ISRAEL JOURNAL OF ECOLOGY & EVOLUTION, Vol. 52, 2006, pp. 263–279

ADAPTIVE MUTATIONS IN RNA-BASED REGULATORY MECHANISMS: COMPUTATIONAL AND EXPERIMENTAL INVESTIGATIONS

Danny Barash,[a,b,*] Johannes Sikorski,[a,e] Elizabeth B. Perry,[d] Eviatar Nevo,[a] and Evgeny Nudler[c]

[a]*Institute of Evolution, University of Haifa, Haifa 31905, Israel*
[b]*Department of Computer Science, Ben-Gurion University of the Negev, Be'er Sheva 84105, Israel*
[c]*Department of Biochemistry, New York University Medical School, New York 10016, New York, USA*
[d]*Department of Biology, Wesleyan University, Middletown 06459, Connecticut, USA*
[e]*Deutsche Sammlung von Mikroorganismen und Zellkulturen GMbH (DSMZ), 8124 Braunschweig, Germany*

ABSTRACT

Recent discoveries of RNA-based regulatory mechanisms have prompted substantial interest in how they formed and the extent to which varying environmental conditions have influenced their evolution. One class of RNA-based regulatory mechanism that has been found in bacteria is the riboswitch, regulating the biosynthesis of certain vitamins by an RNA genetic control element that senses small molecules and responds with a structural change that affects transcription termination or translation initiation without the participation of proteins. By taking the thiamin pyrophosphate (TPP)-riboswitch in *Bacillus subtilis* as a model system, we wish to examine whether beneficial mutations may exist at the level of RNA that will cause an improvement in organism fitness. By computationally analyzing the difference in primary and secondary structure of the *B. subtilis* TPP-riboswitch collected from the xeric "African" south-facing slope (SFS) vs. the mesic, "European", north-facing slope (NFS) in "Evolution Canyon" III at Nahal Shaharut, southern Israel, we wish to experimentally study the environmental effect on transcription termination in these RNA-based regulatory mechanisms that are believed to be of ancient origin in the evolutionary time scale. Computational results, so far, indicate that specific mutations affect the riboswitch conformation by causing a global rearrangement. We would like to check whether such mutations could be adaptive mutations that may have a beneficial fitness effect, taking the TPP-riboswitch as a model system at the micro-scale. Empirical results so far indicate that in the promoter region of the TPP-riboswitch, all mutations increase nucleotide GC content in the xeric SFS, whereas in the mesic NFS they increase AT content. Preliminary examination of termination efficiency

*Author to whom correspondence should be addressed. E-mail: dbarash@cs.bgu.ac.il
Accepted February 2007.

of strains found exclusively on one slope or the other, reveal increased termination efficiency in the presence of TPP and at more moderate temperatures, but only a suggestion of greater termination efficiency from strains found on both slopes. We expect that further results will shed light on the mutational differences of TPP-riboswitch sequences found on opposite slopes of "Evolution Canyon" III at Nahal Shaharut, potentially leading to interesting discoveries that relate to the topic of adaptive, nonrandom mutations.

Keywords: adaptive mutations, RNA regulation

INTRODUCTION

Mutations at the molecular level that will be favored by natural selection are difficult to predict. Their prediction would require knowledge of the traits, how they are coded for genetically, and their fitness consequences. Modeling such a feature space with relationship to the molecular level is obviously a complicated matter in general. However, in certain instances such modeling may prove possible. In those cases, we may be able to make predictions based on first principles. One such case is provided here, using a simple regulation system for which it is possible to predict mutations that intervene in function by dramatically altering the structure. We can then compare these predictions to results derived from exposing organisms with the chosen regulation system to extreme natural environments.

In this work, we present a novel approach for investigating the effect of extreme environmental conditions on an RNA-based regulatory mechanism, called a riboswitch, in "Evolution Canyon" III at Nahal Shaharut, southern Israel. We integrate computational modeling at the level of nucleotides, biochemical experiments that measure transcription efficiency, and evolutionary theory in order to examine and understand mutational behavior in contrasting environmental conditions and their effects on the recently discovered thiamin pyrophosphate (TPP) riboswitch. Our model organism is *Bacillus subtilis*, in which the RNA-based regulation mechanism (riboswitch) acts at the level of transcription termination for the control of biosynthesis of the vitamin.

The riboswitch regulation mechanism was discovered in 2002 in bacteria at the level of transcription termination in *Bacillus subtilis* (Mironov et al., 2002) and translation initiation in *Escherichia coli* (Winkler et al., 2002). Genetic control of transcription termination and translation initiation is typically a sophisticated process, with the participation of many proteins that monitor the environment and selectively bind to targets. In contrast, riboswitches are RNA genetic control elements that are unique in the sense that they are capable of directly sensing small ligands. As a consequence, they can assume a regulatory role by switching between two highly stable states that affect synthesis, without the participation of regulatory proteins. Reviews on riboswitches are available in Nudler and Mironov, 2004 and Winkler and Breaker, 2003.

Mironov et al. (2002) and Winkler et al. (2002) showed that in transcription termination during thiamin (vitamin B1) biosynthesis in *B. subtilis* and in translation initiation in *E. coli*, respectively, an RNA element by itself can respond to a change in the con-

centration of thiamin pyrophosphate (TPP) with a conformational rearrangement. This capability of the mRNA unique element, namely, a "TPP-riboswitch" in the thiamin biosynthesis case, provides an RNA genetic control mechanism in prokaryotic transcription termination and translation initiation. Initial studies that led to these findings (Grundy and Henkin, 1998; Gelfand et al., 1999; Miranda-Rios et al., 2001) described some of these control features. Other riboswitches were found, including flavin mononucleotide (FMN), S-adenosylmethionine (SAM), coenzyme B12, lysine, guanine, and adenine (Winkler and Breaker, 2003).

A substantial amount of information can be inferred about the riboswitch mechanism by examining its primary sequence and secondary structure. This is because certain highly stable secondary structural motifs (e.g., terminators/anti-terminators and sequestors/anti-sequestors) have been located (Rodionov et al., 2002), in addition to the conserved box (Grundy and Henkin, 1998), or aptamer (Winkler and Breaker, 2003) that can identify riboswitches in a manner analogous to fingerprinting. Furthermore, it is speculated that riboswitches are the remains of an ancient regulation mechanism from the RNA World, dating back to very early events in the evolutionary timescale, since riboswitches work as genetic control systems without the involvement of proteins.

Theoretical studies of RNA secondary structure and evolution (Schuster et al., 1994; Higgs, 2000; Cowperthwaite et al., 2005) have attracted considerable attention in recent years. The "switching" mechanism of the riboswitches can be modeled using computational RNA secondary structure prediction methods (Zuker and Sankoff, 1984; Hofacker et al., 1994; Mathews et al., 1999; Zuker, 2003), as was first used for illustration in Mironov et al., 2002. This provides the potential for identifying a priori specific point mutation substitutions that may affect gene regulation. Furthermore, it may eventually be possible to predict based on first principles which mutations will be favored in specified environments. The works of Mironov et al. (2002) on the biology of riboswitches and Barash (2003) on the computational modeling of deleterious mutations serve here as the starting point for investigating how the TPP-riboswitch will be affected by the opposing extreme environments of the xeric "African" south-facing slope (SFS) vs. the mesic "European" north-facing slope (NFS) at "Evolution Canyon" III, Nahal Shaharut, Israel.

The paper is organized as follows. We begin with a description of the computational and experimental methods that are being used for analyzing the *Bacillus* strains collected from the Nahal Shaharut "Evolution Canyon" III (Grishkan et al., 2007). Computational methods are then used for predicting deleterious and beneficial mutations in the TPP-riboswitch sequence (Barash, 2003; Churkin and Barash, 2006) by measuring the differences between mutants and predicting the wildtype RNA secondary structure by energy minimization using tree-graphs. Experimental methods to isolate and subsequently sequence the *Bacillus subtilis* strains of interest are also described. Next, the results obtained from the biology experiments are reported in the Results Section. The main results are arranged as a sequence alignment from sequencing a few strains from the SFS vs. the NFS. Additional results are the preliminary outputs from the readout experiments to determine termination efficiency with and without TPP. Finally, the Discussion pro-

vides a description of the analysis carried out thus far on results that have been obtained, and suggestions for future directions such as the possibility of observing evidence for adaptive and nonrandom mutations in the *B. subtilis* TPP-riboswitch.

MATERIALS AND METHODS

COMPUTATIONAL METHODS

Besides the primary RNA sequence in riboswitches, the secondary structure plays an important role in determining how the riboswitch mechanism functions and generates a conformational switching (Winkler and Breaker, 2003; Nudler and Mironov, 2004). This can be viewed in Fig. 1, which illustrates how a riboswitch regulates gene expression in a schematic way. Figure 1 depicts how RNA secondary structure elements such as the terminator, anti-terminator, sequestor, and anti-sequestor can perform respectively to achieve gene regulation. The salient feature is that the metabolite itself interacts directly with the RNA to alter secondary structure and promote or suppress transcription. The important role of secondary structural elements in riboswitch control of bacterial metabolism can serve as the basis for their analysis by use of methods that have been devised to predict deleterious and beneficial mutations in the secondary structure of RNAs (i.e., those nucleotide substitutions that have the largest effects on secondary structure; Barash, 2003; Churkin and Barash, 2006). Available secondary structure prediction computer packages that seek to minimize the total free energy as a cause of the formed base pairings have evolved over the years, starting with work by Michael Zuker at the beginning of the 1980s (Zuker and Sankoff, 1984). For example, using one of the most widely used computer programs, called mfold (Zuker, 2003), to initially predict the secondary structure of the wildtype depicted in Fig. 2(A) and then to computationally examine the riboswitch alterations as a response to several randomly selected point mutations, it is clearly seen by trial and error that while other parts of the riboswitch are responding with structural changes, the terminator hairpin remains unaltered. Thus, it is expected that at best, only a few point mutations will be able to cause a switch from the terminator structure to an anti-terminator structure. In order to predict the locations of such mutations, the first step is to represent the RNA secondary structure (in our case, predicted by energy minimization rather than experimentally derived) in a simplified representation that will enable the processing of many folds in a computationally tractable manner. This can be achieved by using coarse-grained tree-graphs (see Appendix), in which the graph nodes represent secondary structure motifs such as loops, bulges, and hairpins, whereas the graph edges represent stems in the secondary structure. For example, the anti-sequestor in Fig. 1(D) corresponds to the "star" tree-graph on the right of Fig. A1 in the Appendix, since the 5′–3′ end in Fig. 1(D) is represented by node #3 (Fig. A1) at the root, and the three other loops in Fig. 1(D) are represented by nodes #1,2,4 (Fig. A1). The second step is to transform each tree-graph into a matrix, called the Laplacian (see Appendix), which has special favorable properties to facilitate the computational process. The third step is to extract the eigenvalues of the Laplacian matrix

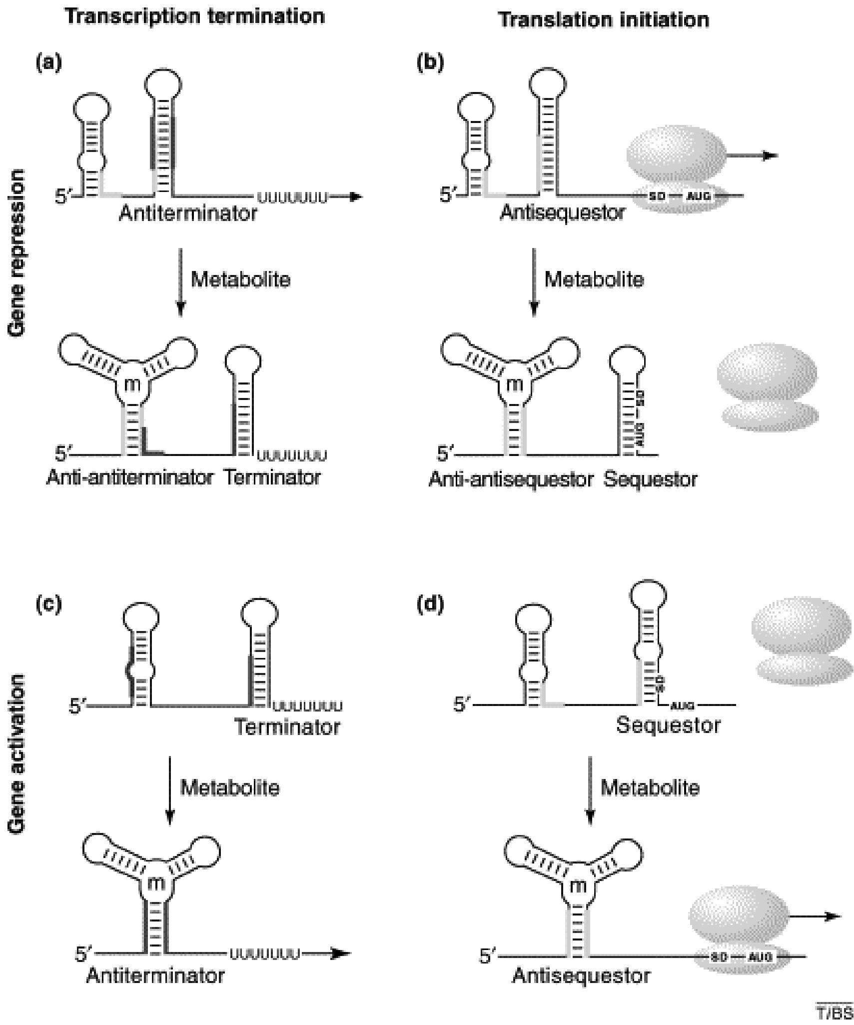

Fig. 1. Riboswitch-mediated control of gene expression. Bacterial riboswitches repress or activate their genes depending on the configuration of the corresponding leader RNA sequence (shown by the line growing from the 5′ end). In response to the change of the metabolite concentration the riboswitches control transcription termination (A,C) or translation initiation (B,D), or both if the stem-loop structure of the terminator also serves as a sequester of the ribosome binding site (RBS). In each case, the binding of a specific metabolite (m) to the conserved RNA sensor domain stabilizes the riboswitch structure (shown as a hypothetical three-stem structure), thus preventing the formation of an alternative RNA structure that could be an anti-terminator (A), anti-sequestor (B), terminator (C), or sequestor of RBS (D). The ribosome is shown in pale gray. The complementary RNA regions are indicated by bold lines.

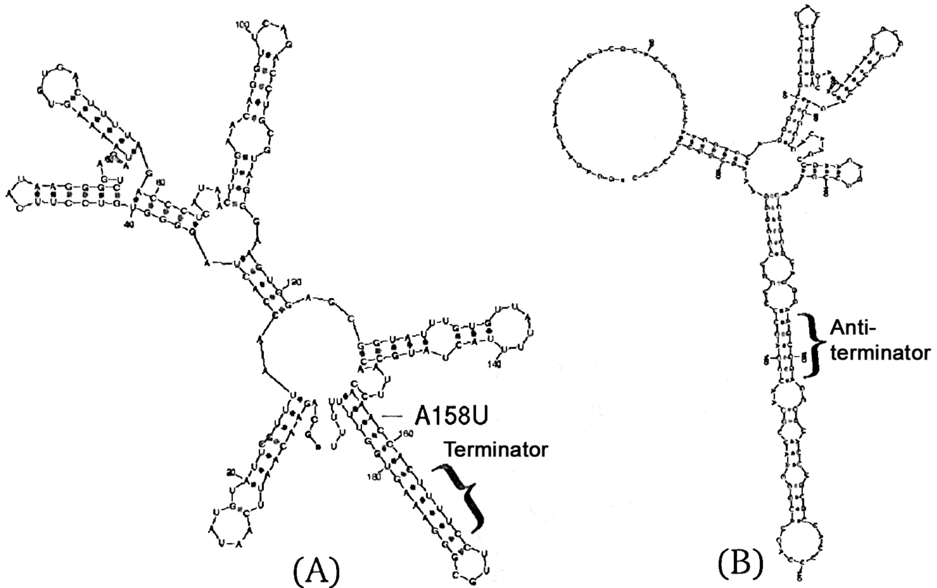

Fig. 2. (A) The predicted wildtype structure of the TPP-riboswitch in transcription termination in *Bacillus subtilis*, using Zuker–Turner's mfold with the latest thermodynamic parameters (Zuker and Sankoff, 1984; Mathews et al., 1999; Zuker, 2003), exhibiting a terminator structure. (B) The predicted structure after the insertion of point mutation A158U, where the riboswitch assumes an anti-terminator structure.

to obtain numerical values indicative of the topology of the graphs. Using mathematical theorems, it can be shown that the second smallest eigenvalue of the Laplacian matrix, corresponding to the tree-graph representation (see Appendix), is a floating point number between 0.0 and 1.0 that captures the compactness of the RNA secondary structure (see Appendix and Barash, 2003, for more detailed explanation). This number has a clear intuitive meaning: the closer it is to zero, the sooner the tree-graph shape becomes spread and linear, whereas the closer it is to 1, the sooner the tree-graph becomes more compact—until it reaches a star shape. As a fourth step, starting with the wildtype, it is possible to computationally introduce all possible single-point mutations one at a time into the RNA sequence and predict the secondary structure of each mutated sequence by energy minimization (as was mentioned in the first step). For each mutated fold predicted by mfold (Zuker, 2003), the smallest second eigenvalue of the Laplacian matrix is calculated and the results are organized in a table (Barash, 2003; Churkin and Barash, 2006). Using the RNAMute software (Churkin and Barash, 2006), one can scan through the table, double click on a desired eigenvalue, obtain another table with all point mutations corresponding to this eigenvalue, and double click on each one of the mutations to view a screen shot of the secondary structure. This allows the identification of terminator and anti-terminator regions in a manual and tractable manner that is convenient for the

user. We examine the implementation of these methods to the case of the TPP-riboswitch sequence, results of which are shown in Fig. 2.

In Fig. 2(B), the mfold-predicted terminator structure in the expression platform domain (the riboswitch part that includes the terminator and anti-terminator hairpins, respectively) was disrupted by introducing a predicted selective mutation that may interfere with or enhance the riboswitch function. In Fig. 2(A), the selective mutation is labeled; upon introducing the mutation A158U to the wildtype sequence, the mutated sequence in Fig. 2(B) switches to an mfold-predicted anti-terminator structure based on the thermodynamic energy rules. In our case this is without manually constraining the relevant base-pairings to simulate a change in TPP concentration, as opposed to a manual forcing of some base-pairings that was performed in Mironov et al. (2002) and obviously resulted in a switch from terminator to anti-terminator structure for the purpose of illustrating how the riboswitch mechanism functions. Preliminary experimental verification at Nudler's laboratory indeed confirmed the computer prediction that the mutation A158U disrupts termination of transcription. This indicates that the mutation A158U is deleterious and interferes with the riboswitch control of vitamin biosynthesis. In our work, it would be extremely interesting if traces for such computational predictions of selective mutations were independently observed in strains extracted from the xeric NFS or the mesic SFS of "Evolution Canyon" III. An even more ambitious long-term goal would be to be able to predict the characteristics of mutations found on the opposite slopes.

The computational methodology for locating these predicted deleterious mutations, some of which have already been confirmed in laboratory experiments, was outlined in Barash, 2003. These mutations reside in the expression platform domain (which extends approximately from nucleotide 1 to 40 and 110 to 189 in Fig. 2, the domain of the terminator/anti-terminator; for nomenclature see Winkler and Breaker, 2003), leaving the aptamer domain (which extends approximately from nucleotide 40 to nucleotide 110 in Fig. 2, the domain that binds the ligands) intact. Interestingly, compensatory mutations (restoring the secondary structure) that correspond to the deleterious mutations reside at locations far away from the expression platform. Such mutations, if detected in nature, may potentially lead to a discovery of nonrandom mutations in the TPP-riboswitch because these mutations are very specific and cause the appearance/disappearance of the terminator, as opposed to random mutations that have no significance on the global structure. Thus, besides examining the sequence alignments of strains collected at various stations in "Evolution Canyon" III, we also checked whether the mutations that are found are computationally predicted mutations that alter the TPP-riboswitch secondary structure dramatically, or alter stability of the riboswitch, thus shifting the regulatory equilibrium between the terminator and anti-terminator structures.

EXPERIMENTAL METHODS

"Evolution Canyon" III (EC III) is located in the southern Negev Desert, at Nahal Shaharut, a tributary of Nahal Hiyyon (29º55′N, 34º58′E). EC III has a south-facing

slope (SFS, about 35° rise) and a north-facing slope (NFS, about 30° rise), separated by about 150 m at the bottom (see description in Grishkan et al., 2007). The soil samples (each ~100 g, from the top 1–3-cm layer) were taken on 25 March 2003, at the transition phase, from winter to spring. Soil was collected from each of the three SFS and NFS stations and one bottom station, with three collecting sites per station. Altogether, 76 strains from the *B. subtilis–B. licheniformis* clade were isolated from different stations along the microclimate gradient of this canyon system and were distinguished by metabolic tests (Cohan et al., 1991).

In order to obtain a fine-scaled genetic differentiation of the strains, we applied a sensitive genetic fingerprinting method (Random Amplified Polymorphic DNA—PCR, RAPD), as described previously (Sikorski et al., 2001a). In order to identify the taxonomic affiliation of prominent RAPD clusters to either *B. subtilis* or *B. licheniformis*, we determined the 16S sequence of representative strains (Sikorski et al., 2001b). An initial PCR of the thi-box using *B. subtilis* 168 primers (as described in Mironov et al., 2002) failed. We therefore developed, on the basis of available genome sequences from the genus *Bacillus*, a new set of forward and reverse primers that encompassed all combinations in the thi-box region spanned by the primer pair described previously (Mironov et al., 2002). Using different combinations of our newly developed primers, we succeeded in determining the thi-box sequence of several strains representing dominant genetic groups from EC III.

A few selected strains were tested for their termination efficiency, which is a measure of how effective the riboswitch mechanism is in controlling transcription termination (Mironov et al., 2002). Some of the strains were not chosen, in particular 2C1 and 6B1, since their sequences are the same in both the NFS and SFS. We present results from strains that so far have been found on one slope or the other only. These are most likely to represent strains with traits adapted to the different conditions. Termination experiments are biochemical experiments designed to measure the efficiency of transcription termination. If the biosynthesis of vitamin is suppressed, the termination efficiency will drop as a consequence. Thus, the efficiency of transcription termination is indicative of the gene expression (whether the vitamin is synthesized or not).

Experimental procedures of in vitro transcription were employed to validate predictions regarding differences in termination efficiencies of strains adapted to different environmental conditions. All transcription templates were generated by PCR and purified from low-melting agarose and diluted in TE to 2 pmol/μl. The *E. coli* His6-β′-tagged RNAPσ70 holoenzyme was purified. The start-up elongation complex on the thi-leader templates was prepared as follows: 2 pmol DNA (1 μl) and 1 pmol RNAP (1 μl) were mixed in the 10 μl of the transcription buffer [TB: 100 mM KCl, 3 mM $MgCl_2$, 50 mM tris-HCl (pH 7.5)] and incubated 5 min at 37 ºC. Next, 1μl of the starting cocktail was added to give the final concentration of GpGpU primer of 20 μM, ATP and GTP of 30 μM each, and [α-32P] CTP (3000 Ci/mmol) of 0.3 μM. After 8 min at 37 ºC, the sample was diluted with TB and the aliquots were taken to the tubes containing all four NTP (250 μM final concentration). The chase reactions were stopped after ~15 min at 25 ºC by adding an equal volume of the sequencing gel loading buffer containing EDTA

(200 mM) and formamide (95%). Positions of the terminated products were determined by the transcription sequencing reaction with 3′-dNTPs (TriLink). Relative amounts of [32P] RNA were determined using PhosphorImager and software from Molecular Dynamics. Termination efficiency (%T) was calculated by dividing the amount of radioactivity in a particular terminated band by the total radioactivity present in that and all read-through bands.

RESULTS

The thi-box aligned sequences are presented in Fig. 3. The sequence at the top row is the original *Bacillus subtilis* sequence that was used in Mironov et al., 2002 and serves as a reference for our purpose. Below this, sequences belonging to various stations in EC III are reported, indicating their change from the reference (a dot is used by default if no change in the nucleotides occurred with respect to the reference). More details are found in the caption to Fig. 3.

Preliminary results from termination experiments on four strains (1C2, 2C3, 4B2, and 6A1) are already available at both room temperature and 37 ºC. The summary of the results given by measuring termination efficiency (%T) are reported in Fig. 4. Termination efficiency is greater in the presence of added TPP (M.S. = 10811.25, $F_{1,16}$ = 111.174, $p < 0.001$) and at room temperature compared to 37 ºC (M.S. = 661.25, $F_{1,16}$ = 6.80, $p = 0.019$). Trends weakly suggest that termination efficiency may be higher in strains that originate from the slopes compared to strains from the wadi bottom (M.S. = 214.938, $F_{1,16}$ = 2.206, $p = 0.157$), and even higher termination efficiency from strains from the SFS than from the NFS, but we will need much more sampling to adequately test this. A priori, while we believe that the differential levels of termination efficiencies convey selective advantage or disadvantage, we hope to find out the direction from the experiments. We do not know from a predictive standpoint under which environmental condition the up- or down-regulation of termination confers a selective advantage.

DISCUSSION

All sequences obtained in Fig. 3 were analyzed using the RNAMute package (Churkin and Barash, 2006) and checked for whether they retain the same terminator/anti-terminator structure as in the reference and whether any deviation of interest could be found in their mutational robustness to conformational rearranging point mutations. So far, in examining the RNA secondary structure itself without taking into consideration primary sequence composition, no differences have been found, and therefore further examinations are needed when more data become available. If this pattern will continue as more strains are processed, this will suggest that the terminator/anti-terminator structures are strongly conserved, and we should look for differences in the nucleotide composition of the sequence rather than a dramatic change in structure. However, only 10 out of 77 strains that are already available have been sequenced and more strains will be collected before reaching a definite conclusion.

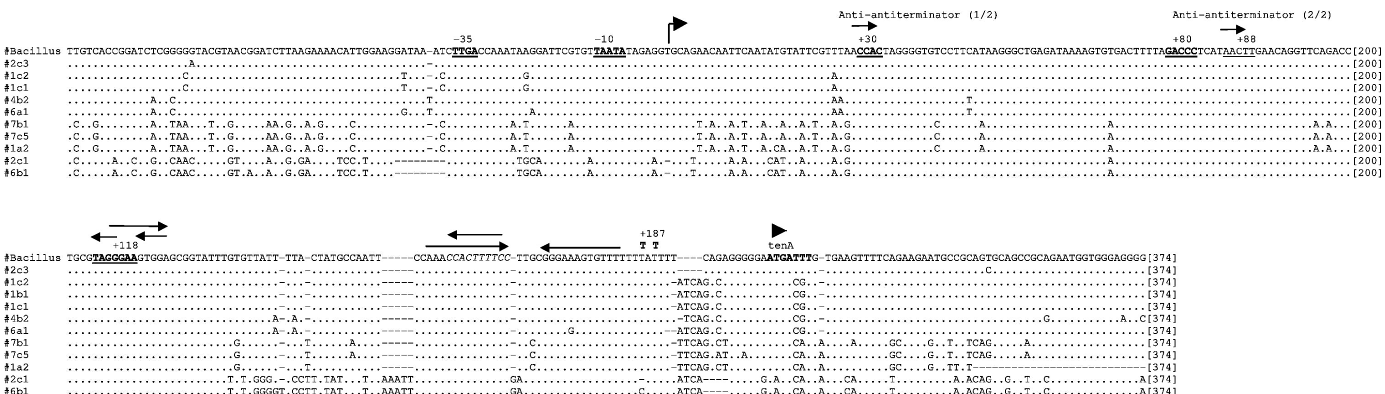

Fig. 3. The thi-box alignment of *Bacillus subtilis* strains from Nahal Shaharut, "Evolution Canyon" III. Following the "#" sign, the numbers represent stations, where stations 1,2 belong to the xeric, "African", south-facing slope (SFS) and stations 6,7 belong to the mesic, "European", north-facing slope (NFS); station 4 is from the valley bottom. Points in the sequence represent the same nucleotides as in the top row sequence. Bottom arrows show the terminator hairpin, top arrows denote the anti-terminator, and middle arrows denote the anti-anti-terminator (a stem structure that does not allow the anti-terminator to form; see Mironov et al., 2002). For reference, location of the anti-anti-terminator (Mironov et al., 2002) is indicated above the top row sequence.

Examining Fig. 3, the promoter region that is the most important region to inspect, as far as the riboswitch functionality in our context is concerned, is located between locations –35 and +15 (taken with the indexing scheme of the reference sequence). The fact that all the mutations are to C and G in the strains unique to the xeric south-facing slope (increasing GC content) and all the mutations are to A and T in the strains unique to the mesic north-facing slope (increasing AT content) may lead to an interesting interpretation that relates to the TPP-riboswitch functionality through nucleotide composition in the sequence while the secondary structure is preserved. In order to further explore this possibility, more strains are currently being exposed to termination experiments: a reference *B. subtilis* taken from “Evolution Canyon” III (in addition to the reference that is currently used in Fig. 3), 7B1, and additionally either 6B1 or 2C1 strains. It should also be noted that termination efficiency results for the control of *B. subtilis* and 2C3 sequences are expected to be highly similar since their nucleotide composition is the same, though with the current reference there is a substantial difference between these two strains. Termination efficiency measurements for the above sequences are essential for proceeding further with the observation on GC content.

From the termination efficiency results at hand, one could set forth a hypothesis that will be put to the test on a significantly larger sample size. Examining Fig. 4, there is a pattern of difference between room temperature (25 ºC) and body temperature (37 ºC) in that in the absence of TPP, the termination efficiency is larger in height and range at 25 ºC than at 37 ºC. Vice versa, at higher temperature, the termination across all strains is more uniform with far lower range than at cold temperature. One could interpret this in a way that at high temperature the strains are quite well able to proceed, in the absence of TPP, towards biosynthesis of TPP (TPP being the active derivative of vitamin B1). Thus, at warmer temperatures, induction of TPP synthesis seems to be more efficient and

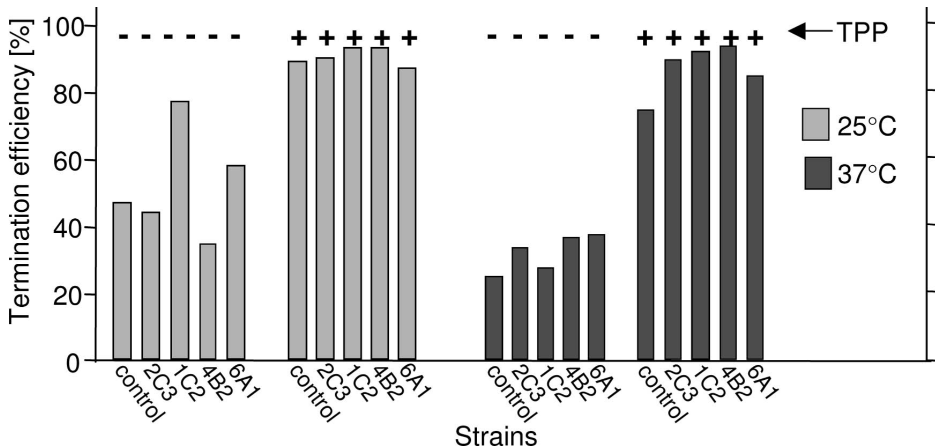

Fig. 4. Termination efficiency (% T) of *Bacillus subtilis* representative strains from “Evolution Canyon” III at room (25 ºC) and body (37 ºC) temperatures, with and without thiamin pyrophosphate (TPP).

under selective pressure. In contrast, at cooler temperatures the strains in the absence of TPP are more diverse in their possibility to explore the space of this phenotype and are associated with a relatively high termination efficiency, which one would not expect if the strains were under stress for TPP biosynthesis. Therefore, higher temperatures seem to trigger a more streamlined and efficient induction of TPP biosynthesis under conditions of absence of TPP. This may indicate that the "African" strains have, in general, a more streamlined and efficient biosynthesis regulation. This could have evolved under energetically and perhaps more stressful conditions on the "African" south-facing slope. This hypothesis can now be examined when increasing dramatically the sample size.

Traditional Darwinian thinking in evolutionary biology has been that mutations occur at random, regardless of fitness consequences to the resulting mutants (Darwin, 1859; Sabeti et al., 2006). If this holds true, then in nature it is only selection that decides whether mutations will be preserved or not. This so-called random mutation hypothesis has been repeatedly challenged and defended since early in the last century. However, experiments conducted during the 1980s presented decisively new findings, providing stronger support to an alternative hypothesis that mutations are more likely to occur when the environment favors the survival of the resulting mutants. These new findings challenged conventional evolutionary thinking, leading to some long-lasting debates. Additional experiments appeared postulating various instances of "adaptive mutations" or "directed mutations", whereby genetic variation apparently occurs in response to the environment rather than occurring independently of it.

The seminal works that have reported these findings in *Escherichia coli*, namely, Shapiro, 1984, followed by Cairns et al., 1988 and Hall, 1990, described genetic systems in which specific conditions seemed to increase the mutation rate. Basically, these studies suggested that organisms can respond to environmental stresses by reorganizing their genes in a purposeful way. Other such systems have recently been described in Hendrickson et al., 2002 and Slechta et al., 2003, as well as in *Saccharomyces* yeast (Fidalgo et al., 2006), and recent comprehensive reviews on the subject are available (Derzhavets et al., 1997; Lamb et al., 1998; Rosenberg, 2001; Roth et al., 2006). The existence of "directed mutations" shifts the course of evolution in a nonrandom way and evokes revolutionary thinking. For example, the problem mentioned by Slechta and coworkers (Slechta et al., 2003) raises an intriguing question whether general mutagenesis is a programmed response to stress, as we found in "Evolution Canyon" I in soil fungi (Lamb et al., 1998) and in *Drosophila* (Derzhavets et al., 1997).

A hypothesis stating that under extreme stresses, such as in the hypersaline Dead Sea, general mutagenesis may indeed be a general programmed response to stress can be put forth. The higher the stress and more specialized and unique (e.g., salt in the Dead Sea (Nevo et al., 2003) or the more extreme the drought in extreme desert conditions, such as in the "Evolution Canyon" at Nahal Shaharut (Grishkan et al., 2007) in the southern Negev Desert), the more favorable the conditions to detect directed, nonrandom, mutations since the organisms are under sharply canalized stress to resist salt or drought, respectively. Thus, the motivation for our current research efforts is to detect such nonrandom mutations in the "Evolution Canyon" of Nahal Shaharut, which serves as

an excellent test bed to model environmental stress (Nevo, 2001; Sikorski and Nevo, 2005). Numerous organisms and biological systems have been studied in "Evolution Canyons" in order to better understand biodiversity evolution, adaptation, and speciation. Environmental-dependent mutation rate has been detected in "Evolution Canyons" in *Drosophila melanogaster* (Derhavets at al., 1997) as well as in the soil fungus *Sordaria fimicola* (Lamb et al., 1998); in both organisms, the fly and the fungus, there is a significantly higher mutation rate on the xeric, "African", south-facing slope (SFS) of an "Evolution Canyon" relative to the mesic, "European", north-facing slope (NFS). We plan to considerably increase the sample size of our sequences and biochemical experiments in order to identify such behaviors within our model system.

Future research work should address several concerns, some of which stem from the variety of approaches being used (computational, experimental, and theoretical) and the need to tie them up coherently. For example, even with the data that we are planning to generate (termination efficiency for the aforementioned strains, in addition to the ones in Fig. 4), we will not have sufficient evidence, from the point of view of population genetics, to test for directed mutations. Also, we need to increase our sample sizes and to generate appropriate statistical tests. Another issue of concern is whether the laboratory results at room temperature adequately mimic the "Evolution Canyon" III conditions, since, during the summer, the soil temperature in "Evolution Canyon" III may reach well above 50 ºC. Moreover, the physiological conditions in an in vitro assay differ from the ones in the cell. Despite these limitations, we plan to overcome some of the technical details and highlight the effect of extreme environmental conditions on the recently discovered RNA-based regulation mechanisms while studying their adaptation to the environment. In particular, we hope to answer the question of whether, indeed, adaptive mutations are selected for at the mutational level, thereby supporting the hypothesis that even at the first evolutionary stage generating variation, mutations may be, at least partly, nonrandom, directed, or adaptive according to environmental stress. The novel interplay between sequence/structure computational modeling at the level of nucleotides, biochemical experiments that measure the functionality of riboswitches, and the extreme physiological conditions in the xeric, "African", north-facing slope (NFS) and the mesic, "European", south-facing slope (SFS) of "Evolution Canyon" III at Nahal Shaharut may well lead us to the observation of peculiar phenomena.

ACKNOWLEDGMENTS

The research was supported by a grant from the United States-Israel Binational Science Foundation (BSF 2004291) and the Ancell–Teicher Research Foundation for Genetics and Molecular Evolution.

REFERENCES

Barash, D. 2003. Deleterious mutation prediction in the secondary structure of RNAs. Nucleic Acids Res. 31(22): 6578–6584.

Cairns, J., Overbaugh, J., Miller, S. 1988. The origin of mutants. Nature 335: 142–145.

Churkin, A., Barash, D. 2006. RNAMute: RNA secondary structure mutation analysis tool. BMC Bioinformatics 7: 221.

Cowperthwaite, M.C., Bull, J.J., Meyers, L.A. 2005. Distributions of beneficial fitness effects in RNA. Genetics 170(4): 1449–1457.

Cohan, F.M., Roberts, M.S., King, E.C. 1991. The potential for genetic exchange by Transformation within a natural population of *Bacillus subtilis*. Evolution 45: 1393–1421.

Darwin, C. 1859. The origin of species by means of natural selection. John Murray, London.

Derhavets, E., Korol, A., Pavlicek, T., Nevo, E. 1997. Adaptation to stressful environment and mutation rate: a case study in *Drosophila melanogaster*. Drosophila Information Service 80: 53.

Fidalgo, M., Barrales, R.R., Ibeas, J.I., Jimenez, J. 2006. Adaptive evolution by mutations in FLO11 gene. Proc. Natl. Acad. Sci. USA 103(30): 11228–11233.

Gelfand, M.S., Mironov, A.A., Jomantas, J., Kozlov, Y.I., Perumov, D.A. 1999. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. Trends Genet. 15 (11): 439–442.

Grishkan, I., Beharav, A., Nevo, E. 2007. Adaptive biodiversity of soil microfungi in extreme Negev Desert, Israel. Biol. J. Linn. Soc., in press.

Grundy, F.J., Henkin, T.M. 1998. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. Mol. Microbiol. 30: 737–749.

Hall, B.G. 1990. Spontaneous point mutations that occur more often when advantageous than when neutral. Genetics 126: 5–16.

Hendrickson, H., Slechta, E.S., Bergthorsson, U., Andersson, D.I., Roth, J.R. 2002. Amplification-mutagenesis: evidence that "directed" adaptive mutation and general hypermutability result from growth with a selected gene amplification. Proc. Natl. Acad. Sci. USA 99(4): 2164–9.

Higgs, P.G. 2000. RNA secondary structure: physical and computational aspects. Q. Rev. Biophys. 33: 199–253.

Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., Schuster, P. 1994. Fast folding and comparison of RNA secondary structures. Monatsh. Chem. 125: 167–188.

Lamb, B.C., Saleem, M., Scott, W., Thapa, N., Nevo, E. 1998. Inherited and environmentally-induced differences in mutation frequencies between wild strains of *Sordaria fimicola* from "Evolution Canyon". Genetics 149: 87–99.

Mathews, D.H., Sabina, J., Zuker, M., Turner, D.H. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288(5): 911–940.

Miranda-Rios, J., Navarro, M., Soberon, M. 2001. A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. Proc. Natl. Acad. Sci. USA 98: 9736–9741.

Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A., Nudler, E. 2002. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. Cell 111: 747–756.

Nevo, E. 2001. Evolution of genome-phenome diversity under environmental stress. Proc. Natl. Acad. Sci. USA 98: 6233–6240.

Nevo, E., Oren, A., Wasser, S. 2003. Fungal life in the Dead Sea. Koeltz Scientific Books, Konigstein, Germany.

Nudler, E., Mironov, A.S. 2004. The riboswitch control of bacterial metabolism. Trends Biochem. Sci. 29(1): 11–17.

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., Gelfand, M.S. 2002. Comparative genomics of thiamin biosynthesis in procaryotes. J. Biol. Chem. 277: 48949–48959.

Rosenberg, S.M. 2001. Evolving responsively: adaptive mutation. Nat. Rev. Genet. 2: 504–14.

Roth, J.R., Kugelberg, E., Reams, A.B., Kofoid, E., Andersson, D.I. 2006. Origins of mutations under selection: the adaptive mutation controversy. Annu. Rev. Microbiol. 60: 477–501.

Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D., Lander, E.S. 2006. Positive natural selection in the human lineage. Science 312 (5780): 1614–1620.

Schuster, P., Fontana, W., Stadler, P.F., Hofacker, I.L. 1994. From sequences to shapes and back: a case study in RNA secondary structures. Proc. Biol. Sci. 255(1344): 279–284.

Shapiro, J.A. 1984. Observations on the formation of clones containing araB-lacZ fusions. Mol. Gen. Genet. 194: 79–90.

Sikorski, J., Nevo, E. 2005. Adaptation and incipient sympatric speciation of *Bacillus simplex* under microclimatic contrast at "Evolution Canyons" I and II, Israel. Proc. Natl. Acad. Sci. USA 102(44): 15924–9.

Sikorski, J., Jahr, H., Wackernagel, W. 2001a. The structure of a local population of phytopathogenic *Pseudomonas brassicacearum* from agricultural soil indicates development under purifying selection pressure. Environ. Microbiol. 3(3): 176–186.

Sikorski, J., Stackebrandt, E., Wackernagel, W. 2001b. *Pseudomonas kilonensis* sp. nov., a bacterium isolated from agricultural soil. Int. J. Syst. Evol. Microbiol. 51: 1549–1555.

Slechta, E.S., Bunny, K.L., Kugelberg, E., Kofoid, E., Andersson, D.I., Roth, J.R. 2003. Adaptive mutation: general mutagenesis is not a programmed response to stress but results from rare coamplification of dinB with lac. Proc. Natl. Acad. Sci. USA 100(22): 12847–52.

Winkler, W.C., Nahvi, A., Breaker, R.R. 2002. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419(6910): 952–956.

Winkler, W.C., Breaker, R.R. 2003. Genetic control by metabolite-binding riboswitches. Chem. Bio. Chem. 4: 1024–1032.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucl. Acids Res. 31: 3406–3415.

Zuker, M., Sankoff, D. 1984. RNA secondary structures and their prediction. Bull. Math. Biol. 46(4): 591–621.

APPENDIX

GRAPHS AND MATRIX REPRESENTATION

The RNA secondary structure can be represented as a coarse-grained tree-graph, in which the graph nodes represent secondary structure motifs such as loops, bulges, and hairpins whereas the graph edges represent stems. More details can be found in Barash, 2003. A simple labeled tree-graph of RNA secondary structure is illustrated in Fig. A1. An equivalent way to represent tree-graphs is by a special matrix called the Laplacian. The Laplacian matrix corresponding to a graph is a symmetric matrix, with one row and column for each node on the graph. It is constructed as follows: in the diagonal of the matrix, the degree of the vertex (number of incident edges) is specified, while in the off-diagonals the value "–1" is inserted if there is an edge at that location, or "0" if there is no connecting edge. Note that the Laplacian matrix is symmetric, and its rows and columns add up to zero. The complete set of eigenvalues of the Laplacian matrix is called the spectrum of the graph and is independent of how graph vertices are labeled. The following properties characterize the Laplacian matrix eigenvalues: (A) The eigenvalues of the matrix are nonnegative, and the first eigenvalue is always zero; (B) The second smallest eigenvalue is a measure of the tree-graph compactness; (C) For a "star-shaped" tree-graph, the second smallest eigenvalue is unity. The eigenvalues indicate the shape of the tree-graph. If a change in the eigenvalue when comparing the tree-graph corresponding to the wildtype and the tree-graph corresponding to the mutant occurred, it means that the nucleotide substitution has a large effect on the secondary structure.

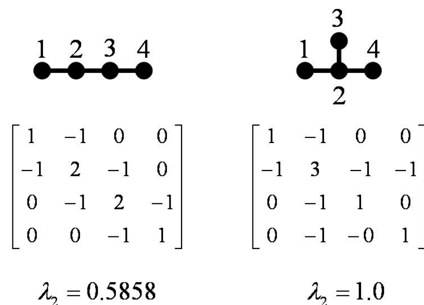

Fig. A1. A labeled tree-graph (as illustrated above with 4 nodes) can in turn be equivalently represented by a matrix called the Laplacian. The diagonal entries of this matrix contain the number of neighbors of each node (the degree of the graph), for example the value "3" in entry (2,2) of the matrix on the right corresponds to the three neighbors of node #2 in the labeled tree-graph above it. The off-diagonal entries are either "1" or "–1". The second eigenvalue of the Laplacian matrix (denoted by λ_2) is a number between 0.0 and 1.0, which indicates how compact the tree-graph is (for example, the tree-graph on the right is in its most compact form, called a "star", and hence its second eigenvalue is 1.0).

The eigenvalue does not tell anything about how the function of the molecule changes; it is only a floating point number between 0.0 and 1.0 that captures the compactness of the RNA secondary structure. The eigenvalue has a clear intuitive meaning: the closer it is to a zero, the sooner the tree-graph shape becomes spread and linear, whereas the closer it is to the number one, the sooner the tree-graph becomes more compact—until it reaches a star shape in which the eigenvalue is 1.0.