Published online 6 March 2013 Nucleic Acids Research, 2013, Vol. 41, No. 9 e98
doi:10.1093/nar/gkt151

Spanning high-dimensional expression space using ribosome-binding site combinatorics

Lior Zelcbuch[1], Niv Antonovsky[1], Arren Bar-Even[1], Ayelet Levin-Karp[1], Uri Barenholz[1], Michal Dayagi[1], Wolfram Liebermeister[1,2], Avi Flamholz[1], Elad Noor[1], Shira Amram[1], Alexander Brandis[1], Tasneem Bareia[1], Ido Yofe[1], Halim Jubran[3] and Ron Milo[1,*]

[1]Department of Plant Sciences, Weizmann Institute of Science, Rehovot 76100, Israel,
[2]Charité - Universitätsmedizin Berlin, Institut für Biochemie, Charitéplatz 1, 10117 Berlin, Germany and
[3]Department of Genetics, The Alexander Silberman Life Science Institute, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Received August 7, 2012; Revised February 7, 2013; Accepted February 14, 2013

Downloaded from http://nar.oxfordjournals.org/ at Weizmann Institute of Science on July 3, 2016

ABSTRACT

Protein levels are a dominant factor shaping natural and synthetic biological systems. Although proper functioning of metabolic pathways relies on precise control of enzyme levels, the experimental ability to balance the levels of many genes in parallel is a major outstanding challenge. Here, we introduce a rapid and modular method to span the expression space of several proteins in parallel. By combinatorially pairing genes with a compact set of ribosome-binding sites, we modulate protein abundance by several orders of magnitude. We demonstrate our strategy by using a synthetic operon containing fluorescent proteins to span a 3D color space. Using the same approach, we modulate a recombinant carotenoid biosynthesis pathway in *Escherichia coli* to reveal a diversity of phenotypes, each characterized by a distinct carotenoid accumulation profile. In a single combinatorial assembly, we achieve a yield of the industrially valuable compound astaxanthin 4-fold higher than previously reported. The methodology presented here provides an efficient tool for exploring a high-dimensional expression space to locate desirable phenotypes.

INTRODUCTION

Native protein abundance in bacteria spans over four orders of magnitude, from a handful of molecules per cell to tens of thousands (1). The intricate balance of protein expression levels is at the heart of proper functioning of biological systems (2) and is often critical for the efficient activity of metabolic pathways (3). In contrast to native biological systems, where the balancing of protein levels is selected for during evolution (4), the expression of a synthetic system can lead to imbalances in protein concentrations (5). As a result, non-native metabolic pathways rarely function optimally when first introduced, and the enzyme levels must be fine-tuned (6). The ability to achieve a balanced expression of multiple genes is often critical for metabolic engineering efforts, where an optimization process is needed to increase the productivity of non-natural biosynthetic pathways towards industrial-scale biochemical production.

Figure 1 schematically depicts the major challenges associated with imbalanced enzyme concentrations, based on a two enzyme metabolic pathway model (Supplementary Results). Low enzyme expression can limit the pathway flux and, therefore, product synthesis rate (blue region). At the other extreme, excessive expression will lead to protein burden (7), resulting in the depletion of cellular resources that limit growth (purple region). Finally, imbalances between enzymes producing and consuming an intermediate metabolite can result in a metabolic bottleneck and a high concentration of potentially toxic pathway intermediates (6) (green region). Only a small region of enzyme expression space is considered to be balanced regarding all three criteria.

How can one arrive at this balanced combination *in vivo*? In the vast majority of cases, it is infeasible to *a priori* determine the expression level required for an optimal activity of a system. Thus, it is vital to develop a method to scan the protein expression space. Several experimental approaches for controlling the intracellular abundance of proteins exist, such as altering the promoter (8) and the ribosome-binding site (RBS) (9,10) sequences, modulating the stability of transcripts (11,12) or varying

*To whom correspondence should be addressed. Tel: +972 8 9344540; Fax: +972 8 9344540; Email: Ron.Milo@weizmann.ac.il

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2013. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

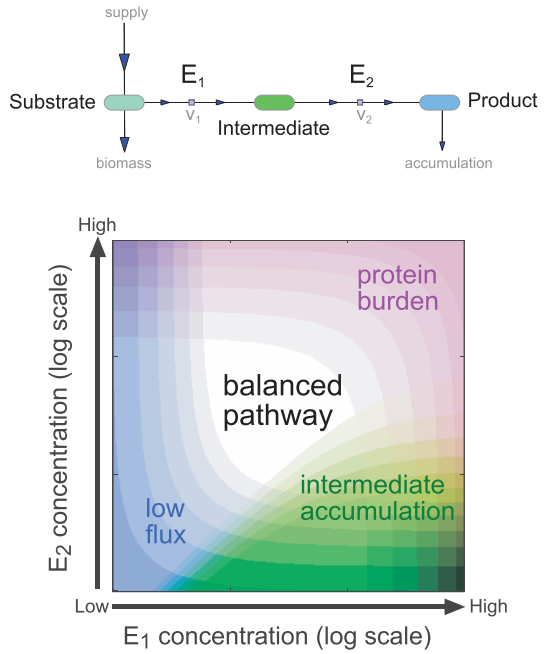

Figure 1. Modulation of enzyme expression levels is required for balanced pathway function. We used a simple quantitative model based on reversible Michaelis–Menten kinetics to depict the outcomes of an unbalanced enzyme expression (see Supplementary Results and Supplementary Figure S13 and S14 for full details). Considering two enzymes in a multi-step metabolic pathway, only a small region of the enzyme expression space (shown in white) sustains optimal production.

the degradation rate of the mature protein (13). Yet, how can modulation of protein expression be systematically used to span a high-dimensional space?

Here, we present a method that uses the relatively short RBS sequence in a combinatorial manner to explore protein expression space to locate desired phenotypes. Previous studies, most notably by Salis *et al.*, demonstrated that changes in the RBS sequence can affect protein expression levels by several orders of magnitude (9). Rather than randomly mutating the RBS sequence, an approach that often requires the screening of an extremely large number of mutants, most of which showing knocked down expression, we use a small set of RBS sequences to modulate in parallel the protein expression levels of multiple genes over several orders of magnitude. Using this approach, we are able to efficiently scan a large fraction of the expression space with a manageable set of tested genotypes. Our methodology involves combinatorially pairing target genes with a small set of RBS sequences and assembling them into a library of synthetic operons. In this manuscript, we demonstrate that such an approach is a fast and efficient way to explore a multi-dimensional expression space and achieve diverse phenotypes.

MATERIALS AND METHODS

Bacterial strains and reagents

The bacterial strain used for cloning and constructs assembly was *Escherichia coli DH5α* unless stated otherwise. All primers were synthesized by Sigma Aldrich, Israel. Polymerase chain reaction reactions were performed using Phusion polymerase (Finnzymes, Finland). Restriction enzymes were purchased from New England BioLabs (Beverly, MA, USA) unless stated otherwise.

Iterative operon assembly process

For the construction of RBS-modulated synthetic operons, we used an approach that adheres to the principles described in 'Idempotent Vector Design for Standard Assembly of Biobrick', also known as the BioBrick (14) standard. To facilitate the assembly process, we first concatenate a chloramphenicol resistance marker (CmR) to each of the DNA sequences designated for assembly. The CmR cassette is paired to the DNA sequence using polymerase chain reaction overlap extension before the assembly process. Next, the target DNA is assembled into a vector using a standard restriction–ligation process. On transformation, only clones that contain properly assembled constructs are able to form colonies on agar plates supplemented with Cm. As the resistance cassette is flanked by restriction sites, it can be easily removed when preparing the vector for the next assembly cycle. In this manner, it is possible to perform multiple assembly rounds while using a single resistance marker (Supplementary Methods and Supplementary Figure S1).

Combinatorial assembly of RBS-modulated operons

For any coding sequence designated for assembly, we first sub-cloned the coding sequence into a linearized vector mixture where each vector contains a distinct RBS sequence upstream to the cloning site (Supplementary Methods and Supplementary Figures S2–S5). To assemble a combinatorial library of operons, where each variant contains the same order of genes but with a different combination of RBS sequences, we construct in parallel the RBS-modulated mixture for each of the desired genes. We then perform iterative assembly steps, where at each step an additional RBS-modulated coding sequence is added along the operon, as described earlier in the text. Therefore, by repeating this process for N rounds, where in each round an additional RBS-modulated coding sequence is added to the combinatorial operon library, we sequentially assemble a combinatorial mixture of plasmids. Where each plasmid contains the same N coding sequences in a pre-defined order but which are driven by a varying combination of the six RBS sequences (Supplementary Figures S6–S10).

Fluorescence measurements of reporter proteins

E. coli K12 MG1655 cells transformed with a tri-color reporter operon (Supplementary Figure S11) were grown in M9 media supplemented with 0.2% glucose and chloramphenicol (34 μg/ml). After overnight incubation, cells were diluted (1:2500) in a 96-well plate containing the same media and incubated in an automated robotic platform (Evoware II, Tecan). Every 15 min the plate was transferred by a robotic arm into a multi-well

Downloaded from http://nar.oxfordjournals.org/ at Weizmann Institute of Science on July 3, 2016

fluorometer (Infinite M200-pro, Tecan). In each measurement, OD was measured at 600 nm, mCherry was measured by excitation at 587 nm and emission measurement at 620 nm and yellow fluorescent protein (YFP) was measured by excitation at 520 nm and emission measurement at 555 nm. We used the fluorescent protein accumulation rate as a measurement for RBS strength, a detailed description of the algorithm used for data analysis can be found in section 5.4.2 of the Supplementary Methods. For flow cytometry measurements, cells were grown in M9 media supplemented with 0.2% glucose until mid-exponential phase (OD $\approx$ 0.3). Fluorescence was quantified using BD LSR II Flow Cytometer. A blue laser (488 nm) and a 530 ± 30 nm emission filter were used to measure YFP fluorescence, and a yellow laser (560 nm) and a 610 ± 20 nm emission filter were used to measure mCherry fluorescence. In all, $\approx$100 000 cells were recorded in each flow cytometry experiment (Supplementary Methods and Supplementary Figure S6).

Microscopy and fluorescence imaging

Fluorescence images of colonies expressing RBS-modulated variants of the tri-color reporter operon were taken using a Nikon ECLIPSE E800 microscope equipped with a Nikon Intensilight (C-HGFIE) for illumination. Chroma filter cubes were used to image fluorescence proteins: mCherry (excitation filter 530–560 nm, emission filter 590–650 nm, 30 ms exposure), cyan fluorescent protein (mCFP: excitation filter 426–446 nm, emission filter 460–500 nm, 60 ms exposure) and YFP (mYFP: excitation filter 490–510 nm, emission filter 520–550 nm, 800 ms exposure). Images were captured with a Nikon DS-5M-L1 digital Sight Camera System using the NIS-Elements BR3.22 software. Different channels were overlaid to give the figures shown. Images of colonies appearing in Figure 5B were taken using a binocular microscope (WILD M8; Heerbrugg, Switzerland) with Schott Ace Fiber Optic Light Source 150W Microscope Illuminator. Images were captured using a Nikon Digital Sight Camera System. Stitching of adjacent fields was done using AutoStitch software (http://www.cs.bath.ac.uk/brown/autostitch/autostitch.html).

Carotenoid extraction and analysis

To screen the combinatorial library (Supplementary Figure S12) for astaxanthin producing variants, we visually inspected ~25 000 colonies representing ~10% of the total number of the 6^7 possible RBS combinations (seven genes in the carotenoid biosynthesis pathway that are modulated, each with six distinct RBS sequences). We re-streaked 500 colonies whose color resembled that of astaxanthin. Out of this sub-group, we chose 50 clones that exhibited the most intense color for thin-layer chromatography carotenoid profile analysis. The thin-layer chromatography allowed examining the carotenoid profile and estimating astaxanthin accumulation levels. The highest astaxanthin producing colonies were selected for further analysis using high-performance liquid chromatography (HPLC). Selected clones were grown in shake flasks containing 100 ml of LB medium. After 48 h incubation at 37°C, 20-ml samples were taken from the culture, and cells were harvested by centrifugation. Cell pellet was washed with cold water, and carotenoids were extracted by vigorous shaking with 1:1 volume of acetone. Insoluble components of the extract were removed by centrifugation (15 000*g*), and supernatant was transferred into a glass round-bottom flask for evaporation in a rotary evaporator. Dried extract was re-solvated in 1.5 ml of acetone, and 50 µl of samples was injected for HPLC analysis. HPLC analysis was performed using Jasco platform with high pressure mixing installed with a Borwin software, P4987 pumps and an MD-915 photodiode array detector. Samples were analyzed by injecting 50 µl on a YMC pack ODS-A column (250 × 4.6 mm, 5 µm, 12 nm). Solvent A: 75% aqueous methanol, Solvent B: ethylacetate. Solvent flow rate of 0.6 ml/min was used with the following gradient: 15–85% of B (0–24 min), 85% (24–30 min), 85–15% (30–34 min) and 15% (34–40 min). The spectra of the eluted carotenoids were recorded online with the photodiode array detector (300–900 nm). Carotenoid compounds were identified by co-chromatography with authentic standard compounds and by analysis of their UV-Vis spectra. For the quantification of the carotenoid compounds, the integrated peak areas were compared with those of authentic standards. The concentration of the standard solutions was determined spectrophotometrically (15) (Jasco V-570 instrument). For additional identification, HPLC peaks were collected and directly injected into a mass-spectrometer (Micromass Quattro Ultima tandem quadrupole instrument equipped with a Z-spray ESI interface and Waters Masslynx v4.1 software). The corresponding masses were analyzed from obtained full-scan [ESI(+), m/z 100–1000] mass spectra (16). To measure dry cell weight, 20 ml from cell culture was taken for each sample. Cells were centrifuged, and cell pellet was transferred into pre-weighted tubes. Cell pellet was lyophilized for 24 h before dry cell weight measurement.

RESULTS

RBS combinatorics is an efficient method to explore a high-dimensional expression space

As RBS expression modulators we selected six sequences that were previously demonstrated by Salis *et al.* to span several orders of magnitude in protein expression levels (9) (Figure 2A). These synthetic DNA sequences are composed of three distinct functional parts: (i) a spacer sequence (5′-UTR), (ii) RBS core, which can also be referred to as a modified Shine–Dalgarno sequence, and (iii) the ATG starting codon followed by 6His-tag encoded in the N-terminus of the protein. Distinct RBS expression modulators contain a modified Shine–Dalgarno sequence as an RBS core, but they are flanked by identical upstream and downstream insulator sequences.

First, we quantified the effect of RBS sequences on protein expression levels by placing each sequence upstream to a YFP reporter and measuring the fluorescence signal using flow cytometry (see Supplementary Methods). The six RBS sequences were labeled 'A' to

Downloaded from http://nar.oxfordjournals.org/ at Weizmann Institute of Science on July 3, 2016

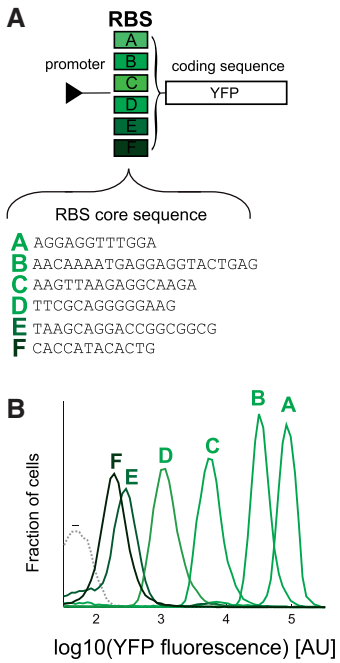

Figure 2. (**A**) A small set of RBS sequences was designed to span several orders of magnitude of protein expression. The RBS set was composed of six pre-characterized RBS core sequences flanked by constant upstream and downstream insulators that were paired to the genes of interest, as detailed in Supplementary Figure S4 and Supplementary Methods. (**B**) Flow cytometry fluorescence measurements of cells expressing YFP, where in each clone a different RBS sequence (A–F) was located upstream to the coding sequence. (–) represents the autofluorescence of cells when no fluorescent protein is expressed.

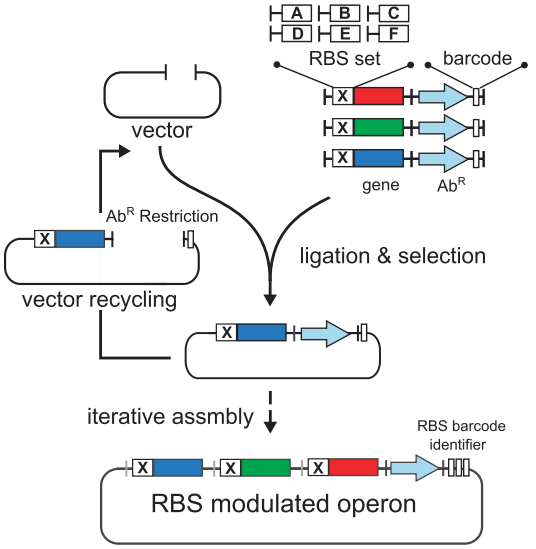

Figure 3. A modular cloning strategy for combinatorial assembly of multi-gene constructs. During the iterative assembly process, each gene of interest is joined with a chloramphenicol (Cm) resistance cassette and paired with the library of RBS sequences. To incorporate the next target gene, the marker is discarded, and the additional part is assembled into the vector. The newly formed construct contains the two RBS-modified genes and a resistance marker, enabling once again direct selection for positive constructs. This sequence of steps can be repeated to easily assemble a combinatorial library of RBS-modulated multi-gene operons.

'F' in descending order of expression level. As shown in Figure 2B, a small set of RBS sequences can span ~100-fold in protein expression levels. Next, we asked whether it is possible to use this compact set of RBS sequences to assemble a library that spans a high-dimensional expression space. We aimed to assemble a combinatorial library of operons in which each variant contains the same genes but under the translational regulation of different RBS sequences.

To facilitate the library construction process, we used an augmented BioBrick (14) cloning strategy designated for the assembly of synthetic operons (Figure 3). In this procedure, genetic parts are iteratively assembled using a positive-selection stage that bypasses the need for time-consuming screening steps. Briefly, a chloramphenicol (Cm) resistance cassette is joined to all of the genetic parts that are to be assembled. In each step, an additional genetic part is incorporated into the construct while the resistance cassette enables a direct selection for properly assembled constructs. The vector is then 'recycled' for the next iteration by excising the resistance cassette. The resulting library of operons is then transformed into cells and screened for a desired phenotype. Inference of the RBS composition across the operon in a specific clone is performed by sequencing a barcode located at the 3′-UTR of the operon. The barcode is generated during the assembly process by iteratively concatenating a short identifying sequence onto the 3′-UTR of the operon (Supplementary Methods and Supplementary Figures S9 and S10). Each genetic variant in the library contains a distinct barcode sequence from which the RBS composition of all the genes in the operon can be inferred in a single sequencing reaction (Supplementary Tables S1 and S2).

RBS-modulated expression of multiple genes: a tri-color fluorescent system

To test whether RBS combinatorics can successfully span a multi-dimensional expression space, we constructed a tri-color reporter system. CFP, YFP and mCherry were randomly paired with three representatives of our RBS set (RBS sequences 'A', 'C' and 'E') and assembled together to yield a library of synthetic operons. The resulting library, therefore, contained $3^3 = 27$ genetic variants, in which all three genes are in the same order but under the regulation of different RBS sequences (Figure 4A). On transformation, colonies display distinct color patterns (Figure 4B), resulting from differential expression of the fluorescent reporters.

The observed color space indicates that the combinatorial assembly of RBS sequences can significantly modulate expression level of multiple genes within the operon (Figure 4C). We verified that the different colors observed are attributed to different combinations of RBS sequences by sequencing sample clones and quantifying their fluorescence levels. In addition, we measured the fluorescence protein accumulations rates in an operon consisting of YFP and mCherry and found a grid of nine clusters (Figure 4D), where each distinct cluster contains variants with an identical RBS composition. Moreover, the

Downloaded from http://nar.oxfordjournals.org/ at Weizmann Institute of Science on July 3, 2016

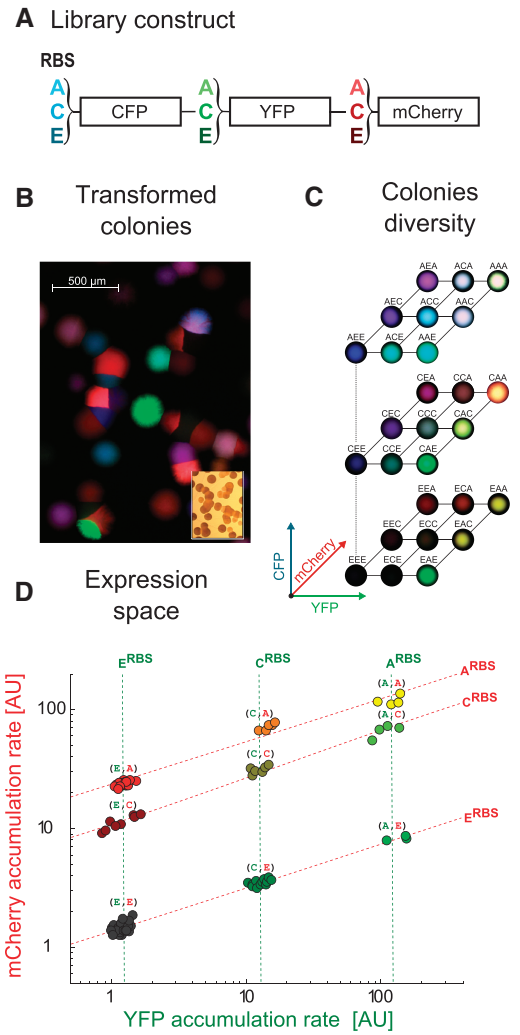

Figure 4. RBS modulation of three fluorescent proteins spans a color space. (**A**) We combinatorially joined CFP, YFP and mCherry with three representatives of our RBS set (sequences 'A', 'C' and 'E') and assembled the genes together into a synthetic operon. The resulting operon library differs only in the RBS sequences regulating gene expression. (**B**) Fluorescence microscopy imaging of *E. coli* colonies, transformed with the operon library. The observed colors represent additive combinations of the three primary colors, assigned to each of the fluorescent proteins. Irregular colony shapes are the result of touching boundaries of adjacent colonies. Some colonies harboring weak RBS for all three fluorescent reporters appear black. Inset: a bright-field microscopy image. (**C**) Fluorescence imaging of *E. coli* colonies containing the tri-color RBS-modulated operon. The images are arranged on a 3D grid where the position on each axis corresponds to the RBS strength of the fluorescent protein. (**D**) YFP and mCherry accumulation rates of clones sampled from a two-color operon library. RBS composition, as determined by barcode sequencing, is shown. Identical genotypes (each labeled by a distinct color) cluster together in the fluorescence space. The effect of translational coupling is also evident, where higher protein accumulation rate of YFP modulates the accumulation rate of mCherry.

spread of clusters demonstrates that RBS modulation spans ≈100-fold in each dimension of the expression space. Notably, the accumulation rate of YFP is dependent only on the RBS sequence regulating it. All colonies with the same RBS upstream to YFP (located first in this operon) exhibit similar accumulation rates as can be observed from the vertical alignment in expression space. However, mCherry fluorescence (located downstream to YFP) depends both on its RBS and on the identity of the RBS upstream to the YFP gene (Figure 4D). We find such coupling over more than two orders of magnitude with a power law exponent of ≈ ⅓, showing an increase by ~2-fold for every 10-fold increase in the accumulation rate of the upstream gene. The dependency is not because of cross-fluorescence (Supplementary Results), and we suggest it is a manifestation of translational coupling (17) between adjacent genes in the operon. The tri-color reporter system is further analyzed in the supplementary section and is an ideal synthetic platform for extending the quantitative analysis of translational coupling (Supplementary Figure S15).

RBS-modulated expression of a metabolic pathway: carotenoid biosynthesis in *E. coli*

Our tri-color reporter system demonstrates that a combinatorial assembly of RBS sequences can span a large fraction of a multi-dimensional expression space. Yet, what is the effect of such expression modulation on the operation of a metabolic pathway? To address this question, we cloned seven genes that compose the carotenoid biosynthesis pathway into *E. coli* (18–20). The end product of this exogenous metabolic pathway is astaxanthin, a high-value xanthophyll (21) known for its potent antioxidant properties. To explore the effect of combinatorial RBS modulation on the biosynthesis of astaxanthin in *E. coli*, each of the genes in the carotenoid pathway was randomly paired with our RBS set and assembled into a library of synthetic operons (Figure 5A). As shown in Figure 5B, the transformed *E. coli* colonies display a variety of colors and intensities. The color pattern of each colony is attributed to differential accumulation of distinct carotenoid intermediates, each colony having a unique color correlated with the carotenoid composition.

For each sampled clone, we determined the RBS composition in the operon by sequencing, and quantified the carotenoid profile using HPLC. Figure 5C shows that clones differing in their RBS composition exhibit diverse carotenoid profiles: some clones accumulate mainly a single product, whereas others produced significant levels of a variety of carotenoids. By screening the combinatorial RBS library for astaxanthin producing variants, we were able to isolate clones yielding up to 2.6 mg/g cell dry weight. Based on previous attempts to optimize astaxanthin production (22), we aimed to further enhance astaxanthin production by overexpressing *dxs*—a gene feeding into the carotenoid pathway. The incorporation of an RBS-modulated *dxs* gene into the synthetic operon led to further increase in astaxanthin production from 2.6 to 5.8 mg/g of cell dry weight. This astaxanthin production yield is ~4-fold higher than the best previously reported results (22). Although the carotenoid profile and astaxanthin accumulation levels varied significantly among sampled clones, we did not observe a consistent correlation between astaxanthin productivity and the growth rate of the cells (Supplementary Table S3). Similarly, the yield of the high astaxanthin clones was not any lower than the other clones.

Downloaded from http://nar.oxfordjournals.org/ at Weizmann Institute of Science on July 3, 2016

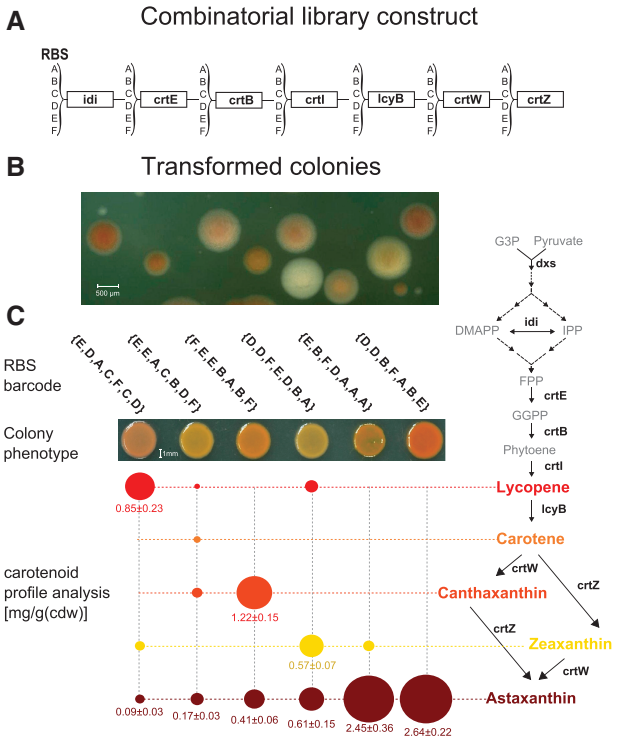

Figure 5. Carotenoid accumulation profile varies with the RBS sequences of biosynthetic genes. (**A**) We assembled a library of synthetic operons differing in the RBS sequences regulating each of the seven genes of the carotenoid biosynthesis pathway. (**B**) A binocular microscopy imaging of *E. coli* colonies transformed with the operon library. The color of the colony corresponds to the composition of the accumulated carotenoids, each having a characteristic color. Image was constructed by stitching multiple adjacent fields. (**C**) The carotenoid accumulation profile and RBS composition of clones isolated from the transformed library. The RBS composition of each clone was determined by sequencing (RBS encoding in barcode refers to the order of genes as illustrated in Figure 3A), and the carotenoid profile of each clone was analyzed using HPLC. Different genotypes result in distinct phenotypes, i.e. distinct carotenoid accumulation profiles. Circle area indicates the production yield of major carotenoid intermediates (>10% of total carotenoids), according to the metabolic pathway described on the right.

To compare the results obtained by our combinatorial approach with 'rational' conventional manual assembly methods, we constructed several pre-designed operons. As expected, an operon modulated uniformly by the weak 'E' RBS sequences produces little astaxanthin and mostly intermediate carotenoids (Supplementary Figure S16). At the other extreme, although we were able to assemble an operon composed solely from the strong 'A' RBS on a promoterless assembly vector, we did not succeed in sub-cloning this operon into our expression vector containing the strong Ptac promoter. A similar situation occurred while attempting to construct other operons composed of strong RBS sequences (several 'B's or 'C's followed by 'A's). These results are consistent with the situation depicted in Figure 1 where the construct viability or genetic stability can be lost at high expression levels, for example, because of protein burden or metabolic imbalance. Therefore, by constructing a combinatorial library containing a range of RBS sequences with varying strengths, our method allows to sample the expression space and locate an RBS combination where the desired phenotype is achieved without hampering viability.

DISCUSSION

Tuning the expression of recombinant enzymes is essential for the optimization of multi-step metabolic pathways (Figure 1). There are two main strategies to achieve balanced expression levels. On one hand, rational design involves the calculation or estimation of the relative and absolute amount of each of the pathway's components (23). However, such attempts are often limited by the lack of sufficient information regarding the kinetics, energetics and regulation of pathway components. Alternatively, strategies based on random mutagenesis of regulatory elements can sample the expression space (24). Yet, even a large pool of mutants do not ensure adequate coverage of the expression space: often, the vast majority of genotypes are clustered in a small portion of the phenotypic space. Moreover, random mutagenesis often yields large libraries, in which screening for a desired phenotype can be challenging or even infeasible. In this study, we introduce a strategy that facilitates the exploration of the phenotypic space using a compact set of regulatory elements. By using a small set of well-characterized RBS sequences to regulate the expression of multiple genes in a synthetic operon, we were able to efficiently sample the multi-dimensional expression space across several orders of magnitude in each axis. The small size of the RBS set can limit the number of genetic variants in the library and enable exhaustive screening. In addition, even in cases in which the screening throughput is limited and the magnitude of the combinatorial library prevents an exhaustive screen, our approach allows to sample the combinatorial space in a way that covers different options in terms of expression levels. Our aim was not to check each and every possible combination but rather to effectively use the limited number of clones that can be screened to locate desired phenotypes.

As a metabolic test case, RBS modulation of carotenoid biosynthesis in *E. coli* suggests that the accumulation of metabolic products of a pathway varies significantly according to the RBS sequences regulating its constituent enzymes. The relationship between the genotype (i.e. the regulating RBS sequences across the operon) and the phenotype (i.e. the accumulation of specific products) is not trivial (Figure 5), suggesting that a rational design would probably fail to achieve optimal metabolic production. As exemplified in our construction efforts, an operon 'rationally' designed for high levels of expression modulation could not be functionally expressed, presumably because of either protein or metabolic burden issues. In contrast, the combinatorial assembly of the astaxanthin biosynthetic pathway resulted in a 4-fold yield increase over conventional assembly and selection methods (22) without any significant decrease in the growth rate or yield of the cells. In conclusion, our results demonstrate the need for an efficient method sampling the expression space to locate an ideally balanced pathway.

Downloaded from http://nar.oxfordjournals.org/ at Weizmann Institute of Science on July 3, 2016

Beyond the optimization of biosynthetic pathways, combinatorial modulation of gene expression can be a valuable research tool for analysis of operons the dominant regulatory unit in prokaryotes. Our combinatorial approach enables us to give a quantitative characterization of effects, such as translational coupling between consecutive genes in an operon, showing that the protein accumulation rate of an upstream gene modulates the accumulation rate of the next gene by up to an order of magnitude.

The strategy we presented here can be expanded in various ways. Specifically, other regulatory elements can be combined in the same approach to further modulate gene expression. For example, by using a small library of promoters to control the transcription of an operon, the span of the overall expression space can in principle be further increased by several orders of magnitude. Finally, although we considered mostly metabolic applications, our approach can be harnessed for numerous biological fields, such as signal transduction and genetic circuits. It is our belief that the strategy we presented here can become a powerful tool for both research efforts and industrial applications.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1–16, Supplementary Methods, Supplementary Results and Supplementary References [25–31].

ACKNOWLEDGEMENTS

Yossi Hirschberg supplied strains and important guidance for the carotenoid biosynthesis project. The authors are grateful to the following people who provided them with vital feedback on this study: Uri Alon, Hernan Garcia, Daniel Koster, Sivan Navon, Rob Phillips, Dave Savage, Eran Segal, Rotem Sorek, Michael Springer and Oren Tzfadia. They also thank Ido Shani, Orel Rivni, Shai Kaplan and Ronen Levy for their valuable technical support.

FUNDING

European research council [260392—SYMPAC]; The Leona M. and Harry B. Helmsley Charitable Trust; Israel Science Foundation [750/09]; Anna and Maurice Boukstein career development chair (to R.M.); Azrieli Foundation award and an Azrieli Fellowship (to E.N.). Funding for open access charge: European Research Council [260392—SYMPAC]; Israel Science Foundation [750/09].

Conflict of interest statement. None declared.

REFERENCES

1. Lu,P., Vogel,C., Wang,R., Yao,X. and Marcotte,E.M. (2006) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.*, **25**, 117–124.
2. Scott,M., Gunderson,C.W., Mateescu,E.M., Zhang,Z. and Hwa,T. (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science*, **330**, 1099–1102.
3. Na,D., Kim,T.Y. and Lee,S.Y. (2010) Construction and optimization of synthetic pathways in metabolic engineering. *Curr. Opin. Microbiol.*, **13**, 363–370.
4. Dekel,E. and Alon,U. (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature*, **436**, 588–592.
5. Koffas,M.A.G., Jung,G.Y. and Stephanopoulos,G. (2003) Engineering metabolism and product formation in *Corynebacterium glutamicum* by coordinated gene overexpression. *Metab. Eng.*, **5**, 32–41.
6. Pitera,D.J., Paddon,C.J., Newman,J.D. and Keasling,J.D. (2007) Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab. Eng.*, **9**, 193–207.
7. Glick,B.R. (1995) Metabolic load and heterologous gene expression. *Biotechnol. Adv.*, **13**, 247–261.
8. Hammer,K., Mijakovic,I. and Jensen,P.R. (2006) Synthetic promoter libraries—tuning of gene expression. *Trends Biotechnol.*, **24**, 53–55.
9. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950.
10. Wang,H.H., Isaacs,F.J., Carr,P.A., Sun,Z.Z., Xu,G., Forest,C.R. and Church,G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894–898.
11. Pfleger,B.F., Pitera,D.J., Smolke,C.D. and Keasling,J.D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat. Biotechnol.*, **24**, 1027–1032.
12. Babiskin,A.H. and Smolke,C.D. (2011) A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. *Mol. Syst. Biol.*, **7**, 471.
13. McGinness,K.E., Baker,T.A. and Sauer,R.T. (2006) Engineering controllable protein degradation. *Mol. Cell*, **22**, 701–707.
14. Shetty,R.P., Endy,D. and Knight,T.F. (2008) Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.*, **2**, 5.
15. Britton,G., Liaaen-Jensen,S. and Pfander,H. (2004) Birkhäuser. *Carotenoids handbook*, Basel; Boston.
16. Chu,F.L., Pirastru,L., Popovic,R. and Sleno,L. (2011) Carotenogenesis up-regulation in *Scenedesmus sp.* using a targeted metabolomics approach by liquid chromatography–high-resolution mass spectrometry. *J. Agric. Food Chem.*, **59**, 3004–3013.
17. Løvdok,L., Bentele,K., Vladimirov,N., Müller,A., Pop,F.S., Lebiedz,D., Kollmann,M. and Sourjik,V. (2009) Role of translational coupling in robustness of bacterial chemotaxis pathway. *PLoS Biol.*, **7**, e1000171.
18. Wang,C., Oh,M.K. and Liao,J.C. (2000) Directed evolution of metabolically engineered *Escherichia coli* for carotenoid production. *Biotechnol. Prog.*, **16**, 922–926.
19. Farmer,W.R. and Liao,J.C. (2000) Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nat. Biotechnol.*, **18**, 533–537.
20. Kim,S.W. and Keasling,J.D. (2001) Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnol. Bioeng.*, **72**, 408–415.
21. Cunningham,F.X. Jr and Gantt,E. (2011) Elucidation of the pathway to astaxanthin in the flowers of *Adonis aestivalis*. *Plant Cell*, **23**, 3055–3069.
22. Lemuth,K., Steuer,K. and Albermann,C. (2011) Engineering of a plasmid-free *Escherichia coli* strain for improved in vivo biosynthesis of astaxanthin. *Microb. Cell Fact.*, **10**, 29.
23. Zhu,X.-G., de Sturler,E. and Long,S.P. (2007) Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm. *Plant Physiol.*, **145**, 513–526.
24. Holátko,J., Elisáková,V., Prouza,M., Sobotka,M., Nesvera,J. and Pátek,M. (2009) Metabolic engineering of the L-valine biosynthesis pathway in *Corynebacterium glutamicum* using promoter activity modulation. *J. Biotechnol.*, **139**, 203–210.
25. Horton,R.M., Hunt,H.D., Ho,S.N., Pullen,J.K. and Pease,L.R. (1989) Engineering hybrid genes without the use of

Downloaded from http://nar.oxfordjournals.org/ at Weizmann Institute of Science on July 3, 2016

restriction enzymes: gene splicing by overlap extension. *Gene*, **77**, 61–68.
26. Davis,J.H., Rubin,A.J. and Sauer,R.T. (2010) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.*, **39**, 1131–1141.
27. Shaner,N.C., Steinbach,P.A. and Tsien,R.Y. (2005) A guide to choosing fluorescent proteins. *Nat. Methods*, **2**, 905–909.
28. Rydzanicz,R., Zhao,X.S. and Johnson,P.E. (2005) Assembly PCR oligo maker: a tool for designing oligodeoxynucleotides for constructing long DNA molecules for RNA production. *Nucleic Acids Res.*, **33**, W521–W525.
29. Baughman,G. and Nomura,M. (1983) Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in *Escherichia coli*. *Cell*, **34**, 979–988.
30. Schümperli,D., McKenney,K., Sobieski,D.A. and Rosenberg,M. (1982) Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. *Cell*, **30**, 865–871.
31. Rex,G., Surin,B., Besse,G., Schneppe,B. and McCarthy,J.E. (1994) The mechanism of translational coupling in *Escherichia coli*. Higher order structure in the atpHA mRNA acts as a conformational switch regulating the access of de novo initiating ribosomes. *J. Biol. Chem.*, **269**, 18118–18127.

Downloaded from http://nar.oxfordjournals.org/ at Weizmann Institute of Science on July 3, 2016