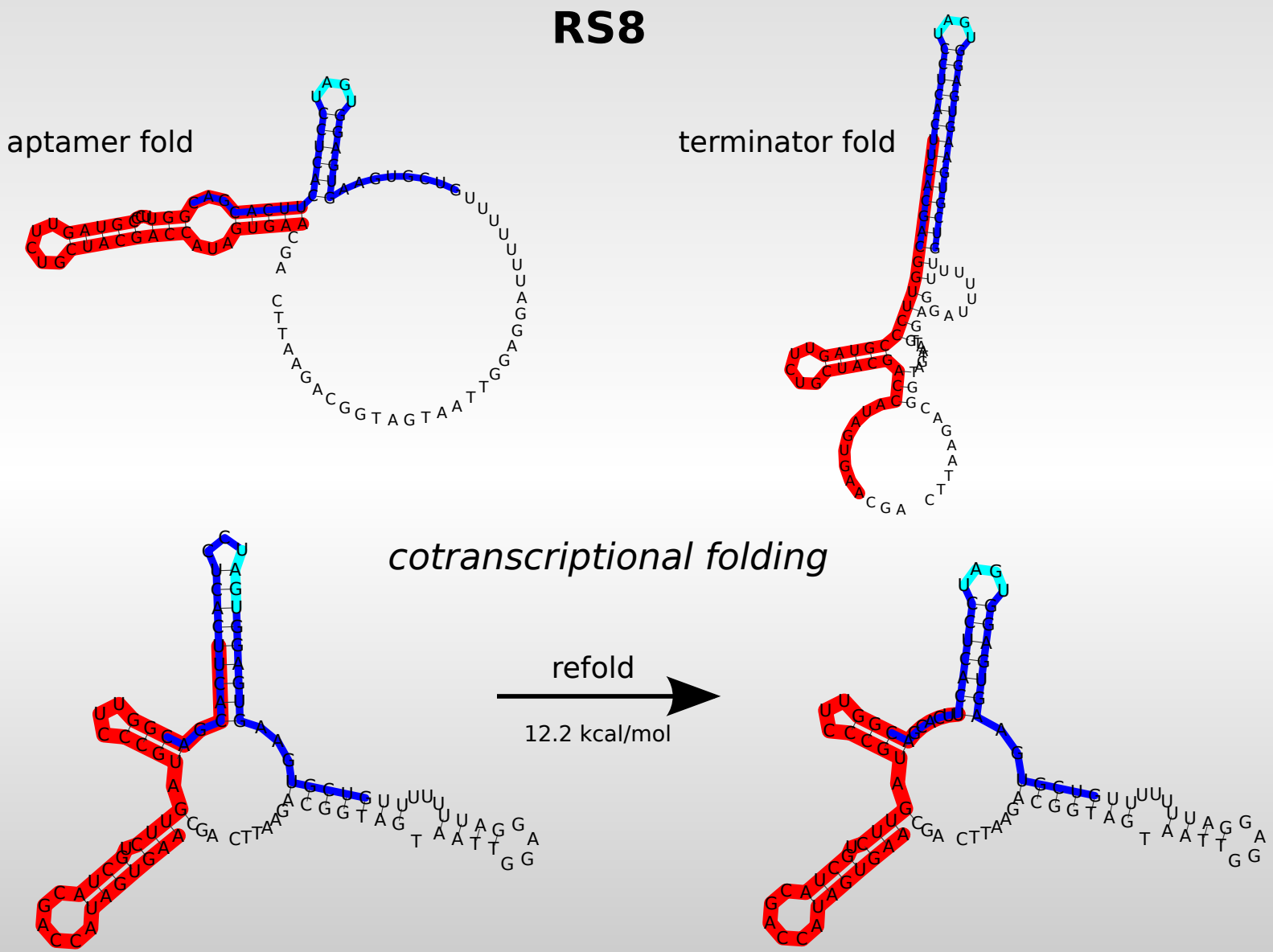
Results:

- `Cofold` output is the same as `RNAfold`
- `kinwalker` predicts cotranscriptional traps in 4 cases RS8, RS10loop2, RS8CCDel, RS8CUDel
- use terminator formation barrier as parameter

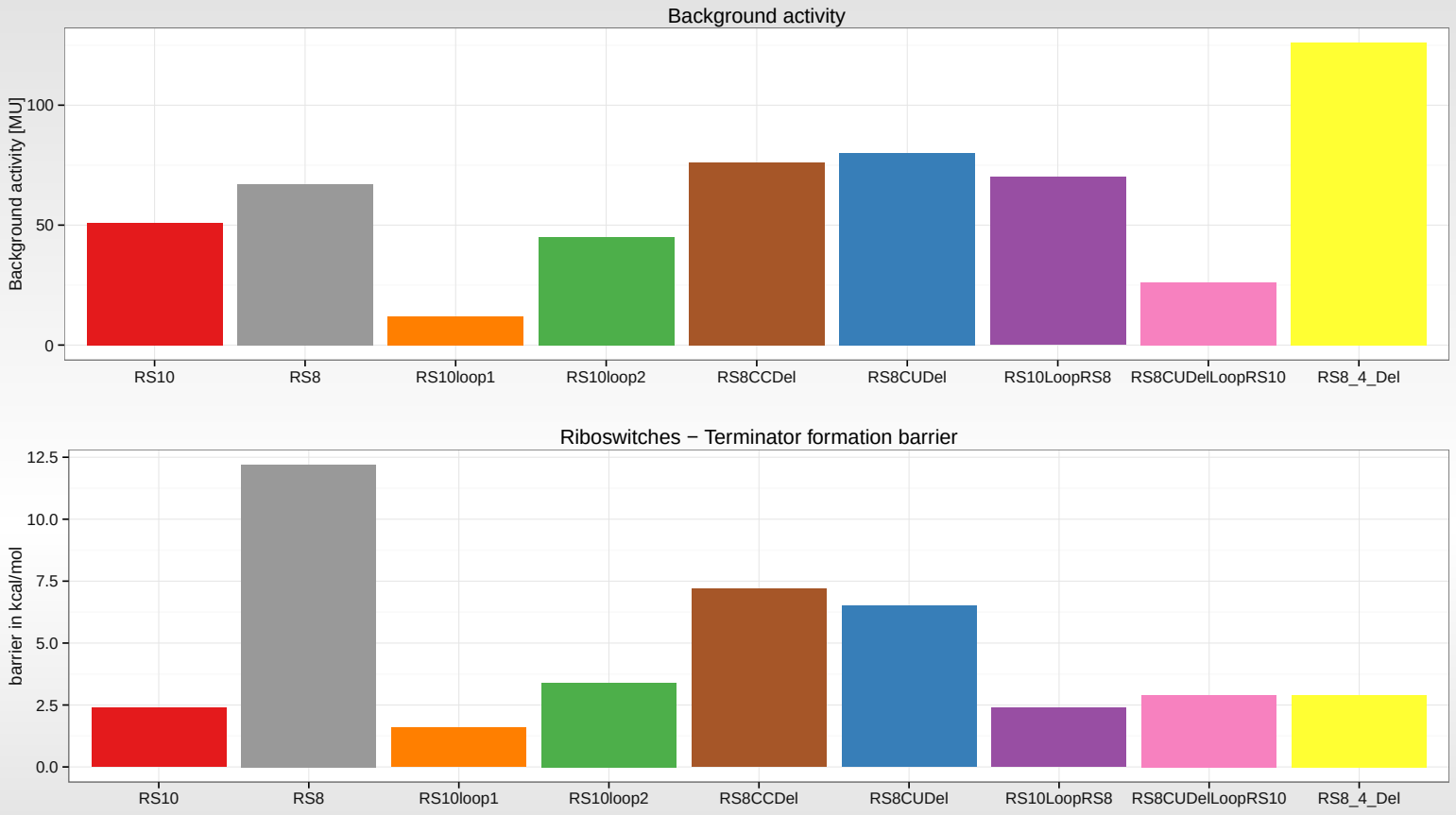
[5] Procter et al. 2013

[6] Geis et al 2008

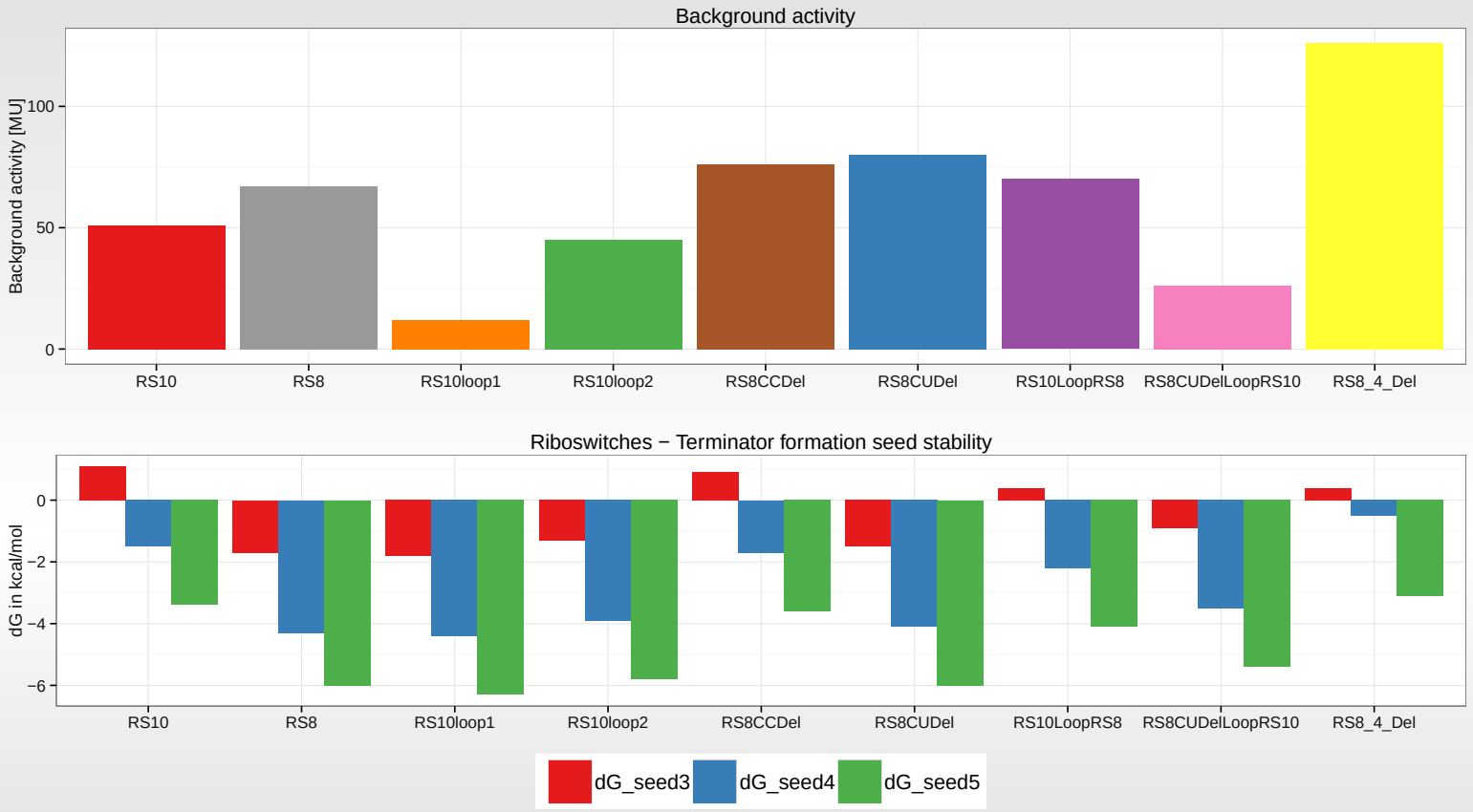
Hairpin formation barrier - RS8



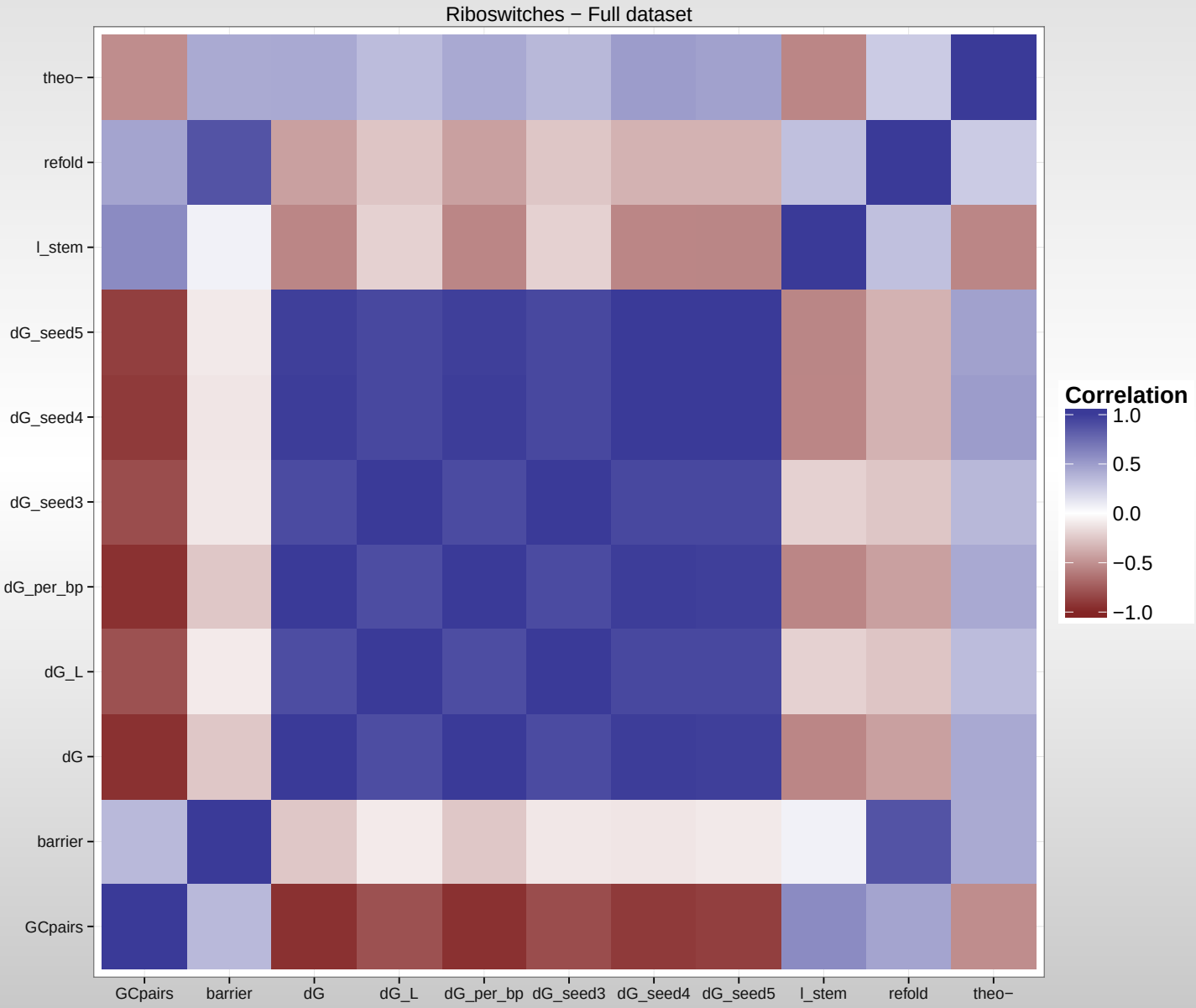
Hairpin formation barrier



Hairpin seed stabilities



Correlation of all parameters



Conclusion

- pure thermodynamic design is insufficient
- terminator seed performance seems to matter

  Best design: Tetra-loops (`GAAA`, `UUCG`) and strong closing pairs
- cotranscriptional effects have to be taken care of

  Design must exclude hairpins attenuating the terminator
- available terminator efficiency scores may be misleading
- construct new measure that incorporates the above parameters

Thanks to

- Manja Wachsmuth
- Sven Findeis
- Mario Mörl
- Peter F Stadler

Thank You for your attention!