Supporting Information

Belitsky and Sonenshein 10.1073/pnas.1300428110

SI Materials and Methods

Construction of Fragmented Chromosomal DNA Library. Seven micrograms of total *Bacillus subtilis* chromosomal DNA was fragmented by sonication. The DNA fragments were blunted and phosphorylated at the 5′ end using the Quick Blunting kit (NEB), followed by addition of dAMP to the 3′ fragment end using Klenow Fragment (exo-) (NEB). The resulting fragments were ligated (Quick Ligation kit; NEB) to a universal nonbarcoded partially double-stranded adapter with a 5′ dT overhang formed by annealing oligonucleotides olj543 and olj331. The ligated products were gel fractionated to isolate fragments of 250- to 300 nt in length and purified using the Gel Extraction kit (Qiagen). The size-fractionated fragments were amplified by PCR using oligonucleotides olj139 and olj331 as primers.

Additional Details of the in Vitro DNA Affinity Purification Sequencing Procedure. The in vitro DNA affinity purification (IDAP) buffer was identical to the conventional binding buffer (1), except for the absence of EDTA and DTT, which would interfere with the downstream steps of IDAP, and salmon sperm DNA. The absence of EDTA and DTT was shown in a separate gel-shift experiment not to affect the efficiency of CodY–DNA interaction. Salmon sperm DNA, used to prevent nonspecific low-affinity binding, was omitted so that its presence would not affect competition for CodY binding among various DNA fragments.

The enrichment of fragments containing known CodY-binding sites during IDAP was followed by real-time PCR using gene-specific oligonucleotides as primers, and the amount of PCR products was normalized to the amount of total recovered DNA determined with olj139 and olj331 or olj528 as primers. PCR products that contained no CodY-binding sites served as negative controls for real-time PCR results.

The Galaxy suite (2) was used for the analysis of sequencing results. Total Illumina reads were filtered for quality, and the coverage maps were generated by aligning the reads to the sequence of *B. subtilis* chromosome (3) using the Bowtie program. The CodY-binding regions ≥61 nt in length and having coverage at each position at least threefold higher than average were determined using the custom Peak Finder script set for 140-nt (70 nt each way) shift and 6-nt smoothing. Importantly, only 5′ nucleotides of each read were counted for coverage determination. For further analysis, the regions were extended by 70 nt on each side.

The double-stranded coverage of each extended region (i.e., strength) was recalculated as the percentage of the coverage of total recovered DNA multiplied by 100. The boundaries of the 323 strongest regions were adjusted manually. The background, determined from the no-CodY sample and normalized for the recovery of non-CodY binding DNA in each sample, was subtracted for each region.

Construction of Transcriptional *lacZ* Fusions. To create transcriptional fusions, the *ispA*, *rapA*, and *rapE* PCR products containing the entire upstream intergenic region and the 5′ parts of the coding regions that included the CodY-binding sites were synthesized by using gene-specific oligonucleotides as the forward and reverse primers and chromosomal DNA of *B. subtilis* strain SMY as the template. The 0.88-, 0.44-, and 0.69-kb fragments of these PCR products were cut with XbaI and HindIII enzymes and cloned in the integrative plasmid pHK23 (1) to create pBB1755 (*ispA-lacZ*), pBB1752 (*rapA-lacZ*), and pBB1754 (*rapE-lacZ*).

The m1 mutations (A to C) in the *ispA*, *rapA*, and *rapE* fusions at positions 638, 248, and 233 with respect to the start of the coding region (pBB1760, pBB1761, and pBB1759, respectively) were introduced by using mutagenic reverse primers. The p1 mutation (A to C) in the *ispA* regulatory region (position −138 with respect to the start of the *ispA* coding region) was introduced by two-step overlapping PCR. In the first step, a product containing the 5′ part of the *ispA* regulatory region was synthesized by using oligonucleotide oBB592 as the forward primer and mutagenic oligonucleotide oBB615 as the reverse primer. In a similar manner, a product containing the 3′ part of the *ispA* regulatory region was synthesized by using mutagenic oligonucleotide oBB616 as the forward primer and oligonucleotide oBB593 as the reverse primer. The two PCR products were used in a second, splicing step of PCR mutagenesis as overlapping templates to generate a modified fragment containing the entire *ispA* regulatory region; oligonucleotides oBB592 and oBB593 served as the forward and reverse PCR primers, respectively. The spliced PCR product was digested with XbaI and HindIII and cloned in pHK23, to create pBB1767 (*ispAp1-lacZ)*. *Escherichia coli* strain JM107 (4) was used for isolation of plasmids. All cloned PCR-generated fragments were verified by sequencing by the Tufts University Core Facility.

B. subtilis strains carrying various *lacZ* fusions at the *amyE* locus were isolated after transforming strain BB2511 (*amyE::spc lacA::tet*) with the appropriate plasmids, by selecting for resistance to erythromycin conferred by the plasmids, and screening for loss of the spectinomycin-resistance marker, which indicated a double crossover, homologous recombination event. Strain BB2511 and all its derivatives have very low endogenous β-galactosidase activity due to a null mutation in the *lacA* gene (5).

Labeling of DNA Fragments. The 928-bp PCR product containing the part of the *ispA* gene corresponding to positions −222 to + 658 with respect to the translation start (plus a 48-bp vector- and primer-specific extension) was synthesized using gene-specific oligonucleotide oBB592 as the forward primer and *lacZ*-specific oligonucleotide oBB102 as the reverse primer and pBB1755 (*ispA-lacZ*) as the template. oBB102 (which would prime synthesis of the template strand of the PCR product) was labeled using T4 polynucleotide kinase and [γ-^{32}P]-ATP. oBB102 starts 36 downstream of the HindIII site that serves as a junction between the regulatory regions and the *lacZ* part of the fusions.

1. Belitsky BR, Sonenshein AL (2008) Genetic and biochemical analysis of CodY-binding sites in *Bacillus subtilis*. *J Bacteriol* 190(4):1224–1236.
2. Goecks J, Nekrutenko A, Taylor J, Team G; Galaxy Team (2010) Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11(8):R86.
3. Barbe V, et al. (2009) From a consortium sequence to a unified sequence: The *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* 155(Pt 6):1758–1775.
4. Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33(1):103–119.
5. Daniel RA, Haiech J, Denizot F, Errington J (1997) Isolation and characterization of the *lacA* gene encoding beta-galactosidase in *Bacillus subtilis* and a regulator gene, *lacR*. *J Bacteriol* 179(17):5636–5638.

Fig. S1. Coverage maps of the *rapA* (*A* and *B*), *rapC* (*C* and *D*), *nupN* (*E-G*), and *putB* (*H* and *I*) CodY-binding regions. 40 nM CodY was used for affinity purification. Black arrows indicate a gap in coverage (*A*, *C*, and *E*); colored arrows indicate gap boundaries (*B*, *D*, *F*, and *G*) or a drop in coverage (*I*). No gap is present in *H*. The upstream CodY-binding site of the *putB* gene is weak and was not included in Dataset S1.

Fig. S2. Diagrams illustrating the strand-specific counting of 5′ nucleotides of Illumina reads for two closely spaced CodY-binding sites. (*A*) Alignment of (+) strand reads with a genome sequence containing two idealized 16-nt binding sites (black rectangles), separated by 24 nt. All fragments are shown to be of a uniform 60-nt size. The 5′ nucleotides are depicted as circles, which are filled in gray for fragments encompassing an entire binding site. Fragments containing black circles overlap both binding sites and should have higher coverage (not reflected in *B*). (*B*) Strand-specific coverage map generated by counting only the 5′ nucleotide of each read. The abundance of each fragment that copurifies with the binding protein was assumed to be 20; the abundance of fragments that do not contain the protein-binding site was assumed to be zero. (*C*) Coverage map of the *bcaP* CodY-binding region using strand-specific counting of 5′ nucleotides. 40 nM CodY was used for affinity purification. Arrows indicate the downstream boundary of the upstream binding site and the upstream boundary of the downstream binding site.

Fig. S3. Motif logos for the subsets of CodY-binding sites. The logos were generated by the MEME function of the Galaxy suite. The 354 strongest core sites, extended by 20 nt at each end, were separated into three equal groups (118 sites each) corresponding to the strongest, next strongest, and least strong sites according to the approximate strength ranking of the binding regions purified in the presence of 8 nM CodY (Dataset S1*B*). (*A*) Sites 1–354; (*B*) sites 1–118; (*C*) sites 119–236; and (*D*) sites 237–354. The canonical 15-nt motif (AATTTTCWGAAAATT) appears to be the main determinant of CodY binding to the strongest binding sites, although the first nucleotide of the canonical motif is apparently less conserved (*B*). However, for less strong CodY-binding sites additional downstream nucleotides are required, and the extent of this requirement increases as the strength of the sites diminishes (*C* and *D*).

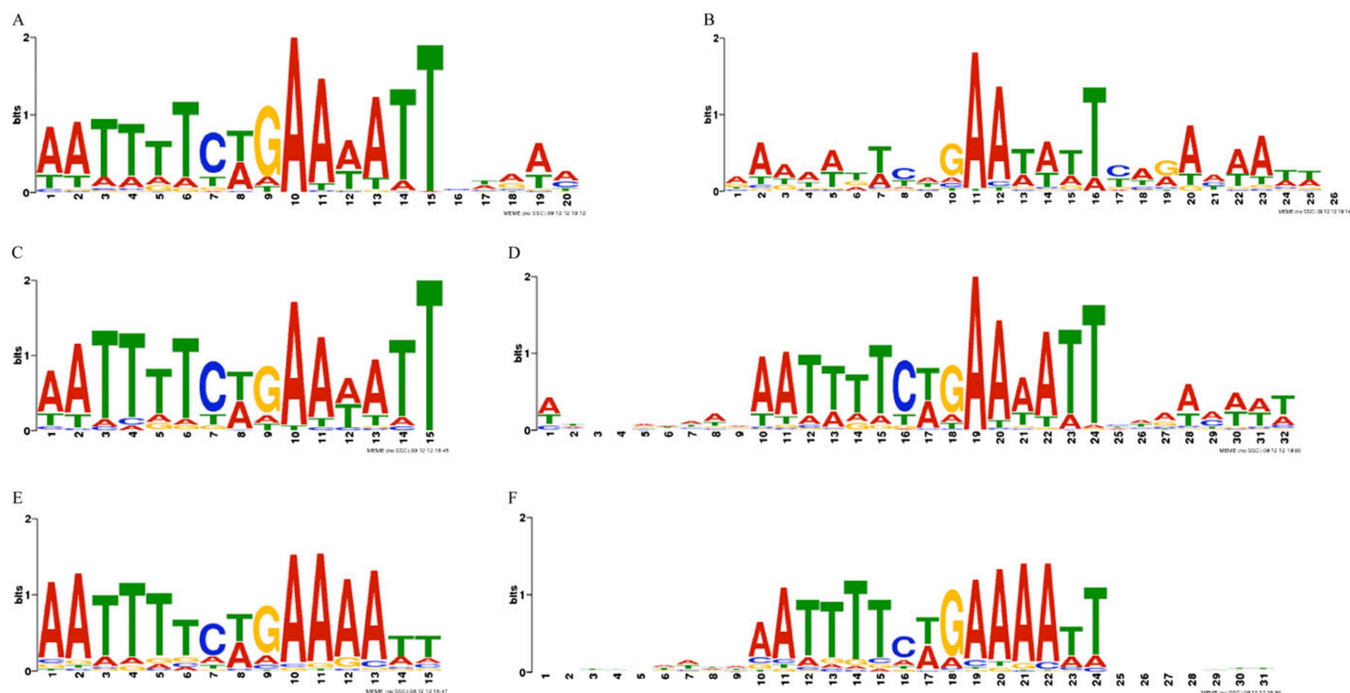

Fig. S4. Motif logos for the subsets of CodY-binding sites and chromosomal motif-like sequences. The logos were generated by the MEME function of the Galaxy suite. (*A* and *B*) The 354 strongest core sites, extended by 20 nt at each end, were separated into two groups of 89 sites, overlapping a 15-nt canonical motif with one, two, or three mismatches with respect to the consensus (*A*), and 265 sites lacking such an overlap (*B*). (*C–F*) The total of 417 15-nt chromosomal sequences with three or fewer mismatches with respect to the 15-nt canonical consensus motif were separated into two groups of 89 sequences overlapping CodY-binding sites detected by IDAP-Seq (*C*) and 318 sequences that do not overlap and are at least 20 nt away from such a site (*D*). (*E*) Motif logo for the 89 sequences, which were extended each by 9 nt on either side. (*F*) Motif logo for 318 sequences, which were extended each by 9 nt on either side. The MEME motif for 89 CodY-binding sites that are associated with strong canonical motifs (*A*) was similar to the MEME motif obtained for the entire dataset of 354 sites (Fig. 5). The MEME motif for the 265 sites that are not associated with strong 15-nt canonical motifs appeared to be composed of two degenerate 15-nt canonical motifs overlapping by 6 nt as suggested previously (1) (*B*). The 15-nt canonical motif-like sequences found in chromosomal regions that do not bind CodY have poorer conservation of residues T14 and T15 and higher occurrences of C and especially G residues at several positions than do 89 sequences that overlap CodY binding sites (*C* and *D*). Moreover, only sequences associated with actual binding sites had an additional signature of conserved nucleotides lying downstream of the 15-nt canonical motif as revealed by the reanalysis of all 15-nt chromosomal sequences (with three or fewer mismatches to the canonical motif) that were extended with nine additional flanking nucleotides on each side (*E* and *F*). Thus, the 15-nt canonical motif with three or fewer mismatches may be sufficient for binding in some cases, presumably those in which the most highly conserved residues are present. Other 15-nt sequences, especially with particularly deleterious mismatches or more than three mismatches, are unlikely to bind CodY unless additional nucleotides contribute to CodY binding or a second canonical motif is present in the vicinity.

1. Wray LV, Jr., Fisher SH (2011) *Bacillus subtilis* CodY operators contain overlapping CodY binding sites. *J Bacteriol* 193(18):4841–4848.

Table S1. CodY-mediated regulation of *lacZ* fusions

Strain	Relevant genotype	Fusion type	β-Galactosidase activity, MU
BB1605	WT	*ald-lacZ*	46.3
BB1820	*codY*		74.1
BB831	WT	*azlB-lacZ*	4.67
BB3101	*codY*		2.25
BB834	*azlB*	*azlB-lacZ*	232.0
BB3102	*azlB codY*		276.8
BB2871	WT	*codV-lacZ*	6.06
BB2874	*codY*		6.66
HJS11	WT	*glnR-lacZ*	12.1
BB2872	*codY*		6.8
BB224	*glnR*	*glnR-lacZ*	64.6
BB2873	*glnR codY*		66.0
HKB194	WT	*odhA-lacZ*	7.74
BB3649	*codY*		6.68
BB3625	WT	*oppA-lacZ*	55.9
BB3627	*codY*		36.7
BB2279	WT	*pdxS-lacZ*	59.2
BB3643	*codY*		81.7

ald-, *azlB-*, *codV-*, and *glnR-lacZ* fusions were described (1–4). Other fusions were constructed using integrative plasmids pJPM82 or pHK23 as described previously (4, 5), and the details will be provided on request. All strains were derivatives of strain SMY (6). Strain construction, cell growth in TSS glucose–ammonium medium containing 16 amino acids, and β-galactosidase-specific activity determination (in Miller units) were as described (5).

1. Schreier HJ, Brown SW, Hirschi KD, Nomellini JF, Sonenshein AL (1989) Regulation of *Bacillus subtilis* glutamine synthetase gene expression by the product of the *glnR* gene. *J Mol Biol* 210(1):51–63.
2. Siranosian KJ, Ireton K, Grossman AD (1993) Alanine dehydrogenase (*ald*) is required for normal sporulation in *Bacillus subtilis*. *J Bacteriol* 175(21):6789–6796.
3. Slack FJ, Serror P, Joyce E, Sonenshein AL (1995) A gene required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon. *Mol Microbiol* 15(4):689–702.
4. Belitsky BR, Gustafsson MC, Sonenshein AL, Von Wachenfeldt C (1997) An *lrp*-like gene of *Bacillus subtilis* involved in branched-chain amino acid transport. *J Bacteriol* 179(17): 5448–5457.
5. Belitsky BR, Sonenshein AL (2008) Genetic and biochemical analysis of CodY-binding sites in *Bacillus subtilis*. *J Bacteriol* 190(4):1224–1236.
6. Zeigler DR, et al. (2008) The origins of 168, W23, and other *Bacillus subtilis* legacy strains. *J Bacteriol* 190(21):6983–6995.

Table S2. Oligonucleotides used in this work

Name	Sequence	Specificity
Library construction		
olj139	5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGA	
olj331	5′-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC* T	
olj527	5′-CAAGCAGAAGACGGCATACGAGATAAAAAAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj528	5′-CAAGCAGAAGACGGCATACGAGATACACACGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj529	5′-CAAGCAGAAGACGGCATACGAGATAGAGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj530	5′-CAAGCAGAAGACGGCATACGAGATATATATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj531	5′-CAAGCAGAAGACGGCATACGAGATCACACAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj532	5′-CAAGCAGAAGACGGCATACGAGATCCCCCCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj533	5′-CAAGCAGAAGACGGCATACGAGATCGCGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj534	5′-CAAGCAGAAGACGGCATACGAGATCTCTCTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj535	5′-CAAGCAGAAGACGGCATACGAGATGAGAGAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj536	5′-CAAGCAGAAGACGGCATACGAGATGCGCGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj537	5′-CAAGCAGAAGACGGCATACGAGATGGGGGGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj538	5′-CAAGCAGAAGACGGCATACGAGATGTGTGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj539	5′-CAAGCAGAAGACGGCATACGAGATTATATAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj540	5′-CAAGCAGAAGACGGCATACGAGATTCTCTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj541	5′-CAAGCAGAAGACGGCATACGAGATTGTGTGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj542	5′-CAAGCAGAAGACGGCATACGAGATTTTTTTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	
olj543	5′-pGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT	
Flanking forward primers		
oBB587	5′-CTGATAAGCTTCGATGTCTTCTAAC	*rapA*
oBB590	5′-TCAGTTCTAGAATCCACCTCCGAAGGC	*rapE*
oBB592	5′-CAGTTTCTAGACACTGCTCCTTTCCCC	*ispA*
Flanking reverse primers		
oBB102	5′-CACCTTTTCCCTATATAAAAGC	*lacZ* (pHK23)
oBB588	5′-CTGATAAGCTTCGATGTCTTCTAac	*rapA*
oBB591	5′-CTTATAAGCTTCAATTTTATTTAAATATTCTGAAAG	*rapE*
oBB593	5′-GTCAAAAGCTTTATTCGCGTTAGAAAATTC	*ispA*
oBB602	5′-CTTATAAGCTTCAATTTTATTTAAATAT**g**CTGAAAGTTC	*rapEm1*
oBB603	5′-GTCAAAAGCTTTATTCGCGTTAGAAAAT**g**CTGATAATTC	*ispAm1*
oBB604	5′-CTGATAAGCTTCGATGTCTTCTAACAAT**g**CTGAAAAC	*rapAm1*
Internal mutagenic forward primer		
oBB616	5′-CAATATTTTG**c**AGATTACTAATTCTTTG	*ispAp1*
Internal mutagenic reverse primer		
oBB615	5′-CAAAGAATTAGTAATCT**g**CAAAATATTG	*ispAp1*

Barcodes and XbaI and HindIII restriction sites are underlined. Altered nucleotides are in bold lowercase. Asterisk (*) denotes a phosphorothioate linkage.

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)