Supporting Information:

Exploring the non-conserved sequence space of synthetic expression modules in *Bacillus subtilis*

Christopher Sauer [1, 2], Emiel Ver Loren van Themaat [2], Leonie M. Boender [2], Daphne Groothuis [2], Rita Cruz [1, 2], Leendert W. Hamoen[1, 3], Colin R. Harwood [1], Tjeerd van Rij [2, 4]

[1] Centre for Bacterial Cell Biology, Institute for Cell and Molecular Biosciences, Newcastle University Newcastle upon Tyne, United Kingdom

[2] DSM Biotechnology Center, P.O. Box 1, 2600 MA Delft, The Netherlands

[3] Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

[4] Corresponding author: Tjeerd van Rij

Content:

Table S1

Synthetic expression elements used for the SEM library.

Element	Sequence	Feature	#nt
A1+-35	ACGTTGATATAATTTAAATTTTATTTGACA	83% AT, native	30
A2+-35	TCGAGGCCTACGCGACCTCGGCGGTTGACA	25% AT	30
A3+-35	GGGGAGGCTGCATGCCCCTGACCCTTGACA	25% AT	30
A4+-35	GCTGCGCGCACCGAGGCCTTCAGGTTGACA	25% AT	30
A5+-35	GGCCACGCGGGTAGTGCCGCCTCATTGACA	25% AT	30
A6+-35	GTCAGGGCGGCTACGCGCTGCCCATTGACA	25% AT	30
A7+-35	TTTCACCGTCTGATAACGGCGGAATTGACA	50% AT	30
A8+-35	TCTAGGGAATTTGCCCGCAGCATATTGACA	50% AT	30
A9+-35	ACTGTTGGCGACCTTCGATGAAACTTGACA	50% AT	30
A10+-35	TTGGTATAGTCCGTACAACCACGGTTGACA	50% AT	30
A11+-35	CTTGGGTAGACTCAGGCAACTCATTTGACA	50% AT	30
A12+-35	GACTCATTAATTAACAGATTTGTATTGACA	75% AT	30
A13+-35	GAGTAATAAGTATCTTCATATCTATTGACA	75% AT	30
A14+-35	ATAGCCTATTATGTTCTTGAAAAATTGACA	75% AT	30
A15+-35	TCATTTGATGTACATAGCATAATATTGACA	75% AT	30
A16+-35	GAGTTTAGATAATCATATACCATTTTGACA	75% AT	30
A17+-35	TATTTTTAATTTATATATAAAAAATTGACA	100% AT	30
A18+-35	AATATTTTATTAATTTATAAAATATTGACA	100% AT	30
A19+-35	TTATATATTAAAATATAATTATATTTGACA	100% AT	30
A20+-35	TAAATTATAATATTTATTAAATTATTGACA	100% AT	30
A21+-35	TTAATATATATTAAATTAATATATTTGACA	100% AT	30
B1+-10	AAAATGGGCTCGTGTTGTATAAT	53 % AT, native	23
B2+-10	ACTGAAGATATATTCTATATAAT	75% AT	23
B3+-10	TTAAAATAGACTTTGCATATAAT	75% AT	23
B4+-10	CACTTGAGAAATATTTATATAAT	75% AT	23
B5+-10	ACAACTAGTAATTTTAGTATAAT	75% AT	23
B6+-10	TAATAGACAACTTATGTTATAAT	75% AT, TATG -17	23
C1+SD	AAATGTAGTGAAGGAGG	58% AT, native	17
C2+SD	GGACTTTACAAAGGAGG	53% AT	17
C3+SD	AAGATCTTCGAAGGAGG	53% AT	17
C4+SD	CGTTCTAGAAAAGGAGG	53% AT	17
C5+SD	AAAGTCGTCTAAGGAGG	53% AT	17
C6+SD	ATTGATCAGCAAGGAGG	53% AT	17
C7+SD	CCTAACTAGTGTATGTAACGAAGGAGG	56% AT	27
C8+SD	TTGAGAGCTCCGATTATCAAAAGGAGG	56% AT	27
C9+SD	ACTTGACGTTCTGAGCAATAAAGGAGG	56% AT	27
C10+SD	TGAGTTAGGTCCTCTCAAAAAAGGAGG	56% AT	27
C11+SD	TCAGAAATAGCTAGTTCTGCAAGGAGG	56% AT	27
C12+SD	GATCCTTTAGAGATGCTAACTGCTCAAGATAAGGAGG	57% AT	37
C13+SD	ACTAAGCATAAGGTTTACTTCGAGCGACTTAAGGAGG	57% AT	37
C14+SD	ATAGTGAGCGTATTCATCCTATCAGAAGCTAAGGAGG	57% AT	37
C15+SD	CCCTGTAGAAGCGAGTAATACAATCTTGTTAAGGAGG	57% AT	37
C16+SD	TACAATACGTTTCAGTTCTTCGGGCGAAAAAAGGAGG	57% AT	37
D1	ATGCA	60% AT, native	5
D2	GATTACTA	75% AT	8

D3	TAGATATC	75% AT	8
D4	ATGTCTAA	75% AT, AA-ATG	8
D5	CTAATGTA	75% AT	8
D6	TGTACTAA	75% AT, AA-ATG	8

Elements A are shown including the -35 sequence, Elements B are shown including the -10 sequence and Elements C are shown including the Shine Dalgarno sequence.

Table S2

Average fluorescence intensities of the twelve bins.

Bin	Average GFP expression (log10)	Average GFP expression
168	2.36	226.98
Bin I	2.81	644.16
Bin II	3.33	2112.23
Bin III	3.45	2824.74
Bin IV	3.59	3854.66
Bin V	3.92	8236.86
Bin VI	3.97	9409.57
Bin VII	4.05	11193.08
Bin VIII	4.23	16985.93
Bin IX	4.41	25749.92
Bin X	4.45	28377.92
Bin XI	4.65	44997.55
Bin XII	4.73	53642.36

Table S3

PCR primers used in this study.

Name	Sequence	Function
oCS001	GCGCCTCGAGAGGTCGACTCTAGAGGATCC	Fwd *Xho*I (pAPNC213)
oCS002	AATTACTAGTGTGAGCGCAACGCAAGCTTC	Rev *Spe*I (pAPNC213)
oCS003	AATTTCTAGAAAACCGGTAGATCTCACGTG	Fwd *Xba*I (pPG49)
oCS004	AATTACTAGTATTTGTAGAGCTCATCCATG	Rev *Spe*I (pPG49)
oCS007	AATTGAATTCCGGCGGAGACGCTTGCAGTGGGCTTACATGG	Fwd *EcoR*I,*BsmB*I (kan^R)
oCS008	AATTGGATCCCCAAGGAGACGTCAGAAGAACTCGTCAAGAAG	Rev *BamH*I,*BsmB*I (kan^R)
oCS009	GGAGACGTCAGAAGAACTCG	Fwd stargate
oCS010	TTGGATTCAATAGAAAAGGTAAGCG	Rev stargate
oCS024	AATTGGATCCGCTCCGTTAGAAACAGCGTC	Rev *BamH*I *nprE*
oCS025	AATTGCGGCCGCGACGGTTCATTCTTCTCTCC	Fwd *Not*I *nprE*
oCS026	AATTCCGCGGCGTGACCTGTAGCAGAATTC	Rev *Sac*II *nprE*
oCS027	AATTTCTAGAATAGGATCCGGAATTGACTCAAGCTTCAC	Fwd stargate pCS05
oCS028	AATTCCGCGGATAGCGGCCGCGCTAACGCCCGAATTCCG	Rev stargate pCS05
oCS029	AATTCCGCGGTATGATTCTTCTCGCTTCCG	Fwd *Sac*II amp^R
oCS030	AATTTCTAGATAGGTATATCATCTCTCGCC	Rev *Xba*I amp^R
oCS045	AATTCTCGAGGGTGGCAAATTCGCTTACC	Fwd *Xho*I *gfp* out
oCS048	GGTGGCAAATTCGCTTACC	Fwd pCS12
oCS049	ATGGTCAGCAAGGGAGAGG	Rev pCS12
oCS062	AATTCTCGAGCCGTTAGCGTTTAAGTACATCC	Fwd *spec* out (*Xho*I)
oCS069	AATAACTAGTGTGAGCGCAACGC	Rev *Spe*I *gfp* out
oCS071	AATTGGTACCCCTCTTGTGAAATTAGAGAACGC	Rev *spec* out (*Kpn*I)
oCS077	ATCCAAGGAGACGTCAGAAG	Rev *mCherry*
oCS492	GGTCATTACAGTGAGGTACG	Fwd SEM library
oCS493	ACTGACAGACTGGTAACTCG	Rev SEM library
oCS523	AATTCGTCTCCCGGCGGCCGCTTCTTTGTATTCTG	Fwd *mCherry*
oCS524	AATTCGTCTCCCCAATTATTACTCGAGTAAGGATCCTTTG	Rev *mCherry*
oCS545	ATGTTTAAGTTTAAAAAGAATTTCTTAG	Fwd *xynA*
oCS546	CTACCACACTGTTACGTTAG	Rev *xynA*
oCS568	GTTCCAATACGGAGAAATCG	Fwd sequencing SEM library
oCS569	TCTTCTCCTTTGCTCATCTG	Rev sequencing SEM library
oCS586	GTATAGCATACATTATACGAACGGTAAAGCTTGCAAGCAAGTGCATATCCTG	Rev T_{amyM}
oCS587	AGCGCGTCTCCATAAGTTTAAACAAATCTTTTTCGAAAAAAGGCCGCC	Fwd T_{amyM}

oCS588	GCTTCACAAGCTTTACCGTTCGTATAATG	Fwd *spec*^{R}
oCS589	GCTATCACCGCCCAGCCTAAAC	Rev *spec*^{R}
oCS590	ATCTGCAGAGTTCCAATACGGAGAAATCGCGGCCGCGGAG ACGCGCTCCTGTGACGGAAGATCACTTCGCAG	Fwd *cap*^{R}
oCS591	CGAAAAAGATTTGTTTAAACTTATGGAGACGCGCTTTAC GCCCCGCCCTGCCACTCATCGC	Rev *cap*^{R}
oCS592	GTCTGGGCGGCCGCGATTTCTCCGTATTGG	Fwd pBest4
oCS593	CGAACGGTACCTGCAGGATCCGTTTAGGCTGGGCGGTG	Rev pBest4
oCS594	AATTTCTAGAAAACCGGTAGATCTCACGTG	Fwd *gfp*
oCS595	AATTACTAGTATTTGTAGAGCTCATCCATG	Rev *gfp*
oCS596	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTAAGGT AACGTTCCAATACGGAGAAATCG	Fwd Bin I
oCS597	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAAGGAG AACGTTCCAATACGGAGAAATCG	Fwd Bin II
oCS598	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAAGAGGA TTCGTTCCAATACGGAGAAATCG	Fwd Bin III
oCS599	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTACCAAG ATCGTTCCAATACGGAGAAATCG	Fwd Bin IV
oCS600	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCAGAAGG AACGTTCCAATACGGAGAAATCG	Fwd Bin V
oCS601	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGCAAG TTCGTTCCAATACGGAGAAATCG	Fwd Bin VI
oCS602	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCGTGAT TCGTTCCAATACGGAGAAATCG	Fwd Bin VII
oCS603	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTTCCGATA ACGTTCCAATACGGAGAAATCG	Fwd Bin VIII
oCS604	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTGAGCGG AACGTTCCAATACGGAGAAATCG	Fwd Bin IX
oCS605	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCTGACCG AACGTTCCAATACGGAGAAATCG	Fwd Bin X
oCS606	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCTCGA ATCGTTCCAATACGGAGAAATCG	Fwd Bin XI
oCS607	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTAGGTGG TTCGTTCCAATACGGAGAAATCG	Fwd Bin XII
oCS608	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTCTTCTC CTTTGCTCATCTG	Rev Bin I-XII

Figure S1

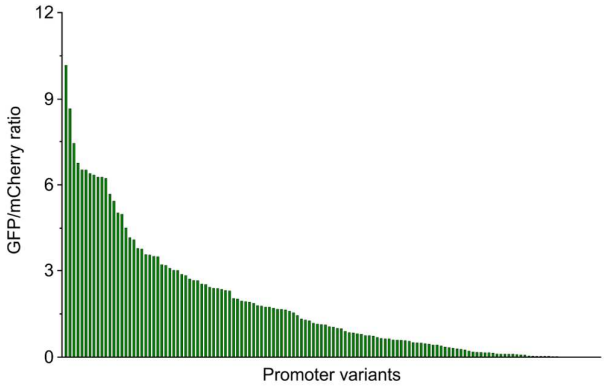

Figure S1: GFP/mCherry ratio of library variants after overnight growth. 135 colonies containing the synthetic expression modules were picked randomly from agar plates and grown overnight. Subsequently, end point measurements of GFP and mCherry were performed in a plate reader to determine expression strength, as reflected by the GFP/mCherry ratio.

Figure S2

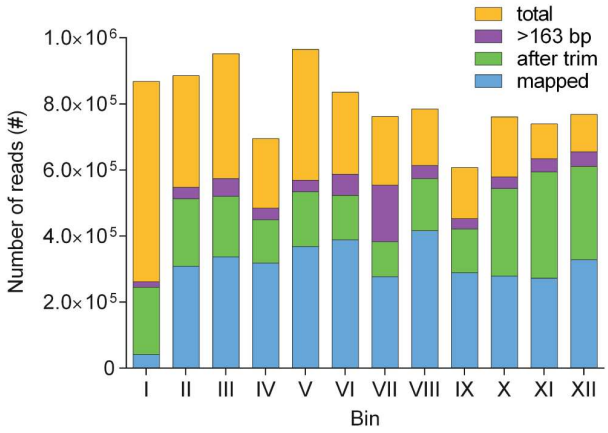

Figure S2: Illumina HiSeq 2500 sequence statistics of twelve bins containing SEMs. The yellow bars represent the total number of reads, followed by the length filtering >163 bp (purple) and the filtering for after trimming 5' *amyE* and *gfp* sequences (green). The cyan bars represent the total number of reads that were mapped to the design library. 91.37% of the 12,000 designed SEMs were identified in the sequence data.

Figure S3

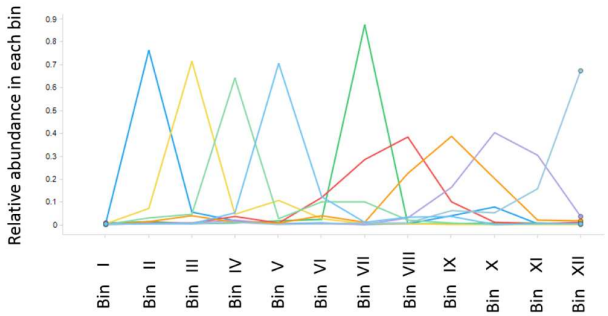

Figure S3: Relative abundance of nine arbitrary SEMs over the twelve bins. The differential sequence distribution of nine arbitrary sequences over all bins is shown.

Figure S4

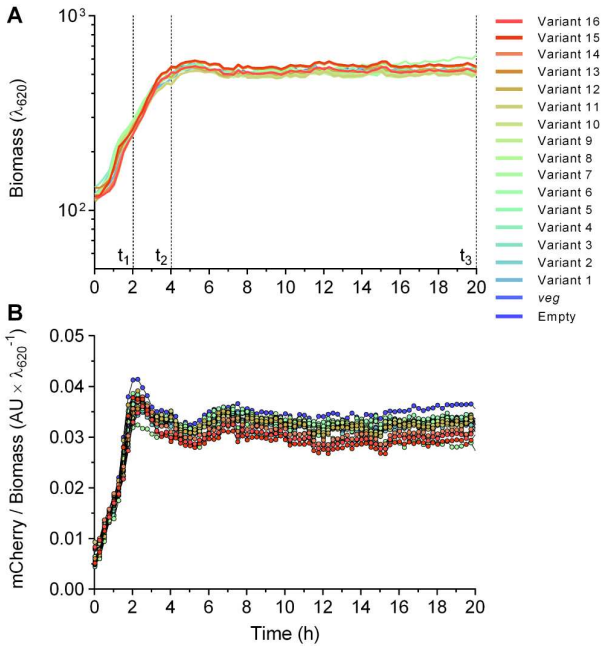

Figure S4: Growth and mCherry production of strains expressing XynA over time. Strains containing sixteen selected SEMs, no expression module (empty) or the native expression module (*veg*) were analyzed during growth. (A) Biomass increase and (B) production of mCherry per unit biomass are shown. Strains were grown in quadruplicate in a BioLector® for 20 hours in LB medium and the average values are shown.

Dataset S1: SEM Constructs

SEM ID, identification of SEM within library; Total counts, total DNA counts of this SEM; Bin 1 through Bin 12, DNA counts in each bin; Element A, variants used in the library (A1-A21); Element B, variants used in the library (B1-B6); Element C, variants used in the library (C1-C16); Element D, variants used in the library (D1-D6); Rel.bin1 through Rel.bin12, relative DNA counts in each bin; Score, GFP expression score (as in Materials and Methods); Average GFP expression (log10), average GFP expression (log10) across all constructs, not filtered; Average GFP expression, average GFP expression across all constructs, not filtered.

Dataset S2: 32 SEMs (high GFP)

Number, number of the 32 selected SEMs fused to *gfp*; SEM ID, identification of SEM within library; Elements, library elements used in these SEMs; Sequence SEM, DNA sequence; Sequence RBS, DNA sequence of the RBS; Sequence *gfp*, first 35 nt of *gfp*; Sequence *gfp* full, sequence of complete *gfp* gene; Sequence RBS+*gfp*, DNA sequence of the 5'UTR and the first 35 nt of *gfp*, used to calculate the translation initiation rate; Translation initiation rate (TIR) of RBS+*gfp* calculated with the RBS calculator V2.1; ΔG total (kcal/mol), total free energy exchange between ribosome and mRNA; Score, GFP expression score (as in Materials and Methods); Average GFP expression (log10), average GFP expression (log10) across all constructs, not filtered.

Dataset S3: RBS

Number, number of the RBS in the library; Element C, variants used in the library minus 6 nt at 5'end resulting in the 5'UTR (C1-C16); Element D, variants used in the library (D1-D6); Sequence Element C; Sequence Element D; first 35 nt *gfp*; RBS and first 35 nt of *gfp* sequence; Full sequence *gfp*, sequence of complete *gfp* gene; Translation initiation rate (TIR), calculated with the RBS calculator V2.1; ΔG total (kcal/mol), total free energy exchange between ribosome and mRNA; Average GFP expression (log10), average GFP expression (log10) across all constructs, not filtered; s.t.d. GFP expression (log10), standard deviation of GFP expression (log10) across all constructs, not filtered; Average GFP expression, average GFP expression across all constructs, not filtered.

Dataset S4: Promoters

Number, number of promoter in the library; Element A, variants used in the library (A1-A21); Element B, variants used in the library (B1-B6); Sequence, sequence of the promoter fusing Element A and B; Average GFP expression (log10), average GFP expression (log10) across all constructs, not filtered; s.t.d. GFP expression (log10), standard deviation of GFP expression (log10) across all constructs, not filtered; Average GFP expression, average GFP expression across all constructs, not filtered. Correlation GFP and TIR (R^2), correlation between measured and predicted translation initiation of *gfp.*

Dataset S5: 16 SEMs fused with *xynA*/*gfp*

Number, number of the 16 selected SEMs fused to *xynA*/*gfp*; SEM ID, identification of SEM within library; Elements, library elements used in these SEMs; Sequence SEM, DNA sequence; Sequence RBS, DNA sequence of the RBS; Sequence *xynA* (35nt), first 35 nt of *xynA*; Sequence RBS+*xynA*, Sequence *xynA*, sequence of complete *xynA* gene; DNA sequence of the 5'UTR and the first 35 nt of *xynA*; XynA (AU), 2 hours, XynA activities after 2 hours (t_1); 4 hours, XynA activities after 4 hours (t_2); 20 hours, XynA activities after 20 hours (t_3); Sequence *gfp* (35nt), first 35 nt of *gfp*; Sequence *gfp*, sequence of complete *gfp* gene; Sequence RBS+*gfp*, DNA sequence of the 5'UTR and the first 35 nt of *gfp*; GFP (AU), 2 hours, GFP fluorescence after 2 hours; 4 hours, GFP fluorescence after 4 hours; 20 hours, GFP fluorescence after 20 hours; Score, GFP expression score (as in Materials and Methods).

Dataset S6: Oligo designs

SEM ID, identification of SEM within library; Sequence, DNA sequence of SEM.