Title page

Title:

Solving the riddle of the evolution of Shine-Dalgarno based translation in chloroplasts

Authors:

Iddo Weiner[1,2]^, Noam Shahar[1]^, Pini Marcu[1], Iftach Yacoby[1]*, Tamir Tuller[2,3]*

* Corresponding authors ^ Contributed equally

Author Affiliations:

1. School of Plant Sciences and Food Security, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel.

2. Department of Biomedical Engineering, The Iby and Aladar Fleischman Faculty of Engineering, Tel Aviv University, Tel Aviv 6997801, Israel

3. The Sagol School of Neuroscience, Tel Aviv University, Tel Aviv 6997801, Israel

Author to whom correspondence should be addressed:

Prof. Tamir Tuller[2,3]: +97236405836, tamirtul@post.tau.ac.il

Dr. Iftach Yacoby[1]: +97236407891, iftachy@tau.ac.il

Key words

Shine-Dalgarno, Translation initiation, chloroplasts, cyanobacteria, plants, 16S rRNA

Abstract

The chloroplast, a photosynthetic organelle found in all plant and algae species, originates from an ancient event in which a cyanobacterium was engulfed by a larger eukaryote[1,2]. Thus, modern chloroplasts still harbor a bacterial-like genome and carry out all stages of gene expression, including mRNA translation by a 70S ribosome[3,4]. However, the Shine-Dalgarno model[5,6], which predominantly regulates translation initiation by base-pairing between the ribosomal RNA and the mRNA in model bacteria genera[7–9], was reported to have ambiguous effects on chloroplast gene expression[4,10,11]. Here we show that while the Shine-Dalgarno motif is clearly conserved in proteobacterial mRNAs, its general absence from chloroplast mRNAs is common to cyanobacteria as well, promoting the idea that the evolutionary process of reducing the centrality of the Shine-Dalgarno mechanism began well before chloroplast endosymbiosis. As plastid ribosomal RNA anti-Shine-Dalgarno elements are highly similar to their bacterial counterparts[4,12], these sites alone cannot explain the decline in plastid Shine-Dalgarno generality. However, by computational simulation we show that upstream point mutations render the ribosomal RNA local folding in chloroplasts, creating significantly tighter complexes around the anti-Shine-Dalgarno nucleotides, which in turn reduce the probability of ribosome binding via the Shine-Dalgarno mechanism. To validate our model, we expressed a mCherry gene harboring a Shine-Dalgarno motif in the *Chlamydomonas reinhardtii* chloroplast. We show that co-expressing it with a 16S ribosomal RNA, modified according to our model, significantly enhances its expression compared to co-expression with an endogenous 16S gene.

Main text

Chloroplasts are the intracellular plant organelles in which photosynthesis takes place. The origin of these organelles is an early event of endosymbiosis in which an ancient cyanobacterium was engulfed by a eukaryotic host[1,2]. As the cell adapted to this new symbiosis, the majority (~95%) of plastid genes have been horizontally transferred to the nucleus. However, modern chloroplasts have maintained a small circular genome consisting of 70S ribosomal RNAs (rRNA), transfer RNAs (tRNA), and open reading frames coding for ribosomal proteins, elongation factors, photosynthetic complexes and other housekeeping functions[3]. On average, roughly 100 coding sequences (CDSs) are translated within chloroplasts by a bacterial-like ribosome and a matching set of gene expression machinery[4].

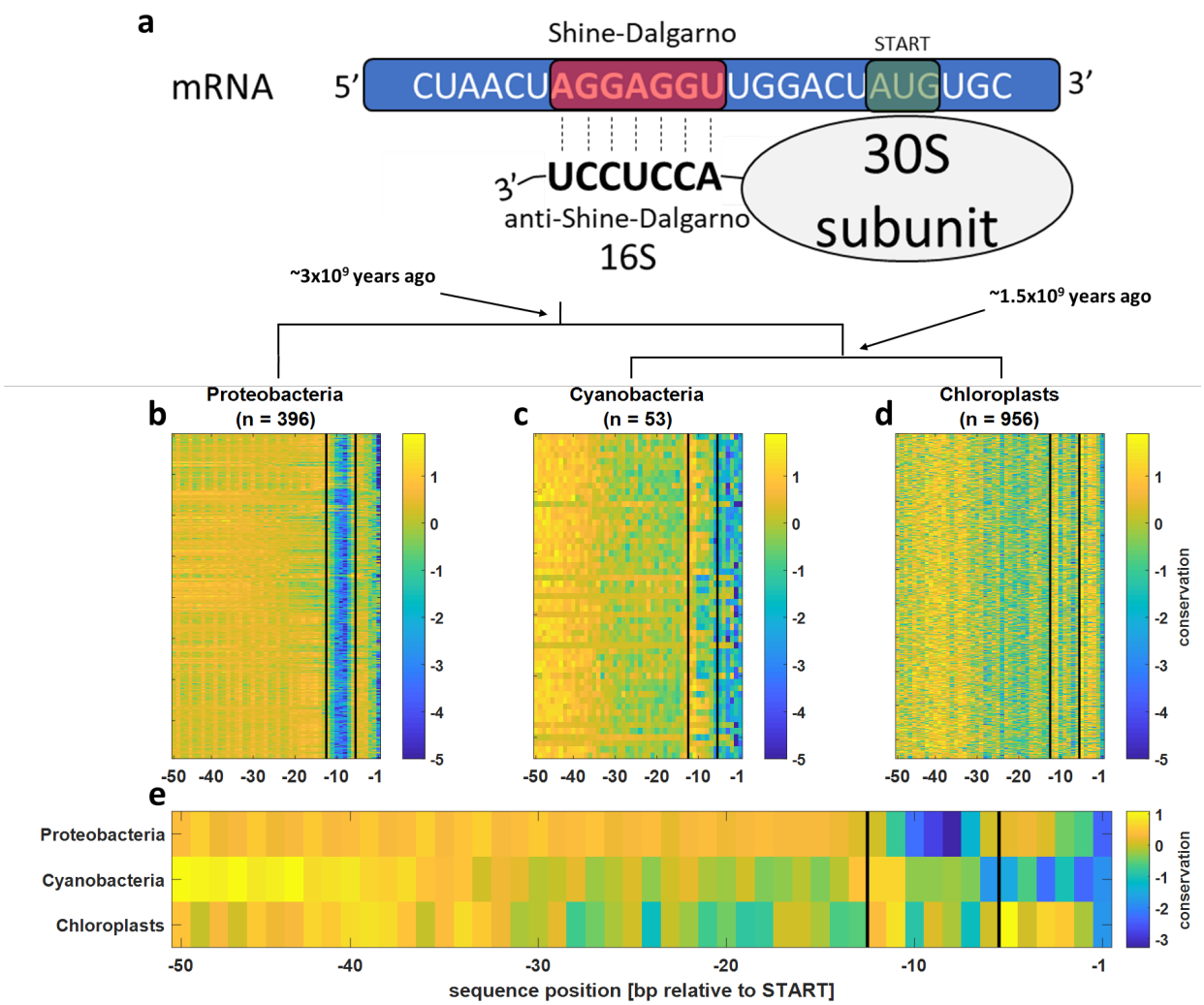
Bacterial gene expression is generally believed to have two canonical hallmarks; the first is short mRNA half-life which establishes transcription as rate-limiting for overall expression[13–17]. The second is the Shine-Dalgarno (SD) mechanism for translation initiation (TI), in which the 30S ribosomal subunit is recruited to the mRNA through base-pairing between a mRNA motif (SD sequence) found just upstream of the START site and an anti-SD (aSD) conserved sequence found at the 3' edge of the 16S rRNA[5,6]. In this TI model, the proximity of the binding site to the START codon (roughly 10 base-pairs upstream) is a key feature necessary for proper initiation[7–9,18] (Fig. 1a).

Since the availability of plastid sequences first began rising, ample research has attempted to examine the relevance of these bacterial models to the chloroplast genome. The common conclusion drawn from these works was that chloroplasts surprisingly differ from model bacteria in these aspects; chloroplast transcripts showed significantly longer half-lives (hours[19]) than their bacterial counterparts (minutes[15,20,21]), and unlike in bacteria – the protein abundance of plastid encoded genes was only mildly affected by changes in mRNA levels[22]. These observations led to the formulation of the concept that translation is the dominant rate-limiting step in chloroplast gene expression and that it is uncoupled from transcription[4,23], differentiating chloroplasts from model bacteria.

Trying to decipher the role of SD proved confusing as well; while the aSD seemed intact in all chloroplast 16S rRNA molecules, the effect of adding and removing a SD motif from different 5' untranslated regions (UTRs) proved unpredictable, thus quickly abolishing the idea that the canonical TI mechanism in bacteria applies neatly to chloroplasts[4,10,11]. The common conclusion from these observations was that since the ancient event of endosymbiosis, chloroplasts have evolved significantly and have shed their bacterial traits in favor of unique novel characteristics. To examine this hypothesis, we calculated the 5' UTR position-wise conservation for a representative from each sequenced chloroplast-containing genera, and compared the results to the equivalent cyanobacteria group of representatives. As an outgroup we used a similarly built database of proteobacteria – which are a large and diverse group containing key model organisms (*e.g. Escherichia coli, Heliobacter pylori, Pseudomonas aeruginosa etc.*) on which these bacterial gene expression concepts were learned. While proteobacteria show a clear SD signal centered roughly eight bases upstream from the START site (Fig. 1b), both chloroplasts and cyanobacteria show a noisier pattern with no clear signal in the SD area (Fig. 1c,d) , in accordance with previous reports[9,24–27]. As positioning plays a key role in the SD mechanism, this observation suggests that the TI mechanisms in the ancient

cyanobacterium engulfed during the formation of chloroplasts have most likely already diverged significantly from those of proteobacteria. This conclusion might apply to the more dominant translation regulation regime as well, as suggested in a recent report which found transcriptome dynamic similarities between cyanobacteria and chloroplasts[28].

Fig. 1: The Shine-Dalgarno motif does not show a clear signal in both Chloroplast and Cyanobacterial mRNAs.

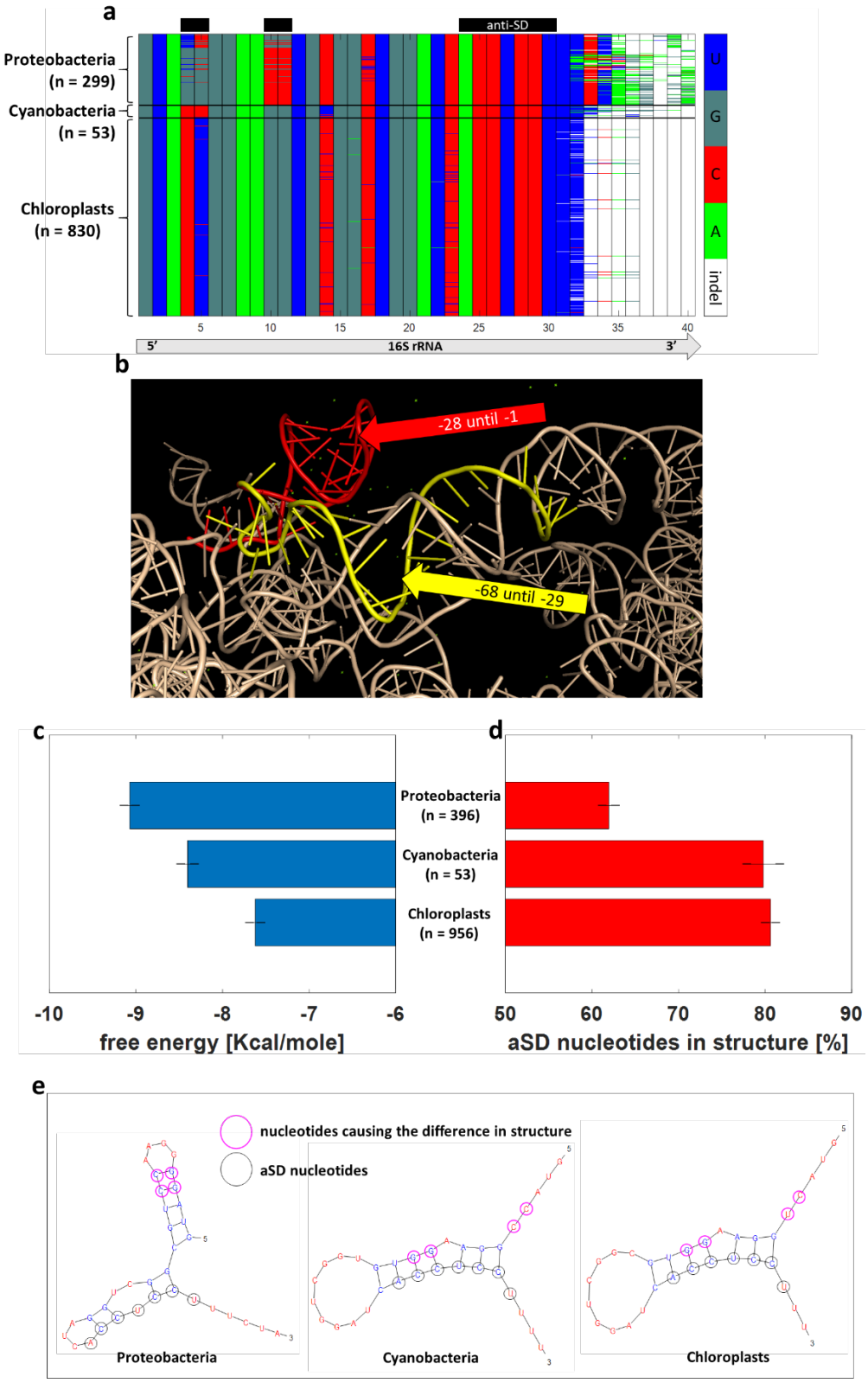


a, Schematic illustration of the Shine-Dalgarno mechanism. A conserved region near the 3' edge of the rRNA spontaneously binds a mRNA motif located slightly upstream of the START site by base-pairing. This interaction properly positions the small ribosomal subunit on the start codon to initiate translation. **a**,**b**,**c,** Sequence conservation along the 5' UTR. In each group the rows represent a single organism, selected to represent its genus. Lower values report on higher conservation in the specific position. The space between the horizontal black lines represents the canonical location of the Shine-Dalgarno motif on the mRNA. The dendogram depicts the

phylogenetic relations between the group. Node dating is based on phylogenetic reconstruction evaluations[2,29]. **d**, The mean value for each position across all rows in **b**, **c**, and **d**.

To investigate this dramatic decline in the presence of the SD motif in chloroplast mRNAs, we examined the 3' edges of all 16S rRNAs in our database. While the aSD is clearly conserved in chloroplasts (and cyanobacteria), they contain a typical pattern of point mutations, clustered in two main loci upstream of the aSD sequence (Fig. 2a). While these base substitutions do not directly affect the aSD sequence, we hypothesized that they might change the local rRNA secondary structure, and thus indirectly complicate the aSD:SD binding event. To examine this theory, we simulated the local folding of the 16S rRNA 3' edge in all organisms in the database. To define the relevant region for local folding simulation, we examined the 30S ribosomal subunit PDB structures from *Escherichia coli*[30] and the *Spinacia oleracea* chloroplast[31]. We chose the region from nucleotide -28 relative to the aSD sequence until the 3' edge, since unlike upstream regions which bind proteins and distant rRNA parts, in both organisms the RNA only interacts with itself in this area (Fig. 2b, Extended data Fig. 3). Interestingly, our folding simulations of this region show that the proteobacteria rRNA structures differ significantly from those of chloroplasts and cyanobacteria in several aspects. To broadly quantify these effects across our database, we computed the rRNA folding that minimizes free energy for each organism and counted the number of aSD nucleotides found in structure. Our results clearly show that these upstream point mutations have rendered the chloroplast and cyanobacterial rRNA structure more stable (Fig. 2c). Additionally, due to this structural change, a significantly larger portion of aSD nucleotides are paired in chloroplasts and cyanobacteria (Fig. 2d). Since ribosome binding in the SD mechanism requires base pairing between the mRNA SD motif and the rRNA aSD sequence, the occupation of the latter by other bonds reduces its affinity to the mRNA and lowers the probability of binding; thus explaining the less canonical role that SD plays in chloroplast TI. Examining the dominant structures received from each group reveals the loss of the well described bacterial hairpin loop[32] in cyanobacteria and chloroplasts, and clearly emphasizes the structural differences in the aSD complex between groups (Fig. 2e). Four point mutations clustered in two loci (Fig. 2a, positions 4-5, 10-11) are the main drivers of this conformational change. Other point mutations and the slightly longer proteobacterial tails (Fig. 2a) have negligible effects on this observation (Fig. 2e).

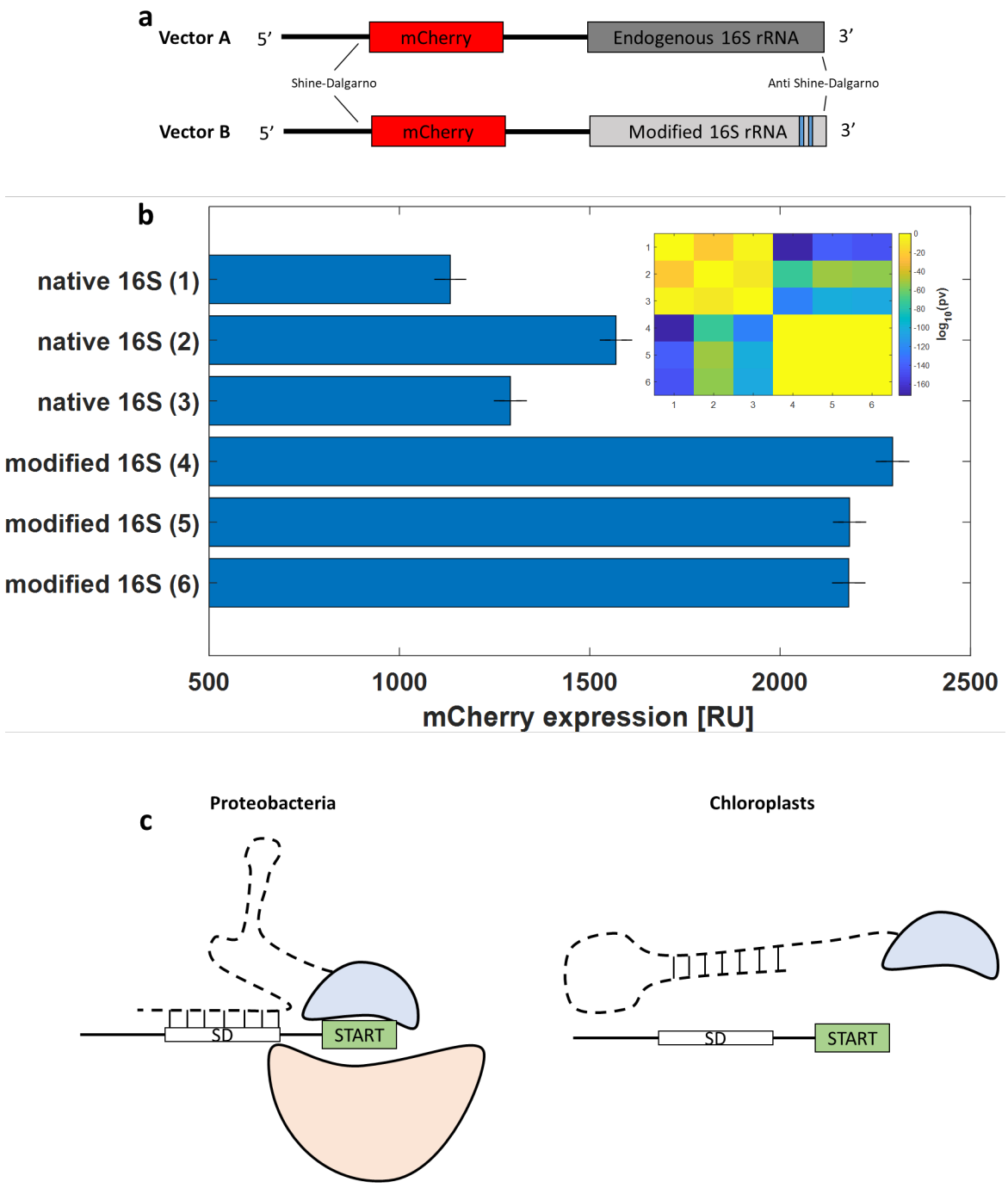
Fig. 2: Local folding at the 16S 3' ribosomal RNA edge affects the aSD structure



a, Multiple sequence alignment of the 16S rRNA 3' edge database, composed of a single representative species from each genus. The black boxes above the matrix highlight important regions: the anti-SD box represents the conserved region where this sequence is found, whereas empty boxes represent columns with high variability between the different groups. Several outlying species which introduce indels in the body of the alignment were excluded for visualization purposes. **b,** Examination of the 3' edge of the 16S rRNA in the 30S ribosomal subunit from the chloroplast of *S. oleracea*. The reference point is the anti-Shine-Dalgarno sequence (*e.g.* -1 means one nucleotide upstream from the aSD sequence). The 3' edge of the molecule (red) only interacts with itself, while the region immediately upstream (yellow) interacts with distant parts of the rRNA. c, Free energy of the 16S rRNA secondary structure. **d,** The percent of nucleotides within the anti-SD sequence found in base-pairing interactions. **c,d,** Data shown is mean ± SE. **e,** Typical rRNA secondary structure. Anti-SD nucleotides are marked in dark green whereas the nucleotides that cause the structure discrepancy between groups are circled in pink.

According to this model, simply changing these two clusters of altered nucleotides into their proteobacterial form will modify the local folding at the 3' edge of the 16S rRNA, and expose the aSD sequence to facilitate easier binding to SD motifs. To test this theory, we designed two plasmids for chloroplast expression in the model green alga *Chlamydomonas reinhardtii*. Both coded for a codon optimized mCherry reporter protein with a clear SD motif properly situated upstream from the START codon. Plasmid A also coded the natural *C. reinhardtii* 16S rRNA gene, whereas plasmid B coded for a slightly modified version of this 16S gene in which the two aforementioned nucleotide clusters were mutated to match the canonical proteobacterial sequence (Fig. 3a). We expressed each of these plasmids in the *C. reinhardtii* chloroplast and brought three separate clones from each group to homoplasmicity[33]. Subsequently, we measured mCherry expression for each of these clones by flow cytometry, in which 10^4 single cells from each mid-log phase culture were examined. We observed that our reporter gene was significantly more highly expressed in the clones transformed with plasmid B (Fig. 3b), thus confirming that point mutations upstream of the aSD enhance the expression of a chloroplast gene with a SD motif. Alongside our computational analysis, this observation suggests that loose aSD structures are important for spontaneous aSD:SD base-pairing (Fig. 3c).

Fig. 3: Introducing a modified 16S rRNA into the chloroplast genome enhances the expression of a reporter gene with a SD motif.



a, Schematic diagram of the chloroplast expression vectors. Vectors A and B were completely identical, except for four base substitutions centered in two clusters at the end of 3' edge of the 16S gene in vector B. **b**, mCherry expression in three individual clones from each transformation group, as measured by flow cytometry. Data shown is mean ± SE. P-values in logarithmic scale are shown on the upper right. All p-values were computed using the two-

tailed Wilcoxon rank-sum test. **c**, Schematic illustration of the model representing the altered role of Shine-Dalgarno in chloroplast translation initiation; while bacterial 16S rRNAs are open and ready to base-pair with the mRNA, chloroplast 16S 3' edges have gained several mutations during evolution, thus strengthening the structure of the aSD sequence and hampering spontaneous interactions with the mRNA.

Recently, it has been shown that SD-mediated ribosome binding still occurs in the tobacco chloroplast[34]. However, unlike in proteobacteria where SD is the canonical TI model, Scharff et al. have shown that it is a dominant TI mechanism for only a specific subset of genes. Importantly, this work shows that genes in which SD is essential for TI also strongly tend to have strong mRNA secondary structures in the vicinity of their START site. These observations correlate nicely with our findings in this work; we have shown that while the aSD is intact, the local folding of the 16S rRNA 3' edge lowers its binding affinity. Thus, mRNAs that are still under SD regulation would require a certain 5' UTR secondary structure to unfold the 16S edge and reveal the aSD sequence.

It has been previously shown that even *E. coli* mRNAs can undergo translation without having a SD signal[11], and alternative TI mechanisms are known to exist[27,35]. Some of these mechanisms are believed to have become more dominant in chloroplasts[11], at the expense of the SD mechanism. However, all the machinery required for proper SD-mediated ribosome binding remained intact in plastids (Fig. 3), and still play a role in regulating chloroplast TI[34]. While chloroplast gene expression regulation has clearly evolved significantly since endosymbiosis, the SD role in cyanobacteria TI is most likely similar to that presented for chloroplasts. Although SD has been reported to initiate translation of specific genes in cyanobacteria[36], it does not leave a strong global trace on its mRNAs (Fig. 1, Extended data Fig. 1,2) and the 3' edge of its 16S rRNA folds to conceal the aSD nucleotides, as in chloroplasts (Fig. 2). Since this phenomenon is common to both groups, symbiosis with the nucleus (which applies to chloroplasts alone) might not be a main driver towards the reduced role of SD in chloroplasts. Alternatively, these drivers could include adaptation to oxygenic photosynthesis or to slower growth rates[9,37].

Finally, our work highlights a generalist method for enhancing heterologous expression in chloroplasts. Our baseline plasmid in this work, pLM21, is known to drive high chloroplast expression in *C. reinhardtii*[33,38,39], yet by adding a slightly modified copy of the 16S rRNA (Fig. 3a), we were able to roughly double the amount of protein achieved (Fig. 3b). As the aSD sequence is concealed in nearly all chloroplasts (Fig. 2), we expect that adding a modified 16S rRNA to any plastid expression vector would similarly enhance the translation of a target transcript harboring a SD motif in any chloroplast.

Methods

Data Base Construction

To build the database used for the analyses described in this work, we retrieved the accessions for all organisms in NCBI (January, 2018) that are classified as either proteobacteria, cyanobacteria or chloroplasts. The full list of accessions is given at: https://github.com/iddoweiner/Solving-the-riddle-of-the-evolution-of-Shine-Dalgarno-based-translation-in-chloroplasts. Within each group, we randomly selected one representative for each genus to appear in the final database. In Fig. 2a alone, several organisms were removed from the alignment as they introduced sporadic indels into the alignment matrix and interfered with data visualization.

Conservation (Entropy Z-score)

To compute the organism-specific position conservation scores presented in Fig. 1 we initially obtained the 50 base-pairs (bp) upstream from the start site in all the CDSs (Extended data Fig. 1) and computed the position specific scoring matrix. For each matrix column we computed the Shannon Entropy by:

$$S_{column} = -\sum_{i=1}^{4} p_i \times log_2\,(p_i)$$

More specifically, given the frequencies p(A), p(C), p(G), p(U), the entropy S of a column (position) is:

$$-p(A)\cdot\log_2(p(A)) -p(C)\cdot\log_2(p(C)) -p(G)\cdot\log_2(p(G)) -p(U)\cdot\log_2(p(U))$$

We computed the same entropy score for 100 different random models of the same matrix, in which the columns were permutated together to maintain the alignment. The final value shown in Fig. 1 is the margin between the entropy computed on the native 5' UTRs and the average of those computed for the randomized UTRs, divided by the standard deviation of the random model distribution (Z-score).

16S rRNA local folding simulation

To simulate the local folding at the 3' edge of 16S rRNAs, we first examined the 30S ribosomal subunit PDB structures from *Escherichia coli*[30] and the *Spinacia oleracea* chloroplast[31]. While using the conserved aSD sequence as a reference point, we observed that from roughly nucleotide -28 until the 3' edge, base pairing is scarce and importantly – only interacts within itself (Fig. 2b, Extended data Fig. 3). Regions slightly upstream from this point interact with distant self-regions (*e.g.* with regions located 500 bp upstream on the 16S), ribosomal proteins and other rRNAs. Thus, we selected this area for our simulations. The folding structure was computed using the ViennaRNA package[40].

Shine–Dalgarno consensus sequence

While there is some debate whether the exact consensus SD sequence is AGGAGGU or just AGGAGG (thus the aSD is either ACCUCCU or CCUCCU, respectively), previous studies lean towards the former[41,42]. The high level of

conservation of the A in position 24 (Fig. 2a) could strengthen this notion. While we have indeed used the 7-base consensus sequence in this work, the only result that depends upon the exact consensus sequence is 3b. A similar analysis based upon the 6-base consensus sequence is shown in (Extended data Fig. 4), where the conclusion remains the same.

Algal cultures and chloroplast transformation

Chlamydomonas reinhardtii wild-type (D66) strains were grown in Tris-Acetate-Phosphate medium at 25°C under continuous cool daylight and cool white fluorescent lights (90 µE m^{-2} s^{-1}) stirring in 100 mL Erlenmeyers capped with silicone sponge enclosures, as previously described[43,44]. Vectors A and B (Fig. 3a) were cloned into the pLM21[38] and transformed into the *C. reinhardtii* wild-type chloroplast using a biolistic gun. Several rounds of PCR were used to isolate homoplasmic clones[33].

Flow Cytometery

The mCherry fluorescence was quantified from a 200 µL sample in mid-log phase (10^6 cells mL^{-1}) from each positive clone, using the S1000EXI Benchtop Flow Cytometer (https://stratedigm.com/s1000exi-flow-cytometer). The mCherry parameters were selected according to the manufacturer's instructions. The natural fluorescence of the *C. reinhardtii* wild-type was measured in the same experiment and subtracted from the clones' intensity.

Code availability

The code generated during the current study is available from the corresponding author on reasonable request.

Additional remarks

Funding:

This study was supported in part by:

- The Israeli Ministry of Science, Technology and Space
- Israel Science foundation 1646/16
- NSF-BSF Energy for Sustainability 2016666
- A fellowship from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University
- A fellowship from the Manna center for Plant Biosciences
- A fellowship from the Rieger foundation for Environmental studies

Acknowledgements

We would like to thank:

- The G. Peltier lab for kindly sharing their pLM21 vector
- Mrs. Barel Weiner for reviewing figure esthetics
- Mrs. Orit Sagi-Assif for Flow Cytometry support

Declaration of author contributions

IW, IY and TT designed the research; IW and NS performed the experimental procedures; IW and NS performed the computational procedures; PM analyzed the PDB ribosome structures; IW, NS, IY and TT wrote the paper.

Declaration of interest

All authors declare no conflict of interest

Statement of Informed Consent

No conflicts, informed consent, human or animal rights applicable

declaration of authors agreement to authorship

All authors declare agreement to authorship and submission of the manuscript for peer review

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Accession numbers are provided at: https://github.com/iddoweiner/Solving-the-riddle-of-the-evolution-of-Shine-Dalgarno-based-translation-in-chloroplasts

Abbreviations

Transfer RNA	tRNA
Ribosomal RNA	rRNA
Shine-Dalgarno	SD
Anti-Shine-Dalgarno	aSD
Translation initiation	TI
Coding sequence	CDS
Untranslated region	UTR
Base-pairs	bp

References

1. Gray, M. W. Origin and evolution of organelle genomes. *Curr. Opin. Genet. Dev.* **3,** 884–90 (1993).

2. Reyes-Prieto, A., Weber, A. P. M. & Bhattacharya, D. The Origin and Establishment of the Plastid in Algae and Plants. *Annu. Rev. Genet.* **41,** 147–168 (2007).

3. Wakasugi, T., Tsudzuki, T. & Sugiura, M. The genomics of land plant chloroplasts: Gene content and alteration of genomic information by RNA editing. *Photosynth. Res.* **70,** 107–118 (2001).

4. Peled-zehavi, H. & Danon, A. Translation and translational regulation in chloroplasts. *Top. Curr. Genet.* **19,** 249–281 (2007).

5. Shine, J. & Dalgarno, L. The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proc. Natl. Acad. Sci.* **71,** 1342 LP-1346 (1974).

6. Shine, J. & Dalgarno, L. Determinant of cistron specificity in bacterial ribosomes. *Nature* **254,** 34–38 (1975).

7. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. **27,** (2009).

8. Jacob, W. F., Santert, M. & Dahlberg, A. E. A single base change in the Shine-Dalgarno region of 16S rRNA of Escherichia coli affects translation of many proteins. **84,** 4757–4761 (1987).

9. Ma, J., Campbell, A. & Karlin, S. Correlations between Shine-Dalgarno Sequences and Gene Features and Operon Structures. **184,** 5733–5745 (2002).

10. Sakamoto, W., Xuemei, C., Karen, K. & David, S. Function of the Chlamydomonas reinhardtii petD 5' untranslated region in regulating the accumulation of subunit IV of the cytochrome b6/f complex.

11. Fargo, D. C., Zhang, M., Gillham, N. W. & Boynton, J. E. Shine-Dalgarno-like sequences are not required for translation of chloroplast mRNAs in Chlamydomonas reinhardtii chloroplasts or in Escherichia coli. *Mol. Gen. Genet.* **257,** 271–282 (1998).

12. Zoschke, R. & Bock, R. Chloroplast Translation: Structural and Functional Organization, Operational Control and Regulation. *Plant Cell* (2018).

13. Gowrishankar, J. & Harinarayanan, R. MicroReview Why is transcription coupled to translation in bacteria ? **54,** 598–603 (2004).

14. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. Real-Time Kinetics of Gene Activity in Individual Bacteria. 1025–1036 (2005). doi:10.1016/j.cell.2005.09.031

15. Selinger, D. W., Saxena, R. M., Cheung, K. J., Church, G. M. & Rosenow, C. Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation. 216–223 (2003). doi:10.1101/gr.912603.

16. Reppas, N. B., Wade, J. T., Church, G. M. & Struhl, K. The Transition between Transcriptional Initiation and Elongation in E . coli Is Highly Variable and Often Rate Limiting. 747–757 (2006). doi:10.1016/j.molcel.2006.10.030

17. Hijum, S. A. F. T. Van, Medema, M. H. & Kuipers, O. P. Mechanisms and Evolution of Control Logic in Prokaryotic Transcriptional Regulation. **73,** 481–509 (2009).

18. Shaham, G. & Tuller, T. Genome scale analysis of Escherichia coli with a comprehensive prokaryotic sequence-based biophysical model of translation initiation and elongation. *DNA Res.* **25,** 195—205 (2018).

19. Kim, M., Christopher, D. A. & Mullet, J. E. Direct evidence for selective modulation of psbA , rpoA , rbcL and 16S RNA stability during barley chloroplast development. 447–463 (1993).

20. Dierstein, R. Synthesis of pigment-binding protein in toluene-treated Rhodopseudomonas capsulata and in cell-free systems. **518,** 509–518 (1984).

21. Brawerman, G. Determinants of Messenger RNA Stability Minireview. **48,** (1987).

22. Drapier, D. *et al.* The chloroplast atpA gene cluster in Chlamydomonas reinhardtii. Functional analysis of a polycistronic transcription unit. *Plant Physiol.* **117,** 629–641 (1998).

23. Manuell, a, Beligni, M. V, Yamaguchi, K. & Mayfield, S. P. Regulation of chloroplast translation: interactions of RNA elements, RNA-binding proteins and the plastid ribosome. *Biochem. Soc. Trans.* **32,** 601–5 (2004).

24. Scharff, L. B., Childs, L., Walther, D. & Bock, R. Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-Binding Sites. **7,** (2011).

25. Starmer, J., Stomp, A., Vouk, M. & Bitzer, D. Predicting Shine – Dalgarno Sequence Locations Exposes Genome Annotation Errors. **2,** (2006).

26. Voigt, K. *et al.* Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity. 2056–2068 (2014). doi:10.1038/ismej.2014.57

27. Nakagawa, S., Niimura, Y., Miura, K. & Gojobori, T. Dynamic evolution of translation initiation mechanisms in prokaryotes. *Proc. Natl. Acad. Sci.* **107,** 6382 LP-6387 (2010).

28. Shi, C. *et al.* Full transcription of the chloroplast genome in photosynthetic eukaryotes. *Sci. Rep.* **6,** 1–10 (2016).

29. Schirrmeister, B. E., Vos, J. M. De, Antonelli, A. & Bagheri, H. C. Evolution of multicellularity coincided with increased diversi fi cation of cyanobacteria and the Great Oxidation Event. **110,** (2013).

30. Svidritskiy, E., Demo, G. & Korostelev, A. A. Mechanism of premature translation termination on a sense codon. (2018). doi:10.1074/jbc.AW118.003232

31. Boerema, A. P. *et al.* chl-RRF and hibernation-promoting factor. **4,** (2018).

32. Harry F. Noller and Carl R. Woese. Secondary Structure of 16S Ribosomal RNA. *Science (80-.).* **212,** 403–411 (1981).

33. Weiner I, Shahar N, Feldman Y, Landman S, Milrad Y, Ben-Zvi O, Avitan M, Dafni E, Schweitzer S, Eilenberg H, Atar S, Diament A, Tuller T, Y. I. Overcoming the expression barrier of the ferredoxin-hydrogenase chimera in Chlamydomonas reinhardtii supports a linear increment in photosynthetic hydrogen output. *Algal Res.* **33,** 310–315 (2018).

34. Scharff, L. B. *et al.* Shine-Dalgarno Sequences Play an Essential Role in the Translation of Plastid

mRNAs in Tobacco. *Plant Cell* tpc.00524.2017 (2017). doi:10.1105/tpc.17.00524

35. Moll, I., Grill, S., Gualerzi, C. O. & Bläsi, U. Leaderless mRNAs in bacteria : surprises in ribosomal recruitment and translational control. **43,** 239–246 (2002).

36. Asada, Y. & Koike, Y. Heterologous expression of clostridial hydrogenase in the cyanobacterium Synechococcus PCC7942. **1490,** 269–278 (2000).

37. Karlin, S., Mrázek, J., Campbell, A. & Kaiser, D. Characterizations of Highly Expressed Genes of Four Fast-Growing Bacteria. **183,** 5025–5040 (2001).

38. Tibiletti, T., Auroy, P., Peltier, G. & Caffarri, S. Chlamydomonas reinhardtii PsbS protein is functional and accumulates rapidly and transiently under high light. *Plant Physiol.* **171,** 2717–2730 (2016).

39. Sawyer, A., Bai, Y., Lu, Y., Hemschemeier, A. & Happe, T. Compartmentalisation of [FeFe]-hydrogenase maturation in Chlamydomonas reinhardtii. *Plant J.* **90,** 1134–1143 (2017).

40. Package, V. *et al.* ViennaRNA Package 2.0. 1–14 (2011).

41. Schneider, T. D. & Stephens, R. M. Sequence logos: A new way to display consensus sequences. *Nucleic Acids Res.* **18,** 6097–6100 (1990).

42. Tompa, M. An exact method for finding short motifs in sequences, with application to the ribosome binding site problem. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 262–271 (1999).

43. Weiner, I. *et al.* Enhancing heterologous expression in Chlamydomonas reinhardtii by transcript sequence optimization. *Plant J.* **94,** (2018).

44. Eilenberg, H. *et al.* The dual effect of a ferredoxin-hydrogenase fusion protein in vivo: successful divergence of the photosynthetic electron flux towards hydrogen production and elevated oxygen tolerance. *Biotechnol. Biofuels* **9,** 182 (2016).