

(10) International Publication Number
WO 2013/148321 A1(43) International Publication Date
3 October 2013 (03.10.2013)

(51) International Patent Classification:

C07K 14/195 (2006.01) *C40B 40/10* (2006.01)
C12N 15/63 (2006.01) *G06N 3/12* (2006.01)

(21) International Application Number:

PCT/US2013/032145

(22) International Filing Date:

15 March 2013 (15.03.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/616,176 27 March 2012 (27.03.2012) US

(71) Applicants: **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US). **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 5th Floor, Oakland, California 94607-5100 (US).

(72) Inventors: **RHODIUS, Virgil, A.**; 3166 Keith Drive, El Sobrante, CA 94803 (US). **VOIGT, Christopher**; 57 Elizabeth Road, Belmont, MA 02478 (US). **GROSS, Carol, A.**; 1700 Jaynes Street, Berkeley, CA 94703 (US).

(74) Agent: **JOHNSTONE, Oona, M.**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

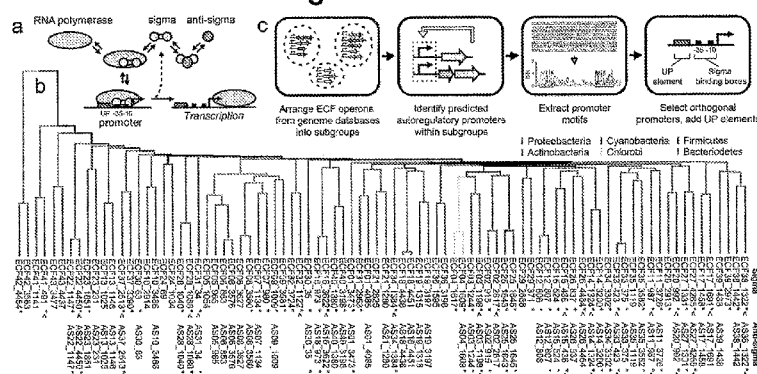
(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: ANTI-SIGMAS FOR PROGRAMMABLE TRANSCRIPTIONAL REGULATION

Figure 1



(57) Abstract: The invention relates to anti-sigma factors ("anti-sigmas") that bind to sigma factors and block activation of transcription. Anti-sigmas and their cognate sigma factors provide a highly effective mechanism for regulating gene expression in genetic circuits.

ANTI-SIGMAS FOR PROGRAMMABLE TRANSCRIPTIONAL REGULATION

RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application Serial No. 61/616,176, entitled ANTI-SIGMAS FOR PROGRAMMABLE TRANSCRIPTIONAL REGULATION, filed on March 27, 2012, the entire disclosure of which is herein incorporated by reference.

GOVERNMENT INTEREST

This invention was made with Government support under Grant No. EEC0540879 awarded by the National Science Foundation. The government has certain rights in this invention.

FIELD OF INVENTION

The invention relates to recombinant genetic circuits comprising anti-sigma factors.

BACKGROUND OF INVENTION

Synthetic biology is limited by the small number of available and well-characterized transcription factors from which to program large genetic circuits. Bacterial sigma factors (σ s), the promoter recognition subunits of RNA polymerase (RNAP), are modular proteins with domains that recognize DNA sequences in the -10 and -35 regions of their target promoters¹. In addition to the housekeeping σ s (e.g., $\sigma 70$ in *E. coli*) that recognize the thousands of canonical promoters essential for growth, bacteria have a variable number of stress-activated alternative σ s that direct RNAP to distinct promoter sequences. This enables cells to express multiple genes associated with a particular developmental state or stress response² and execute complex gene expression dynamics that implement temporal control and serve as developmental checkpoints³. For example, spore formation in *B. subtilis* requires a cascade of 5 σ s ($\sigma H \rightarrow \sigma F \rightarrow \sigma E \rightarrow \sigma G \rightarrow \sigma K$)⁴. σ s can be embedded in complex webs of partner swapping networks, including anti- σ s, which physically block σ s from interacting with RNAP⁵⁻⁷, and anti-anti- σ s. Such feedback loops and protein-protein interactions generate more complex dynamics for integrating many environmental and cellular signals⁸.

SUMMARY OF INVENTION

Described herein are anti-sigma factors (“anti-sigmas”) that bind to sigma factors and block activation of transcription. Anti-sigmas and their cognate sigma factors provide a highly effective mechanism for regulating gene expression in genetic circuits.

Aspects of the invention relate to a recombinant genetic circuit including an extracytoplasmic function (ECF) sigma factor; an anti-sigma factor that binds to the ECF sigma factor; and a promoter that is recognized by the ECF sigma factor. In some embodiments, the ECF sigma factor, the anti-sigma factor and/or the promoter are genetically engineered. In some embodiments, the ECF sigma factor and/or the promoter are chimeric.

In some embodiments, the recombinant genetic circuit is expressed within a host cell, which can be a prokaryotic cell or a eukaryotic cell. In some embodiments, the ECF sigma factor and/or anti-sigma factor are codon-optimized for expression in the host cell. In certain embodiments, the ECF sigma factor is selected from the group consisting of ECF01-ECF43. In certain embodiments, the anti-sigma factor is selected from the group of anti-sigma factors contained within Table 1.

In some embodiments, the recombinant genetic circuit comprises a combination of logic gates. In some embodiments, the logic gates are selected from the group consisting of AND, NAND, NOR, OR, NOT, XOR, EQUALS, AND, IMPLIES, and ANDN gates. In certain embodiments, the AND gates comprises an ECF sigma factor and a promoter that is recognized by the ECF sigma factor. In some embodiments, the recombinant genetic circuit is a component of a synthetic genetic switch. In certain embodiments, the synthetic genetic switch is bistable. In certain embodiments, the recombinant genetic circuit is a component of a pulse generator.

Further aspects of the invention relate to a system comprising a plurality of recombinant genetic circuits. In some embodiments, the ECF sigma factors and the anti-sigma factors are orthogonal.

Further aspects of the invention relate to a host cell comprising a heterologous genetic circuit including an extracytoplasmic function (ECF) sigma factor; an anti-sigma factor that binds to the ECF sigma factor; and a promoter that is recognized by the ECF sigma factor. In some embodiments, the ECF sigma factor, the anti-sigma factor and/or the promoter are

genetically engineered. In some embodiments, the ECF sigma factor and/or the promoter are chimeric. In some embodiments, the host cell is a prokaryotic host cell.

In some embodiments, the genetic circuit comprises one or more logic gates selected from the group consisting of AND, NAND, NOR, OR, NOT, XOR, EQUALS, AND, IMPLIES, and ANDN gates. In certain embodiments, the AND gates comprises a sigma factor and a sigma factor target DNA sequence. In certain embodiments, two or more logic gates are combined by having the output promoter of an upstream gate serve as the input promoter of a downstream gate.

Further aspects of the invention relate to a library comprising two or more anti-sigma factors, wherein each anti-sigma factor selectively binds to one or more ECF sigma factors. In some embodiments, the library further includes one or more ECF sigma factors. In some embodiments, the library further includes one or more promoters that are recognized by the one or more sigma factors.

In some embodiments, the anti-sigma factors, ECF sigma factor and/or promoter are genetically engineered. In certain embodiments, the ECF sigma factor and/or promoter is chimeric. In some embodiments, the ECF sigma factors and/or anti-sigma factors are codon-optimized for expression in a host cell.

In some embodiments, the ECF sigma factors and the anti-sigma factors are orthogonal. In some embodiments, the ECF sigma factors are under the control of an inducible promoter. In certain embodiments, the ECF sigma factor is selected from the group consisting of ECF01-ECF43. In certain embodiments, the anti-sigma factor is selected from the anti-sigma factors within Table 1.

Further aspects of the invention relate to a non-transitory computer readable storage medium encoded with instructions, executable by a processor, for designing a host cell.

Further aspects of the invention relate to a computer product comprising a computer readable medium encoded with a plurality of instructions for controlling a computing system to perform an operation for designing a host cell.

Each of the limitations of the invention can encompass various embodiments of the invention. It is, therefore, anticipated that each of the limitations of the invention involving any one element or combinations of elements can be included in each aspect of the invention. This invention is not limited in its application to the details of construction and the arrangement of

components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced or of being carried out in various ways.

BRIEF DESCRIPTION OF DRAWINGS

The accompanying drawings are not intended to be drawn to scale. In the drawings, each identical or nearly identical component that is illustrated in various figures is represented by a like numeral. For purposes of clarity, not every component may be labeled in every drawing. In the drawings:

FIG. 1 depicts a strategy for the genomic mining of Extracytoplasmic Function (ECF) σ s, anti- σ s, and promoters. FIG. 1A shows how sigmas recruit core RNA polymerase (RNAP) to promoters; a function that is inhibited by the anti- σ . Sigmas have a two-domain structure that binds to the -10 and -35 regions of the target promoter. FIG. 1B shows the complete libraries of 86 synthesized σ s (top row) and their 62 cognate anti- σ s (bottom row) organized as a phylogenetic tree. Asterisks indicate active σ s (>5-fold activation) or anti- σ s (>2-fold repression). Carets indicate σ s or anti- σ s that appear in the final orthogonal sets. All σ s in the library are named ECFXX_YYYY, where “XX” denotes the ECF subgroup, and “YYYY” denotes the unique σ ID given by Staron and co-workers⁷. The anti- σ s were named ASXX_YYYY, where “XX” and “YYYY” denote the ECF subgroup and unique ID of the cognate σ . Consequently, cognate σ / anti- σ pairs have the same numbering (e.g., ECF11_987 and AS11_987). FIG. 1C shows that for each σ , target promoters are identified through a process of computational search, selection, and design. The first step involves the organization of the ECF operons according to the subgroups defined by Mascher and co-workers⁷.

FIG. 2 reveals promoter models for 29 ECF sigma subgroups. The models contain a sequence logo illustrating the -35 / -10 motifs and intervening spacer sequence. The exact -35 and -10 sequences identified by BioProspector²⁷ are underlined underneath each sequence logo. The bar chart histograms illustrate the number of promoters with different length distances between underlined -35 and -10 motifs. The promoters were organized vertically to cluster similar -35 and -10 motifs, as determined by eye. The bottom three promoter models (ECF5*,

ECF14, ECF27) represent promoters that were not found to be active (>5-fold activation) in the specific embodiments tested. Promoter model ECF5* represents the model for subgroups 5 – 10.

FIG. 3 depicts the activity and orthogonality of ECF σ s. FIG. 3A depicts an embodiment wherein ECF σ s were induced by IPTG via a T7 expression system, and σ -dependent promoter activity was measured by GFP expression and flow cytometry. Plasmid pN565 encodes the IPTG inducible T7* expression system¹⁹; plasmid series pVRa and pVRb encode the ECF σ library and test promoter library, respectively. The specific example shown (ECF11_987 and P11_3726) is highlighted in the following subfigures. FIG. 3B depicts activities of active ECF σ library members titrated against their cognate target promoters. The grey lines show levels of GFP expression for one active ECF σ :promoter pair in each subgroup induced with 0, 10, 20, 50, and 100 μ M IPTG. Highlighted in black is the averaged activity of sigma ECF11_987 against its promoter P11_3726. Data are shown from three independent assays and error bars represent one standard deviation. Plots of the other σ :promoter pairs are shown in more detail in FIG. 12. FIG. 3C depicts the liquid culture growth curves (OD₆₀₀) for each σ shown in part c under high induction (100 μ M IPTG). Highlighted in black is the growth curve of sigma ECF11_987 averaged from three independent growth assays and the error bars represent one standard deviation. The growth curves of two negative controls are shown in dark grey. FIG. 3D depicts the activity of one promoter (P11_3726) for the complete library of active σ s expressed with 100 μ M IPTG. Each bar represents the average promoter activity from two independent assays and error bars represent one standard deviation. The two σ s from subgroup 11 that were expected to activate the promoter are bracketed. FIG. 3E shows all cross-reactions for the 20 most orthogonal σ :promoter pairs. Each σ was induced with 100 μ M IPTG, and the fold-induction was measured as the fluorescence with σ induction divided by the basal activity of the promoter in the absence of any σ . Each square represents the average fold-induction from two independent assays of a unique σ :promoter combination. All promoters were named using the convention Pxx_YYYY, where “XX” and “YYYY” denotes the subgroup and unique ID of the downstream parent σ gene (e.g. P02_2817 is the promoter upstream of σ ECF02_2817). Promoters containing synthetic UP-elements were renamed to Pxx_UPYYYY (e.g., P15_UP436). The σ :promoter pairs were ordered by the absolute amount of off target activity caused by/affecting the pair, with the lowest off-target activity in the upper left and the highest in the lower right. FIG. 3F shows

promoter scores, as calculated from PWMs, compared to the experimental measurements in part e. The promoter scores are calculated using the ECF promoter models (UP+PWM-35+PWM-10+spacer penalty) for the -60 to +20 promoter fragment including 30 nt flanking vector sequence. The ECF11_987:P11_3726 activity is highlighted with an arrow. FIG. 3G shows ECF02_2817 and ECF11_3276 recombined in their flexible linker region between domains 2 and 4 to create chimeric σ s ECF02-11 and ECF11-02 (Figure 20). The promoters activated by the two parental σ s were similarly recombined between the -10 and -35 regions to create chimeric promoters. FIG. 3H shows the activity and orthogonality of the σ s and chimeric σ s against their cognate promoters. All of the σ s were induced with 10 μ M IPTG and the fold-induction was as defined previously. Each square represents the average fold-induction from three independent assays.

FIG. 4 demonstrates that anti-sigma factors can be used to create orthogonal threshold-gated switches. FIG. 4A shows that in addition to the expression and reporter systems shown in FIG. 3A, cells contain the plasmid series pVRc, which allows HSL-inducible independent expression of anti- σ s to bind and sequester σ s. FIG. 4B shows repression of ECF11_987 activity on promoter P11_3726 by different anti- σ s. Each bar represents average fold-repression, as defined by normalizing the fluorescence of cells containing the promoter with both induced σ (induced with 10 μ M IPTG) and induced anti- σ (induced with 50 nM HSL) against the fluorescence of cells containing just the promoter and induced sigma factor. Bar heights represent the average from three independent assays and error bars represent one standard deviation. FIG. 4C shows the cross reactivity of 12 anti-sigma factors on the set of 12 orthogonal σ s targeted by the anti- σ s. The activity of each σ paired with its cognate promoter was measured in the absence and presence of different anti- σ s. Sigmas were partially induced (10 μ M IPTG) and anti- σ s maximally induced (50 nM HSL). Fold activity repression by the anti- σ , defined as the activity of the promoter in the absence of the anti- σ divided by the activity in its presence is indicated. The anti- σ : σ pairs were arranged in the same order as the σ s in FIG. 3E. FIG. 4D shows the influence of the expression of the anti- σ for a series of switches. The plots show the transfer function of each σ -dependent promoter for differing expression levels of anti- σ induced by the inducer HSL: (a), no anti sigma plasmid; (b), 10 nM HSL; (c), 50 nM HSL.

FIG. 5 depicts a phylogenetic tree of 1873 ECF sigmas across bacteria divided into 43 subgroups.

FIG. 6 depicts a method of identifying target promoters using bioinformatics. Each ECF sigma subgroup is predicted to recognize different promoter sequences. Previously, promoters were known for only 18 sigma subgroups (Staron et al. 2009). Many ECF sigmas autoregulate their expression; consequently, promoters can be found by searching for conserved motifs upstream of sigma genes in each subgroup. The upstream sequences of all sigmas within each subgroup were extracted. Conserved motifs were identified using a 2-block motif search algorithm (BioProspector). For each ECF sigma subgroup, upstream sequences of ECF sigma genes and ECF sigma operons were searched.

FIG. 7 depicts repression of ECF sigmas by their cognate anti-sigmas, tested *in vivo* at target promoters.

FIG. 8 depicts titration of sigma/anti-sigma pairs (ECF/AS) and measurement of promoter (Pecf) activity (fluorescence was measured by flow cytometry).

FIG. 9 depicts complete promoter models for each ECF subgroup. The models contain a sequence logo illustrating the upstream (UP) sequence, -35 sequence, spacer sequence, -10 sequence, and 10 bases following the -10. The histograms show, from all the analyzed promoters, the distance between the -35 and -10 motifs, and the distance between the -35 motif and the nearest downstream gene. The exact -35 and -10 sequences identified by the 2-block search algorithm, BioProspector, are underlined underneath each sequence logo, and were used to calculate the distances for the -10/-35 spacer histograms.

FIG. 10 depicts predicted orthogonality of ECF sigma promoter models and individual subsite models. FIG. 10A shows a Heatmap of 29 ECF sigma promoter models used to score all 706 identified ECF sigma promoters. Each promoter model, comprised of -35 and -10 PWMs and spacer penalties (Equation S3) was used to score target promoter sequences from approximately -40 to +1 (a 40 nt window starting from 5 nt upstream of the identified -35 motif)

and the highest score for each sequence recorded. A high score represents a similar sequence to the promoter model; a low score (e.g., <0) represents a divergent sequence to the promoter model. Target promoters represent all 706 promoters identified in ECF sigma regulatory regions and are grouped by their parent ECF sigma subgroup. FIG. 10B shows a Heatmap of the 29 -35 subsite models used to score all 706 identified ECF sigma promoters. FIG. 10C shows a Heatmap of the 29 -10 subsite models used to score all 706 identified ECF sigma promoters.

FIG. 11 depicts improvement of promoter activity by adding UP-elements. Promoter sequences were tested for activity against both cognate sigmas from their ECF sigma group. (Sigma 1 denotes the ECF with a lower number in the library, while sigma 2 denotes the higher one. For example, with P02_2817, sigma 1 is ECF02_915 and sigma 2 is ECF02_2817. In some embodiments, inactive promoters tended to contain G/C-rich upstream sequences. These sequences were replaced with synthetic UP-element (CATGACAAAATTTT TAGATGCGTT; SEQ ID NO:64; -60 to -35), improving promoter activity. *In vivo* assays were performed by inducing ECF sigma expression with 100 μ M IPTG for 6 hr during exponential growth and measuring promoter activity from GFP fluorescence using flow cytometry. Each bar represents the average promoter activity from at least two independent assays and error bars represent one standard deviation.

FIG. 12 depicts the activity of all ECF sigmas for all promoters constructed. A Heatmap of *in vivo* activities for the complete library of 86 ECF sigmas against 26 promoters is presented. Assays were performed by inducing ECF sigma expression with 100 μ M IPTG for 6 hr during exponential growth and measuring promoter activity from GFP fluorescence using flow cytometry. Each square represents the average fold induction from at least two independent assays.

FIG. 13 depicts sigma titration assays. Each of the chosen 52 active ECF sigma factors was induced at 0, 10, 20, 50 and 100 μ M IPTG with their most active cognate promoter *in vivo*. Assays were performed with 6 hr induction during exponential growth and promoter activity measured from GFP fluorescence using flow cytometry. Plots represent the average promoter activity from three independent assays and error bars represent one standard deviation. Hash

marks indicate toxicity (as judged by a fall in the 8hr OD₆₀₀ to 80% or lower of wild type; see Example 3 and FIG. 16; growth measurements were taken at 0, 10, 20, and 100 μ M IPTG) occurring at different levels of IPTG induction. Cytometry distributions show typical log-scale histograms at 0 (black) and 100 μ M IPTG (grey). One sigma:promoter pair, ECF03_1244:Pecf03_up1198, showed extreme toxicity at full induction, so three histograms are shown: (a) = 0 μ M, (b) = 20 μ M, (c) = 100 μ M. All cytometry histograms are shown with log scaling and normalized to the mode of the distribution. A portion of the cells falls on the Y-axis and is not visible when plotted in this way. Bracketed titles indicate that that sigma:promoter set was included in the orthogonal subset of the ECF sigma library (FIG. 3E).

FIG. 14 depicts anti-sigma titration assays. Transfer functions are shown for 35 anti-sigma:ECF sigma:promoter sets. Promoter activity was measured across a range of anti-sigma and ECF sigma inductions. For each assay: from top to bottom inducing the anti-sigma, no anti-sigma control, anti-sigma with 0, 10 and 50 nM HSL; from left to right inducing ECF sigma, 0, 5, 20 and 100 μ M IPTG. Assays were performed *in vivo* with a 6 hour induction and promoter activity determined by measuring the fluorescence of sfGFP with flow cytometry. Data is from a single replicate. Bracketed titles indicate that the anti-sigma ECF sigma:promoter set was active and targets one of the ECF sigmas in the orthogonal subset (included in FIG. 4C). Boxed titles indicate that the anti-sigma was one of the 25 tested in the combinatorial assay (FIG. 15).

FIG. 15 demonstrates repression of ECF sigmas by all anti-sigmas. A heatmap of fold repression of sigma activity by different anti-sigmas is shown. Assays were performed by inducing ECF sigma expression with 10 μ M IPTG and anti-sigma expression with 50 nM HSL for 6 hr during exponential growth and fold repression calculated from fluorescence measurements. Each square represents the average fold repression from three independent assays.

FIG. 16 depicts a sigma factor transition phase growth assay. OD₆₀₀ measurements were taken after 8 hours of growth at different levels of expression of each of the ECF sigma factors in the library. Plots represent the average measurement from three independent assays normalized to cells not expressing a sigma factor, and error bars represent one standard deviation. Boxed

titles indicate 'active' ECF sigma factors with more than 5-fold induction of a promoter in the library, and bracketed titles indicate a ECF sigma factor in the final orthogonal set (FIG. 3E).

FIG. 17 depicts an anti-sigma transition phase growth assay. OD₆₀₀ measurements were taken after 8 hours of growth at different levels of expression of each of the anti-sigma factors in the library. Plots represents the average measurement from three independent assays normalized to cells not expressing an anti-sigma factor, and error bars represent one standard deviation. Boxed titles indicate the 25 anti-sigma factors tested for orthogonality (FIG. 15), and bracketed titles indicate anti-sigmas that repress ECF sigma factors in the final orthogonal set (FIG. 4C).

FIG. 18 depicts anti-sigma factor thresholding. Sixteen sets of anti-sigma, sigma factor, and promoter were assayed at varying inductions of both anti-sigma and sigma factor to characterize their ability to implement thresholding. The anti-sigmas were either not present (a), or induced at 0nM HSL (b) or 50nM HSL (c), while the sigma factors were induced at 0, 5, 25, and 100 μ M IPTG. Plots represent the average promoter activity from three independent assays and error bars represent one standard deviation. Histograms show representative distributions from a 0 and 100 μ M induction for each set. Bracketed titles indicate data shown in FIG. 4D.

FIG. 19 depicts Hill fits to AS/ECF20_992 anti-sigma:sigma threshold switch. Threshold-gated switches were built using sigma factor ECF20_992 and its corresponding anti-sigma and promoter. ECF20_992 was induced with varying levels of IPTG (0, 5, 25, 50, 75, 100, 150, and 200 μ M) in the presence of plasmids expressing AS20_992 from varying promoters: No anti-sigma (a), BBa_J23117 (b), BBa_J23105 (c), BBa_J23101 (d), BBa_J23100 (e), pLux induced with 50 nM HSL (f). Equation S5 was fit to the data using a relative least-squares algorithm and the fit is shown. The Hill coefficient n is higher when the anti-sigma is present: $n = 1.7$ (No anti-sigma), $n = 3.3$ (BBa_J23117), $n = 4.1$ (BBa_J23105), $n = 2.6$ (BBa_J23101), $n = 3.9$ (BBa_J23100 and pLux). Plots represent the average promoter activity from three technical replicates (No anti-sigma) or three independent assays (all others) and error bars represent one standard deviation.

FIG. 20 depicts chimeric sigma factor and promoter engineering. FIG. 20A presents three alignments (SEQ ID NOs: 65 & 66) of the flexible linker region connecting domains 2 and 4 of the parental sigma factors shown with the crossover seams derived from each alignment marked. Alignment 1 was generated through protein sequence alignment, hand-adjusted based on crystal structures. Alignment 2 was generated through protein sequence alignment alone. Alignment 3 was created with secondary structure prediction by PredictProtein. Each of the six marked seams was used to engineer one chimeric sigma factor of either orientation, for a total of 12 chimeric sigma factors. FIG. 20B presents three chimeric promoters of each orientation that were similarly engineered. pECF02_rpoHP3 and pECF11_3726 were recombined between -20 and -21 to make p02-110 and p11-020. To correct for any differences in optimal spacing between the chimeras and parental sigma factors, 1 bp was added or removed at the -20/-21 seam to make additional promoters with longer or shorter spacers.

FIG. 21 depicts testing of chimeric sigma factor and chimeric promoter variants. Each of the engineered chimeric sigma factors was tested with each of the variants of its cognate chimeric promoter. Negative controls represent an empty plasmid lacking either a sigma factor coding region, or a sigma promoter:reporter cassette. Arrows point to the sigma chimera : chimera promoter pairs chosen for orthogonality testing. Assays were performed *in vivo* in CAG22216 cells at 30°C, induced for 6 hours with 10 μ M IPTG, and promoter activity was measured by flow cytometry. Each bar represents the average promoter activity from three independent assays, each of which had three technical replicates, and error bars represent one standard deviation taken between the independent assays.

FIG. 22 depicts orthogonality testing of chimeric and parental sigma factors. The data corresponding to FIG. 3H are shown for the error bars (a) and cytometry data (b). Assays were performed *in vivo* in CAG22216 cells at 30°C, induced for 6 hours with 10 μ M IPTG, and promoter activity was measured by flow cytometry. (-) controls represent cells carrying a plasmid identical to the chimeric sigma expression or reporter plasmids, but with the chimeric sigma gene, or promoter/reporter removed. Each bar represents the average promoter activity from three three independent assays, each of which had three technical replicates, and error bars represent one standard deviation taken between the independent assays. Histograms show

representative flow cytometry distributions for each promoter in the presence of all of the sigma factors.

FIG. 23 depicts plasmids used for ECF sigma and anti-sigma characterization. Low processivity T7 RNA polymerase (T7*) was expressed from pN565 using an IPTG-inducible Ptac promoter with a symmetric lac operator (lacOsym). T7* was used to express the ECF sigma library via a T7 and IPTG induced promoter (consisting of the PT7 promoter sequence followed by a symmetric lac operator) from the pVRaXX_XXX plasmid series. Sigma-dependent promoters (PSigma) were carried on the pVRbXX_XXX plasmid series fused to superfolder gfp. The anti-sigma library carried on the pVRcXX_XXX plasmid series was under HSL control via the Plux promoter. XX_XXX in each of the library names represents which anti-sigma / ECF sigma / or sigma-dependent promoter the plasmid carries. For example, the set pVRa20_992, pVRb20_992, pVRc20_992 carries ECF20_992, AS20_992, and pECF20_992, respectively.

FIG. 24 depicts an alternate plasmid set used in anti-sigma threshold experiments. An alternate series of plasmids, pAG_AS20_992_J23XXX, replaces the pVRcXX_XXXX plasmid series used the anti-sigma library testing experiments (FIG. 23). This plasmid series expresses the anti-sigma factor AS20_992 from a range of constitutive promoters instead of from the Plux promoter.

FIG. 25 depicts plasmids used for chimeric sigma characterization. A two plasmid system was used to characterize the chimeric sigma factors. Plasmid library pTSaXX expressed the parental or chimeric sigma factors under the control of an IPTG inducible PTac promoter with a symmetric lac operator (lacOsym). Plasmid library pTSbXX contained the parental and chimeric sigma factor-dependent promoters driving sfGFP.

DETAILED DESCRIPTION

The invention is based, at least in part, on the identification of multiple anti-sigmas from the genomes of diverse bacteria. Surprisingly, anti-sigmas were easily transferable between species and were found to function orthogonally with their cognate sigma factors even when derived from different organisms. Significantly, anti-sigmas serve as highly effective

sequestering molecules in genetic circuits, dramatically expanding available tools for constructing and programming large genetic circuits.

This invention is not limited in its application to the details of construction and the arrangement of components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced or of being carried out in various ways. Also, the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use of “including,” “comprising,” or “having,” “containing,” “involving,” and variations thereof herein, is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

Aspects of the invention relate to sigma factors. As used herein, a sigma (σ) factor refers to a sequence-specific DNA binding protein that recruits RNA polymerase (RNAP) to specific promoter sequences to initiate transcription. Bacterial cells contain several categories of sigma factors, including general or “housekeeping” sigma factors, which are involved in transcriptional activation of many genes. Other sigma factors serve more specialized functions, responding to certain cellular conditions and activating transcription of specific genes. The Extracytoplasmic Function (ECF) sigma factors represent the largest and most diverse group of sigma factors. Approximately 19,314 ECF sigma factors have been annotated (Ulrich et al. (2007) *Nucleic Acids Res.* 35, D386-390), and have been subdivided into 43 subgroups (Staron et. Al (2009) *Mol. Microbiol.* 74(3):557-81). A description of classification of sigma factors, including a listing of sigma factors ECF01-ECF43, from a variety of bacterial species, is incorporated by reference herein from WO 2012/170436, e.g., Table 1 on pages 28-30 of WO 2012/170436.

It should be appreciated that any sigma factor can be compatible with genetic circuits described herein and with methods for making and using such genetic circuits. In some embodiments, the sigma factor is an ECF sigma factor. For example, the ECF sigma factor can be ECF01, ECF02, ECF03, ECF04, ECF05, ECF06, ECF07, ECF08, ECF09, ECF10, ECF11, ECF12, ECF13, ECF14, ECF15, ECF16, ECF17, ECF18, ECF19, ECF20, ECF21, ECF22, ECF23, ECF24, ECF25, ECF26, ECF27, ECF28, ECF29, ECF30, ECF31, ECF32, ECF33, ECF34, ECF35, ECF36, ECF37, ECF38, ECF39, ECF40, ECF41, ECF42 or ECF43, as incorporated by reference from WO 2012/170436, or the ECF sigma factor can be a homolog of any of ECF01-ECF43 from other bacterial species, or can be any other ECF Sigma factor. The sigma factor can be naturally occurring or can be synthetic. As used herein, a naturally occurring

sigma factor means that the sigma factor has the same sequence that it possesses in its natural environment, or a variant thereof, while a synthetic sigma factor refers to a sigma factor that has a different sequence than is possessed by a sigma factor in its natural environment. In some embodiments, genetic circuits can include multiple different sigma factors including different types of sigma factors.

ECF sigma factors bind to specific DNA sequences within their target promoters, referred to as "-10" and "-35" regions, based on measurement from the transcriptional start site. The -10 region is also referred to as a "Pribnow box", and is generally six nucleotides long. In some embodiments, the -10 region has the sequence TATAAT or a variant thereof. The sequence of the "-35 region" ranges from 8-12 nucleotides. In certain embodiments, the sequence of the -35 region is TGTTGACA or a variant thereof. It should be appreciated that any -10 or -35 regions of any sequence can be compatible with aspects of the invention. Sigma factors comprise a region referred to as "Region 2," which is a conserved domain that recognizes -10 regions of promoters, and a region referred to as "Region 4," which is a conserved domain that recognizes -35 regions of promoters.

Promoters recognized by ECF sigma factors can be naturally occurring or can be synthetic. As used herein, a naturally occurring promoter means that the promoter has the same sequence that it possesses in its natural environment, or a variant thereof, while synthetic refers to a promoter that has a different sequence than is possessed by a promoter in a natural environment. In some embodiments an ECF sigma factor recognizes the same promoter that it recognizes in its natural environment, while in other embodiments an ECF sigma factor recognizes a different promoter than the promoter that it recognizes in its natural environment.

ECF sigma factors can be chimeric. As used herein, a chimeric sigma factor refers to a sigma factor formed from portions of two or more sigma factors. For example, in some embodiments, a chimeric sigma factor comprises a "Region 2" from a first sigma factor and a "Region 4" from a second sigma factor, thereby generating a chimeric sigma factor with novel DNA binding activities. Chimeric sigma factors can include any combination of DNA binding domains from any sigma factor. A sigma factor, including a chimeric sigma factor can be a component of a library of DNA binding proteins, such as a library of sigma factors. Further aspects of the invention relate to the generation of a library of sigma factors including chimeric

sigma factors. Example 1 describes a non-limiting example of a library containing 86 ECF sigmas including 2 from each of the 43 identified subgroups.

Promoters that are recognized by sigma factors can be chimeric. As used herein, a chimeric promoter refers to a promoter formed from portions of two or more promoters. For example, in some embodiments, a chimeric promoter includes a -10 region from a first promoter and a -35 region from a second promoter, thereby generating a novel promoter. Chimeric promoters can include any combination of regions that are recognized by sigma factors. A promoter, including a chimeric promoter, can be a component of a library. Further aspects of the invention relate to the generation of a library of promoters including chimeric promoters.

Aspects of the invention relate to a sigma factor and a cognate promoter. As used herein, a cognate promoter for a sigma factor refers to a promoter to which the sigma factor specifically binds. A sigma factor can have more than one cognate promoter.

Aspects of the invention encompass sigma factor-promoter interactions that are orthogonal. As used herein, an orthogonal sigma factor-promoter interaction refers to an interaction that does not exhibit “cross-talk,” meaning that the sigma factor does not interfere with or regulate transcriptional regulatory elements in a system other than the transcriptional regulatory elements containing the cognate promoter of the sigma factor. In some embodiments ECF sigma factors autoregulate, meaning that an ECF sigma binds to a promoter of the gene encoding that ECF sigma, thereby regulating its own transcription.

Examples 1 and 2 describe a computational approach to identifying native promoters for subgroups of ECF sigma factors. Promoter modeling is described that predicts whether a promoter will be orthogonal with a given sigma factor or subgroup of sigma factors. In some embodiments, BioProspector is used, which allows the user to search for two sequence blocks (such as the -10 region and the -35 region) connected by a variable spacer (Liu et al. (2001) *Pac Symp Biocomput* 127-138). Example 1 demonstrates the identification of 706 promoters, including 29 promoter motifs in the 43 ECF sigma subgroups, and describing testing the orthogonality of such promoters. Examples 1 and 2 further demonstrate designing of orthogonal promoters for ECF sigma subgroups. As demonstrated in Examples 1 and 5 and FIGs 3 and 20-22, engineered chimeric sigma factors and promoters were designed and demonstrated to function orthogonally.

Further aspects of the invention relate to the identification of anti-sigmas. As used herein, “anti-sigma factor,” “anti-sigma” and “anti- σ ” are used interchangeably to refer to a polypeptide that binds to a sigma factor and inhibits its ability to activate transcription. Examples 1 and 3 describes identification of 63 anti-sigmas from a variety of bacterial species and the demonstration that many of these anti-sigmas could inhibit their cognate sigma factor in *E. coli*, demonstrating that anti-sigmas exhibit a high level of transferability between species. As used herein a cognate sigma factor for an anti-sigma refers to a sigma factor to which the anti-sigma specifically binds. An anti-sigma can bind to more than one sigma factor.

Aspects of the invention encompass interactions between anti-sigmas and sigma factors that are orthogonal. As used herein, an orthogonal anti-sigma - sigma factor interaction refers to an interaction that does not exhibit “cross-talk,” meaning that the anti-sigma does not interfere with or regulate elements in a system other than the elements containing the cognate sigma factor.

Significantly, anti-sigmas were demonstrated herein to act as sequestering molecules in switches containing sigma factors. Sequestering molecules can lower background, increase cooperativity and allow the threshold to be tuned within switches (Buchler et al. (2009) *Molecular Systems Biology* 5). FIGs 4, 19 and 24 reveal that the inclusion of anti-sigmas in switches comprising sigma factors results in decreased background and increased cooperativity. FIG. 18 reveals anti-sigma thresholding ability. Through sequestration, anti-sigmas serve to tune response characteristics within genetic circuits and facilitate the evolution of complex dynamics (Buchler et al. (2008) *J Mol Biol* 384:1106-1119).

Tables 1 and 2 describe anti-sigmas identified herein including providing sequence information for each anti-sigma.

Table 1: Sequences of Anti-sigma Factors

Anti-sigma	Plasmid	Cognate ECF sigma	SEQ ID NO.	Sequence
AS01_3 473	pVRc01_ 3473	ECF01_3473	1	MSNYDNAELLALWLEGKLTpkQRDEFEQRCIQDSQ FAEQVDAAAMVKIGADDYSSSEVPQWNKAATFEP QQANNSGSWLSGVSLVTSALAIVLVLTGTQVTTSEG ELRIRFGSGQSEQALVKLVDSKLDafKQNQQDAFTL YAQTLQQQQSESASQLTNYLLSSSRKERREDFaelI KFVNQQRSDDQLFYARQLNQLQQDVYHDATGVAL ESIQQ
AS01_4 085	pVRc01_ 4085	ECF01_4085	2	MNQSPQDLDSLKSLPLEMQPTTDLWPEITAQLSPQS QQKTHLTRPWLIAASVAVLSLLAMLLWQRPDGNLSL LPQTATITTTVPGATLNTEATLAESTLVELVDQIALT HQTQLDVFNQNYQTVSWQLSSTDAPQQIQSDISQAL AELDTASKQVQAALKQQPTNQQMWQLWRWIMQR QITLLQQGQKLpFTSKRTSQGNTI
AS02_9 15	pVRc02_ 915	ECF02_915	3	MEKTGQEWVSASVDGETDRQTlaELAADTDShak WQRYHLMGDAMRDELPQAMPLDLSARIAAALEDE PTILAPKVEREATQAPSRVVVPFMRQLGQYGIAAA VALMAVVGvQNYQSTQDDAPLPVLNTRPLVGTAT PVS LQTGPVANQNGNANDQLLEQRRRINAYLQDH MLQQRlNTGAVVDDNSEVTPIpVNR
AS02_2 817	pVRc02_ 2817	ECF02_2817	4	MQKEQLSALMDGETLDSELLNELAHNPemQKTWE SYHLIRDSMRGDTPEVLHFDISSRVMAAIEEEPVRQP ATLIPEAQpAPHQWQKMPFWQKVRPWAAQLTQMG VAACVSLAVIVGVQHYNQSETSQQPETPVFNTLP MMGKASPvSLGVPSEATANNGQQQQVQEQRrrIN AMLQDYELQRRLHSEQLQFEQAQTQQAaVQVPgIQ TLGTQSQ
AS03_1 198	pVRc03_ 1198	ECF03_1198	5	MDKDFDFDNIGKRTpYRTPDNFFeETQRKILERTVD EQRKKRRLKRIIPTVIAVAaVLagILFTPSLRyMNTD TPSASNILAVDKNNVTDPVDKwIKELSDEELeELV SFSENDIFLN
AS03_1 244	pVRc03_ 1244	ECF03_1244	6	MKQFDPDVNIRPKEDSIRHYSVPDDYFASFTDKLMA QIPAVKEENTVVVPSVALWPKLRPLLYLAASFLMI GMFKAFSLFGVGSdTGRttSTSAGLVALDHGDTRW SEDtdYRDFLHDNCAETVSDEWVLtDFSE
AS04_1 609	pVRc04_ 1609	ECF04_1609	7	MNRTTEQRNAEVMKTIGLLDQMPRVEVDHLFRVR LMQRIEAMEVKKTSWSALPGGAFNPRLAFMALLL MLNIASALMLFMHGTPQATGSSGAIAESLTEDYGGP ALSYYDDQTTIDR

AS05_9 65	pVRc05_ 965	ECF05_965	8	MSDQREPSERMITEAASWLALLQDEPVSAADRHA ERWRKADPGHQLALSRMQSLWGSFDELSDTPARIA LRQTFAPAGSKPTSRTVQALALVGVLCGWMSVER LPVWMADRHTDVGERSEFSLADGSQVQLNSGSALD VKFDGRQRVIELLQGELWVEVAKDVQRPFVVRTDQ GTITALGTRFVVRGEEGTTVSVLESAIAAQANTAD VINVATGQQALLKDGRVQTPHALGSDDPADWTRG VLKVVDDQPLSEVLQTLATYRHGLLRDYDTQALAGLR VSGVRLDDTDAALATLADNLPIKVERFTDLLVIVK PDAR
AS06_8 53	pVRc06_ 853	ECF06_853	9	MSNLPVSSRVLEAAIAWKLSLGESSGTPDERNEFMR WHAASEEHARAWRQLGAMDQRVSAAGPARQAL LQSRASLRRRIGKVGGGLAGTFLGALLAWVGAPS LAPSYWLADQRTATGELRTLRLLEDGTLLSLNHTA VDIEYAGAQRVIVLHQGEISVETGHQDPRPLLVRTE DGRRLRPLGTRFLVRREAGGTRLEVLQASVAAMP HD SGDEQVLREGQQVLMNANGLGEVGTVPAGADAW TRGMLVVDNVRLGDLLATLGQYRSGYLGVDAKVA DLRVTGSFPLTNTDLALASLVPALPVKIERHTQWW VNVTSK
AS06_3 576	pVRc06_ 3576	ECF06_3576	10	MSADDRHSPVRARVLDEAIAWQLLLDSGEAHPDDH RAFHRWYAAHPEHERAWEQLGGLDRHLARAANGP ARNALLSGNARFKRRLRRLGGSALGLVLALGVGLG VANRYVPVRYLLADAYSATGEQRELTPDATHVRL NSRSAIDVRFDGERRQVVLLAGEILVETAHGDP RPF VVSSADGDMRALGTRFLVRREEPGTRLTVLQSA VA ARAETLSEERVIKEGQQVLILPQGLQASEA APALAG AWAQGMLVVENARLADLVAELGRYSP ALLQVDPSI ADLRVTGSFPLKDTRLALQALEPS LPVRSVRHNAW WFEVVPR
AS07_1 134	pVRc07_ 1134	ECF07_1134	11	MNSPQEQQQIRQQA AEWAIRLDGGDLDRSRREALD GWLAADPRHRAALALAQR TWKQLGSLAEPR TMVQ TPVASAPRRAGGRKGW RGWAAAAAVLLALGSA WNERDAGVSWLADHSTGKGEVRILRLVDGSEVEL DAQSAIDVAYDSRERRVRLLEGSAIFRAAPRAGRET RPFVVESAGGSTALGTRFLVSRNDDGSGVQVGVLE HRVAVALAHPRTGTVGRRELGEGESLRYSAEGGVE APLGGRLDDLT SWRRGLLVFDEQPLGEVVARLNRY RPGHLLVAPGALAQRRVSGVFRVADLEASLQ SISDE LGVRSGLAGVTLLY
AS08_3 580	pVRc08_ 3580	ECF08_3580	12	MTASDSAADETGDLRHEAHAWVISLTSGRVTQGD A RAFRQWCARSPQHLRAFVEARDLWQALGSAAALP LEPPAVAQIAPRRFGRRWFGGALAASVALFVLRPS LLDHGLGGADYVTAVGEQRQVQVSGETRIEMNTRT RLNVRNQEQQETIELLGGEAEIIASHPPQSSLRVMA GSAWLSASRARFNVRSSGDVCVVTCLGSSVRLEHL GQRLDLQAGQQLTFDERRNGPPVPFDVAEVM AWR ERMLVFNDVPLATVIDEINRYRPGMLLLLDKALGR RRVQARFSLDQLADVATLIRDAYGIEVTRLPGGV VLLG

AS08_3 627	pVRc08_ 3627	ECF08_3627	13	MNIFSISTPQASADQQLNEARDWLVLLTSGQATAA DARALRQWCGQSPQHAAAFEQTKALWHCLQPA LLEQQARPRHFGRRAFLGGALAASAALFMVRLTVP GGFAGLTADFATEVGEQRRVDLAEGVSLELNTQTRI SRRDLGAGEQGIELLEGEVEVFSQRLQPLKVQAGEG WLSARQARFNLNRTDHQVCVTCIEGSLQVDVAGRS IGLDSGRQLTYDPRSIGEPQVVDIHSVIAWREQVLVF DNASLNTVISEINRYRPGMLVLLNAELGKRKVQARF NLNQLAGVALLIRDAYGAKCTELPGGVVLLS
AS09_1 009	pVRc09_ 1009	ECF09_1009	14	MWQAAMDWLLQCHSAPDDALLQAHARWLAAD ERHAWWRKAKEKVWLLSGGLAPLEPPVPQPLPTPL RARRNRPRRAKALALAAACLLLLAGPTPPTAHTSPA GEHRQVLLSDGSRIELGSDSAIRVDFEPGTRAVTLLR GQAFFEVSHDASRPFTVQAADV KVRVIGTAFDVDL SRTAVVVAVQSGAVQVRDGRGELAVPALGPGDSL RLGLDQGPQRGRLLPGQVAPWRQWQLLVNDRPL SEVVEALQDYYPGVLLLTDPALGERRITASLNLRS VSALQLAIAPLGHLRQWGPYLTIRKEPQVPAKQ
AS10_3 486	pVRc10_ 3486	ECF10_3486	15	MSNIHQFTPDKLILETAANWISAIDRGLNKVEKEQF KLWMLQSNHAHQDAVYELAKLWDELSVLNELSTLF PHKNNTTEEKSKWVFSYGIAASLFAALMICSYLLVNL ETGYNQALAKVNYTKIYKTKVGEQATYVLPDGTIV QLNTNSLLEVAYS KGRRQLLSRGEGRFNVAKD ATRPFSVMAGDKSFTALGTVFNVQRNTSSHEL LVTEGKVMITDPSVAVDANDFKAYQLADNSTQKIRKINA NIVLSGEKAIIEKSVTAPIKRLSADDVQRDLAWQNG MLIFNGEQLSDALNEVSRYTATRFELSSAELANIKV AGVFKAGDVAGLLES LKTNFSIDHERLGEHVVS LKRQTKS
AS11_9 87	pVRc11_ 987	ECF11_987	16	MNKHPDNNLLEAYASGSIDAVSGLVVATHLETCSK CRAVYNQVEASQANTVSESPSEYSPEFDDMLNDIIN AEPVNDNVVIQDTAFVNVAGKSFELPKTLVRFSDLV GSWRSYGGKVFS AQIDLGEDARVSLMYIGENVQIP QHTHRGLSTLV LHGGFSDGQYEEGDLMVRDAS VKHSPFTQEGEDCLCLTVLTEPMIFTQGVARIFNLFG KGLYP
AS11_3 726	pVRc11_ 3726	ECF11_3726	17	MGPLHHPDDATLISYAAGALSQVIAVVTA AHLERC AECRARLRQAEQIGGVLMQQSISRVPPLKSRMAML ARLDEQETSVDSVMHAMPAANHDPDLLPHCMHAH FGRHLSTLKWKTLPQVQVSAQGIEQGNLMMLKIA PGVSMPVHSHESGEMTMVLKGAYHDVQGEFGLND VADLD SHIQHQPIAYPDRECICVLAAESKLRFHGW MARMMPFFGI

AS12_8 07	pVRc12_ 807	ECF12_807	18	MRTDSTLTAAMDCRELERSIDAYLDGEFDERERAE AEAHLATCTPCRAMADRQGALRLALRAKLREAMA SPAAAGCAPPHLRARVRTSLAHERRPLWRRVLQPV PVAAVAACAMGVLVVLAGHRGDTALVDDAIRVHH RALPLEVDAAAMPGWFFAGKLDHFHPALPHFAGAVA RLEGARLSNLREWPAAYVRYQLPRGQAGLFIVDDP DRRFDTPGREVKVGPQVVRVNVNARGYNVAVWRQ DEIVYSLVSDLDEDALFKLVQAAQAEAAAAGR
AS12_8 08	pVRc12_ 808	ECF12_808	19	MNCQDLERLLYPYLDGEFQPEERVLDLETHLSGCAD CRRRAEEEEKMQQALRRAARHSVSGMRAPASLRA GIQVGLKQEQRrvQFGVWLRAGAMALVVVTVGG GWAAFHAEQRLSAARTEAVQRHSSKSKALPFEIASN TPEQVEEWFKDKVDPRITVPQIPKAKPLGGRISILNG REVAYISYETLPDNEGEPSRRLGVFVLPDNEVVIPK FQALQAVEVDSAQGFNVVTWRDDEIVYEMVTDMD ESDIRRMLAERDSGEKLARKSAPEADEPLYSLPPAP RTPHSWPPISVEPVITYPTYPQ
AS13_1 025	pVRc13_ 1025	ECF13_1025	20	MLSCYQATRLMSQALDEKIVLSQHVQLMLHLKICD GCRNFRQQLADLRTMTSAFARGENENQNKVT
AS13_1 146	pVRc13_ 1146	ECF13_1146	21	MNCEHASELISLNCEQKLKVKDSLQLQIHLWLCPKC RHFKKNNNEMRKLLQNYCDPKSNCEKET
AS14_1 324	pVRc14_ 1324	ECF14_1324	22	MSGSRPEPEGHLAEQHLGDRLSALVDGELGHDARE RVLAHVATCPKCKAEVDAQRRLLKNVFAEAAPPAPS ESFLARLQGLPGGSDSDGGGSPFSGPLPGGFASAAS GVFGPRRDERFEFDYVPAGSHTPVLP SATSGRGFRI HEVGRHESDRSASRGLRFAFAAAAGAVSLAAIALGG VTLGTPDTTTEARGSGSGSNVTPLRTPGSAAATGSE SQRRRTAGPLLGGQQRALGDLPVASTTASAPLLPG MPAPAGGDARQQAVRALTPVTAGAAAMSPLIRPL EAVPPLSLSSWSAAPEVRPPGLLAAPDPAPSPYPAAS PAASSSPLR
AS14_3 200	pVRc14_ 3200	ECF14_3200	23	MADPGSVGHVFRRAFSWLPAQFASQSDAPVGAPRQ FRSTEHL SIEAIAAFVDGELRMNAHLRAAHHLSLCA QCAAEVDDQSRARAALRDSHPPIRIPSTLLGLLSEIPR CPPEGPSKGSSGSSQGPPDGAAAGFGDRFADGDG GNRGRQSRVRR
AS15_4 36	pVRc15_ 436	ECF15_436	24	MAQSTEPKASPQEATAKGGDNTKGKSRKELQQQID ENLRRVYEEALVQEVDPDRFAMLLDQLRQKGTGK
AS15_5 24	pVRc15_ 524	ECF15_524	25	MNFGVEDMIEHVP MEDKRGKGAALDEARLRQQAI GVKLRQMFDEVVNVPDEF LAILRKAERPAGGE
AS16_9 73	pVRc16_ 973	ECF16_973	26	MKTDDLIALLAAGEGPVPRHAVGRRMAVAALGGL TAALLLTITLYGVRSDITEVAQTPLFWGKVAFTSL ALIGLWLT SRLARPGGKGAAGWKMLGLPLLLVWC GAAVSIAGAPVDARADLLFGRTWRTALNIALLSVP AFVTVFWALKGLAPTRLRLAGAAGLLAGSSATVA YCLHCPEMGVPFWGVWYVLGMLVPTVLGAWWGP RMLRW

AS16_3 622	pVRc16_ 3622	ECF16_3622	27	MMKTDELISLLATAEGPVDRHALARRLGLALLAGL LGALLLTVALYGVRSDLAEVARTPLFWAKVALPTS LALLGLWLTQRLARPGVRGGALWGLLGVPLLLWV LGAAISLFGAPPEARADLIFGRTWRTCALNITLLSTP VFIAVFWALRGLAPTRLRQAGAAGGLLAGSTATLV YCLHCPMGVPPFWGLWYLLGMLVPTLLGAVLGPR LLRW
AS17_1 458	pVRc17_ 1458	ECF17_1458	28	MQGTPAPNEHETVGAYALGILDDAEATAFEAHLAT CEWCAQQLDELAGEPMMAALADLPGTGTPAVAE SLTVKPSARLSEKLVEVAERRASKRRRNFYLVGT AAALIIGGPFAAVATTGGGGGGGDDGGGRRAEATQ QAASPAESAFAAMPDRVTATDPGTQVSATVALEKK AWGTETVLELKNLKGPKQCSLIAVGKNGERETLTS WSVPDWGYGIPGATTEKAKKPLYVHGGAAPFNQI SHFEVMTFDGKRLVEVDA
AS17_1 691	pVRc17_ 1691	ECF17_1691	29	MTMPLRGLGPPDDTGVREVSTGDDHHYAMWDAA YVLGALSAADRREFEAHLAGCPECRGAVTELCGVP ALLSQLDRDEVAAISESAPTIVASGLSPELLPSLLAA VHRRRRRRLITWVASSAAAVALAIGVLVGVQGH AAPQRAAVSALPMAQVGTQLLASTVSISGEWGTFI NLRCVCLAPPYASHDTLAMVVVGRDGSQTRLATW LAEPGHTATPAGSISTPVDQIAAVQVVAADTGQVLL QRS
AS18_4 438	pVRc18_ 4438	ECF18_4438	30	MSGTMSTPFPIDQEPPRDVLAGEYVLGLLSAEERLA AEQRIATDGGQFAQAVLQWQELLAPLLEEIVAQTPPD QVWVRVRQALGFDTPLRAVPSAAPVSTTAPAAPLW NSVRFWRWASVGGGLATAAVCVLALLNLRTPPAPV QPPHTGEVVQTPVTPPATNPPAATGIAMTSTLATED GRPGYVALMDADKHTITVTPLDRTATADKVPELWL ITPDGKAHSMGTFDDQRARRAQIPDQLMPMLSNEAI LAVTLEPPGGAPGGVATGTVVAKGGISTLAMAP
AS18_4 451	pVRc18_ 4451	ECF18_4451	31	MSGTMSTPFPIDQEPPRDVLAGEYVLGLLSADDRLA VEQRIATDAQFAQAVQWQEHAPLLEEIAAVTPA DQVWTRVRQALGFDTPLHAVPASGVQSPAPPAAVP LWNSVRFWRWASAGGLATAAVCVLALLSVRAPPT APPSGPVQNTPIVQTPPAKPPATGIAMTSTLATADGR PGYVALMDADKQVITVTPLDRTATAGKVPELWLIT PDGKAHSMGVFDDQRARRASIPAPLMPMLSNEAIL AVTLEPPGGAPGGVATGTVVAKGGISTLAMAP
AS19_1 315	pVRc19_ 1315	ECF19_1315	32	MRTEDLHSLTGAYALHALPDDEREAFERHLAGCAT CEQEAREFAAATARLGLAATVVPAPALRDRVLHRV TTVRQVPPGGGTAEKARRVVPRGRGLARWALAAC VAAAAGLGGTAVWQYERAQDAGQRAAQAEERRAE TLAGVLAAPDAESRTARLADGASGTLVVSERQDRA VFLASGMAEPPRGKVYQLWFDDHGTMRASGLMDP GSTSQAVLMDGAVDGAAGVGITVEPAGGSKQPTSD PIALLSMPA

AS19_3 197	pVRc19_ 3197	ECF19_3197	33	MTEHTDFELLELATPYALNAVSDDERADIDRRVAA APSPVAAAFNDEVRAVRETMAVVSAAATTAEP PAHLRTAILDATKPEVRRQSRWRTAAAFASAAIAVGLGA FGLGVLTRPSPPTVAEQVLTAPDVRTVSRPLGAGT ATVVFSRDRNTGLLVMNNVAPPSRGTVYQMWLLG GAKGPRSAGTMGTAAVTPSTTATLTDLGASTALAF TVEPGTGSPQPTGTILAEPLG
AS20_9 92	pVRc20_ 992	ECF20_992	34	MTPERFVHLADAYGADLQRWPSAERAAAQALLDC GDAQAVAALRQAHWLDSQLDRYQVPAPSPALAQ RIIAAAQQPGAPFWSRYAGWLASLGVVGVGLTGVA AGMLAVALSLPLSTSAEALPSVFDQSDAEFVLSINA EEAEQ
AS21_1 280	pVRc21_ 1280	ECF21_1280	35	MEEKELWMNKLKEKLG DYSEPLPASGWEQLEKEL MPPVERKIYPYRKWTVAAA VILLALGSSVSLYFLG TPAADEIRHAKTPALASVPDVLPAQ QPDMTGTTIE PVVRPVVKNRIAKAERNIPQPTANIDEPVKKEEQPSE LNAQTGDRKEKEEVEPVEETKAIRHKPADTEQPRN KPRRPSSRDKLHIPAEKASSQKGTWSMGLSVGNSG GASTELGSGIPSYMSRVSMVSVSNGLLSIPNDQQLV FEDGVPLYLRQANQVVDMEHHQPISFGLSVRKS LAKGFSVETGLTYTLLSSDAKFADSDQKTEQKLH YLGIP LKANWNFLDKKLFTLYVSGGGMIEKCVY GKLGTEKETVKPLQFSVSGAVGAQFNATKRVGIY VEPGVAYFFDDGSDVQTIRKENPFNFNIQAGIR LTY
AS22_1 147	pVRc22_ 1147	ECF22_1147	36	MELDDIAVAWRSLEQRLDQHAALAGQVLGDLRSH AARAHRLPLWLSQSAQLLCAIALSGLVAHSWLA FPEQAAAIVGGVLLQIWCVALAASAARQLWLLS QLDFARPLLQTQRALAQLRRWRTRVAPWLGVAF VWLWVAVADAAWRALTGRSLPYAWLLCNLLVG VLGGIGTWLGYRRLQRS GHPWLERLDTVHAGRS VARTETLLEQIARFQRE
AS22_4 450	pVRc22_ 4450	ECF22_4450	37	MELDDMKHAWQTLEQRLDQQA AQTGQLLG VVHEETVRSSLRPLWVAQTAQLL CALALAI VSARSWIPHTDQPVAVIGGVLLQAWCMALAI SAMVQLQLLTQFN VAGPLLRTQH ARLRR WRTRVAPWLGVAFVWLWIAVADALWRALTG QTLPTDWLVNL LVGVAGGITWLGFRRLQRA QHPWLERIDRAHAGTG VIRAERMLEEIARFQ RD
AS23_2 31	pVRc23_ 231	ECF23_231	38	MDDKNLFRNYNDIVVSEEEIEKYETKKVDD DVL SKMKNKVKKLYMKVNMEEA FEKVEKQ FEDDEKVEYMFWAETYGVRKYQMVC GGY YSLEGAISNDWG TKTGIVLTNKGIFGIET NDAYGV LKIKNFRFKDVEYIESKKIKNFT VFAIKSTSGIEIKVEIYNGDRHIKFLNYIR NNNIKVNIRMIQDRKIQIAYVSIIIMIFIH FVISSSIMRSGITK

AS23_1 851	pVRc23_ 1851	ECF23_1851	39	MKDIYELLNDIDIDEKELEEIEASEIEKEKVKRVKQ SIRTKKKMKSWKKGVAAASILVGLSVTTLGIGFPTY AGGLPIVGDI FRFLDNGR TGLYENYKEFSTELNMTR ESNGVKVTINDVISDGR TLSITYSLESEQDLGDDPIIL GGLDIMDAHGSSSGSGKMTKVTEKKYVGMVTTTHH DSNKKDKVNFRWNIEGIEIPDRKKS IQGHWNFALT V KSMSDKERTIGGSSEKEGIKANMEKVAMSPVSFILY YNQEVSKGARKEWDSVDVELTVKDDLGN DYSGEG NGGSGNDPYNIRWSATFQKL NENATKLIVTPRVHL RVHTPENHGGVEYVNGKEKKIEVPNKEAKKKDIVL DDIVIDLKK
AS25_1 643_clo ne433	pVRc25_ 1643_clon e43	ECF25_1643	40	MTTDSNFNDPCRRQFSRDL P QKMARHTNEATGAM DMVKRDRFELL SAYLGGEV TAAERRQVEDWPAND VAVQRLYSRL LKLRQGIR TMPIPTAQ QSPETTAEQV FAKVNRRSRLAWKLGGAAVAACVIGAVTNWLPGR QTGIPQLAQQPQEPTQAVTTPDALPMIALNNPVIEI PKAAVASPTKFIQPQPQLGEIPPDIN
AS25_1 643_clo ne440	pVRc25_ 1643_clon e44	ECF25_1643	41	MTTDSNFNDPCRRQFSRDL P QKMARHTNEATGAM DMVKRDRSELL SAYLDGEV TAAERRQVEDWLAND VAVQRLYSRL LKLRQGIR TMPIPTAQ QSPETTAEQV FAKVNRRSRLAWKLGGAAVAACVIGAVTNWLPGR QTGIPQLAQQPQEPTQAVTTPDALPMIALNNPVIEI PKAAVASPTKFIQPQPQLGEIPPDIN
AS25_1 645	pVRc25_ 1645	ECF25_1645	42	MNESFSQRQTDGDRRRGGDLEMKKNHTSHQHHTD SETPMGHNF EQFQRL SAYFDGEATPAERKEIQHLLD TDPQVKQYQQLRQLKQALQLLPIPTSISAQYLGQR VLARLRRSQLRTL SLWGSGAIAALFVAGVMGQMPR LNFDRFAKNDPDQQPTAALVETSPQGEEALVVALN RPVLQIPKLATTESP
AS26_8 37	pVRc26_ 837	ECF26_837	43	MNAPSDEQLVAYLDDELDREQRSQLDNL IADDP LL SLRVQWLSRSSLPFKAAYDELAQQAPLDRLQARLD AAPSPQKP VFSRRWFIGAAAAGVALAAVAADRLFL AWQAQQSHNWREL VGDYMAL YVPQTLEHLPTDEA SQLAQLRTVDARLGVSLSPAKLKLPGAQFKRAQLL EYGGVPIAQMTWLD AKYGPLALCVTRTNSGSQPLA HERRHGMNVVYWTEREHAWMLIGHHPASELED M AKMFKTRLNV
AS26_4 464	pVRc26_ 4464	ECF26_4464	44	MKDIDESTLLAYADGALTPDQAGRVEAVLAADPQR AADVRQLQQVKARLRNGYASVLEEPIPAHLLDAAR QRPPSPQTSVVTATAPIQAPATRQHATRRWAVPTSI AAALLIGLWLWQRQPAQPAPSALLAEQGH DASGTL ALALDRQLSGEQQGKIRMGLSFRAHDGRYCRS FSL QSSHAGLACRQGERWR IEAVSPLQPQRNDSEL RMA SSTLPAALLDAIDARIDGQALDAEGERSARARHWR

AS27_1 331	pVRc27_ 1331	ECF27_1331	45	MSAADKDPDKHSADADPPLTVELLADLQAGLLDD ATAARIRSRVRS DPQAQQILRALNRVRRDVAAMGA DPAWGPAA RPAVVDSISAALRSARPNSSPGA AHAA RPHVHPVRMIAGAAGLC AVATAIGVGAVVDAPPPA PSAPTTAQHITVSKPAPVIPLSRPQVLDLLHHTPDY G PPGGPLGDPSRRTSCLSGLGYPASTPVLGAQPIDIDA RPAVLLVIPADTPDKLAVFAVAPHCSAADTGLLAST VVPRA
AS27_4 265	pVRc27_ 4265	ECF27_4265	46	MTGHPDVAEIADLAEGLLPTTRTTEVRQHLESCELC ADVYASL TEIQGLLGTL PAPAMPDDVAARIDAALA AEPLGIADGTRVSRETSTPADRPAGHARPSSTGPGR KDRRRGGRRRIAVLGAVAAAAAIGIGSVVVSSLTED SSSGNTAREQQTAIADTFSEGRLKDRVTNLVADGSA ENGSRTPRSFGMESENGGETAENHVFKQPTVPECIR KGIGRDDAVIATEPGVYKGREALLVVLPA DTNDTQ VTAYIVETACVDQPAVGKAKILLEHSYARS
AS28_1 040	pVRc28_ 1040	ECF28_1040	47	MVQNTTRSSKMDELEFRRKVMSPKQRDNDTLDM MTSSEANAKFVDDVLQLDKQIAQAFKVDVPDDL A DKILFKQTTLVEDEKVIRPQFVRKAMAIAASVAFTA GLLVGQIQWGNLLISPAQASLSDMAVQHVIHEEGFV NRLDEQADMQQINAKMRPFAYKMEGDFPYHVYYL NHC GFKDN AVH MVFQGEKGKVTLFFTPIHSAQSS L FKQEGMAGHIEPVGNASLILVGEKDENLTNIANKL MPMIQSSI
AS28_1 088	pVRc28_ 1088	ECF28_1088	48	MDDLQFRRHAYGDPNNQADDFLAHLAENEDDAKF VKDLQAFDHKL TQALNISVPDGLADKLILRQQLSQH QKSKKQTRYLMAMAASVAFIVGVSFSLLRFTP VNL GENSLAHVHHETKALVMEQDIGFNDVNFKLASLEG LSDSKFIQQPGRVFYTSYCDFQG VKSLHLMADEN GNKVTLFIVPVESRIVLEEA FADNQYKGQSFQTADA YMLVVGEPASDLEFVKKEVENTFI
AS30_3 5	pVRc30_ 35	ECF30_35	49	MDKRLQQLREEYKNVQIPKELDIIVEKALQQEPKKK RIVMWPTSA AIAAAILFTALVNINPDAAQAMSKIPVI GKIVKAITFIEIKEEKDQSSIDVKTPALSGLSNKELEN SINEKYLKESQQLYKEFIQSTSKNKKGHLSIYSDYET VTDTPDLLSIRRNIE TTQASSYTQSRYTIDKKNDILL TLKSLFKDERYIKVISQNIKEQMKQMKEDPNKIYW LTDEDAEPFKTILPDQTFYITEDHKLVISFDEYEVAP GYMGVTEFTIPTGVISNLLVGERYIR

AS30_8 3	pVRc30_ 83	ECF30_83	50	MIDDFDRKLFEMARESKVKEPNALKYKVDYTFKKL KKNKFNFRHLGSI AAILIFCILSVGIYFPTYAMNIPILG DVVEILSNKFNLSGYEINAQNLNYQVSNEYDTLTIES AYYNGLETTFFFKIKGNAKLNKSGQYFFEANFKYN EDISYEGGLEKGEFIDYTFAGMMTFYINPYSESKLP EKFNIKFSIPNIIADSEILAVNSDTLNL SFDITDLNVKE TKINKEIQANENSILISSIKKYPT SIVIDYDEKFNNPEN KLSFILWHETLGQINYLLPSTPGKLFIVPKTRSKLNV DNLIKESMPLSIGETKTFGKVGVKVSINIETKDGKTY ISIRKTGDINSYDFNIIKENINSKLNMYEYETKVIGI LDTLT TYVIPDFTSDIDYLLEYEYISNDDIEILYDQIIE IN
AS31_3 4	pVRc31_ 34	ECF31_34	51	MNKEKLS DHLKSEWKKIDQTANPSIPNQKELLHQLS QMKA EYRKLLQEILFVFCALMVVSAAILAFTQAP AVFIVLQVCVLAVLPILIAAEKKRHLGECEVKRG
AS33_3 75	pVRc33_ 375	ECF33_375	52	MMALSKKMLEQEPSEIELLPWHAAGTLNARDARR VEDALARDPELAKQYAAIRGEYEETIHLNESLGAPS ARAMQKLFGAIDAEPARETGSLPLSARIATFFASLSP RTLAWASASLGAVALVLQAGIIGAVLMKTQPTTFQT ASLSTSAPITRELGA AVAPARALVRFTPEARVADITA LLDSYQASIIGDAKGGMFRLQFDKAMSQDELASLL GRMQREKFVNLA VAAP
AS33_4 23	pVRc33_ 423	ECF33_423	53	MMAMSNTMPDPREP GDVEALLPWYAAGTLNARD ARRVADALDRDPALARQYAVILEEYASTIELNESLG APSSRAMQKLFAAIDAEPARAPGAGQGLGARFAGF FAGLSPKALAWSASVAGLALLLQAGLIGALLTWP AAPVQTAAYQPQREVARAPASSPAPATVSPAAAM ADRAADAGKSTPPMVMAERSGGPVVRS LAPQSGPR VLVKFAPEARASEIAALLDQYNAV VVDSSRGGLFR LQFGTQSLSKQDQETLIGRLQKEPVVSVVLSAP
AS34_1 384	pVRc34_ 1384	ECF34_1384	54	MGHVHPSHLVELALGHASGEADV GALRHAASCPR CREELLRLTRVVT AARGAEASDLPVPPPERVWQRIA LEVLPETDRVPRLRESSAHGSADERVRSQRRWTD HAGEGLLGLALAI AVL LRRWRIRAGSGN
AS34_3 302	pVRc34_ 3302	ECF34_3302	55	MKRKENKAWIDLALENLSEKELKELRQN LASDPQF QEELVSVKELLAAIALNLEPEPPAPELKARLLDGICG KNRFLPFLSRLTELF DLSPREAQAYLERLDDPTAWK TVLPGVQTIKIQAGPATAGAKSNFLRVLP GASFPYH THRGLESSLLQGCCRS EDGVINRAGDLLYQETGTA HSFQVISEQAVIAAVVCFGIDFINPPDRKR
AS35_1 119	pVRc35_ 1119	ECF35_1119	56	MSNSTLSSEEQALHQEVSAWYFRQALEMPPERLDQ DILRLAQTQLSERNVSQLTPSAMP IWRFPWVLSSA ASLVIVVGLVMLNRGQFEEDMGAPAALTMSAPMP AAHVASDVADAKVQEAEMASQARLVEDTAKQNAP KEMMMAQANMAAEENIQA KSRSLPQVARAHPEGD VQATANTDTAALMLSLARLQELIESKQIQEALALEQ TLVKQYPELSHISSAKVAADDAKAI AKFKALQQQL HPLRN

AS35_3 582	pVRc35_ 3582	ECF35_3582	57	MNPQKHSASVAEEQMLAHFRAHAPQQPAPALDQAI LAAARRQAAHVEPARSWRRWLEASRRPRWQAA FASLLGVALVLGLVSHNVLDDAERQARPEVAFSDV PLRDGVAGAAAAKRAMRAPAAPAPLSGEMSEPPAL LASYASSGEAPQLMAEAAPPAPAAALADRPPAQAAQ QAKVQAALAGDFVAQARGKAVAVKPEVLDEALGA VLALREQGKTEQAATQLAELQKRFPGENLVERLER LATIAASARKRP
AS37_2 513	pVRc37_ 2513	ECF37_2513	58	MSSAPATKEEVDFAFKVRRALDERAASLPDATTD LAAARRAALARKKPDAAIVLVPALAGSAGTLELRP PGEPRKSLARRLARAWPLALLLAGLIGIAYWEDMQ RTAELADIDAAMLSDNLPLTAYLDHGFNAYLSHTH
AS38_1 322	pVRc38_ 1322	ECF38_1322	59	MSDGNDARPDARDGHDPGSDARDGHDPGSHARDG HDPGAGPFDGHHGPGSDAGDGDDAGPDARDGHGAP AGPAGREQRKHPHDTHDMHDTHDTHAGKGT VNHPGAAQGPDPADGPATDEAALRAMMQRAVREM EPSDGTLEHLRRAVPARRARKRQALVGAAAAALFL GTAVPALVHVSNTGAGADPSVAGNASQAQGGAS QGKDPAGGQSGVAGTGDTPEDRDKADPKETPGGK EPGAATGAPPSGVPSASSPADVPACAPGSLGPAVAS SAEPDSTGVVYGSFRVTNVSSDGCTVTGPGTVVTAS LGAAEATRIGTARHAAGDAAAGLPDPSLETASLAL APGAAYEVQFAWVPSETCPTTGGTTGGGSGGSPD PSPTADTTAAGGTSAGGGEAGPTTQLITEDGPAEGS VSVTYTPEGGSGSATATVSNACAGTVYWTGLLADS GSGA
AS38_1 442	pVRc38_ 1442	ECF38_1442	60	MSDRTPLGPLDPDPDGDGELSPTARRLREALAARAA GVHPTDRLEEHVTSRADRRRSRTRAVVAAAGVAA VVVVGGGGYALAQRDGGSVRTVAGSPAAAPASST TTAAGAGTPGATAPAAAPAPATASGSTGPATAAT SAPATSTPTGAAAPALPTGAARVPVYWTGGGKLFR EFTPVPGGRDDATNALQVLLGGTAADADYRTSWG VDPAAEVTRDGSAGYVVDVSAAAVSTPLSAPEAEL AVQQLVHTVTAAGGGSAPVRLLDGREGATVFGS HRVPAEVERAPQVDVQAPAWITQVTPGAGSVTVA GVGTAFEGTLLCTLDAAGVEVAREPVQAGANGTF GEFSLAVAAPAGTYTVAVFAPDESGGEGPVAVGDT KTVTVR

AS39_1 438	pVRc39_ 1438	ECF39_1438	61	MLENERQDPFEDRLGTALRDAGDGFADRAALVTA GRARGRRALLRRRAAVVGGVAGVALAGVGGVLVL PADHPAGPDRSGTASAPSAGDATTAAASFTGDDL HELKGLLPPTYGEESARGSDHQLGPTAQLVYDDG AGAAAIGMGFARVEPGSAQVRELMACPDHNTIPYD DCSSDRLPDGSLKLYQGYEYPDLRVDTKRWTADL VTAEGQHVSSEWNSPAEEKGAPVSREEPPLSTERLR ELVTAGVWREVVDAVPKSRKPPRSAAPRTERPEVS GKSVGDTLAALLPRKLDVVSRRGGQSEYAYVVVD DGRGRSLVQINVQHGMADVAGQLYADGETLPDGT RVATRQGPGEKAGSGVVMWTVDTLRPGPAGFRVV ISAFNTGDQNKDATRDAPALTMEQLRKIALSGEWD RLR
AS40_1 380	pVRc40_ 1380	ECF40_1380	62	MTRRLHGGEQDQGEHVKGQLKQLFDDDAFLTDLS RGVDPSEGGDALAGLLDLTKEAQEPPATMPDWST LLPGILDQDQDLPVESTSDTTVMQASNPATQEFAPV SISDTPNTATNSADADESATVVPLAARREKRAKSGS SGVHSLDASATQRKSHPFSLVGAATAATLVIAGG GAAVYNADENSPLYGMNQQLFGNQDSPSVVELAST LEEVDSTASGDVEGARALLEQARAMLDGMAPPR KAPSEATRTVESEPGTQTLTATVTESASPEPPVTETQ TVTSTEVQTVTTTAVAPPVWTPNPEPTTTAAPTSTPS TGGGEGTGNDGDSGLVPPQTPGN
AS40_3 198	pVRc40_ 3198	ECF40_3198	63	MREFGNPLGDRPPLDELARTDLLLDALAEREEVDF ADPRDDALAALLGQWRDDLWPPASALVSQDEAV AALRAGVAQRRRARRSLAAVGSVAAALLVLSGFG AVVADARPGDLLYGLHAMMFNRSRVSDQIVLSA KANLAKVEQMIAQQGWAEAQDELAEVSSTVQAVT DGSRRQDLINENLLNTKVETRDPNATLRPGSPSNP AAPGSVGNSTWPLAPVVEPPTPPTPASAAEPSMSAG VSESPMPNSTSTVAASPSTPSSKPEPGSIDPSLEPADE ATNPAGQPAPETPVSPH

Table 2: Classification information for anti-sigmas presented in Table 1

Anti-sigma	SPECIES	CLASS	PHYLUM	GENOME
AS01_3 473	<i>Pseudoalteromonas atlantica T6c</i>	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudoalteromonas atlantica T6c_uid58283</i>
AS01_4 085	<i>Shewanella frigidimarina NCIMB 400</i>	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Shewanella frigidimarina NCIMB_400_uid58265</i>
AS02_9 15	<i>Shewanella amazonensis SB2B</i>	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Shewanella amazonensis SB2B_uid58257</i>
AS02_2 817	<i>Escherichia coli K12</i>	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Escherichia coli K_12_substr_MG1655_uid57779</i>
AS03_1 198	<i>Bacteroides thetaiotaomicron VPI-5482</i>		<i>Bacteroidetes</i>	<i>Bacteroides thetaiotaomicron_VPI-5482</i>

AS03_1 244	<i>Porphyromonas gingivalis</i> W83		<i>Bacteroidetes</i>	<i>Porphyromonas_gingivalis_W83_uid57641</i>
AS04_1 609	<i>Chlorobium tepidum</i> TLS	<i>Chlorobi</i>	<i>Chlorobi</i>	<i>Chlorobium_tepidum_TLS_uid57897</i>
AS05_9 65	<i>Pseudomonas syringae</i> pv. tomato str. DC3000	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas_syringae_tomato_DC3000_uid57967</i>
AS06_8 53	<i>Pseudomonas putida</i> KT2440	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas_putida_KT2440_uid57843</i>
AS06_3 576	<i>Pseudomonas aeruginosa</i> PAO1	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas_aeruginosa_PAO1_uid57945</i>
AS07_1 134	<i>Pseudomonas aeruginosa</i> PAO1	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas_aeruginosa_PAO1_uid57945</i>
AS08_3 580	<i>Pseudomonas aeruginosa</i> PAO1	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas_aeruginosa_PAO1_uid57945</i>
AS08_3 627	<i>Pseudomonas fluorescens</i> Pf-5	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas_fluorescens_Pf_5_uid57937</i>
AS09_1 009	<i>Pseudomonas fluorescens</i> Pf-5	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas_fluorescens_Pf_5_uid57937</i>
AS10_3 486	<i>Pseudoalteromonas haloplanktis</i> TAC125	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudoalteromonas_haloplanktis_TAC125_uid58431</i>
AS11_9 87	<i>Vibrio parahaemolyticus</i> RIMD 2210633	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Vibrio_parahaemolyticus_RIMD_2210633_uid57969</i>
AS11_3 726	<i>Pseudomonas syringae</i> pv. tomato str. DC3000	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas_syringae_tomato_DC3000_uid57967</i>
AS12_8 07	<i>Anaeromyxobacter dehalogenans</i> 2CP-C	<i>Deltaproteobacteria</i>	<i>Proteobacteria</i>	<i>Anaeromyxobacter_dehalogenans_2CP_C_uid58135</i>
AS12_8 08	<i>Myxococcus xanthus</i> DK 1622	<i>Deltaproteobacteria</i>	<i>Proteobacteria</i>	<i>Myxococcus_xanthus_DK_1622_uid58003</i>
AS13_1 025	<i>Photorhabdus luminescens</i> subsp. laumondii T101	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Photorhabdus_luminescens</i>
AS13_1 146	<i>Haemophilus ducreyi</i> 35000HP	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Haemophilus_ducreyi_35000HP_uid57625</i>
AS14_1 324	<i>Streptomyces coelicolor</i> A3(2)		<i>Actinobacteria</i>	<i>Streptomyces_coelicolor_A3_2_uid57801</i>
AS14_3 200	<i>Mycobacterium tuberculosis</i> H37Rv		<i>Actinobacteria</i>	<i>Mycobacterium_tuberculosis_H37Rv_uid57777</i>
AS15_4 36	<i>Rhodobacter sphaeroides</i> 2.4.1	<i>Alphaproteobacteria</i>	<i>Proteobacteria</i>	<i>Rhodobacter_sphaeroides_2_4_1_uid57653</i>
AS15_5 24	<i>Caulobacter crescentus</i> CB15	<i>Alphaproteobacteria</i>	<i>Proteobacteria</i>	<i>Caulobacter_crescentus_CB15_uid57891</i>
AS16_9 73	<i>Pseudomonas putida</i> KT2440	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas_putida_KT2440_uid57843</i>
AS16_3 622	<i>Pseudomonas entomophila</i> L48	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas_entomophila_L48_uid58639</i>
AS17_1 458	<i>Streptomyces coelicolor</i> A3(2)		<i>Actinobacteria</i>	<i>Streptomyces_coelicolor_A3_2_uid57801</i>
AS17_1 691	<i>Mycobacterium tuberculosis</i> H37Rv		<i>Actinobacteria</i>	<i>Mycobacterium_tuberculosis_H37Rv_uid57777</i>
AS18_4 438	<i>Xanthomonas axonopodis</i> pv. citri str. 306	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Xanthomonas_axonopodis_citri_306_uid57889</i>

AS18_4 451	<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. ATCC 33913	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Xanthomonas campestris</i> _A TCC_33913_uid57887
AS19_1 315	<i>Streptomyces coelicolor</i> A3(2)		<i>Actinobacteria</i>	<i>Streptomyces coelicolor</i> _A3_2_uid57801
AS19_3 197	<i>Mycobacterium tuberculosis</i> H37Rv		<i>Actinobacteria</i>	<i>Mycobacterium tuberculosis</i> _H37Rv_uid57777
AS20_9 92	<i>Pseudomonas fluorescens</i> Pf-5	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas fluorescens</i> _Pf_5_uid57937
AS21_1 280	<i>Bacteroides thetaiotaomicron</i> VPI-5482		<i>Bacteroidetes</i>	<i>Bacteroides thetaiotaomicron</i> _VPI-5482
AS22_1 147	<i>Xanthomonas axonopodis</i> pv. <i>citri</i> str. 306	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Xanthomonas axonopodis</i> _citri_306_uid57889
AS22_4 450	<i>Xanthomonas campestris</i> pv. <i>campestris</i> str. ATCC 33913	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Xanthomonas campestris</i> _A TCC_33913_uid57887
AS23_2 31	<i>Clostridium acetobutylicum</i> ATCC 824	<i>Firmicutes</i>	<i>Firmicutes</i>	<i>Clostridium acetobutylicum</i> _ATCC_824_uid57677
AS23_1 851	<i>Bacillus anthracis</i> str. Ames		<i>Firmicutes</i>	<i>Bacillus anthracis</i> _Ames_uid57909
AS25_1 643_clone433	<i>Nostoc</i> sp. PCC 7120		<i>Cyanobacteria</i>	<i>Nostoc</i> _PCC_7120_uid57803
AS25_1 643_clone440	<i>Nostoc</i> sp. PCC 7120		<i>Cyanobacteria</i>	<i>Nostoc</i> _PCC_7120_uid57803
AS25_1 645	<i>Synechococcus</i> sp. PCC 7002		<i>Cyanobacteria</i>	<i>Synechococcus</i> _PCC_7002_uid59137
AS26_8 37	<i>Pseudomonas fluorescens</i> PfO-1	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Pseudomonas fluorescens</i> _PfO_1_uid57591
AS26_4 464	<i>Xanthomonas oryzae</i> pv. <i>oryzae</i> KACC10331	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Xanthomonas oryzae</i> _KACC10331_uid58155
AS27_1 331	<i>Mycobacterium bovis</i> AF2122/97		<i>Actinobacteria</i>	<i>Mycobacterium bovis</i> _AF2122_97_uid57695
AS27_4 265	<i>Streptomyces coelicolor</i> A3(2)		<i>Actinobacteria</i>	<i>Streptomyces coelicolor</i> _A3_2_uid57801
AS28_1 040	<i>Vibrio cholerae</i> O1 biovar eltor str. N16961	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Vibrio cholerae</i> _O1_biovar_El_Tor_N16961_uid57623
AS28_1 088	<i>Shewanella frigidimarina</i> NCIMB 400	<i>Gammaproteobacteria</i>	<i>Proteobacteria</i>	<i>Shewanella frigidimarina</i> _NCIMB_400_uid58265
AS30_3 5	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168		<i>Firmicutes</i>	<i>Bacillus subtilis</i> _168_uid57675
AS30_8 3	<i>Clostridium perfringens</i> str. 13		<i>Firmicutes</i>	<i>Clostridium perfringens</i> _13_uid57681
AS31_3 4	<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168		<i>Firmicutes</i>	<i>Bacillus subtilis</i> _168_uid57675

AS33_3 75	<i>Bradyrhizobium japonicum</i> USDA 110	Alphaproteobacteria	Proteobacteria	<i>Bradyrhizobium japonicum</i> _USDA_110_uid57599
AS33_4 23	<i>Rhodopseudomonas palustris</i> CGA009	Alphaproteobacteria	Proteobacteria	<i>Rhodopseudomonas palustris</i> _CGA009
AS34_1 384	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria	<i>Streptomyces coelicolor</i> _A3_2_uid57801
AS34_3 302	<i>Nitrosococcus oceani</i> ATCC 19707	Gammaproteobacteria	Proteobacteria	<i>Nitrosococcus oceani</i> _ATC_C_19707_uid58403
AS35_1 119	<i>Shewanella oneidensis</i> MR-1	Gammaproteobacteria	Proteobacteria	<i>Shewanella oneidensis</i> _MR_1_uid57949
AS35_3 582	<i>Pseudomonas aeruginosa</i> PAO1	Gammaproteobacteria	Proteobacteria	<i>Pseudomonas aeruginosa</i> _PAO1_uid57945
AS37_2 513	<i>Burkholderia thailandensis</i> E264	Betaproteobacteria	Proteobacteria	<i>Burkholderia thailandensis</i> _E264_uid58081
AS38_1 322	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria	<i>Streptomyces coelicolor</i> _A3_2_uid57801
AS38_1 442	<i>Kineococcus radiotolerans</i> SRS30216		Actinobacteria	<i>Kineococcus radiotolerans</i> _SRS30216_uid58067
AS39_1 438	<i>Streptomyces coelicolor</i> A3(2)		Actinobacteria	<i>Streptomyces coelicolor</i> _A3_2_uid57801
AS40_1 380	<i>Corynebacterium glutamicum</i> ATCC 13032		Actinobacteria	<i>Corynebacterium glutamicum</i> _ATCC_13032_Bielefeld
AS40_3 198	<i>Mycobacterium tuberculosis</i> H37Rv		Actinobacteria	<i>Mycobacterium tuberculosis</i> _H37Rv_uid57777

Aspects of the invention relate to recombinant genetic circuits. As used herein “recombinant” and “heterologous” are used interchangeably to refer to a relationship between a cell and a polynucleotide wherein the polynucleotide originates from a foreign species, or, if from the same species, is modified from its original (native) form. As used herein, a genetic circuit refers to a collection of recombinant genetic components that responds to one or more inputs and performs a specific function, such as the regulation of the expression of one or more genetic components and/or regulation of an ultimate output of the circuit. In some embodiments, genetic circuit components can be used to implement a Boolean operation in living cells based on an input detected by the circuit.

Aspects of the invention relate to recombinant cells that comprise logic functions that influence how each cell responds to one more input signals. In some embodiments, a logic function can be a logic gate. As used herein, a “logic function,” “logic gate” or “logic operation” refers to a fundamental building block of a circuit. Several non-limiting examples of logic gates compatible with aspects of the invention include AND, OR, NOT (also called INVERTER), NAND, NOR, IDENTITY, XOR, XNOR, EQUALS, IMPLIES, ANDN and N-IMPLIES gates.

The use of Logic Gates is known to those of skill in the art (see, e.g. Horowitz and Hill (1990) *The Art of Electronics*, Cambridge University Press, Cambridge). Genetic circuits can comprise any number of logic gates. In some embodiments, NOR gates can comprise a transcriptional repressors and a transcriptional repressor target DNA sequence, while AND gates can comprise a transcriptional activator and a transcriptional activator target DNA sequence.

Genetic circuits can be comprised of one or more logic gates that process one or more input signals and generate an output according to a logic design. In some embodiments, genetic components respond to biological inputs and are regulated using combinations of repressors and activators. Non-limiting examples of logic gates using genetic components have been described (Tamsir et al. (2011) *Nature* 469(7329):212-215). In some embodiments, the genetic circuit functions as, for example, a switch, oscillator, pulse generator, latch, flip-flop, feedforward loop, or feedback loop.

Genetic circuits can comprise other components such as other transcriptional activators and transcriptional repressors. Non-limiting examples of transcriptional activators and transcriptional repressors are disclosed in and incorporated by reference from WO 2012/170436 (see, e.g., pages 27-40; Table 1 on pages 28-30; and Tables 2 and 3 on pages 36-38, of WO 2012/170436).

Aspects of the invention relate to recombinant host cells that express genetic circuits. It should be appreciated that the invention encompasses any type of recombinant cell, including prokaryotic and eukaryotic cells. As used herein, a "host cell" refers to a cell that is capable of replicating and/or transcribing and/or translating a recombinant gene. A host cell can be a prokaryotic cell or a eukaryotic cell and can be *in vitro* or *in vivo*. In some embodiments, a host cell is within a transgenic animal or plant.

In some embodiments the recombinant cell is a bacterial cell, such as *Escherichia* spp., *Streptomyces* spp., *Zymonas* spp., *Acetobacter* spp., *Citrobacter* spp., *Synechocystis* spp., *Rhizobium* spp., *Clostridium* spp., *Corynebacterium* spp., *Streptococcus* spp., *Xanthomonas* spp., *Lactobacillus* spp., *Lactococcus* spp., *Bacillus* spp., *Alcaligenes* spp., *Pseudomonas* spp., *Aeromonas* spp., *Azotobacter* spp., *Comamonas* spp., *Mycobacterium* spp., *Rhodococcus* spp., *Gluconobacter* spp., *Ralstonia* spp., *Acidithiobacillus* spp., *Microtholunatus* spp., *Geobacter* spp., *Geobacillus* spp., *Arthrobacter* spp., *Flavobacterium* spp., *Serratia* spp., *Saccharopolyspora* spp., *Thermus* spp., *Stenotrophomonas* spp., *Chromobacterium* spp., *Sinorhizobium* spp.,

Saccharopolyspora spp., *Agrobacterium* spp. and *Pantoea* spp. The bacterial cell can be a Gram-negative cell such as an *Escherichia coli* (*E. coli*) cell, or a Gram-positive cell such as a species of *Bacillus*.

In other embodiments, the cell is an algal cell, a plant cell, an insect cell or a mammalian cell. In certain embodiments, the mammalian cell is a human cell.

In some embodiments, multicellular systems described herein contain cells that originate from more than one different type of organism.

Aspects of the invention relate to recombinant expression of one or more genes encoding components of genetic circuits. It should be appreciated that some cells compatible with the invention may express an endogenous copy of one or more of the genes associated with the invention as well as a recombinant copy. In some embodiments, if a cell has an endogenous copy of one or more of the genes associated with the invention, then the methods will not necessarily require adding a recombinant copy of the gene(s) that are endogenously expressed.

According to aspects of the invention, cell(s) that recombinantly express one or more components of genetic circuits are provided. It should be appreciated that the genes associated with the invention can be obtained from a variety of sources. As one of ordinary skill in the art would be aware, homologous genes for any of the genes described herein could be obtained from other species and could be identified by homology searches, for example through a protein BLAST search, available at the National Center for Biotechnology Information (NCBI) internet site (ncbi.nlm.nih.gov). Genes associated with the invention can be PCR amplified from DNA from any source of DNA which contains the given gene. In some embodiments, genes associated with the invention are synthetic. Any means of obtaining a gene associated with the invention are compatible with the instant invention. Aspects of the invention encompass any cell that recombinantly expresses one or more components of a genetic circuit as described herein.

One or more of the genes associated with the invention can be expressed in a recombinant expression vector. As used herein, a “vector” may be any of a number of nucleic acids into which a desired sequence or sequences may be inserted, such as by restriction and ligation, for transport between different genetic environments or for expression in a host cell. Vectors are typically composed of DNA, although RNA vectors are also available. Vectors include, but are not limited to: plasmids, fosmids, phagemids, virus genomes and artificial chromosomes.

A cloning vector is one which is able to replicate autonomously or integrated in the genome in a host cell, and which can be further characterized by one or more endonuclease restriction sites at which the vector may be cut in a determinable fashion and into which a desired DNA sequence may be ligated such that the new recombinant vector retains its ability to replicate in the host cell. In the case of plasmids, replication of the desired sequence may occur many times as the plasmid increases in copy number within the host cell such as a host bacterium or just a single time per host before the host reproduces by mitosis. In the case of phage, replication may occur actively during a lytic phase or passively during a lysogenic phase.

An expression vector is one into which a desired DNA sequence may be inserted, for example by restriction and ligation, such that it is operably joined to regulatory sequences and may be expressed as an RNA transcript. Vectors may further contain one or more marker sequences suitable for use in the identification of cells which have or have not been transformed or transfected with the vector. Markers include, for example, genes encoding proteins which increase or decrease either resistance or sensitivity to antibiotics or other compounds, genes which encode enzymes whose activities are detectable by standard assays known in the art (e.g., β -galactosidase, luciferase or alkaline phosphatase), and genes which visibly affect the phenotype of transformed or transfected cells, hosts, colonies or plaques (e.g., green fluorescent protein). Preferred vectors are those capable of autonomous replication and expression of the structural gene products present in the DNA segments to which they are operably joined.

As used herein, a coding sequence and regulatory sequences are said to be “operably” joined when they are covalently linked in such a way as to place the expression or transcription of the coding sequence under the influence or control of the regulatory sequences. If it is desired that the coding sequences be translated into a functional protein, two DNA sequences are said to be operably joined if induction of a promoter in the 5' regulatory sequences results in the transcription of the coding sequence and if the nature of the linkage between the two DNA sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the promoter region to direct the transcription of the coding sequences, or (3) interfere with the ability of the corresponding RNA transcript to be translated into a protein. Thus, a promoter region would be operably joined to a coding sequence if the promoter region were capable of effecting transcription of that DNA sequence such that the resulting transcript can be translated into the desired protein or polypeptide.

When the nucleic acid molecule that encodes any of the genes associated with the claimed invention is expressed in a cell, a variety of transcription control sequences (e.g., promoter/enhancer sequences) can be used to direct its expression. The promoter can be a native promoter, i.e., the promoter of the gene in its endogenous context, which provides normal regulation of expression of the gene. In some embodiments the promoter can be constitutive, i.e., the promoter is unregulated allowing for continual transcription of its associated gene. A variety of conditional promoters also can be used, such as promoters controlled by the presence or absence of a molecule.

The precise nature of the regulatory sequences needed for gene expression may vary between species or cell types, but shall in general include, as necessary, 5' non-transcribed and 5' non-translated sequences involved with the initiation of transcription and translation respectively, such as a TATA box, capping sequence, CAAT sequence, and the like. In particular, such 5' non-transcribed regulatory sequences will include a promoter region which includes a promoter sequence for transcriptional control of the operably joined gene. Regulatory sequences may also include enhancer sequences or upstream activator sequences as desired. The vectors of the invention may optionally include 5' leader or signal sequences. The choice and design of an appropriate vector is within the ability and discretion of one of ordinary skill in the art.

Expression vectors containing all the necessary elements for expression are commercially available and known to those skilled in the art. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Fourth Edition, Cold Spring Harbor Laboratory Press, 2012. Cells are genetically engineered by the introduction into the cells of heterologous DNA (RNA). That heterologous DNA (RNA) is placed under operable control of transcriptional elements to permit the expression of the heterologous DNA in the host cell. A nucleic acid molecule that comprises a gene associated with the invention can be introduced into a cell or cells using methods and techniques that are standard in the art.

In some embodiments, it may be advantageous to use a cell that has been optimized for expression of one or more polypeptides. As used herein, "optimizing expression" of a polypeptide refers to altering the nucleotide sequences of a coding sequence for a polypeptide to alter the expression of the polypeptide (e.g., by altering transcription of an RNA encoding the polypeptide) to achieve a desired result. In some embodiments, the desired result can be optimal

expression, but in other embodiments the desired result can be simply obtaining sufficient expression in a heterologous host cell to test activity (e.g., DNA sequence binding) of the polypeptide.

In other embodiments, optimizing can also include altering the nucleotide sequence of the gene to alter or eliminate native transcriptional regulatory sequences in the gene, thereby eliminating possible regulation of expression of the gene in the heterologous host cell by the native transcriptional regulatory sequence(s). Optimization can include replacement of codons in the gene with other codons encoding the same amino acid. The replacement codons can be those that result in optimized codon usage for the host cell, or can be random codons encoding the same amino acid, but not necessarily selected for the most "preferred" codon in a particular host cell.

In some embodiments, it may be optimal to mutate the cell prior to or after introduction of recombinant gene products. In some embodiments, screening for mutations that lead to enhanced or reduced production of one or more genes may be conducted through a random mutagenesis screen, or through screening of known mutations. In some embodiments, shotgun cloning of genomic fragments can be used to identify genomic regions that lead to an increase or decrease in production of one or more genes, through screening cells or organisms that have these fragments for increased or decreased production of one or more genes. In some instances, one or more mutations may be combined in the same cell or organism. Recombinant gene expression can involve in some embodiments expressing a gene on a plasmid and/or integrating the gene into the chromosomal DNA of the cell. For example, nucleic acid molecules can be introduced by standard protocols such as transformation including chemical transformation and electroporation, transduction, particle bombardment, etc. Expressing the nucleic acid molecule can also be accomplished by integrating the nucleic acid molecule into the genome.

Optimization of protein expression may also require in some embodiments that a gene be modified before being introduced into a cell such as through codon optimization for expression in a bacterial cell. Codon usages for a variety of organisms can be accessed in the Codon Usage Database (<http://www.kazusa.or.jp/codon/>).

Protein engineering can also be used to optimize expression or activity of a protein. In certain embodiments a protein engineering approach could include determining the three dimensional (3D) structure of a protein or constructing a 3D homology model for the protein

based on the structure of a related protein. Based on 3D models, mutations in a protein can be constructed and incorporated into a cell or organism, which could then be screened for increased or decreased production of a protein or for a given feature or phenotype.

A nucleic acid, polypeptide or fragment thereof described herein can be synthetic. As used herein, the term “synthetic” means artificially prepared. A synthetic nucleic acid or polypeptide is a nucleic acid or polypeptide that is synthesized and is not a naturally produced nucleic acid or polypeptide molecule (e.g., not produced in an animal or organism). It will be understood that the sequence of a natural nucleic acid or polypeptide (e.g., an endogenous nucleic acid or polypeptide) may be identical to the sequence of a synthetic nucleic acid or polypeptide, but the latter will have been prepared using at least one synthetic step.

Aspects of the invention thus involve recombinant expression of genes encoding sigma factors and anti-sigmas, functional modifications and variants of the foregoing, as well as uses relating thereto. Homologs and alleles of the nucleic acids associated with the invention can be identified by conventional techniques. Also encompassed by the invention are nucleic acids that hybridize under stringent conditions to the nucleic acids described herein. The term “stringent conditions” as used herein refers to parameters with which the art is familiar. Nucleic acid hybridization parameters may be found in references which compile such methods, e.g. *Molecular Cloning: A Laboratory Manual*, J. Sambrook, et al., eds., Fourth Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2012, or *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York. More specifically, stringent conditions, as used herein, refers, for example, to hybridization at 65°C in hybridization buffer (3.5 x SSC, 0.02% Ficoll, 0.02% polyvinyl pyrrolidone, 0.02% Bovine Serum Albumin, 2.5mM NaH₂PO₄(pH7), 0.5% SDS, 2mM EDTA). SSC is 0.15M sodium chloride/0.015M sodium citrate, pH 7; SDS is sodium dodecyl sulphate; and EDTA is ethylenediaminetetracetic acid. After hybridization, the membrane upon which the DNA is transferred is washed, for example, in 2 x SSC at room temperature and then at 0.1 - 0.5 x SSC/0.1 x SDS at temperatures up to 68°C.

There are other conditions, reagents, and so forth which can be used, which result in a similar degree of stringency. The skilled artisan will be familiar with such conditions, and thus they are not given here. It will be understood, however, that the skilled artisan will be able to manipulate the conditions in a manner to permit the clear identification of homologs and alleles

of nucleic acids of the invention (e.g., by using lower stringency conditions). The skilled artisan also is familiar with the methodology for screening cells and libraries for expression of such molecules which then are routinely isolated, followed by isolation of the pertinent nucleic acid molecule and sequencing.

In general, homologs and alleles typically will share at least 75% nucleotide identity and/or at least 90% amino acid identity to the sequences of nucleic acids and polypeptides, respectively, in some instances will share at least 90% nucleotide identity and/or at least 95% amino acid identity and in still other instances will share at least 95% nucleotide identity and/or at least 99% amino acid identity. In some embodiments, homologs and alleles share at least 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or more than 99% nucleotide identity and/or at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more than 99% amino acid identity. The homology can be calculated using various, publicly available software tools developed by NCBI (Bethesda, Maryland) that can be obtained through the NCBI internet site. Exemplary tools include the BLAST software, also available at the NCBI internet site (www.ncbi.nlm.nih.gov). Pairwise and ClustalW alignments (BLOSUM30 matrix setting) as well as Kyte-Doolittle hydrophobic analysis can be obtained using the MacVector sequence analysis software (Oxford Molecular Group). Watson-Crick complements of the foregoing nucleic acids also are embraced by the invention.

The invention also includes degenerate nucleic acids which include alternative codons to those present in the native materials. For example, serine residues are encoded by the codons TCA, AGT, TCC, TCG, TCT and AGC. Each of the six codons is equivalent for the purposes of encoding a serine residue. Thus, it will be apparent to one of ordinary skill in the art that any of the serine-encoding nucleotide triplets may be employed to direct the protein synthesis apparatus, *in vitro* or *in vivo*, to incorporate a serine residue into an elongating polypeptide. Similarly, nucleotide sequence triplets which encode other amino acid residues include, but are not limited to: CCA, CCC, CCG and CCT (proline codons); CGA, CGC, CGG, CGT, AGA and AGG (arginine codons); ACA, ACC, ACG and ACT (threonine codons); AAC and AAT (asparagine codons); and ATA, ATC and ATT (isoleucine codons). Other amino acid residues may be encoded similarly by multiple nucleotide sequences. Thus, the invention embraces degenerate nucleic acids that differ from the biologically isolated nucleic acids in codon

sequence due to the degeneracy of the genetic code. The invention also embraces codon optimization to suit optimal codon usage of a host cell.

The invention also provides modified nucleic acid molecules which include additions, substitutions and deletions of one or more nucleotides. In preferred embodiments, these modified nucleic acid molecules and/or the polypeptides they encode retain at least one activity or function of the unmodified nucleic acid molecule and/or the polypeptides, such as enzymatic activity. In certain embodiments, the modified nucleic acid molecules encode modified polypeptides, preferably polypeptides having conservative amino acid substitutions as are described elsewhere herein. The modified nucleic acid molecules are structurally related to the unmodified nucleic acid molecules and in preferred embodiments are sufficiently structurally related to the unmodified nucleic acid molecules so that the modified and unmodified nucleic acid molecules hybridize under stringent conditions known to one of skill in the art.

For example, modified nucleic acid molecules which encode polypeptides having single amino acid changes can be prepared. Each of these nucleic acid molecules can have one, two or three nucleotide substitutions exclusive of nucleotide changes corresponding to the degeneracy of the genetic code as described herein. Likewise, modified nucleic acid molecules which encode polypeptides having two amino acid changes can be prepared which have, e.g., 2-6 nucleotide changes. Numerous modified nucleic acid molecules like these will be readily envisioned by one of skill in the art, including for example, substitutions of nucleotides in codons encoding amino acids 2 and 3, 2 and 4, 2 and 5, 2 and 6, and so on. In the foregoing example, each combination of two amino acids is included in the set of modified nucleic acid molecules, as well as all nucleotide substitutions which code for the amino acid substitutions. Additional nucleic acid molecules that encode polypeptides having additional substitutions (i.e., 3 or more), additions or deletions (e.g., by introduction of a stop codon or a splice site(s)) also can be prepared and are embraced by the invention as readily envisioned by one of ordinary skill in the art. Any of the foregoing nucleic acids or polypeptides can be tested by routine experimentation for retention of structural relation or activity to the nucleic acids and/or polypeptides disclosed herein.

The invention embraces variants of polypeptides. As used herein, a “variant” of a polypeptide is a polypeptide which contains one or more modifications to the primary amino acid sequence of the polypeptide. Modifications which create a variant can be made to a polypeptide

1) to reduce or eliminate an activity of a polypeptide; 2) to enhance a property of a polypeptide; 3) to provide a novel activity or property to a polypeptide, such as addition of an antigenic epitope or addition of a detectable moiety; or 4) to provide equivalent or better binding between molecules (e.g., an enzymatic substrate). Modifications to a polypeptide are typically made to the nucleic acid which encodes the polypeptide, and can include deletions, point mutations, truncations, amino acid substitutions and additions of amino acids or non-amino acid moieties. Alternatively, modifications can be made directly to the polypeptide, such as by cleavage, addition of a linker molecule, addition of a detectable moiety, such as biotin, addition of a fatty acid, and the like. Modifications also embrace fusion proteins comprising all or part of the amino acid sequence. One of skill in the art will be familiar with methods for predicting the effect on protein conformation of a change in protein sequence, and can thus “design” a variant of a polypeptide according to known methods. One example of such a method is described by Dahiyat and Mayo in *Science* 278:82-87, 1997, whereby proteins can be designed *de novo*. The method can be applied to a known protein to vary only a portion of the polypeptide sequence. By applying the computational methods of Dahiyat and Mayo, specific variants of a polypeptide can be proposed and tested to determine whether the variant retains a desired conformation. In general, variants include polypeptides which are modified specifically to alter a feature of the polypeptide unrelated to its desired physiological activity. For example, cysteine residues can be substituted or deleted to prevent unwanted disulfide linkages. Similarly, certain amino acids can be changed to enhance expression of a polypeptide by eliminating proteolysis by proteases in an expression system (e.g., dibasic amino acid residues in yeast expression systems in which KEX2 protease activity is present).

Mutations of a nucleic acid which encode a polypeptide preferably preserve the amino acid reading frame of the coding sequence, and preferably do not create regions in the nucleic acid which are likely to hybridize to form secondary structures, such as hairpins or loops, which can be deleterious to expression of the variant polypeptide.

Mutations can be made by selecting an amino acid substitution, or by random mutagenesis of a selected site in a nucleic acid which encodes the polypeptide. Variant polypeptides are then expressed and tested for one or more activities to determine which mutation provides a variant polypeptide with the desired properties. Further mutations can be made to variants (or to non-variant polypeptides) which are silent as to the amino acid sequence

of the polypeptide, but which provide preferred codons for translation in a particular host. The preferred codons for translation of a nucleic acid in, e.g., *E. coli*, are well known to those of ordinary skill in the art. Still other mutations can be made to the noncoding sequences of a gene or cDNA clone to enhance expression of the polypeptide. The activity of variant polypeptides can be tested by cloning the gene encoding the variant polypeptide into a bacterial or mammalian expression vector, introducing the vector into an appropriate host cell, expressing the variant polypeptide, and testing for a functional capability of the polypeptides as disclosed herein.

The skilled artisan will also realize that conservative amino acid substitutions may be made in polypeptides to provide functionally equivalent variants of the foregoing polypeptides, i.e., the variants retain the functional capabilities of the polypeptides. As used herein, a “conservative amino acid substitution” refers to an amino acid substitution which does not alter the relative charge or size characteristics of the protein in which the amino acid substitution is made. Variants can be prepared according to methods for altering polypeptide sequence known to one of ordinary skill in the art such as are found in references which compile such methods, e.g. *Molecular Cloning: A Laboratory Manual*, J. Sambrook, et al., eds., Fourth Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2012, or *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York. Exemplary functionally equivalent variants of polypeptides include conservative amino acid substitutions in the amino acid sequences of proteins disclosed herein. Conservative substitutions of amino acids include substitutions made amongst amino acids within the following groups: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, T; (f) Q, N; and (g) E, D.

In general, it is preferred that fewer than all of the amino acids are changed when preparing variant polypeptides. Where particular amino acid residues are known to confer function, such amino acids will not be replaced, or alternatively, will be replaced by conservative amino acid substitutions. Preferably, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 residues can be changed when preparing variant polypeptides. It is generally preferred that the fewest number of substitutions is made. Thus, one method for generating variant polypeptides is to substitute all other amino acids for a particular single amino acid, then assay activity of the variant, then repeat the process with one or more of the polypeptides having the best activity.

Conservative amino-acid substitutions in the amino acid sequence of a polypeptide to produce functionally equivalent variants of the polypeptide typically are made by alteration of a nucleic acid encoding the polypeptide. Such substitutions can be made by a variety of methods known to one of ordinary skill in the art. For example, amino acid substitutions may be made by PCR-directed mutation, site-directed mutagenesis according to the method of Kunkel (Kunkel, *Proc. Nat. Acad. Sci. U.S.A.* 82: 488-492, 1985), or by chemical synthesis of a gene encoding a polypeptide.

Genetic circuits described herein can contain elements other than sigma factors and anti-sigma factors. For example, genetic circuits can comprise transcriptional regulatory elements. As used herein, a "transcriptional regulatory elements" refer to any nucleotide sequence that influences transcription initiation and rate, or stability and/or mobility of a transcript product. Regulatory sequences include, but are not limited to, promoters, promoter control elements, protein binding sequences, 5' and 3' UTRs, transcriptional start sites, termination sequences, polyadenylation sequences, introns, etc. Such transcriptional regulatory sequences can be located either 5', 3', or within the coding region of the gene and can be either promote (positive regulatory element) or repress (negative regulatory element) gene transcription.

Aspects of the invention encompass a non-transitory computer readable storage medium encoded with instructions, executable by a processor, for designing a host cell and a computer product comprising a computer readable medium encoded with a plurality of instructions for controlling a computing system to perform an operation for designing a host cell. As used herein, "computer-readable medium" refers to any media that is involved in providing one or instructions to a processor for execution. Computer-readable media can be anything that a computer is able to read, such as, for example, disks, magnetic tape, CD-ROMs, any other optical medium, punch cards, paper tape, any other physical medium with patterns of holes, a RAM, PROM, and EPROM, a FLASH-EPROM, any other memory chip or cartridge or a carrier wave.

The present invention is further illustrated by the following Examples, which in no way should be construed as further limiting. The entire contents of all of the references (including literature references, issued patents, published patent applications, and co pending patent

applications) cited throughout this application are hereby expressly incorporated by reference, including the entire contents of WO 2012/170436.

EXAMPLES

Example 1: ECF σ factors, anti- σ s, and promoters: a complete map of crosstalk and application to synthetic switches

Cells react to their environment through gene regulatory networks, the integrity of which requires the minimization of undesired crosstalk between their component biomolecules. Here, a comprehensive map was established of the promoter specificities for ECF σ s, which have a central role in prokaryotic gene expression, as well as their interaction with anti- σ s. DNA synthesis was used to build 86 ECF σ s (two from every subgroup), their promoters, and 63 anti- σ s identified from the genomes of diverse bacteria. A subset of 20 σ s and promoters were found to be highly orthogonal. The set was expanded by swapping the -35 and -10 promoter binding domains from different species to recognize new chimeric promoters. The orthogonal σ s, anti- σ s, and promoters were used to build synthetic genetic switches in *E. coli*. This demonstrated transferability of these regulators, which can be gleaned from diverse genomic contexts and made to function in a new host with minimal re-engineering.

ECF σ s are the smallest and simplest alternative σ s, as well as the most abundant and phylogenetically diverse^{7,9}. Possessing just the two domains that bind the promoter -10 and -35 regions² (FIG. 1A), they provide cells with a highly modular means to respond to their environment^{7,10}, often responding to a signal through the action of an anti- σ . ECF σ s can autoregulate their own expression and that of their anti- σ ^{11,12}. This organization can lead to diverse dynamical phenomena, including ultrasensitive bistable switches and pulse generators^{13–15}. Moreover, promoters of an ECF σ are highly conserved, facilitating identification, modeling, and rational design^{7,16}. Promoter specificity also results in a large dynamic range of output, where the OFF state is very low in the absence of the σ and the ON state produces a high level of expression.

Their aggregate properties suggest that ECF σ s may be ideal for implementing programs of gene expression for applications in biotechnology. Individual synthetic genetic circuits have been constructed using ECF σ s to implement memory and timer functions^{17,18}. Such circuits can be connected to form programs that implement control over metabolic pathways and cellular

functions^{19–21}. The size and sophistication of such programs has been growing, but have been limited by a lack of regulatory parts that are orthogonal; that is, can be simultaneously used without interference²². In the case of ECF σ s, crosstalk can arise by binding to off-target promoters or anti- σ s.

ECF σ s may provide a large reservoir of orthogonal regulators with 19,314 currently annotated in the MiST database²³. Bioinformatic analysis of the sequence relationships among ~2700 ECF σ s by Mascher and colleagues⁷ identified 43 phylogenetically distinct ECF σ subgroups, thought to have similar promoter binding sequences within subgroups, but with significant variation between subgroups. It was hypothesized that if a significant fraction of these subgroups had σ s with orthogonal promoter recognition, such that they target their cognate promoter with no detectable cross-reactivity to another σ , then there would be the potential to build large programs in single cells based on σ s. This diversity can be physically accessed from the sequence databases by using high-throughput DNA synthesis and screening, in a process referred to as “part mining”²⁴. Here, this approach was applied to construct a comprehensive library that encompasses the phylogenetic diversity of σ s, anti- σ s, and promoters from the 43 phylogenetically distinct ECF σ groups, generating a library of 86 ECF σ s (2 from each subgroup) and their corresponding 63 anti- σ s. Both sets of genes were optimized for expression in *E. coli*²⁵ and obtained via DNA synthesis (Example 6). The library was then functionally screened to identify an orthogonal subset (FIG. 1B).

When mining regulators, a challenge is to determine the DNA sequence to which they bind in order to build a responsive promoter. To this end, a computational approach was developed that identifies native promoters for each subgroup and uses their sequences to build a promoter model. This model was used to select a promoter out of the genome and predict whether the promoter is specific to that subgroup (orthogonal). Native promoters were identified by exploiting that fact that most ECF σ s autoregulate by targeting a promoter upstream of their own gene^{7,9}. Consequently, promoter motifs can be found by searching for over-represented conserved motifs in the regulatory regions upstream of the σ genes in each subgroup. Using this information, Mascher identified motifs for 18/43 subgroups⁸. Using an automated procedure, all regulatory regions upstream of the σ s and their putative operons were culled from their 329 cognate genomes. The 6 Fec-I like subgroups (ECF05-10) were excluded as they do not autoregulate^{7,26}. Conserved promoter-like motifs were identified from the upstream sequences

using BioProspector, which can search for two sequence blocks (i.e., the -10 and -35 regions) connected by a variable spacer²⁷. This approach confirmed and improved the motifs identified by Mascher and co-workers⁸. These combined efforts identified 706 promoters and 29 promoter motifs in the 43 ECF σ subgroups. Promoter models were constructed for the promoter motifs based on position weight matrixes (PWMs)²⁸ for each ECF subgroup and a spacer penalty for suboptimal motif spacing (FIGs 2 and 9).

Using these promoter models, all 706 promoter sequences were scored for orthogonality, demonstrating that most promoters are highly orthogonal, with remarkably little crosstalk across subgroups (FIG. 10A). Surprisingly, the -10 and -35 sequences alone showed considerably less orthogonality than the entire promoter (FIG. 10B,C), indicating that high specificity is achieved by combining both promoter regions. It also implies that new promoter specificities can be achieved by swapping the protein domains that bind the -10 and -35 promoter regions.

Candidate orthogonal promoters were designed for each of the 29 ECF σ subgroups, based on the predictions of the promoter models across all 706 promoters. The promoters selected were predicted to score highly against their cognate σ and poorly against other σ s. Preference was given to promoters that occur immediately upstream from their σ . Promoters were also screened against having sequences similar to those recognized by *E. coli* σ 70 and FecI. This was particularly relevant for promoters from AT-rich genomes that often contain σ 70-like promoter sequences. The candidate promoters contained natural promoter sequences from -60 to +20, and were tested with their cognate σ s (see below). Promoters from GC-rich genomes were often found to be non-functional and this was corrected by replacing the -35 to -60 region with a synthetic UP-element designed to enhance promoter recognition by the α -subunits of RNAP^{29,30} (FIGs 1C and 11). This process yielded a set of 18 promoters that were functional with σ s from their cognate ECF subgroups. The promoters were retained from the remaining 11 subgroups in case they were activated by σ s from different subgroups.

A test system was devised to measure promoter activity and orthogonality. A phylogenetically diverse library of 86 σ s comprised of 2 σ s from each subgroup was built (FIG. 1B), and a multi-plasmid system was developed to control expression and enable transformation of different combinations of σ s, anti- σ s, and promoters to rapidly determine activity (FIG. 3A). The toxicity of each σ was measured. Under high induction, 80% exhibited near wild-type growth levels during exponential phase (>90% wild-type *E. coli* DH10b carrying an empty

vector) (FIG. 3C). Each promoter was assayed against the complete set of σ s, resulting in an exhaustive activity map based on measurements of 29 promoters x 86 σ s and identifying 26 active promoters (FIG. 12). Interestingly, 8 promoters that were non-functional with their cognate σ s, were activated by σ s from different ECF subgroups. This implies that these σ s are regulated by other σ s as part of a cascade rather than autoregulating themselves. Among the inactive σ s, usually both examples from a subgroup were nonfunctional, suggesting that their promoter motifs were incorrect or a shared property of these σ s prevents function in *E. coli*. In total, 58 of the 86 σ s activated at least one promoter >5-fold. The transfer function was measured for each of the 52 most active σ s against its most active promoter (FIG. 3B and 13). This induction has a large dynamic range, from 9- to 270-fold. Many of the σ / promoter pairs are highly orthogonal and the 20 most orthogonal are shown in FIG. 3E. Some crosstalk was observed between different subgroups (e.g., ECF02, 07, 11, 14, 15, 17, 25, 27, 33) and these interactions can be predicted using the promoter models (FIG. 3F).

The σ s in the library spanned 6 bacterial classes, but were biased towards organisms phylogenetically related to *E. coli* to increase their likelihood of binding *E. coli* core RNAP. Interestingly, functional or non-functional σ s were found to exhibit no bias in their phylogenetic distance from *E. coli*. For example non-functional σ s were observed from γ -proteobacteria, the same subclass as *E. coli*, and functional σ s were observed from *Firmicutes*, the most distant class. This lack of bias underscores the surprising success rate of importing foreign σ s from different genomes into *E. coli*. This also implies a high degree of conservation for the σ -RNAP binding interface required for function.

The ECF promoter models demonstrate that promoter specificity is generated through the combination of -10 and -35 motif recognition (FIG. 10). This implies that new promoters could be engineered by combining different -10 and -35 motifs. Indeed, a synthetic hybrid σ that combined the -10 DNA binding domain of σ 70 with the -35 DNA binding domain of σ 32 was able to recognize a cognate hybrid promoter containing a consensus σ 32 -35 motif and consensus -10 motif³¹. The ECF σ s are simple, consisting two domains separated by a flexible linker. The N-terminal domain (Domain 2) binds the -10 motif and the C-terminal domain (Domain 4) binds the -35 motif 32–34. The question of whether these domains could be swapped between different ECF subgroups to create chimeric σ s that activate chimeric promoters was investigated. Two σ s were selected from different subgroups that recognize different -10 and -35 motifs: ECF02_2817

(*E. coli* σ E) and ECF11_3726 (FIG. 3G). For each orientation, six chimeric σ s were constructed by making crossovers in the disordered linker region and in helices near the domain boundary (FIG. 20). Similarly, a library of three chimeric promoters was constructed based on the -35 and -10 motifs from P02_rpoHP3 and P11_3726 that represent a range of spacer lengths. From these small libraries, chimeric σ s were identified that activate their chimeric promoters at a level equivalent to wild-type (FIG. 3G, 21 and 22). Further, the chimeric σ s are orthogonal, exhibiting negligible activity against the opposite promoter chimeras and the parental promoters. Interestingly, while *E. coli* σ E (ECF02_2817) is toxic at high concentrations³⁵, neither of the chimeras based on this σ exhibit toxicity. Extending this concept further, the promoter models roughly predict that there are 16 σ domains across subgroups that bind to different -35 sequences and 10 σ domains that bind to different -10 sequences. This estimates that, considering only DNA sequence specificity, there is an upper limit of ~160 potential ECF σ s that could be orthogonal and potentially operate within one cell.

Anti- σ s bind to σ s and inhibit them by blocking their interaction with core RNAP. A library of the 63 cognate anti- σ s under the control of a 3-O-C6-HSL inducible Plux promoter was constructed. Of these, 46 were associated with an active σ , and the most promising 35 were tested to determine whether they repressed activity of their cognate σ on its most active promoter. 32/35 anti- σ s were able to repress the activity of their target σ more than 2-fold (FIG. 14), indicating that most anti- σ s from different organisms were able to repress their target σ in *E. coli*. Compared to the σ s, a larger fraction of the anti- σ s exhibited toxicity when expressed in *E. coli*; this could occur by the interaction of the anti- σ s with essential host σ s such as ECF σ E (FIG. 17).

To determine the orthogonality of the anti- σ / σ interactions, the activity of each σ was measured in the presence of the 25 most active anti- σ s. An example of this screen is shown in FIG. 4B where the anti- σ AS11_987 represses the activity of its cognate σ ECF11_987 by over 70-fold while the other anti- σ s have a significantly smaller effect. Measurements of repression by all anti- σ s against all σ s reveal that most anti- σ s specifically repress their cognate σ , but there is a higher background of cross-reactivity against non-cognate σ s (FIG. 15). FIG. 4C shows a subset of this data for the 20 most orthogonal σ s for which functional anti- σ s were identified. When serving as switches, the majority of σ s yield a graded induction of the output promoter (FIG. 3B). This induction occurs at a similar threshold of input activity. There is also a 100-fold

range in the basal OFF state. All of these features pose a challenge for assembling the switches into larger programs: more complex dynamical behaviors require a cooperative response, connecting circuits requires matching their thresholds, and a high basal level can trigger the next circuit in series. Previously, it was demonstrated that the addition of a sequestering molecule into a switch lowers the background, increases cooperativity and allows the threshold to be tuned³⁶. The question of whether the anti- σ s serving as a sequestering molecule would improve these properties of the switches was tested. A series of switches was constructed based on controlling the expression level of anti- σ (FIG. 4A and 24). In each case, as the expression level of the anti- σ increased, the basal level of expression decreased (up to 10-fold) and the cooperativity increased (FIG. 4D). For example, when the data for ECF20_992 is fit to a Hill function, the cooperativity goes from $n = 1.7$ to 4.1 as a function of the expression of the anti- σ (FIG. 19). There was also an exquisite capacity for the threshold of the circuit to be tuned, in several cases, by over two orders of magnitude.

A surprising outcome of this work is the ease by which ECF σ s, as well as their anti- σ s and promoters, can be moved between diverse organisms and retain function. Many of these parts could be moved “as-is,” but even those that could not only required modest engineering techniques to achieve functionality. This has implications for the horizontal transfer of these regulatory units between genomes. There is evidence for the horizontal transfer of ECF σ s across species³⁷ and they also appear in mobile genomic islands and plasmids associated with resistance to environmental stress³⁸. By providing a sequestration function, anti- σ s can tune the response characteristics and facilitate the evolution of more complex dynamics³⁹. Transferability and orthogonality are also essential features for applications in genetic engineering, which require the functional movement of such circuits into in a foreign host. Here, these properties of ECF σ s across many species were characterized via part mining, where DNA synthesis is used to access large sets of genes identified from the growing sequence databases. This approach enables taking a comprehensive view towards the biochemical characterization of entire classes of genes, as well as providing a rich resource for the harnessing of these functions in biotechnology.

References for Example 1

1. Hook-Barnard, I. G. & Hinton, D. M. Transcription Initiation by Mix and Match Elements: Flexibility for Polymerase Binding to Bacterial Promoters. *Gene Regul Syst Bio* 1, 275–293 (2007).
2. Gruber, T. M. & Gross, C. A. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annual Review of Microbiology* 57, 441–466 (2003).
3. Chater, K. F. Regulation of sporulation in *Streptomyces coelicolor* A3(2): a checkpoint multiplex? *Current Opinion in Microbiology* 4, 667–673 (2001).
4. Stragier, P. & Losick, R. MicroReview Cascades of sigma factors revisited. *Molecular Microbiology* 4, 1801–1806 (1990).
5. Helmann, J. D. Anti-sigma factors. *Current Opinion in Microbiology* 2, 135–141 (1999).
6. Campbell, E. A., Westblade, L. F. & Darst, S. A. Regulation of bacterial RNA polymerase σ factor activity: a structural perspective. *Curr Opin Microbiol* 11, 121–127 (2008).
7. Staroń, A. et al. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. *Molecular Microbiology* 74, 557–581 (2009).
8. Marles-Wright, J. & Lewis, R. J. Stress responses of bacteria. *Current Opinion in Structural Biology* 17, 755–760 (2007).
9. Helmann, J. D. in *Advances in Microbial Physiology* Volume 46, 47–110 (Academic Press, 2002).
10. Lonetto, M. A., Brown, K. L., Rudd, K. E. & Buttner, M. J. Analysis of the *Streptomyces coelicolor* sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *PNAS* 91, 7573–7577 (1994).
11. Rhodius, V. A., Suh, W. C., Nonaka, G., West, J. & Gross, C. A. Conserved and Variable Functions of the σ E Stress Response in Related Genomes. *PLoS Biol* 4, e2 (2005).
12. Rouvière, P. E. et al. rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*. *EMBO J* 14, 1032–1042 (1995).
13. Voigt, C. A., Wolf, D. M. & Arkin, A. P. The *Bacillus subtilis* sin Operon An Evolvable Network Motif. *Genetics* 169, 1187–1202 (2005).
14. Tiwari, A., Ray, J. C. J., Narula, J. & Igoshin, O. A. Bistable responses in bacterial genetic networks: Designs and dynamical consequences. *Mathematical Biosciences* 231, 76–89 (2011).

15. Locke, J. C. W., Young, J. W., Fontes, M., Jimenez, M. J. H. & Elowitz, M. B. Stochastic Pulse Regulation in Bacterial Stress Response. *Science* 334, 366–369 (2011).
16. Rhodius, V. A. & Mutalik, V. K. Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, σ^E . *PNAS* 107, 2854–2859 (2010).
17. Shin, J. & Noireaux, V. An *E. coli* Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. *ACS Synth. Biol.* 1, 29–41 (2012).
18. Chen, D. & Arkin, A. P. Sequestration-based bistability enables tuning of the switching boundaries and design of a latch. *Molecular Systems Biology* 8, (2012).
19. Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F. & Voigt, C. A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucl. Acids Res.* (2012).doi:10.1093/nar/gks597
20. Solomon, K. V., Sanders, T. M. & Prather, K. L. J. A dynamic metabolite valve for the control of central carbon metabolism. *Metab. Eng.* 14, 661–671 (2012).
21. Zhang, F., Carothers, J. M. & Keasling, J. D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology* 30, 354–359 (2012).
22. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology* 10, 410–422 (2009).
23. Ulrich, L. E. & Zhulin, I. B. MiST: a microbial signal transduction database. *Nucleic Acids Res.* 35, D386–390 (2007).
24. Bayer, T. S. et al. Synthesis of Methyl Halides from Biomass Using Engineered Microbes. *J. Am. Chem. Soc.* 131, 6508–6515 (2009).
25. Raab, D., Graf, M., Notka, F., Schödl, T. & Wagner, R. The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. *Syst Synth Biol* 4, 215–225 (2010).
26. Braun, V., Mahren, S. & Ogierman, M. Regulation of the FecI-type ECF sigma factor by transmembrane signalling. *Current Opinion in Microbiology* 6, 173–180 (2003).
27. Liu, X., Brutlag, D. L. & Liu, J. S. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput* 127–138 (2001).
28. Staden, R. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res* 12, 505–519 (1984).

29. Gourse, R. L., Ross, W. & Gaal, T. UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Molecular Microbiology* 37, 687–695 (2000).
30. Rhodius, V. A., Mutalik, V. K. & Gross, C. A. Predicting the strength of UP-elements and full-length E. coli σ E promoters. *Nucl. Acids Res.* 40, 2907–2924 (2012).
31. Kumar, A. et al. A hybrid sigma subunit directs RNA polymerase to a hybrid promoter in *Escherichia coli*. *Journal of Molecular Biology* 246, 563–571 (1995).
32. Campbell, E. A. et al. Crystal Structure of *Escherichia coli* σ E with the Cytoplasmic Domain of Its Anti- σ RseA. *Molecular Cell* 11, 1067–1078 (2003).
33. Lane, W. J. & Darst, S. A. The Structural Basis for Promoter –35 Element Recognition by the Group IV σ Factors. *PLoS Biol* 4, e269 (2006).
34. Campbell, E. A. et al. A Conserved Structural Module Regulates Transcriptional Responses to Diverse Stress Signals in Bacteria. *Molecular Cell* 27, 793–805 (2007).
35. Nitta, T., Nagamitsu, H., Murata, M., Izu, H. & Yamada, M. Function of the ζ E Regulon in Dead-Cell Lysis in Stationary-Phase *Escherichia coli*. *J. Bacteriol.* 182, 5231–5237 (2000).
36. Buchler, N. E. & Cross, F. R. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Molecular Systems Biology* 5, (2009).
37. Schmidt, T. R., Scott, E. J. & Dyer, D. W. Whole-genome phylogenies of the family Bacillaceae and expansion of the sigma factor gene family in the *Bacillus cereus* species-group. *BMC Genomics* 12, 430 (2011).
38. Van Houdt, R. et al. Variation in genomic islands contribute to genome plasticity in *Cupriavidus metallidurans*. *BMC Genomics* 13, 111 (2012).
39. Buchler, N. E. & Louis, M. Molecular Titration and Ultrasensitivity in Regulatory Networks. *Journal of Molecular Biology* 384, 1106–1119 (2008).

Example 2: Promoter Identification and Design

Identification of ECF promoters using genomic information

A three-step search strategy was used to maximize the identification of promoter sequences for each ECF sigma subgroup. First, based on the observations that many ECF sigma groups autoregulate their own gene expression¹ and that some sigmas may regulate their own anti-sigma², promoter motifs were searched for in sequences directly upstream of the sigma

gene, sigma operon, and cognate anti-sigma gene. Second, upstream regulatory regions were extracted for all ECF sigmas within each subgroup to maximize the ability to find over-represented motifs. Third, BioProspector³ was used to identify over-represented motifs in these upstream regulatory regions. BioProspector is a 2-block motif search algorithm that is ideally suited for bacterial promoters with variable length spacers between the -10 and -35 motifs. All ECF sigmas in subgroups 01-43 and their cognate anti-sigmas were identified from Staron and co-workers (incorporated by reference from Table S5 of Staron et al.)¹. To enable efficient retrieval of their upstream regulatory sequences, all 1232 complete bacterial genome sequences and annotations were downloaded from the NCBI FTP site (11/1/2010). Both sigmas and anti-sigmas were identified from these genomes based on the annotation supplied by Staron et al: source genome, gene ID (GI) and their listed amino acid sequence (sigmas only). From the 1736 listed ECF sigmas and cognate 1203 anti-sigmas listed by Staron et al., 1329 sigmas and 880 anti-sigmas were successfully identified from the NCBI annotated genomes. The remaining sigmas and anti-sigmas were from genomes not listed in the NCBI database and therefore were not used in this analysis.

For each ECF sigma subgroup, three libraries of upstream regulatory sequences were extracted from: 1) directly upstream of the sigma gene; 2) directly upstream of the sigma gene operon (sigma operons were defined as all consecutive genes adjacent to the sigma gene, in the same orientation and separated by less than 50 nt from each other); 3) directly upstream of the cognate anti-sigma gene (if known). Most promoters occur near the start of genes but can be difficult to detect when searching long upstream regulatory sequences for over-represented motifs. To facilitate identification, different length upstream regulatory sequences were extracted for each library, from the start codon to 100, 150, 200 and 300 nt upstream. For each library, searches for over-represented motifs were performed using BioProspector with the short 100 nt upstream sequences first and then repeated with the successively longer sequences. Motif searches with BioProspector were performed only on the forward strand and the highest scoring motifs selected from 100 reinitializations. The search for 2-block motifs was typically of the form, W7 w5 G18 g15: where W and w denotes the length (nt) of the upstream and downstream blocks, respectively; and G and g denotes the maximum and minimum distances (nt) separating the two blocks, respectively. These parameters were varied iteratively to optimize the searches for different promoter motifs. From all the library, sequence and motif search combinations, the

highest scoring 2-block motif was selected as the representative promoter motif for each ECF sigma. These were typically from the 100 or 200 nt sequences upstream of the ECF sigma gene or operon.

Promoters for ECF subgroups 05-10, 19, 27 and 32 listed in Staron et al. were not identified in our search. Subgroups 05-10 are not autoregulated¹ and the remaining subgroups only had a few sigmas with highly related upstream sequences, making it difficult to search for over-represented motifs. For all of these cases, the promoter sequences were obtained from Staron et al. and Bioprospector was used to redefine the -35 and -10 motifs. Promoter sequences and their -10/-35 motifs are incorporated by reference from WO 2012/170436 (see, for e.g., page 58 of WO 2012/170436).

ECF sigma promoter modeling and prediction

For each ECF sigma subgroup, the highest scoring 2-block motif identified by BioProspector was used to construct promoter models following the method described by Rhodius et.al.⁴ The upstream and downstream motif sequences were used to compile Position Weight Matrices (PWMs)⁵ for the -35 and -10 motifs, respectively. Specifically, for the regions identified by BioProspector, the weights ($W_{b,i}$) for each position (i) and base (b) were computed as

(S1)

$$W_{b,i}^{motif} = \ln \left[\frac{(n_{b,i} + 0.5)/(N + 2.0)}{P_b} \right]$$

where $n_{b,i}$ is the number of times that the base b is found at position i in the promoter set, N is the number of promoters in the promoter set, and P_b is the probability of finding a specific base at any given position (assumed to be 0.25). Bayesian pseudocounts of 50% were added to each base to represent the relative uncertainty in the promoter sequences. To evaluate a motif in a promoter, the appropriate weights can be summed for a given sequence of bases b at positions i to obtain a complete -35 or -10 score. Additionally, the variable distances between the -35 and -10 motifs were used to construct spacer length histograms and to calculate a penalty score S for suboptimal spacer lengths

(S2)

$$S = \ln[(f + 0.0005F)/(F + 0.0005F)]$$

where F is the frequency of the most commonly observed (assumed to be optimal) spacer length in the promoter set, and f is the frequency of the spacer length in the promoter being evaluated. Bayesian pseudocounts of 0.5% of the frequency of the optimal spacer length were added to account for uncertainty.

The total promoter score was calculated as a sum of the -35 and -10 motifs evaluated with PWMs and the spacer length penalty

(S3)

$$Score = \sum_{i=1}^{L_{-35}} W_{b,i}^{-35} + \ln \left[\frac{(f + 0.0005F)}{(F + 0.0005F)} \right] + \sum_{i=1}^{L_{-10}} W_{b,i}^{-10}$$

When visualizing motifs, the sequence logos of aligned promoter sequences were generated using WebLogo 3 (available at the WebLogo website; Composition set to 50% GC;⁶). For the Weblogos to compensate for the variable spacing between the -35 and -10 motifs for each promoter model, the distances between them was fixed to the most commonly observed spacer length. FIG. 2 focuses on the -35 and -10 regions. FIG. 9 contains the complete information for the promoter models, including more of the sequence flanking the -35 and -10 motifs, as well showing how far downstream of these promoters genes are found.

Predicted orthogonality of the promoters in the library, as well as of their individual -35 and -10 regions

The 29 generated promoter models were used to analyze all 706 promoters in the promoter library (FIG. 10A). This analysis revealed a high level of predicted orthogonality between the ECF subgroups. A similar analysis was performed on just the -10 or -35 subsites, revealing far less predicted orthogonality (FIG. 10B,C). Equation S3 was used to evaluate the full promoters, and the first or third term of that equation was used for the -35 or -10 subsite analysis as appropriate.

Sigma 70 promoter modeling and prediction

A sigma 70 promoter model was built to screen promoter constructs for potential overlapping sigma 70 promoter sequences. The sigma 70 promoter model was constructed from 674 known sigma 70 promoter sequences with experimentally determined transcription starts obtained from RegulonDB 7.0 (available at the website of CCG (Centro de Ciencias Genomicas). Since the -10 and -35 motifs of sigma 70 promoters are poorly conserved, work by Shultzaberger et al.⁸ was used as a guide for identifying the motifs. A 2-step search using the 1 block function of BioProspector was used. First, the -10 motif was identified as a 6 mer between positions -16 to -5 (a large window was used to allow for inaccuracies mapping the start site). Next, the -35 motif was identified as a 6 mer 15-20 nt upstream of the identified -10 motif. Four PWMs were constructed using the method of Rhodius and Mutalik⁴. As discussed above, a PWM-35 was built for the -35 motif (aTTGaca) and a PWM-10 for the -10 motif (TATAaT). In addition, a PWMspacer was built for a 10-mer block aligned from -21 to -13 aligned with the -10. This incorporates the putative Zn finger contact (-21 to -18; 9), -17/-16 dyad and -15/-14 TG motif^{10,11}. Finally, PWMstart was included to capture the transcription start site (-1/+1). All of these PWMs were built using Equation S1. Two spacer penalties were constructed⁴ with Equation S2 based on distance histograms between the -35, -10 and start motifs: a spacer penalty (-35 to -10) and a discriminator penalty (-10 to +1). Upstream sequences were scored using counts of overlapping A- and T-tracts between positions -57 to -37, assuming the 5' end of the -35 motif is at position -3612. From these terms, the total sigma 70 promoter score was calculated as:

(S4)

$$\begin{aligned} \text{Score} = & (UP \text{ model}) + PWM_{-35} + PWM_{\text{spacer}} + PWM_{-10} + PWM_{\text{start}} \\ & + (\text{Spacer penalty}) + (\text{Discriminator penalty}) \end{aligned}$$

Inserting the terms described above in the subsection entitled “ECF sigma promoter modeling” yields,

(S5)

$$\begin{aligned}
Score = (N_{AAA} + N_{TTT}) + \sum_{i=1}^7 W_{b,i}^{-35} + \ln \left[\frac{(f_{spacer} + 0.0005 F_{spacer})}{(F_{spacer} + 0.0005 F_{spacer})} \right] + \sum_{i=1}^6 W_{b,i}^{-10} \\
+ \ln \left[\frac{(f_{discrim} + 0.0005 F_{discrim})}{(F_{discrim} + 0.0005 F_{discrim})} \right] + \sum_{i=1}^2 W_{b,i}^{start}
\end{aligned}$$

where NAAA is the number of AAAs, and NTTT is the number of Ts proceeding the -35 site. Note that this promoter model is more complex than that used for the ECF sigma factors for several reasons. First, the additional PWM_{spacer} term was based on a number of contacts between $\sigma 70$ and the promoter region that are not known to occur with ECF sigma factors⁹⁻¹¹. Second, the discriminator penalty and PWM_{start} scores rely on the correct identification of the transcriptional start site for each promoter. This was experimentally established for the $\sigma 70$ promoters, but is unknown for the ECF sigma promoters. Third, while the UP model could be applied to the ECF sigma promoters, it is not thought to vary between ECF sigma subfamilies and would therefore not affect the orthogonality of promoter recognition. As that was the main goal of modeling the ECF promoters, it was therefore left out of the computational analysis.

Improving promoters with synthetic UP elements

Promoter sequences were initially tested for activity against both cognate sigmas from their own ECF sigma group. Many non-functional promoter constructs had poor upstream sequences with AAA and TTT-tract counts of ≤ 2 . These were scored by counting the number of overlapping AAA- and TTT-tracts within the sequence window -35 to -57 (assuming that the 5' end of the -10 motif is at position -10). For these promoters, the sequence between -60 to -35 was replaced with a synthetic UP-element similar to that region in the Pecf02_2817 promoter; CATGACAAAATTTTTTAGATGCGTT (SEQ ID NO:64), which generates a score of 6. The A- and T-tracts were designed predominantly in the proximal α binding site (-47 to -57) to mimic the location of the observed A- and T-rich sequences of the active ECF sigma promoters. Adding the UP-element greatly increased the function of a number of the nonfunctional promoters (FIG. 12), and the UP-element was added to all promoters except for those that proved functional without it in this test (Pecf02_2817, Pecf11_3726, Pecf16_3622, Pecf20_992, Pecf30_2079, Pecf31_34, Pecf32_1122, Pecf33_375). UP-element modified promoters were used in all following experiments.

Example 3: Sigma and Anti-sigma Library Characterization***Complete σ screening data, including multiple σ s from each subgroup and non-orthogonal data***

After promoter optimization, activity assays were performed combinatorially between all optimized promoters and all members of the ECF sigma library (FIG. 13). For each promoter, cells containing the promoter-gfp construct were transformed with the entire ECF sigma library in 96 well format, recovered, induced, and fluorescence measured with flow cytometry. Fluorescence measurements were compared to controls lacking sigma factors (but including the promoter-gfp construct) to calculate fold-induction (Example 6). This testing was used to identify the active sigma factors and promoters in the library, even in cases where the promoter models did not match their intended subgroup. Additionally, these results allowed the selection of a subset of orthogonal sigma factor:promoter pairs, which could be used in the same engineered system without crosstalk. The orthogonal subset of this data is shown in FIG. 3E.

Full transfer functions and cytometry data for promoter induction

Based on the combinatorial sigma:promoter matrix (FIG. 12), 58 members of the ECF sigma library were found to activate a promoter by at least 5-fold. Of these, 52 were chosen for further testing. Each of these 52 sigma factors was paired with its most active promoter from the combinatorial assay, and induced at multiple levels of IPTG to determine promoter activity at multiple levels of sigma factor (i.e., the induction curve) (FIG. 14). Measurements were performed in a similar manner to the combinatorial assay, at 0, 10, 20, and 50 μ M IPTG in addition to 100 μ M. These induction curves show a wide range of activities. A subset of this induction curve data consisting of one member from each active ECF subgroup is shown in FIG. 3B.

Complete anti- σ screening data

This section outlines the initial screen for anti- σ activity; more detailed titration curves for those deemed active are presented in Example 4. Of the 58 ECF sigma factors shown to activate a promoter by more than 5-fold (FIG. 13), 47 have cognate anti-sigma factors in the synthesized library. Based on the strength and orthogonality of the sigma:promoter interaction,

the most promising 35 anti-sigma factors were chosen for further testing. To check for anti-sigma factor activity, titrations of the sigma and anti-sigma factors were performed with the promoter:reporter construct most activated by the sigma factor. These assays were performed using four levels of induction for the sigma factor (0, 5, 20, and 100 μ M IPTG), and three for the anti-sigma factor (0, 50, and 100 nM HSL) in addition to a control lacking the anti-sigma expression plasmid. This test showed that 32 of the 35 tested anti-sigma factors were able to repress their cognate sigma factor by at least 2-fold (FIG. 14).

The 25 anti-sigmas with the best repression of their cognate sigma factor from the titration assay were chosen for combinatorial orthogonality testing (FIG. 15). In this test, the set of 25 anti-sigmas plus a no anti-sigma control was tested against the 25 sigma:promoter pairs targeted by the anti-sigmas. In order to better see any repression effects, the sigma factors were induced to an intermediate level (10 μ M IPTG), while the anti-sigma factors were induced to a high level (50 nM HSL). This assay shows that a number of the anti-sigma:sigma interactions appear to be fairly orthogonal, while there are also a number that affect many sigma factors. In some cases, the broadly-active anti-sigma factors reduce growth (discussed below).

ECF sigma factor and anti-sigma factor library growth assays

Both the σ and anti- σ libraries were tested for toxic effects occurring with expression in *E. coli* DH10b. Toxicity can be due to aberrant gene expression or titration of host RNAP by the σ s, or by interaction of the anti- σ s with essential host σ s such as ECF σ^E . The effect of expressing the sigmas and anti-sigmas was measured using 3 types of growth assays across a range of inductions: 1) transition phase culture density in liquid LB media; 2) exponential growth rates in liquid LB media; 3) colony size on LB agar plates (FIGs 3C; 21). For each condition, growth assays were performed from at least 3 separate transformations and across a range of inducer concentrations: 0, 10, 20 and 100 μ M IPTG for the ECF sigma library; 0, 10 and 50 nM HSL for the anti-sigma library. The ECF sigma library assay strains were freshly transformed DH10b cells carrying pN565 with the pVRa plasmid library and plasmid pET21a (Novagen, Billerica, MA) as a no sigma control; the anti-sigma library assay strains were DH10b cells freshly transformed with the pVRc plasmid library and pACYC18413 as a no anti-sigma control.

Under low levels of induction (10 μ M IPTG or 10 nM HSL for the σ and anti- σ libraries, respectively) 88% of the σ library and 75% of the anti- σ library exhibited near wild type growth

levels by all metrics (>80% DH10b wild type growth). Under high induction levels (100 μ M IPTG or 50 nM HSL for the σ and anti- σ libraries, respectively) most growth defects were observed during transition phase and by colony size. For the σ library, 98% exhibited near wild type growth levels (>80% DH10b wild type growth) during exponential growth, whilst 72% and 85% exhibited near wild type growth measured in transition phase or by colony size, respectively. A similar pattern was observed with the anti- σ library but with slightly larger defects: 78%, 43% and 43% exhibited near wild type growth levels during exponential growth, in transition phase and by colony size, respectively. In general, transition phase and colony size yielded a similar pattern of growth defects in both states across the σ and anti- σ libraries, ($R = 0.79$ and 0.85 , respectively), likely due to the transition/stationary phase growth properties of cells in the centre of colonies.

Both ECF σ s from subgroup 02 exhibited the highest toxicity. *E. coli* ECF σ^E is also from subgroup 02 and is represented by the candidate ECF02_2817 in the σ library. *E. coli* σ^E is toxic when highly expressed¹⁴; consequently, the toxic effects of high expression of both ECF02 σ members in the library (ECF02_2817 and ECF02_915) suggest similar function. *E. coli* σ^E is also essential^{15,16}; accordingly, high expression of its cognate anti- σ AS02_2817 is lethal due to repression of host σ^E activity. Interestingly, high expression of anti- σ AS02_915 from the same subgroup only gave reduced growth levels, suggesting that this anti- σ has reduced specificity for host σ^E . Both σ pairs from subgroups 03 and 25, and anti- σ pairs from subgroups 19, 33 and 35 were also highly toxic (<50% wild type growth), indicating similar activities of each member within the subgroup. There were also several instances of where just one subgroup member was toxic, indicating different functionality in an *E. coli* host (e.g. ability to bind *E. coli* RNAP). Importantly, the lack of toxicity of most library members suggests that they could have utility as orthogonal regulators in *E. coli*.

Example 4: Quantification of Anti- σ Threshold Control

A subset of the anti-sigma:ECF sigma pairs were assayed in more detail to determine their capability to implement ultrasensitivity through sequestration¹⁷ (FIG. 18). Sixteen of the sigma factor:anti-sigma factor:promoter sets previously tested were selected based on either: targeting one of the ECF sigma factors in the orthogonal subset, or having a promising induction curve in FIG. 16. These sets were induced at four levels (0, 5, 20, and 100 μ M IPTG) of ECF

sigma and three levels (no anti-sigma plasmid, 0 nM HSL induction, 50 nM HSL induction) of anti-sigma factor in triplicate (the assay method varied from that used in Example 3; see Example 6) and the promoter activities were measured. High expression of the anti-sigma factor often significantly reduced the promoter output at all levels of ECF sigma factor induction, in many cases also causing highly toxic effects. In contrast, the lower anti-sigma induction showed the desired threshold effect in many cases. At this level of anti-sigma, the higher levels of sigma factor induction showed promoter activity close to the no anti-sigma control, while the lower levels of sigma factor had much lower activity than the equivalent induction points with no anti-sigma. This differential repression is characteristic of a threshold system, and increases the utility of these proteins in applications where a more digital-like signal response is desired. The 9 anti-sigma:ECF sigma:promoter sets that have the best induction curves and correspond to an orthogonal ECF sigma factor are shown in FIG. 4D.

In addition, a threshold-gated switch was constructed using ECF20_992 and AS20_992 and characterized more thoroughly (FIG. 19). The inducible anti-sigma system was supplemented by a set of plasmids constitutively expressing the anti-sigma AS20_992 at a number of levels (FIG. 24). Changing the strength of a constitutive promoter allowed for finer control over the expression level of anti-sigma. This system was tested at 8 induction levels of the ECF sigma factor (0, 5, 25, 50, 75, 100, 150, 200 μ M IPTG) in triplicate to characterize the transfer function. Finally, the Hill equation

$$\frac{y - y_{min}}{y_{max} - y_{min}} = \frac{x^n}{x^n + K^n}$$

where x is the IPTG induction concentration, y is the output (promoter activity), y_{max} is the maximum output, y_{min} is the minimum output, K is the half-maximum, and n is the Hill coefficient, was used to fit the data using a nonlinear least-squares optimization function in MATLAB. The optimization was weighted inversely to the value at each point to minimize the relative least-squares error so that the model fit both the low and high ends of the data.

Example 5: Creating Chimeric Sigma Factors

Design of chimeric sigma factors and promoters

A combination of protein alignment, structural information, and secondary structure prediction algorithms were used to generate chimeric sigma factors from ECF02_2817 and ECF11_3726 (FIGs 3G and 21A). These parental sigma factors were chosen since they have high activity in *E. coli* and there is protein structural information is available that could be used to guide the construction of the chimeras (ECF02_2817 (*E. coli* σ^F)¹⁸ and *R. sphaeroides* σ^E , which belongs to the same subgroup as ECF11_3726¹⁹). Chimeras of both orientations (N-terminal ECF02_2817 / C-terminal ECF11_3726 and N-terminal ECF11_3726 / C-terminal ECF02_2817) were created by recombining the parental proteins at six ‘crossover seams’ located in the flexible linker region between the conserved domains 2 and 4, which recognize the -10 and -35 promoter subsites, respectively. While domains 2 and 4 play the most important roles in promoter recognition, the linker region between these regions in Group I σ s plays an important role in abortive initiation and promoter escape²⁰, and likely plays a similar role in the ECF σ s. Consequently, the choice of crossover seams within the linkers of the ECF02_2817 and ECF11_3726 σ s may affect the functionality of the resultant chimeras. The structure and precise boundary of the linker region in the ECF σ s is ambiguous for two reasons: 1) in both structures the ECF σ s are bound to their cognate anti- σ , distorting the structure of the linker; 2) the amino acid sequence of the linker region is poorly conserved, making accurate alignments challenging.

In order to select a range of potentially functional crossover seams, the full library of 86 ECF sigma factors was initially aligned using ClustalW (available at the EMBL-EBI website)²¹. The alignment of ECF02_2817 and ECF11_3726 was then tweaked by hand based on the protein structures mentioned previously. Crossover seams 1 and 2 were located at either end of the flexible linker in this alignment. Due to some uncertainties in the structural analysis (specifically, that the linkers were too distorted by binding anti- σ s for proper structural analysis) crossover seams 4, 5, and 6 were based off of the unaltered ClustalW alignment near the beginning, middle, and end of the linker. Finally, a secondary structure prediction algorithm, PredictProtein²², was used to analyze ECF02_2817 and ECF11_3726 for α -helices. Crossover seam 6 was placed one residue before the beginning of the first α -helix after the linker region in both proteins.

Chimeric promoters were similarly created by crossing over cognate promoters for ECF02_2817 and ECF11_3726 between the -10 and -35 boxes (FIG. 21B). The promoter

rpoHP323 from *E. coli* was used as the parental pECF02 promoter, with a 1 bp mutation (T-34G) made from the WT sequence to differentiate it more from ECF11 promoters. (This promoter contains an overlapping $\sigma 70$ promoter²³, which likely accounts for the high background induction level and low dynamic range of activation by ECF02.) The pECF11_3726 promoter from the ECF sigma library was chosen as the parental pECF11 promoter. In each case, to -60 to +20 region of the promoter was used, and these parental promoters were crossed over between -20 and -21 to make chimeric promoters. While the initially engineered chimeric promoters were functional, they were relatively weak when compared to the parental promoters. Without wishing to be bound by any theory, one explanation for reduced activity is that while the -10 and -35 recognition sites are identical to the parental plasmids, the spacing between them may not be optimal for the chimeric proteins. This is made even more likely because of uncertainties in identifying the -10 and -35 sites in the promoter, and because the ECF02 and 11 promoter models have different optimal spacings (FIGs 2 and 9, ECF02 has optimal spacing 14, and ECF11 has optimal spacing 16). For these reasons, additional chimeric promoters were engineered with the -10 and -35 sites moved either 1bp closer or 1bp farther apart.

Chimeric sigma factor characterization

The chimeric sigma factors and promoters were first assayed to determine which crossover seams and promoter variants were most successful (FIG. 22). Each of the six versions of each chimera was paired with each of the three versions of its cognate promoter, and the promoter activity determined *in vivo*. From this assay, it seems as though ECF sigma factors can tolerate chimeragenesis in many different areas and alignments within the linker. Despite the differing alignments used to design the chimeras, seams 1-4 produced very active chimeras for both ECF02-11 and ECF11-02, seams 5 and 6 were slightly active in ECF11-02, and seam 6 was slightly active in ECF02-11. Of these seams, seam 1 was chosen for further experimentation as it was the most active variant of ECF11-02, and one of the more active variants of ECF02-11. In contrast to the flexibility on protein crossover location, the chimeric promoter spacing had an extreme effect on sigma chimera activity. The initially built chimeric promoters had an intermediate level of activity, while pECF02-11-1 and pECF11-02+1 were greatly improved. In contrast, pECF11-02-1 and pECF02-11+1 were inactive, indicating that 1-2 bps of change in the distance between the -10 and -35 sites is enough to abrogate promoter activity. Based on these

results, ECF11-02 #1, ECF02-11 #1, pECF02-11-1 and pECF11-02+1 were chosen as the chimeric sigma factors and promoters to be used for further chimera testing.

Next, using the optimized chimera constructs, the parental and chimeric sigma factors and promoters were tested with each other to check their orthogonality (FIGs 23 and 3E). Each of the two parental sigma factors and chimeras from the most active seam was tested with each of the two parental promoters and best chimeric promoters. Promoter activity was measured *in vivo*, and fold-activation of each promoter was calculated using a negative control plasmid that does not express a sigma factor. This assay demonstrates that both the -10 and -35 site must be recognized to get ECF sigma factor promoter activation. The chimeric sigma factors activated their promoters by more than 50-fold more than the parental sigma factors, and the parental sigma promoters likewise only recognized the parental sigma factors. ECF02_2817:pECF02_rpoHP3 displayed the weakest activation at ~10-fold, however, this is due to extremely high background activation.

Example 6: Materials and Methods

Strains and media

E. coli strain DH10b (MC1061 F- endA1 recA1 galE15 galK16 nupG rpsL ΔlacX74 Φ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) λ-) (Durfee et al. 2008) was used for all manipulations and assays unless otherwise noted. *E. coli* DH10b strains were grown at 37°C with aeration in LB Miller broth for expression assays, and in LB Miller broth, 2YT, SOB (2% Bacto-tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl), SOB +Mg (SOB + 10 mM MgCl₂, 10 mM MgSO₄), SOB + 20mM glucose, or SOC (SOB +Mg + 20 mM glucose) for cloning and CaCl₂ high-throughput transformations. *E. coli* strain CAG22216 (MC1061 λ (rpoH P3-lacZ) rpoE::Cam, CamR)²⁴ was used for expression and testing of chimeric sigma factors. *E. coli* CAG22216 strains were grown at 30°C with aeration in the same media as *E. coli* DH10b. All cultures were supplemented with appropriate antibiotics. Expression of the ECF sigma library, chimeric sigma factors, and ECFs for threshold experiments, was induced with 0-100 μM IPTG (Isopropyl β-D-1-thiogalactopyranoside from Sigma Aldrich, #I6758). The anti-sigma library was induced with 0-50 nM HSL (3-O-C6-HSL (N-(β-ketocaproyl)-L-Homoserine Lactone from Cayman Chemical, #10011207). Cultures were grown in either 14 mL Falcon tubes (BD, cat#352059), shaken at 250 rpm at 37 °C or 30 °C, or 96-well

format in sterile V96 tissue culture plates (NUNC, cat #249935) using an ELMI plate shaker-thermostat (DTS-4 from Elmi Ltd, Riga, Latvia) shaking at 1,000 rpm at 37 °C or 30 °C. Plates were covered with gas-permeable membranes (AeraSeal from EK Scientific, cat #T896100-S).

Plasmids

A 4-plasmid system was used for expressing the ECF sigma, promoter and anti-sigma libraries (FIG. 23). Plasmid pN565 encodes an IPTG-inducible low processivity T7 RNA polymerase enzyme. This was used to weakly express the ECF sigma library under control of a T7-regulated promoter encoded on the pVRa plasmid series. The pVRb plasmid series carries the ECF sigma-dependent promoters fused to the fluorescent reporter, superfolder GFP (Pedelacq et al. 2006). The pVRc plasmid series encodes the anti-sigma library under control of HSL. Plasmid modifications were performed using Type II restriction enzyme cloning, PCR and one-step isothermal DNA assembly (Gibson et al. 2009). The ECF sigma and anti-sigma gene libraries were codon optimized for *E. coli* K12 MG1655, constructed by gene synthesis and assembled into their parent vectors by GeneArt, Life Technologies.

Plasmid pN565 (incW, SpecR) is a variant of the low processivity T7 RNA polymerase expression vector, pN24925 and is tightly regulated by IPTG. The plasmid encodes T7 RNAP with a GTG initiation codon for low translation, an N-terminal degradation tag and the active site mutation R632S. T7 RNAP is expressed from a weak RBS sequence tuned to 50 units using the RBS calculator²⁶ and a modified Ptac promoter with a symmetrical LacO operator sequence (aattgtgagcgctcacaatt; SEQ ID NO:67), enabling near complete promoter repression in the absence of IPTG. The plasmid also encodes LacI.

Plasmid series pVRa (pBR322, AmpR) expresses the ECF sigma library from a T7-lacO promoter. The plasmids are derived from pET15b (Novagen) in which the thrombin cleavage site was replaced with a PreScission protease cleavage site. The series encodes codon optimized ECF sigma genes on NdeI-HindIII fragments in frame with an N-terminal His6 tag and intervening PreScission site. The ECF sigma library is described further in, and incorporated by reference from WO 2012/170436 (see, for e.g., pages 55-56 of WO 2012/170436).

Plasmid series pVRb (SC101, KanR) carries the ECF sigma-dependent promoter library fused to superfolder GFP (sfGFP)²⁷. The plasmids are derived from the GFP expression vector, pUA6628, in which the reporter gene gfpmut2 was replaced with sfGFP on a BamHI-PstI

fragment. Promoter sequences from -60 to +20 with respect to the transcription start site were inserted upstream of *sfgfp* into the BbsI-BamHI sites of pVRb (the 5' end of the -10 motif was assumed to be at position -10). For each promoter, DNA fragments were assembled from 4 overlapping 45-mer DNA oligos that corresponded to native promoter sequence, and 2 flanking vector specific oligos. The oligos were assembled by PCR to generate 120 bp fragments in which the 80 nt promoter sequence is flanked by 20 nt of vector sequence. The fragments were gel purified and assembled into purified pVRb BbsI-BamHI vector using one-step isothermal DNA assembly. The ECF promoter library is described further in, and incorporated by reference from, WO2012/170436 (see, for e.g., paragraph [0235] of WO2012/170436).

Plasmid series pVRc (p15a, CmR) expresses the anti-sigma library from a HSL-regulated Plux promoter. The plasmids contain *cat* and *LuxR* under constitutive control, and replicate via a p15a origin. The plasmids and amino acid sequences of the anti-sigmas are listed in Table 1.

Anti-sigma threshold analysis was performed with a four-plasmid system very similar to that used to characterize the ECF sigma and anti-sigma libraries (FIG. 24). One additional plasmid series was used to supplement pVRc20_992 with a series of plasmids that constitutively express the anti-sigma. Plasmid series pAG_AS20_992_J23XXX (p15a, CmR) expresses AS20_992 from constitutive promoters of varying strengths. The plasmids are derived from pVRc20_992 in which *LuxR* and the Plux promoter were deleted and replaced by a constitutive promoter. Constitutive promoters BBa_J23100, BBa_J23101, BBa_J23105, and BBa_J23117 (in order of decreasing relative strength: 2547, 1791, 623, 162 au) were selected from the Registry of Standard Biological Parts (available on the iGEM Synthetic Biology website).

A 2-plasmid system was used to test the chimeric sigma factors and their cognate promoters (FIG. 25). Plasmid series pTSaXX (p15a*, SpecR) expresses parental (ECF02_2817, ECF11_3726) and chimeric sigma factors under the control of a modified Ptac promoter with a symmetrical LacO operator sequence. These plasmids were derived from pSB3C529, and contain a mutation in the origin that causes them to be maintained at a higher copy number than wild-type p15a. Plasmid series pTSbXX (pSC101, KanR) contains parental and chimeric sigma-dependent promoters driving expression of *sfgfp*. These plasmids are very similar to plasmid series pVRb, with only the promoter region varying. All construction of these plasmid series was done with one-step isothermal DNA assembly or PCRs and blunt ligations.

High-throughput Transformations of sigma and anti-sigma libraries

In vivo assays of strains carrying ECF sigma or anti-sigma libraries were performed from freshly transformed *E. coli* DH10b host cells. This was to reduce the occurrence of potential suppressor mutations from toxicity of some of the sigmas and anti-sigmas by long-term maintenance in a host. A CaCl₂-based high-throughput transformation protocol in 96-well format was employed that enabled convenient transformation of several hundred strains a day. CaCl₂ competent cells were prepared using the method of Hanahan et al.³⁰ for MC1061-based strains. Briefly, 50 ml cultures of cells were grown in SOB (-Mg) media, harvested at OD₆₀₀ = 0.3, pelleted and supernatant discarded, cells resuspended and pelleted in 25 ml ice-cold CaCl₂ buffer (50 mM CaCl₂, 10 mM Tris.HCl pH 7.5) and then finally resuspended in 3.3 ml fresh ice-cold CaCl₂ buffer + 15% glycerol. Plasmid DNA stocks of each library was prepared at 5 ng/μl in 96-well format. For transformation, 10 ng of each plasmid was placed into a sterile 96-well PCR-plate with 25 μl ice-cold CaCl₂ competent cells and incubated on ice for 60 min (for double plasmid transformations, 5 ng each plasmid + 40 μl CaCl₂ competent cells was used). The entire PCR-plate was then heat-shocked at 42°C in a dry-block for 2 min and then placed on ice for 5 min. Afterwards, cells were transferred to a fresh 96-well tissue culture plate containing 100 μl SOC, mixed, sealed with a breathable membrane and incubated at 37°C, 1000 rpm for 2 hr. 30 μl cells were then transferred to a fresh 96-well tissue culture plate containing 130 μl SOB +Mg + appropriate antibiotics for selection, covered with breathable membrane and incubated overnight (~16 hr) at 37°C, 1000 rpm. This liquid selection in the presence of antibiotics was sufficient to prevent growth of no plasmid controls. The fresh overnight transformants grown to saturation were used for all downstream assays by diluting 200-fold into fresh media with antibiotics and inducers, and growing fresh cultures as required.

ECF sigma activity assays

The sigma-promoter gfp assays were performed in *E. coli* DH10b host cells using a 3 plasmid system: pN565 carrying IPTG-inducible T7 RNAP, pVRa plasmid series carrying the ECF-sigma library, and pVRb plasmid series carrying ECF promoters fused to sfgfp (FIG. 23). Titrations of a sigma against a specific promoter (FIGs 3B and 13) were performed at 0, 5, 10, 20, 50, and 100 μM IPTG. Assays of all sigmas against all promoters (FIGs 3E and 12) were also performed in 96-well format with each plate containing the entire ECF sigma library assayed

against a specific promoter. Specifically, *E. coli* DH10b cells carrying pN565 and a specific pVRb promoter::sfgfp plasmid were transformed with the complete pVRa ECF sigma library and pET21a control in 96-well format. Overnight liquid transformants grown to saturation (~16 hr) were diluted 200-fold into fresh prewarmed LB +Spec, Amp, Kan and 100 μ M IPTG in a 96-well cell culture plate and covered with a breathable membrane. Cultures were incubated in an Elmi plate shaker for 6 hr at 37°C, 1000 rpm. After 6 hr, 5 μ l of culture was added to 200 μ l PBS (137 mM NaCl, 2.7 mM KCl, 8 mM Na₂HPO₄, 2 mM KH₂PO₄) and 2 mg/ml Kanamycin. Samples were run on a BD Biosciences LSRII flow cytometer to quantify GFP accumulation.

Anti-sigma library activity assays

The anti-sigma activity assays were performed in *E. coli* DH10b host cells using a 4 plasmid system: pN565 carrying IPTG-inducible T7 RNAP, pVRa plasmid series carrying the ECF-sigma library, pVRb plasmid series carrying ECF promoters fused to sfgfp, and pVRc plasmid series carrying the anti-sigma library under HSL inducible control. Plasmid pACYC184 was used as a no anti-sigma control. Anti-sigma activity was determined by its ability to repress sigma activity. Accordingly, each sigma was paired with its most active promoter to determine fold repression in the presence and absence of anti-sigma. The anti-sigma activity assays were performed exactly as described for the sigma-promoter assays with the following differences: Anti-sigma – sigma titrations assays: In FIGs 4B and 14, the transfer functions are shown for the sigma factor for different levels of anti-sigma expression. For each anti-sigma – sigma titration set, *E. coli* DH10b cells carrying pN565 and a specific pVRc anti-sigma were doubly transformed with pVRa ECF sigma / pVRb promoter plasmid pair. A single overnight transformation was then diluted 200-fold into 12 wells of a 96-well plate containing LB, Spec, Amp, Kan and Cm, and a 2-dimensional grid of inducer concentrations: 0, 5, 20 or 100 μ M IPTG, and 0, 10 or 50 nM HSL. A no anti-sigma control was also included using DH10b pN565 pACYC184 cells doubly transformed with pVRa ECF sigma / pVRb promoter plasmid pair. The control was diluted 200-fold into 4 wells of a 96-well plate containing LB+Spec, Amp, Kan and Cm and 0, 10, 20 or 100 μ M IPTG.

Anti-sigma – sigma activity assays

Each anti-sigma was assayed against all active sigmas paired with an active promoter in 96-well format (FIGs 4C and 15). To maximize the ability of anti-sigmas to repress target sigmas, the sigmas were only partially induced with 10 μ M IPTG and the anti-sigmas maximally induced with 50 nM HSL. DH10b cells carrying pN565 and a specific pVRc anti-sigma (or no anti-sigma control, pACYC184) were doubly transformed with a library of pVRa ECF sigma / pVRb promoter plasmid pairs. Overnight liquid transformants were diluted 200-fold into fresh prewarmed LB +Spec, Amp, Kan, 10 μ M IPTG and 50 nM HSL.

Sigma and anti-sigma exponential phase liquid growth rate assays

These were performed by diluting freshly transformed overnight cultures 200-fold into prewarmed LB media with appropriate antibiotics and inducer. Cultures were in 96-well cell culture plates covered with a clear lid and were grown in a Varioskan plate reader/shaker (Thermo Fisher Scientific) at 37°C, shaking at 480 rpm, 6 mm orbital motion. Cell densities (OD_{600}) were recorded automatically by the Varioskan every 15 min for 2 hr during exponential growth. All OD_{600} measurements on the Varioskan were converted to standard 1 cm pathlength ODs using a calibration curve generated from an exponentially growing 50 ml culture in a 250 ml shake-flask. Samples from the shake-flask were taken every hour throughout the growth curve and the OD_{600} measured in a 1 cm pathlength cell with a standard spectrophotometer (with appropriate dilution so that OD_{600} readings were always between 0.25-0.35) and from 160 μ l samples measured in a 96-well cell culture plate by the Varioskan. The calibration curve generated from the plot of actual 1 cm pathlength OD_{600} values versus 160 μ l Varioskan OD_{600} values was used to normalize all experimental culture ODs measured in the Varioskan. The normalized experimental OD readings were plotted as $\ln(OD_{600})$ versus time (h). Bacterial growth rate μ was calculated from the slope of the linear section of the plot,

(S6)

$$\mu = (\ln N_t - \ln N_0) / (t - t_0)$$

where μ is the growth rate, N is the number of cells (approximated by OD), and t is time. The growth rates of all sigma and anti-sigma libraries were expressed as a percentage of WT (averaged from eight control cultures).

Transition phase liquid cell densities

These were performed exactly as the exponential phase growth rate assays with the following modifications. Assay cultures were induced and grown in the Varioskan for 8 hr and the growth curve monitored from OD₆₀₀ readings performed every hr. Wild-type cultures typically entered transition phase after 2-3 hr. Sick cultures often exhibited a decrease in culture OD₆₀₀ values during transition phase, likely due to cessation of growth and subsequent cell lysis. Transition phase cell densities were recorded from the final 8 hr OD₆₀₀ values, normalized to 1 cm pathlength ODs and presented as a percentage of WT OD₆₀₀ (from eight control cultures).

Colony size measurements

These were performed in 96-colony format from 96-well cultures. Fresh overnight transformants in 96-well format from each library were pinned onto separate LB-agar master plates containing appropriate antibiotics using a Singer Rotor robot and a 96-pin liquid to solid pinner head. Each plate was incubated overnight for 14 hr at 37 °C to grow colonies in 96-format. From each master plate, colonies were pinned onto inducer plates with the Singer robot using 96-pin solid to solid pinners. The inducer plates contained LB-agar plus appropriate antibiotics and IPTG or HSL inducer, and were incubated overnight for 14 hr at 37 °C to grow colonies in 96-format. Colony sizes were recorded using a 6 megapixel camera under controlled lighting³¹, and colony diameter measured using automated image analysis software, HT Colony Grid Analyzer (available at the sourceforge.net website). The sizes of all sigma and anti-sigma expressing colonies were converted to a percentage of WT (from two control colonies).

Anti-sigma threshold assays

The anti-sigma:sigma titrations were repeated in more detail (FIGs 4D and 18). DH10b cells were transformed with pN565 and a set of pVRa, pVRb, and pVRc plasmids corresponding to one of 16 promising anti-sigma : sigma : promoter sets. A negative anti-sigma control for each set was also made that lacked the pVRc plasmid. Glycerol stocks were made of each strain and

stored at -80 °C. For each assay, the glycerol stocks were used to start overnights in LB + Amp, Spec, Kan, and Chl (or Amp, Spec, and Kan for the no-anti-sigma controls). After growing to saturation, these overnights were diluted 1:200 into LB + antibiotics in a 96-well cell culture plate. The four plasmid strains were added to a grid of inducer concentrations: 0, 5, 20 or 100 μ M IPTG, and 0 or 50 nM HSL. The no-anti-sigma strains were added to inducing conditions of 0 nM HSL and 0, 5, 20 or 100 μ M IPTG. Each 96-well plate was shaken at 37°C for 6 hours at 37°C, 1000 rpm. 2 μ L of each induction was added to 198 μ L PBS + 2 mg/mL Kanamycin and stored at 4°C. Samples were run on a BD Biosciences LSRFortessa flow cytometer to quantify GFP accumulation.

Anti-sigma threshold switch testing

For detailed characterization of a threshold switch (FIG. 19), the plasmid series pVRc was partially replaced with the plasmid series AG_AS20_992_J23XXX, which expresses AS20_992 under the control of a series of constitutive promoters (FIG. 24). This allowed for finer and more predictable control of intermediate anti-sigma expression levels in the switch. This switch was characterized in much the same way as the anti-sigma threshold assays. DH10b cells were transformed with pN565, pVRa20_992, pVRb20_992, and pVRc20_992 or one of AG_AS20_992_J23XXX, and glycerol stocks were made. A no-anti-sigma control was also made with only pN565, pVRa20_992, and pVRb20_992. Assays were run by diluting saturated overnights into 96 well cell culture plates with LB + antibiotics + 0, 5, 20 or 100 μ M IPTG. The strain with pVRc20_992 was run with 50 nM HSL for maximum induction, and the other strains had no HSL. 96-well plates were shaken for 6 hours at 37 °C, 1000 rpm. 2 μ L of each induction was added to 198 μ L PBS + 2 mg/mL Kanamycin and stored at 4C. Samples were run on a BD Biosciences LSRFortessa flow cytometer to quantify GFP accumulation.

Chimeric sigma factor assays

All assays (FIGs 3H, 21, and 22) of chimeric sigma function were performed in strain CAG2221624 supplemented with antibiotics to maintain carried plasmids. A two plasmid assay system was used, consisting of the series pTSaXX, which contains the parental and chimeric sigma factors under IPTG inducible control, and pTSbXX, which contains the parental and chimeric sigma factor promoters driving sfgfp (FIG. 25).

Similar to the assays with ECF sigma or anti-sigma libraries, the chimeric sigma factors were transformed into cells directly before assaying. Z-competent (Zymo Research, cat# T3002) cell stocks of CAG22216 carrying plasmids from series pTSbXX were made per manufacturer's instructions. The day before the functional assay, 100ng aliquots of plasmids from series pTSa were added to 50uL Z-competent cells at 4 °C. The cells were kept on ice for 30 seconds, incubated at room temperature for 120 seconds, 150uL SOC was added and the cells were grown at 30 °C, 1000 rpm for 2 hours. These growths were diluted 1:100 into 150uL LB + Spec/Kan, and incubated 16 hrs at 30 °C, 1000 rpm.

For both the assays, transformed overnights were diluted 1:200 into LB + Spec/Kan + 10uM IPTG and grown for 8 hours at 30 °C, 1000 rpm. 5uL of each induction was added to 195uL PBS + 2 mg/mL Kanamycin and stored at 4 °C. Samples were run on a BD Biosciences LSRFortessa flow cytometer to quantify GFP accumulation.

Flow cytometry analysis

GFP fluorescence of the diluted samples was measured using either a BD Biosciences LSRII flow cytometer (UCSF) or a BD Biosciences LSRFortessa flow cytometer (MIT). Initial analysis of the ECF sigma and anti-sigma libraries was performed on the LSRII, while threshold analysis and chimera testing was done with the LSRFortessa. LSRII analysis: For each sample, 50,000 counts were recorded using a 0.5 μ L/s flow rate. All data was exported in FCS2 format and processed using FlowJo (TreeStar Inc., Ashland, OR). Data was gated by forward and side scatter and the geometric mean fluorescence calculated. LSRFortessa analysis: For each sample, at least 5,000 counts were recorded using a 0.5 μ L/s flow rate. All data was exported in FCS3 format and processed using FlowJo (TreeStar Inc., Ashland, OR). Data was gated by forward and side scatter then gated to remove any fluorescence values lower than 0. The geometric mean fluorescence was calculated from this gated population.

Fold calculations

Promoter activity represents the mean fluorescence value obtained from flow cytometry analysis. Fold induction is calculated by dividing the promoter activity from a test population, containing both a sigma factor and a sigma-dependent reporter, by the promoter activity of a population of cells containing just the reporter. Inversely, fold repression is calculated by

dividing the promoter activity of a population of cells containing a sigma factor and sigma reporter by the promoter activity of cells containing an anti-sigma factor, sigma factor, and sigma-reporter.

References for Examples 2-6

1. Staroń, A. et al. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. *Molecular Microbiology* 74, 557–581 (2009).
2. Rhodius, V. A., Suh, W. C., Nonaka, G., West, J. & Gross, C. A. Conserved and Variable Functions of the σ E Stress Response in Related Genomes. *PLoS Biol* 4, e2 (2005).
3. Liu, X., Brutlag, D. L. & Liu, J. S. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput* 127–138 (2001).
4. Rhodius, V. A. & Mutalik, V. K. Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, σ E. *PNAS* 107, 2854–2859 (2010).
5. Staden, R. Computer methods to locate signals in nucleic acid sequences. *Nucleic Acids Res* 12, 505–519 (1984).
6. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* 14, 1188–1190 (2004).
7. Salgado, H. et al. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* 41, D203–D213 (2013).
8. Shultzaberger, R. K., Chen, Z., Lewis, K. A. & Schneider, T. D. Anatomy of *Escherichia coli* σ 70 promoters. *Nucl. Acids Res.* 35, 771–788 (2007).
9. Yuzenkova, Y., Tadigotla, V. R., Severinov, K. & Zenkin, N. A new basal promoter element recognized by RNA polymerase core enzyme. *The EMBO Journal* 30, 3766–3775 (2011).
10. Burr, T., Mitchell, J., Kolb, A., Minchin, S. & Busby, S. DNA sequence elements located immediately upstream of the –10 hexamer in *Escherichia coli* promoters: a systematic study. *Nucl. Acids Res.* 28, 1864–1870 (2000).
11. Mitchell, J. E., Zheng, D., Busby, S. J. W. & Minchin, S. D. Identification and analysis of ‘extended –10’ promoters in *Escherichia coli*. *Nucl. Acids Res.* 31, 4689–4695 (2003).

12. Rhodius, V. A., Mutalik, V. K. & Gross, C. A. Predicting the strength of UP-elements and full-length *E. coli* σ E promoters. *Nucl. Acids Res.* 40, 2907–2924 (2012).
13. Chang, A. C. & Cohen, S. N. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* 134, 1141–1156 (1978).
14. Nitta, T., Nagamitsu, H., Murata, M., Izu, H. & Yamada, M. Function of the ζ E Regulon in Dead-Cell Lysis in Stationary-Phase *Escherichia coli*. *J. Bacteriol.* 182, 5231–5237 (2000).
15. Hiratsu, K., Amemura, M., Nashimoto, H., Shinagawa, H. & Makino, K. The *rpoE* gene of *Escherichia coli*, which encodes sigma E, is essential for bacterial growth at high temperature. *J. Bacteriol.* 177, 2918–2922 (1995).
16. Peñas, A. D. L., Connolly, L. & Gross, C. A. SigmaE is an essential sigma factor in *Escherichia coli*. *J. Bacteriol.* 179, 6862–6864 (1997).
17. Buchler, N. E. & Cross, F. R. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Molecular Systems Biology* 5, (2009).
18. Campbell, E. A. et al. Crystal Structure of *Escherichia coli* σ E with the Cytoplasmic Domain of Its Anti- σ RseA. *Molecular Cell* 11, 1067–1078 (2003).
19. Campbell, E. A. et al. A Conserved Structural Module Regulates Transcriptional Responses to Diverse Stress Signals in Bacteria. *Molecular Cell* 27, 793–805 (2007).
20. Murakami, K. S., Masuda, S. & Darst, S. A. Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 Å Resolution. *Science* 296, 1280–1284 (2002).
21. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948 (2007).
22. Rost, B., Yachdav, G. & Liu, J. The PredictProtein server. *Nucleic Acids Research* 32, W321–W326 (2004).
23. Erickson, J. W., Vaughn, V., Walter, W. A., Neidhardt, F. C. & Gross, C. A. Regulation of the promoters and transcripts of *rpoH*, the *Escherichia coli* heat shock regulatory gene. *Genes Dev.* 1, 419–432 (1987).
24. Rouvière, P. E. et al. *rpoE*, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*. *EMBO J* 14, 1032–1042 (1995).

25. Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F. & Voigt, C. A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucl. Acids Res.* (2012).doi:10.1093/nar/gks597
26. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology* 27, 946–950 (2009).
27. Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology* 24, 79–88 (2005).
28. Zaslaver, A. et al. Just-in-time transcription program in metabolic pathways. *Nature Genetics* 36, 486–491 (2004).
29. Shetty, R., Endy, D. & Knight, T. Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering* 2, 5 (2008).
30. Hanahan, D., Jessee, J. & Bloom, F. R. Plasmid transformation of *Escherichia coli* and other bacteria. *Meth. Enzymol.* 204, 63–113 (1991).
31. Typas, A. et al. High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nature Methods* 5, 781–787 (2008).

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

All references, including patent documents, disclosed herein are incorporated by reference in their entirety.

What is claimed is:

CLAIMS

1. A recombinant genetic circuit comprising:
an extracytoplasmic function (ECF) sigma factor;
an anti-sigma factor that binds to the ECF sigma factor; and
a promoter that is recognized by the ECF sigma factor.
2. The recombinant genetic circuit of claim 1, wherein the ECF sigma factor, the anti-sigma factor and/or the promoter are genetically engineered.
3. The recombinant genetic circuit of claims 1 or 2, wherein the ECF sigma factor and/or the promoter are chimeric.
4. The recombinant genetic circuit of any one of claims 1-3, wherein the recombinant genetic circuit is expressed within a host cell and wherein the host cell is a prokaryotic cell.
5. The recombinant genetic circuit of any one of claims 1-3, wherein the recombinant genetic circuit is expressed within a host cell and wherein the host cell is a eukaryotic cell.
6. The recombinant genetic circuit of any one of claims 1-5, wherein the ECF sigma factor and/or anti-sigma factor are codon-optimized for expression in the host cell.
7. The recombinant genetic circuit of any one of claims 1-6, wherein the ECF sigma factor is selected from the group consisting of ECF01-ECF43.
8. The recombinant genetic circuit of any one of claims 1-6, wherein the anti-sigma factor is selected from the group of anti-sigma factors contained within Table 1.
9. The recombinant genetic circuit of claim 1, wherein the recombinant genetic circuit comprises a combination of logic gates.

10. The method of claim 9, wherein the logic gates are selected from the group consisting of AND, NAND, NOR, OR, NOT, XOR, EQUALS, AND, IMPLIES, and ANDN gates.
11. The method of claim 10, wherein the AND gates comprises an ECF sigma factor and a promoter that is recognized by the ECF sigma factor.
12. The recombinant genetic circuit of any one of claims 1-11, wherein the recombinant genetic circuit is a component of a synthetic genetic switch.
13. The synthetic genetic switch of claim 11, wherein the synthetic genetic switch is bistable.
14. The recombinant genetic circuit of any one of claims 1-13, wherein the recombinant genetic circuit is a component of a pulse generator.
15. A system comprising a plurality of recombinant genetic circuits of any of claims 1-14.
16. The system of claim 15, wherein the ECF sigma factors and the anti-sigma factors are orthogonal.
17. A host cell comprising a heterologous genetic circuit comprising
an extracytoplasmic function (ECF) sigma factor;
an anti-sigma factor that binds to the ECF sigma factor; and
a promoter that is recognized by the ECF sigma factor.
.
18. The host cell of claim 17, wherein the ECF sigma factor, the anti-sigma factor and/or the promoter are genetically engineered.
19. The host cell of claim 17 or 18, wherein the ECF sigma factor and/or the promoter are chimeric.
20. The host cell of any one of claims 17-19, wherein the host cell is a prokaryotic host cell.

21. The host cell of any one of claims 17-20, wherein the genetic circuit comprises one or more logic gates selected from the group consisting of AND, NAND, NOR, OR, NOT, XOR, EQUALS, AND, IMPLIES, and ANDN gates.
22. The host cell of claim 21, wherein the AND gates comprises a sigma factor and a sigma factor target DNA sequence.
23. The host cell of claims 21 or 22, wherein two or more logic gates are combined by having the output promoter of an upstream gate serve as the input promoter of a downstream gate.
24. A library comprising two or more anti-sigma factors, wherein each anti-sigma factor selectively binds to one or more ECF sigma factors.
25. The library of claim 24, further comprising one or more ECF sigma factors.
26. The library of claim 25, further comprising one or more promoters that are recognized by the one or more sigma factors.
27. The library of any one of claims 24-26, wherein the anti-sigma factors, ECF sigma factor and/or promoter are genetically engineered.
28. The library of any one of claims 24-27, wherein the ECF sigma factor and/or promoter is chimeric.
29. The library of any one of claims 24-28, wherein the ECF sigma factors and/or anti-sigma factors are codon-optimized for expression in a host cell.
30. The library of any one of claims 24-28, wherein the ECF sigma factors and the anti-sigma factors are orthogonal.

31. The library of any one of claims 24-30, wherein the ECF sigma factors are under the control of an inducible promoter.

32. The library of any one of claims 24-31, wherein the ECF sigma factor is selected from the group consisting of ECF01-ECF43.

33. The library of any one of claims 24-31, wherein the anti-sigma factor is selected from the anti-sigma factors within Table 1.

34. A non-transitory computer readable storage medium encoded with instructions, executable by a processor, for designing a host cell according to claim 17.

35. A computer product comprising a computer readable medium encoded with a plurality of instructions for controlling a computing system to perform an operation for designing a host cell according to claim 17.

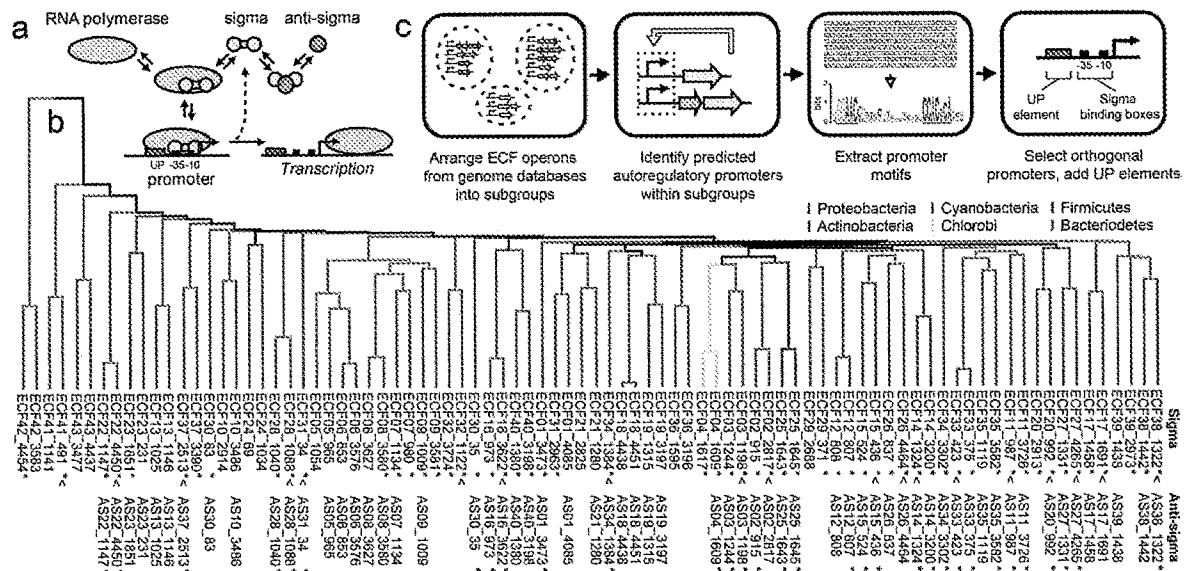


Figure 2

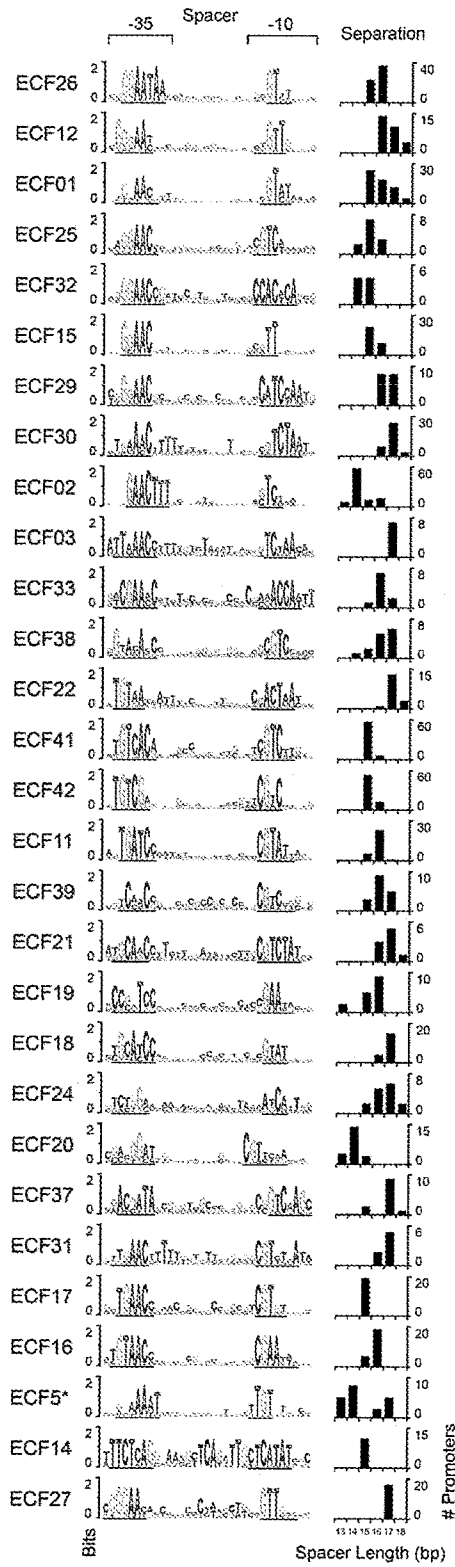


Figure 3

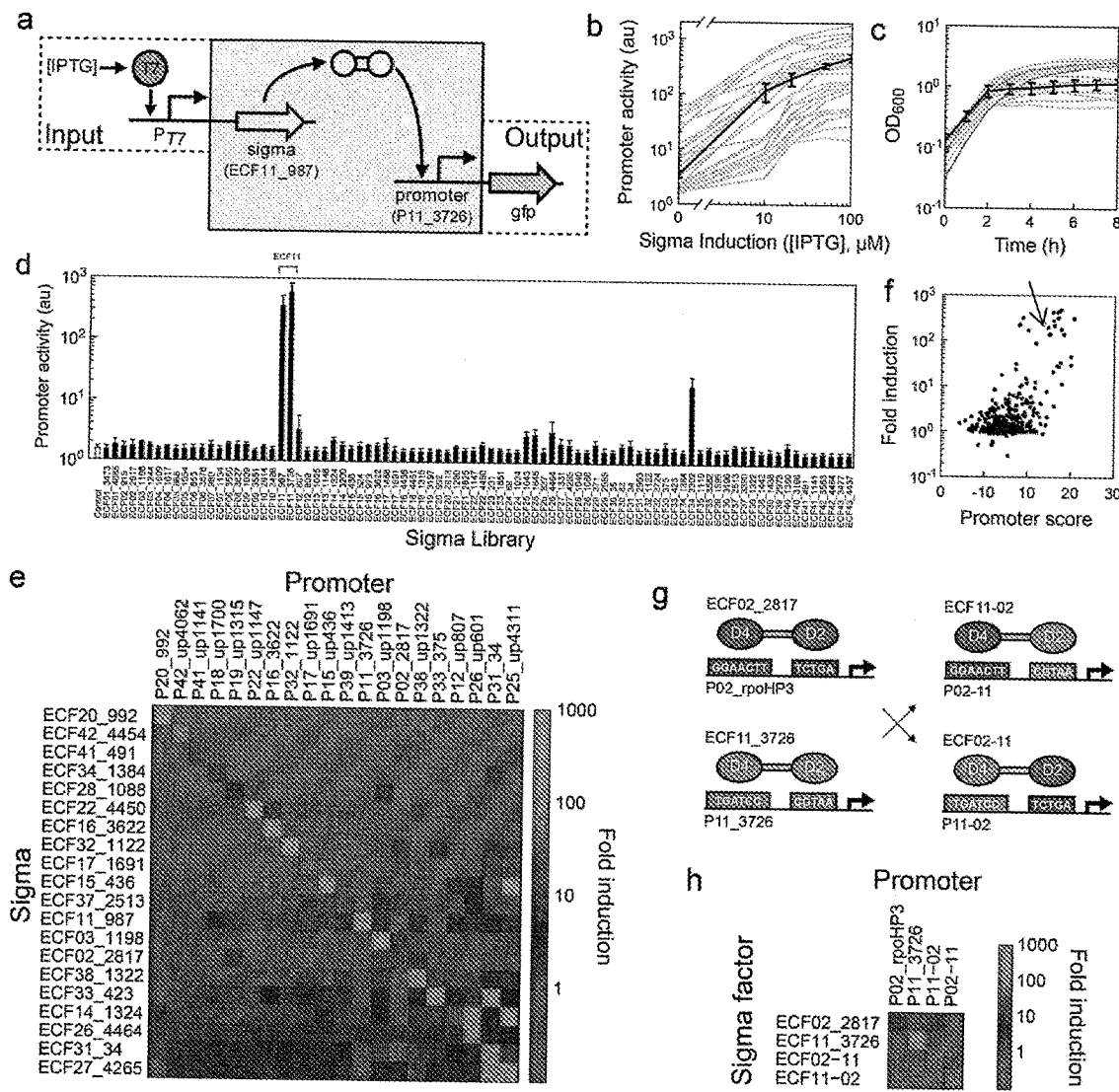
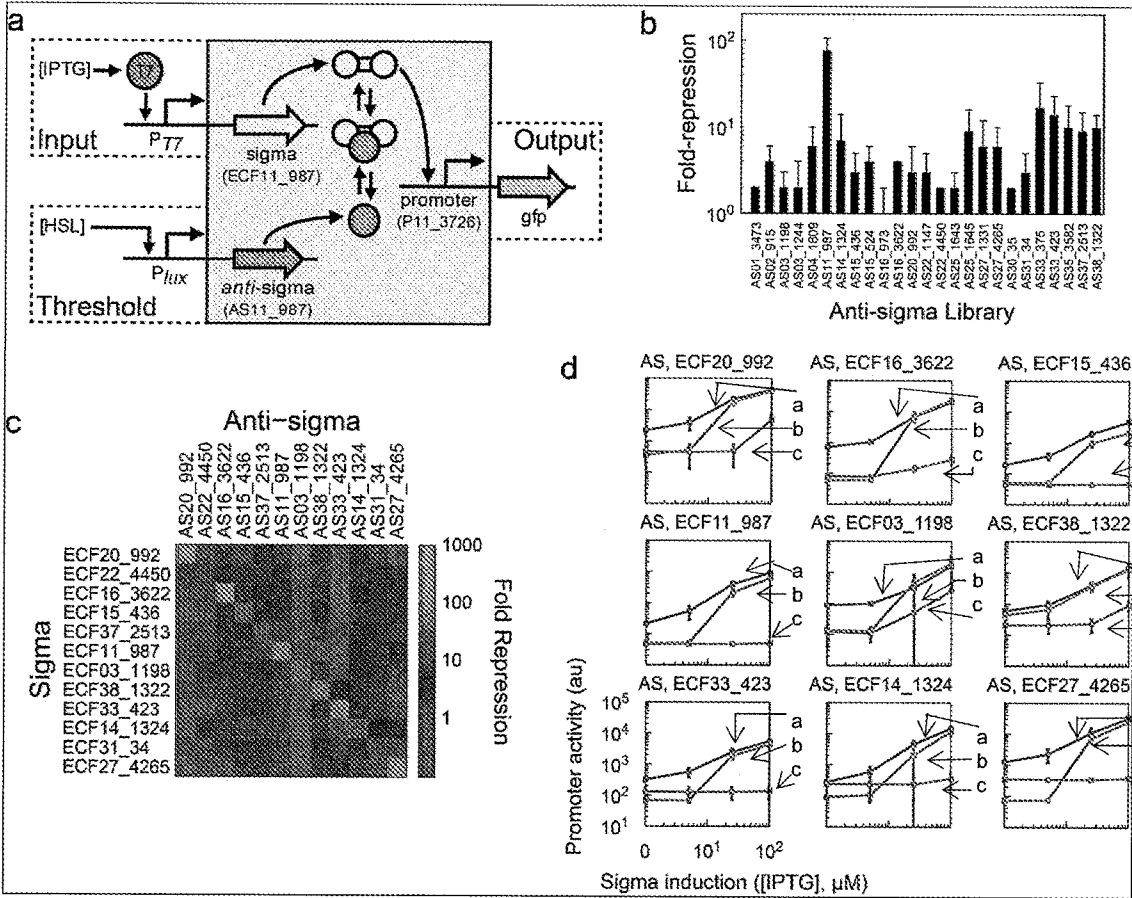


Figure 4



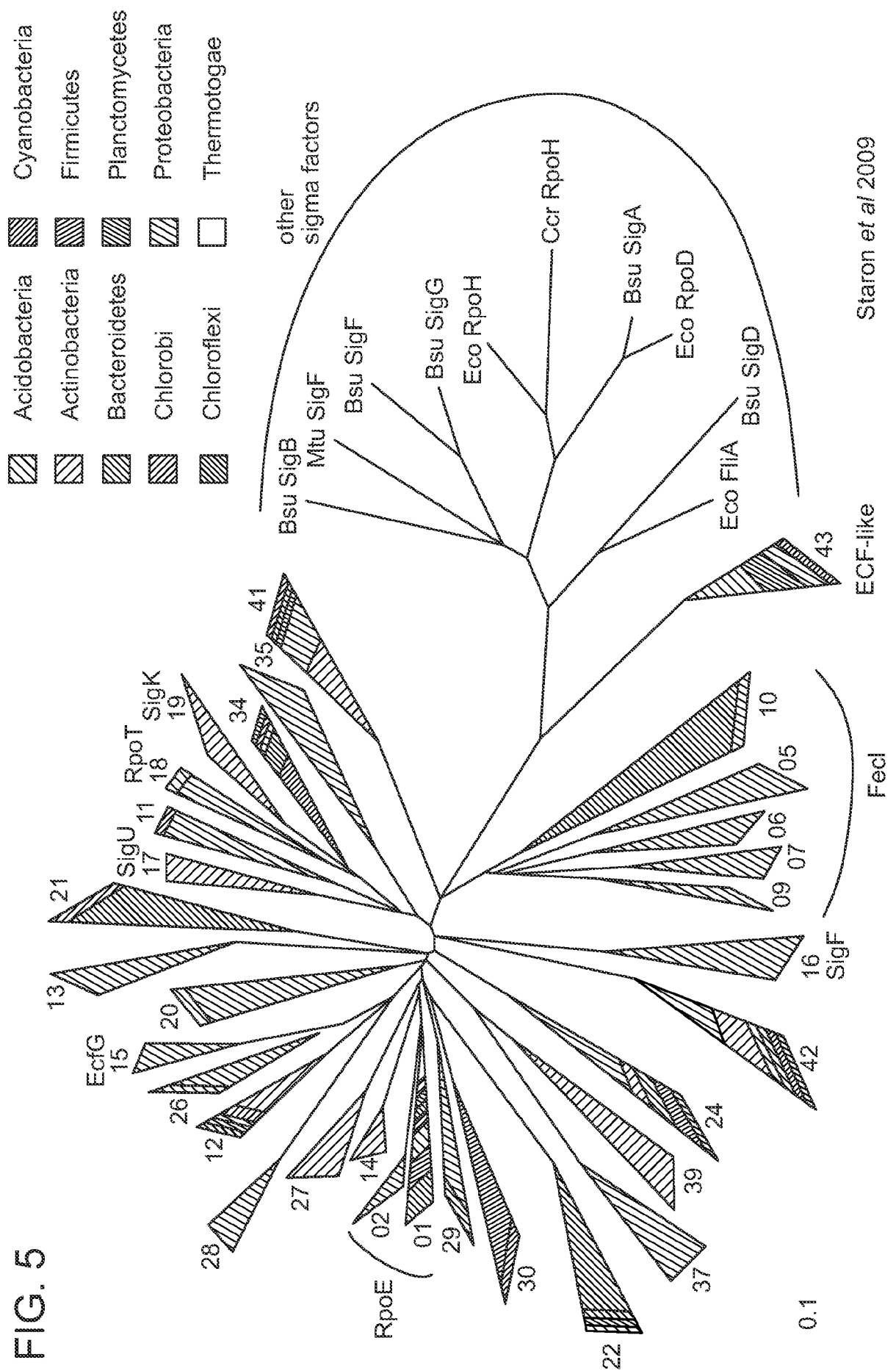


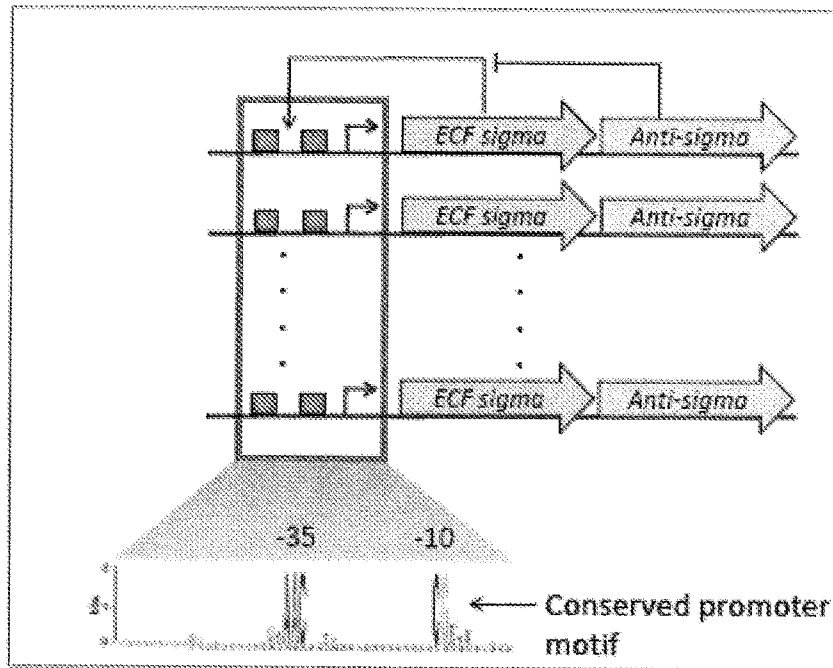
Figure 6

Figure 7

Activity Titrations of 12 sigma (ECF)/anti-sigma (AS) pairs

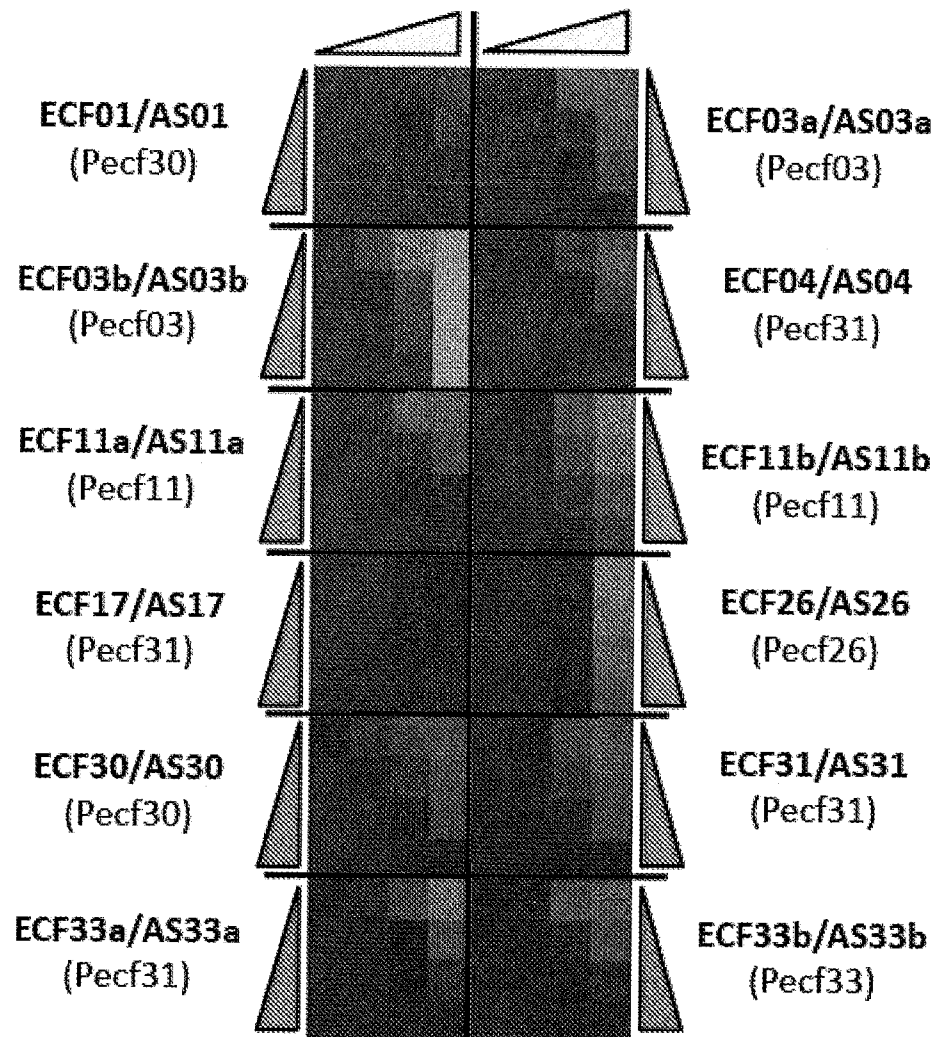


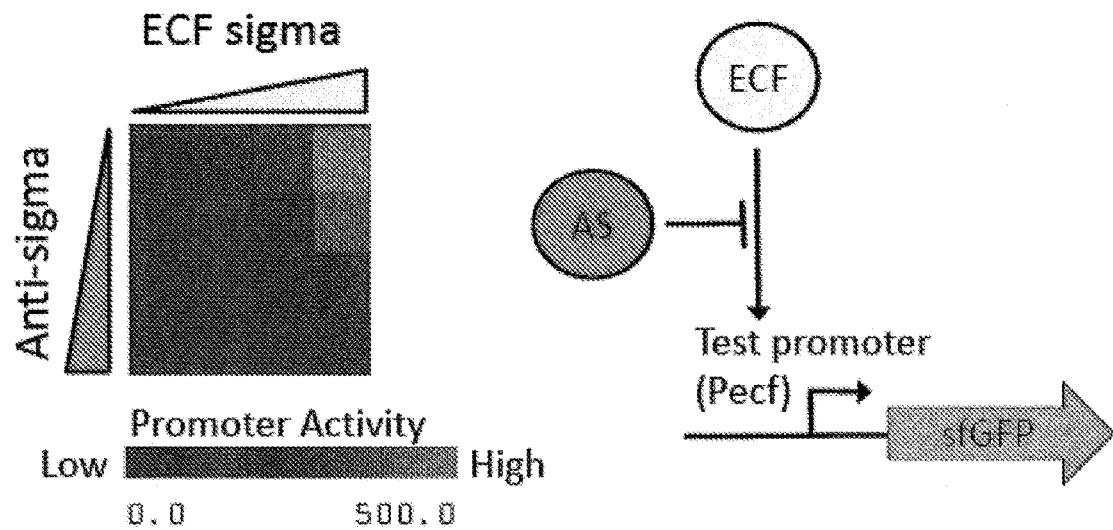
Figure 8

Figure 10

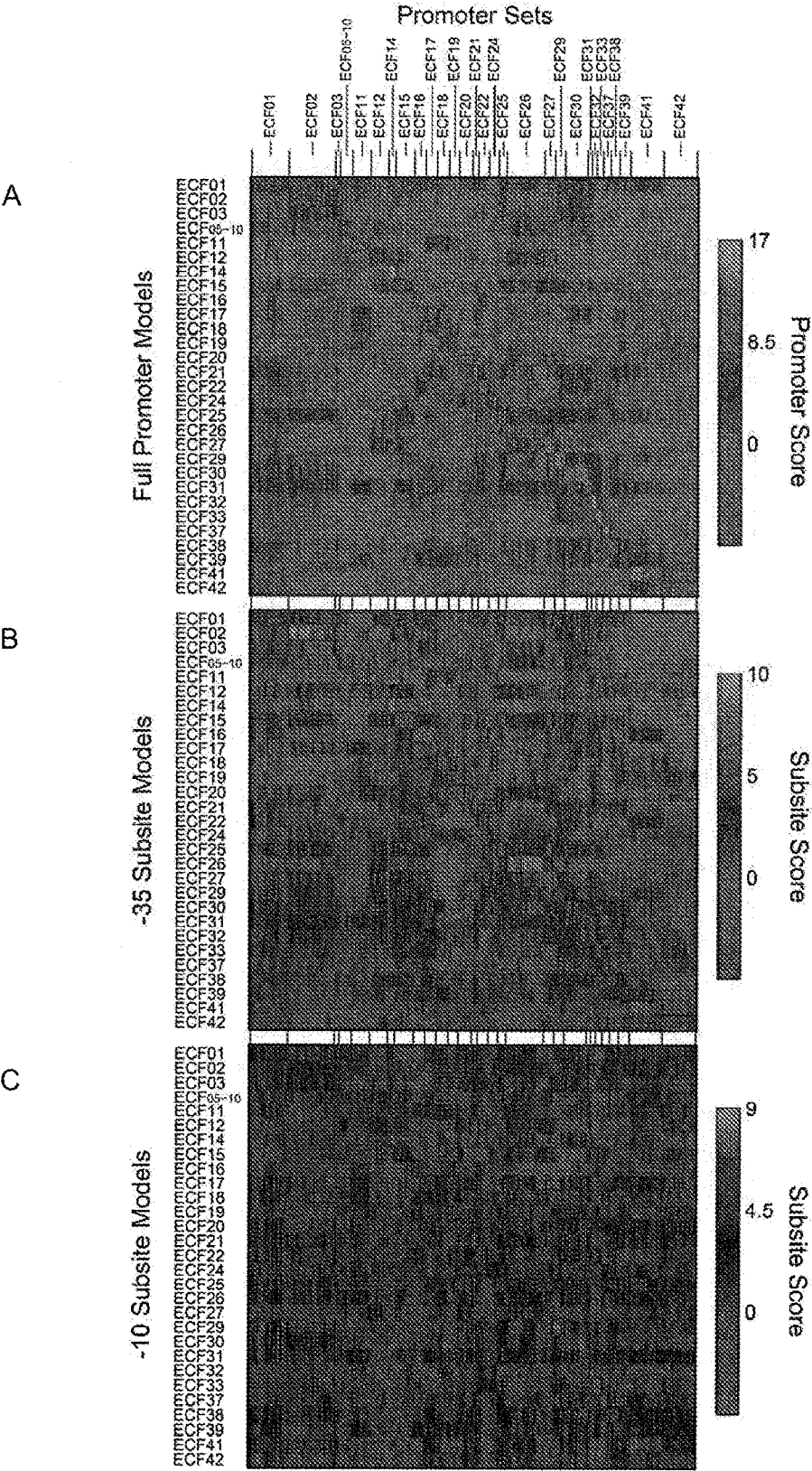


FIG. 11

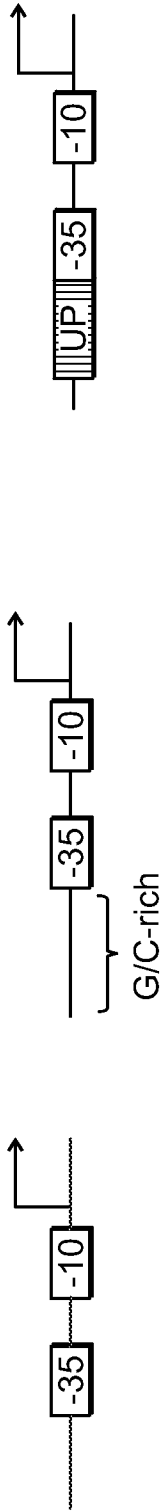
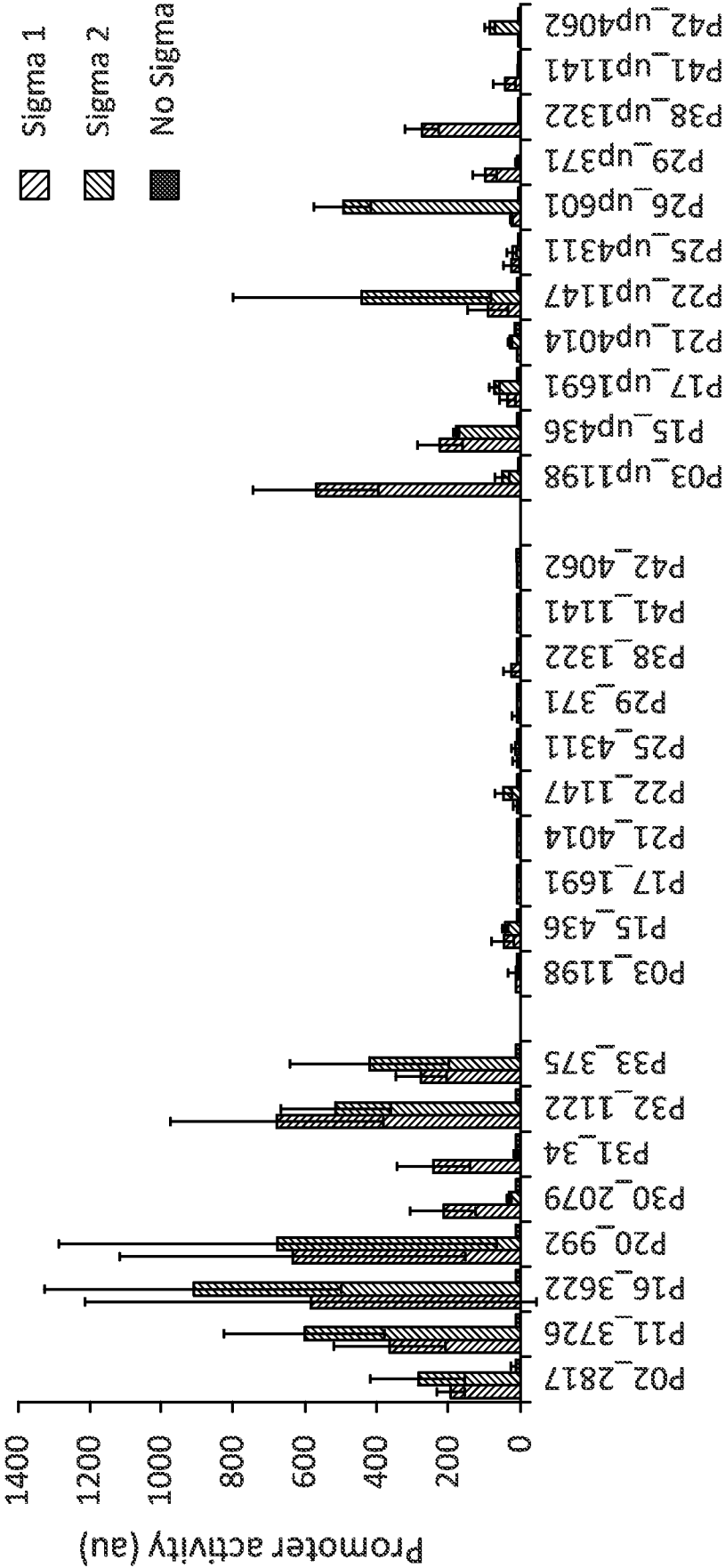


Figure 12

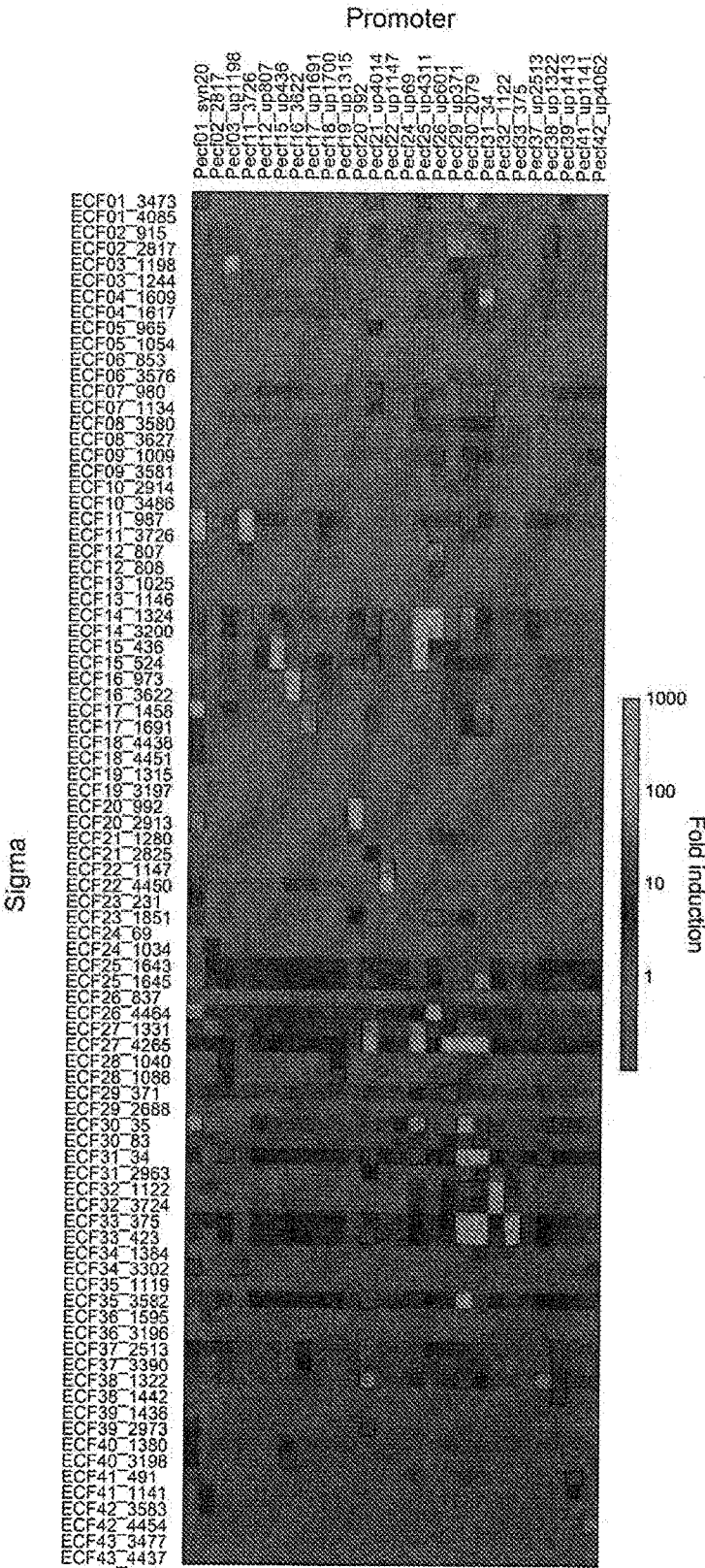


Figure 13

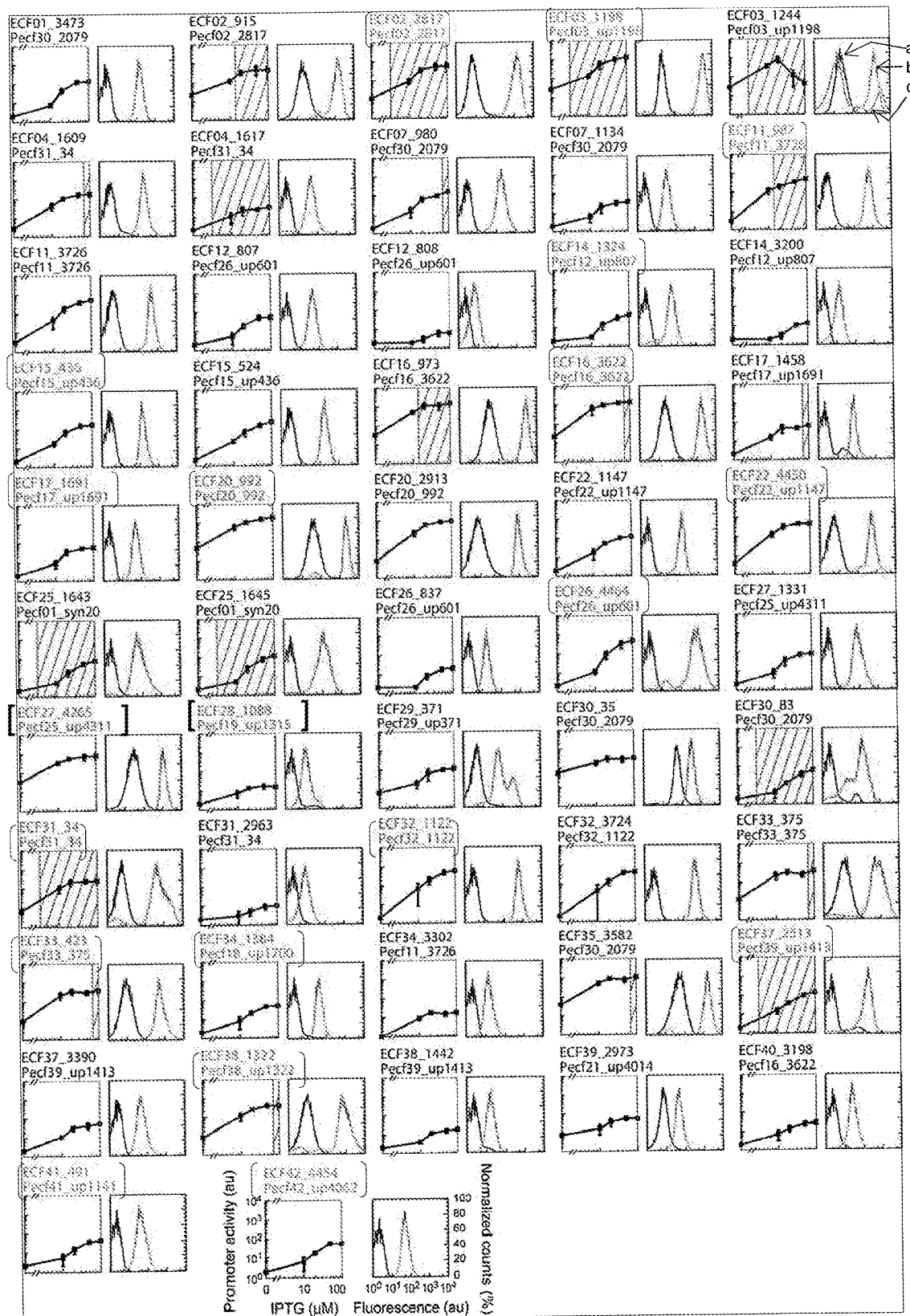


Figure 14

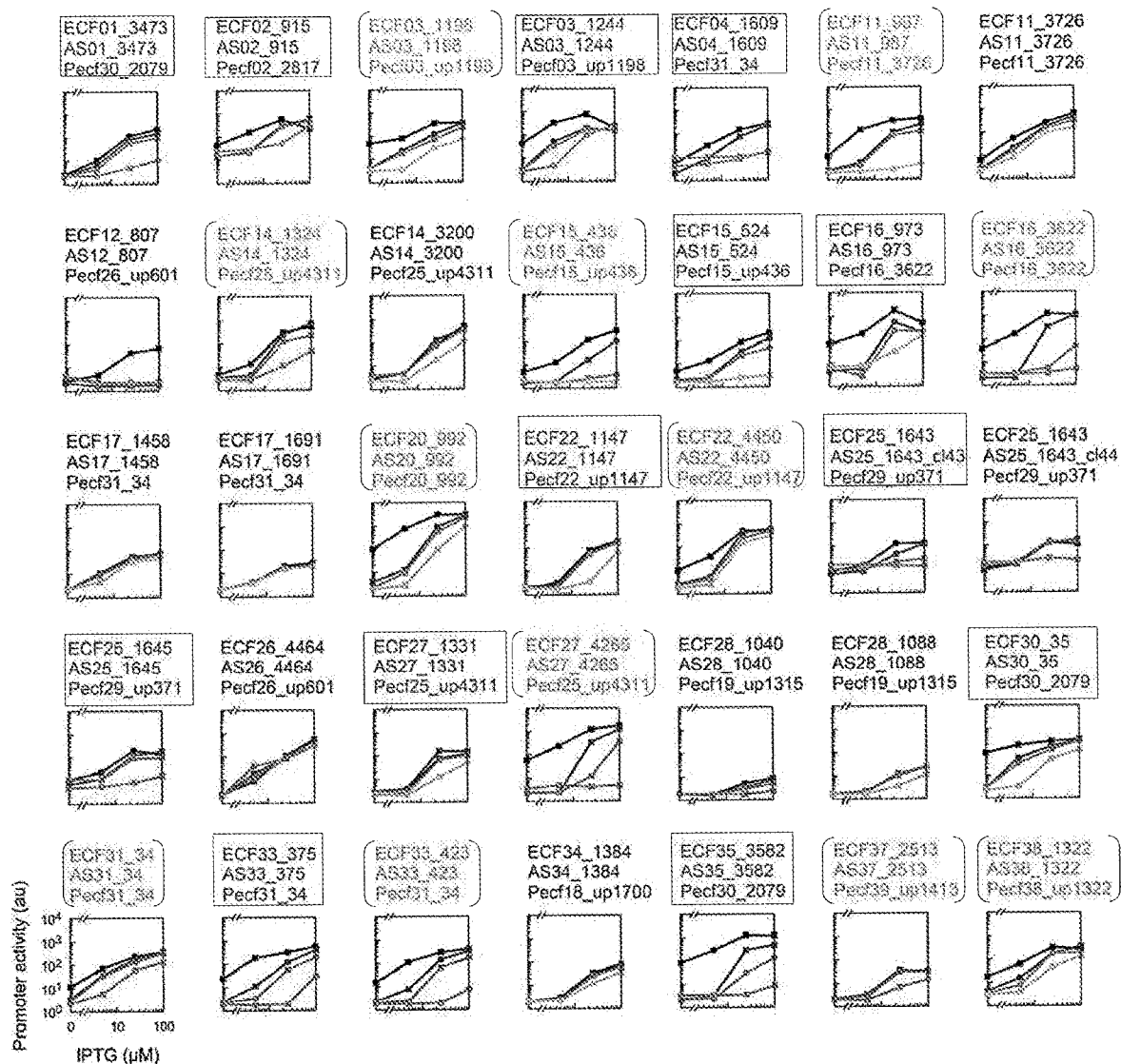


Figure 16



Figure 17

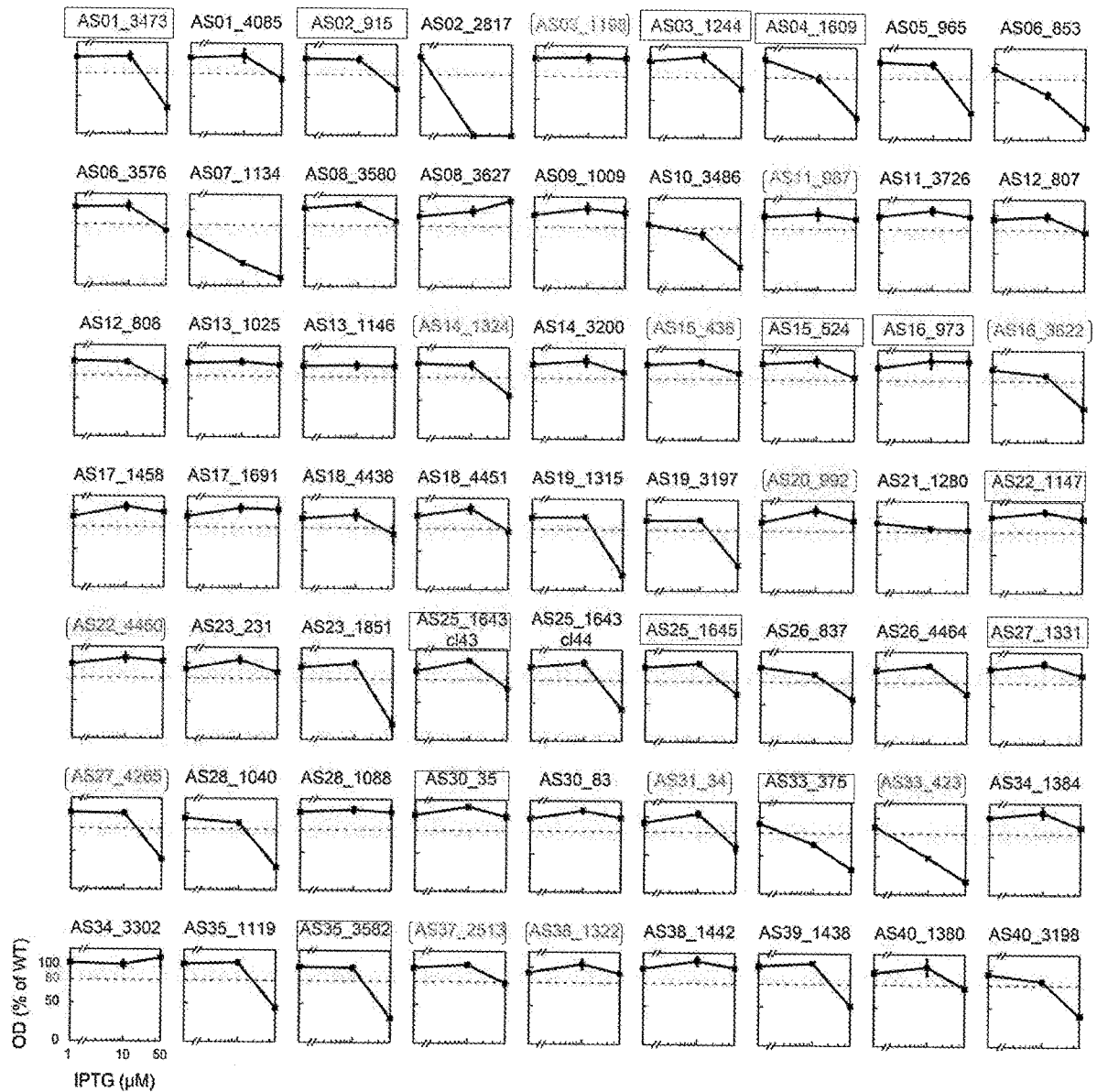


Figure 18

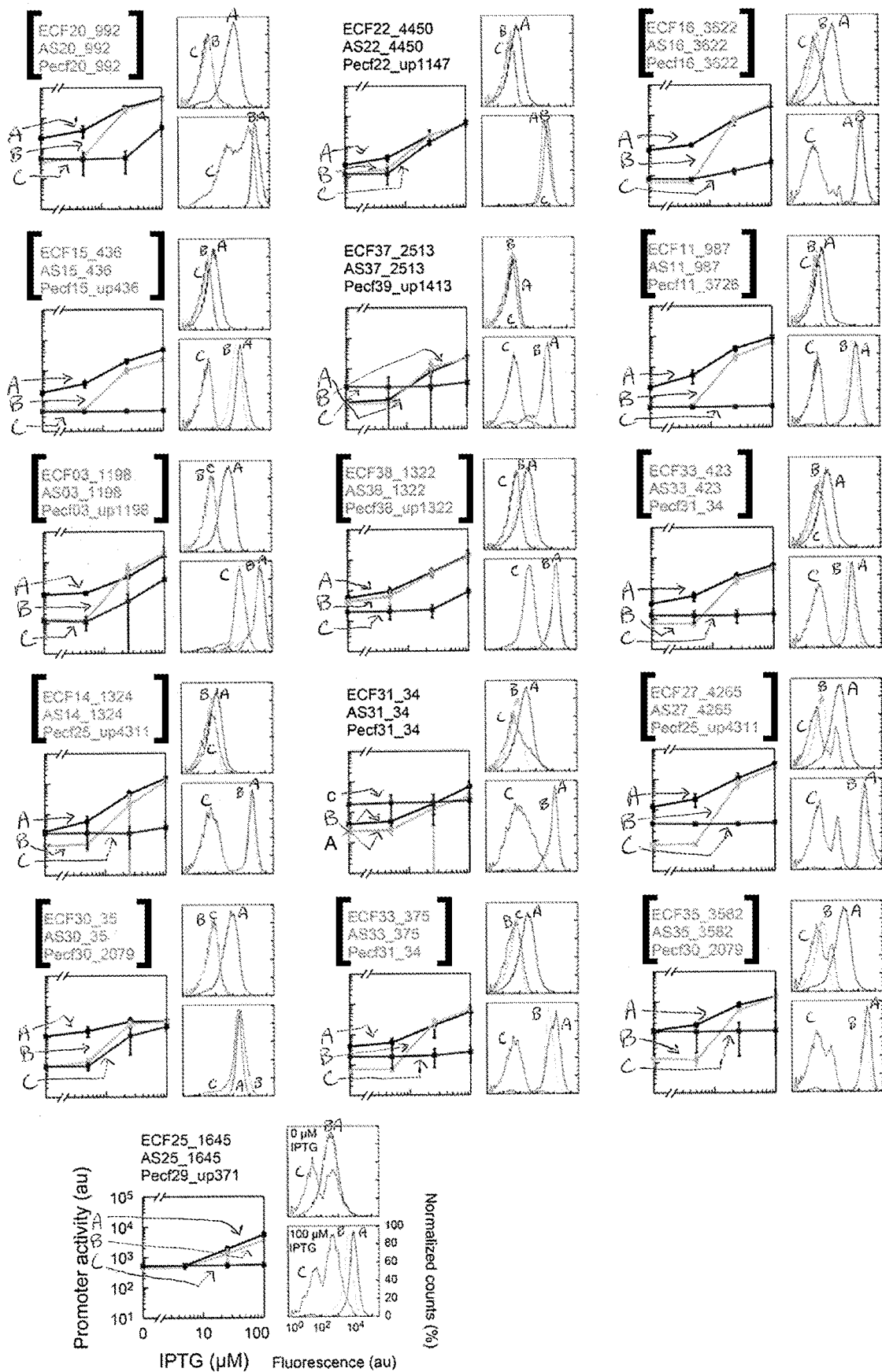


FIG. 19

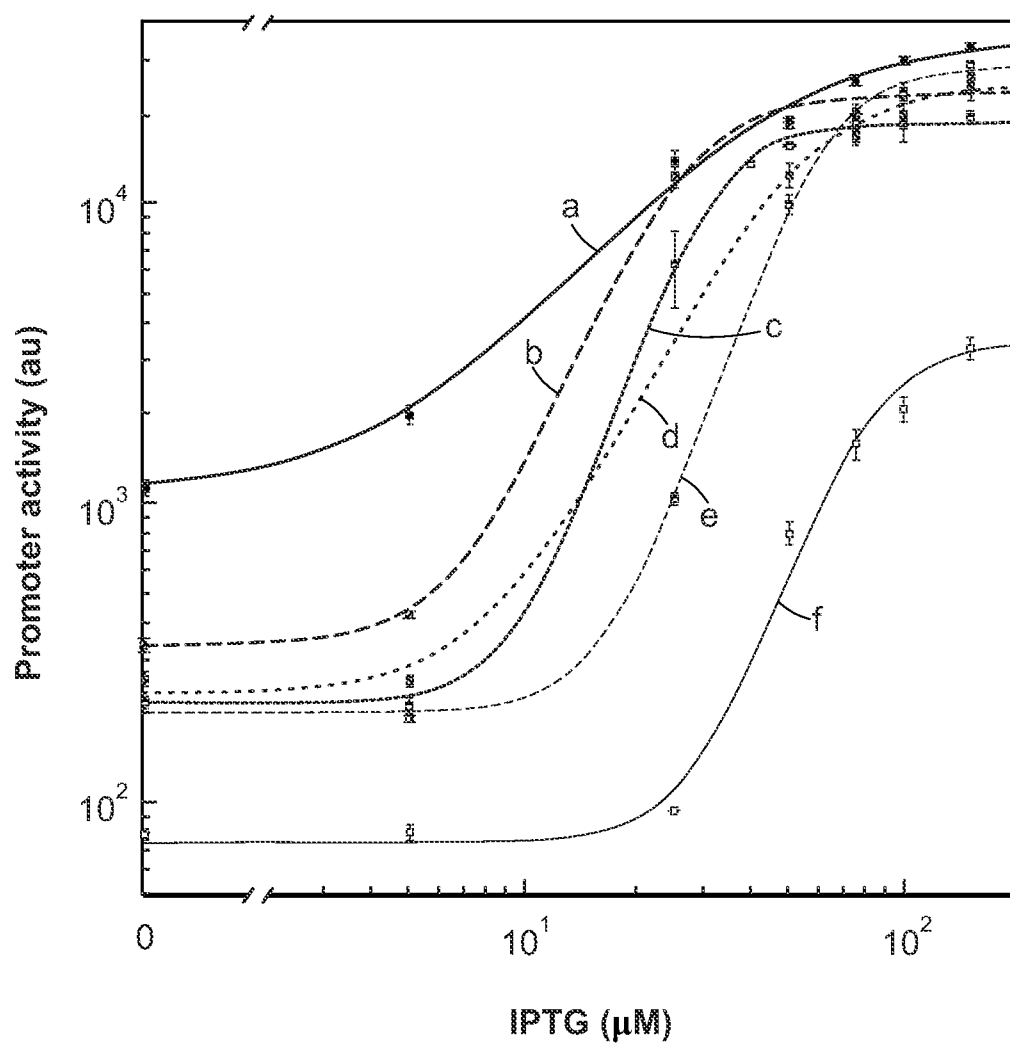


FIG. 20A

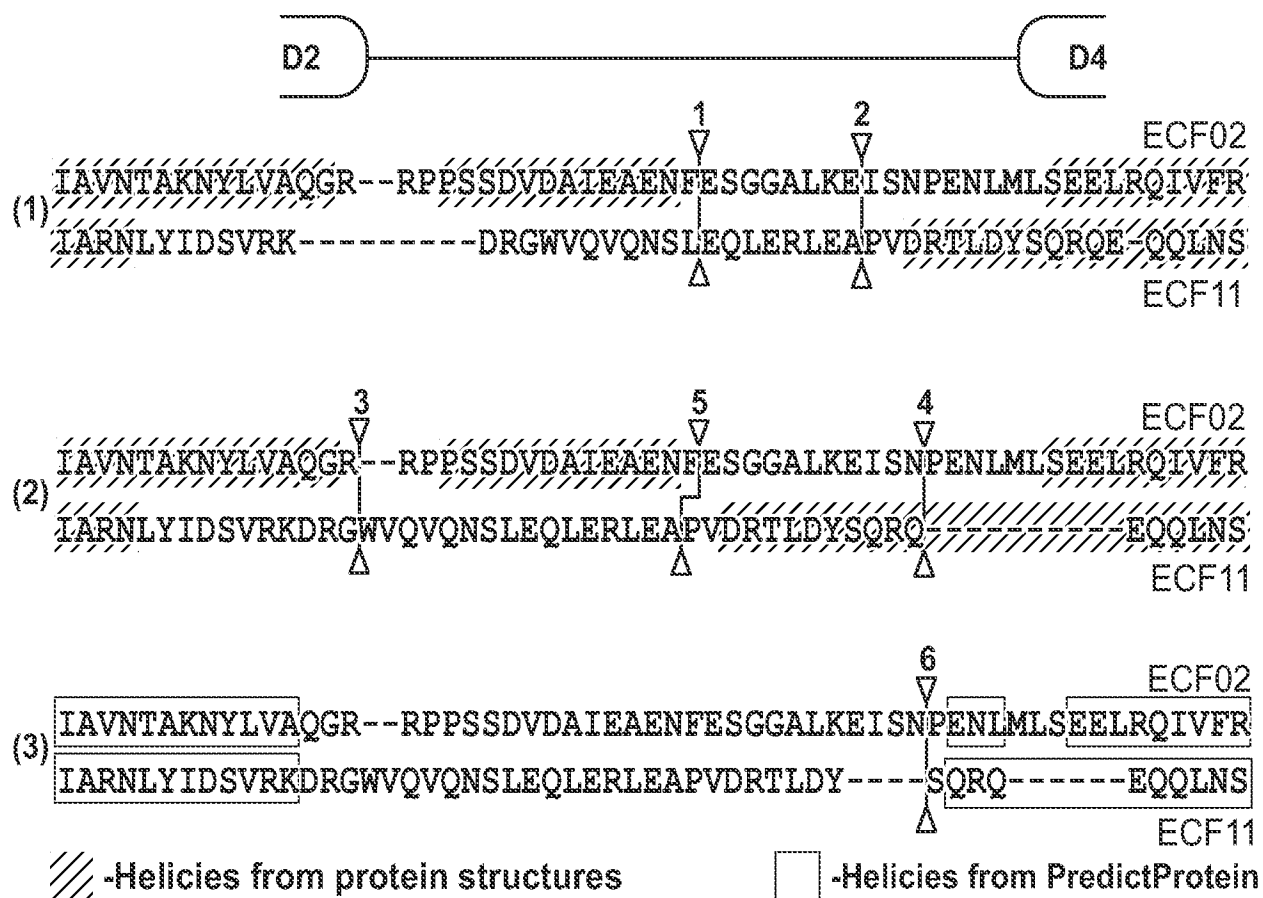


FIG. 20B

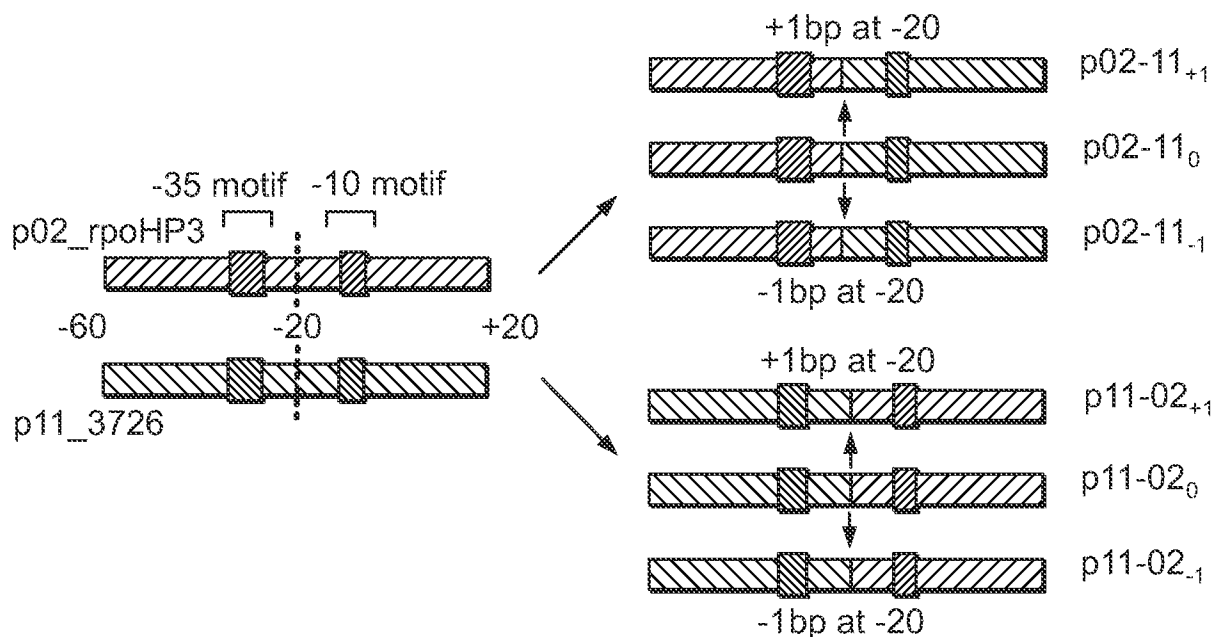


FIG. 21

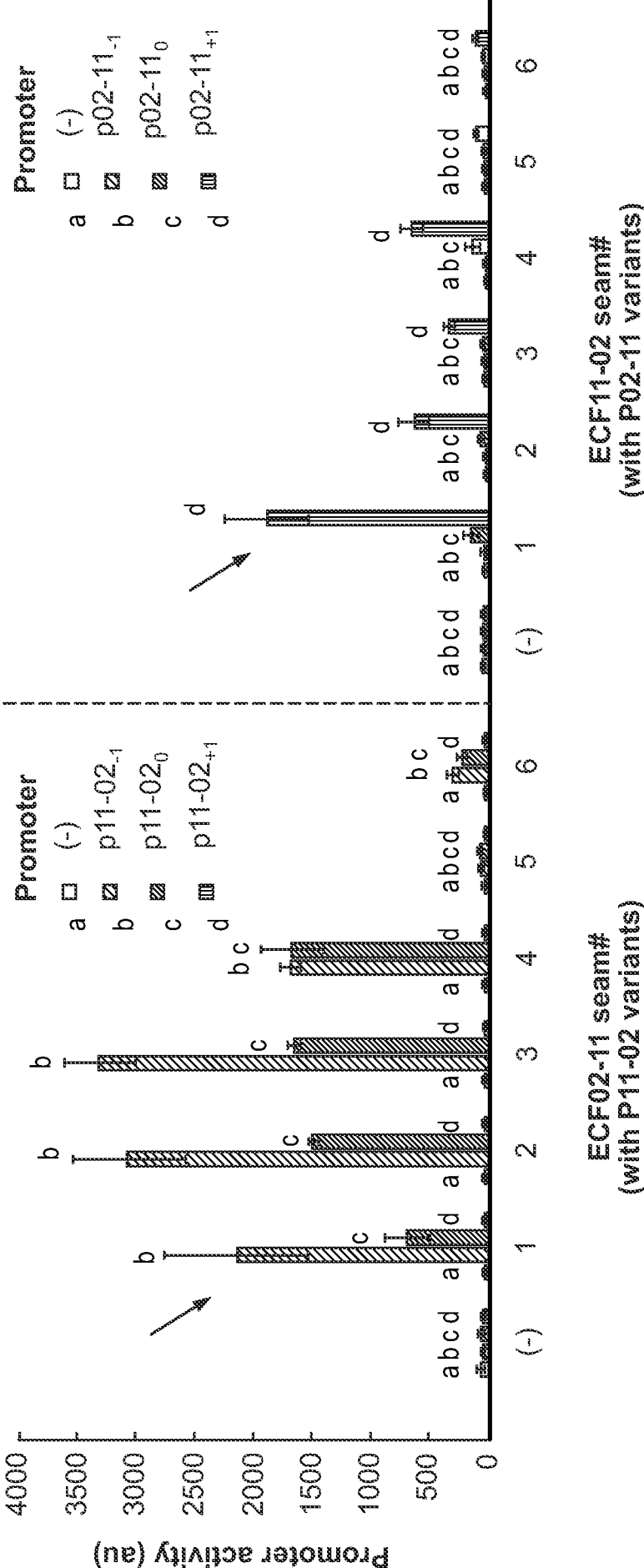


FIG. 22A

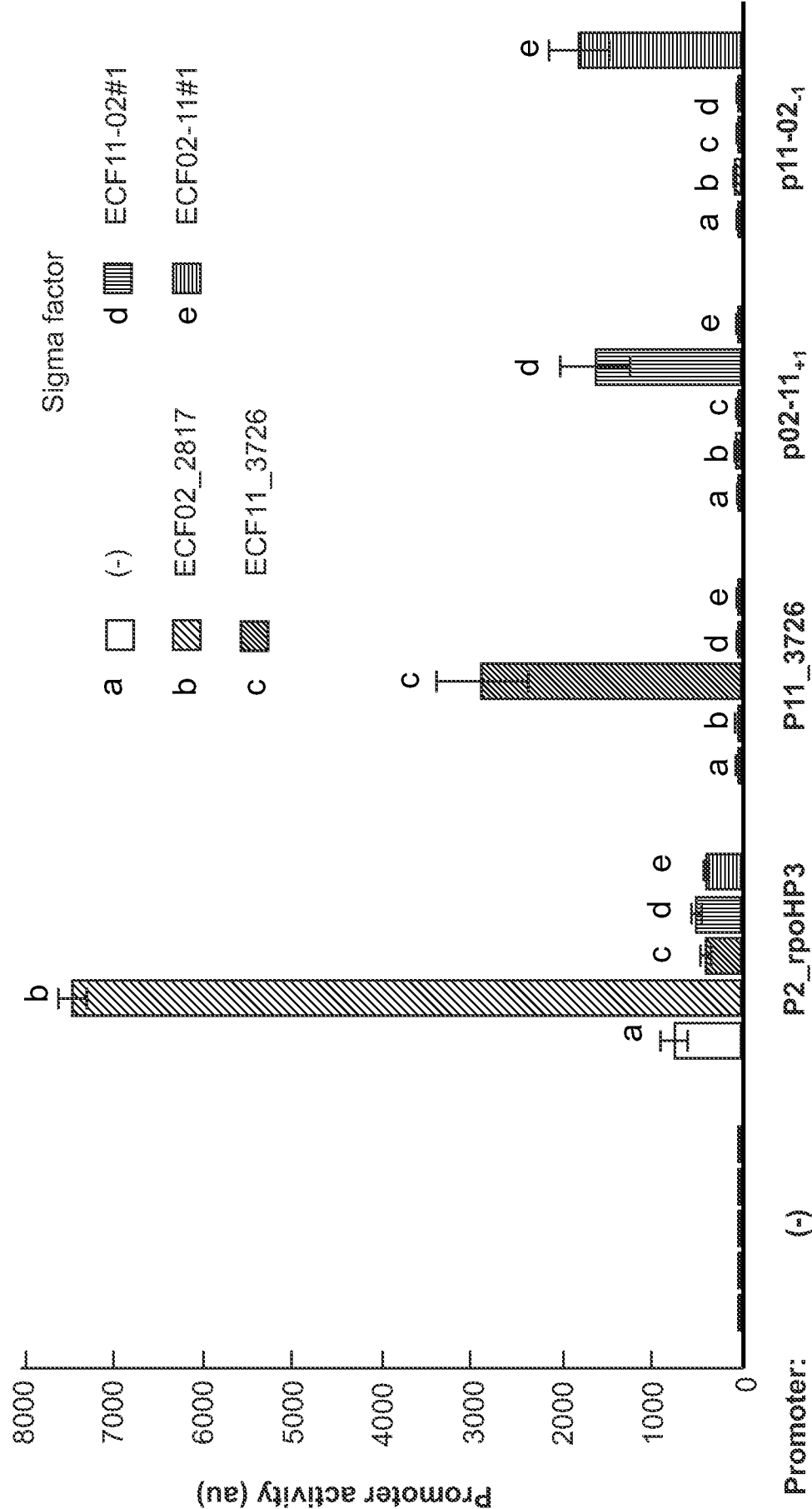


FIG. 22B

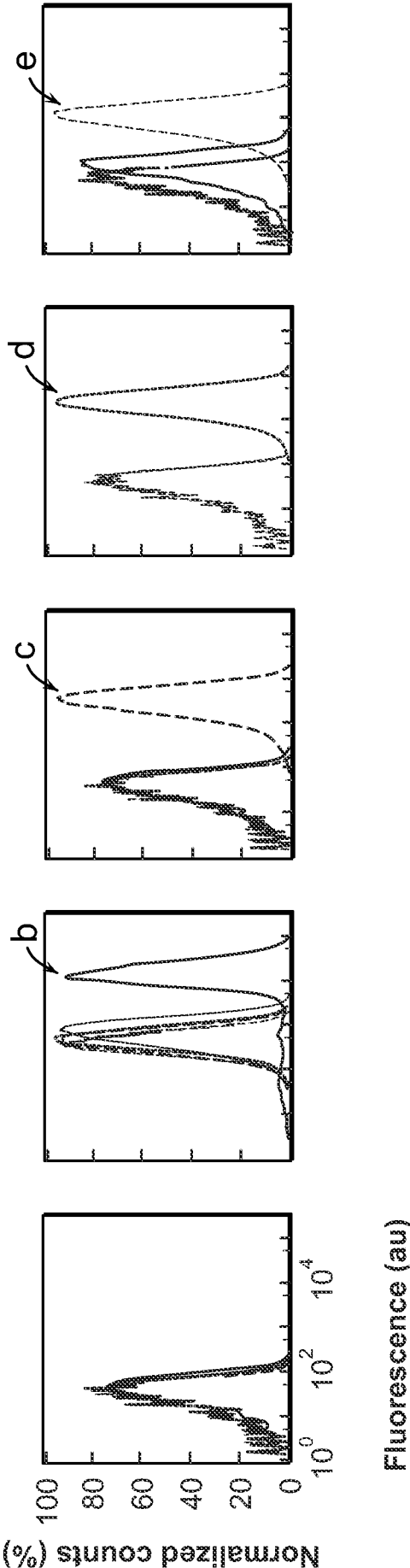


FIG. 23

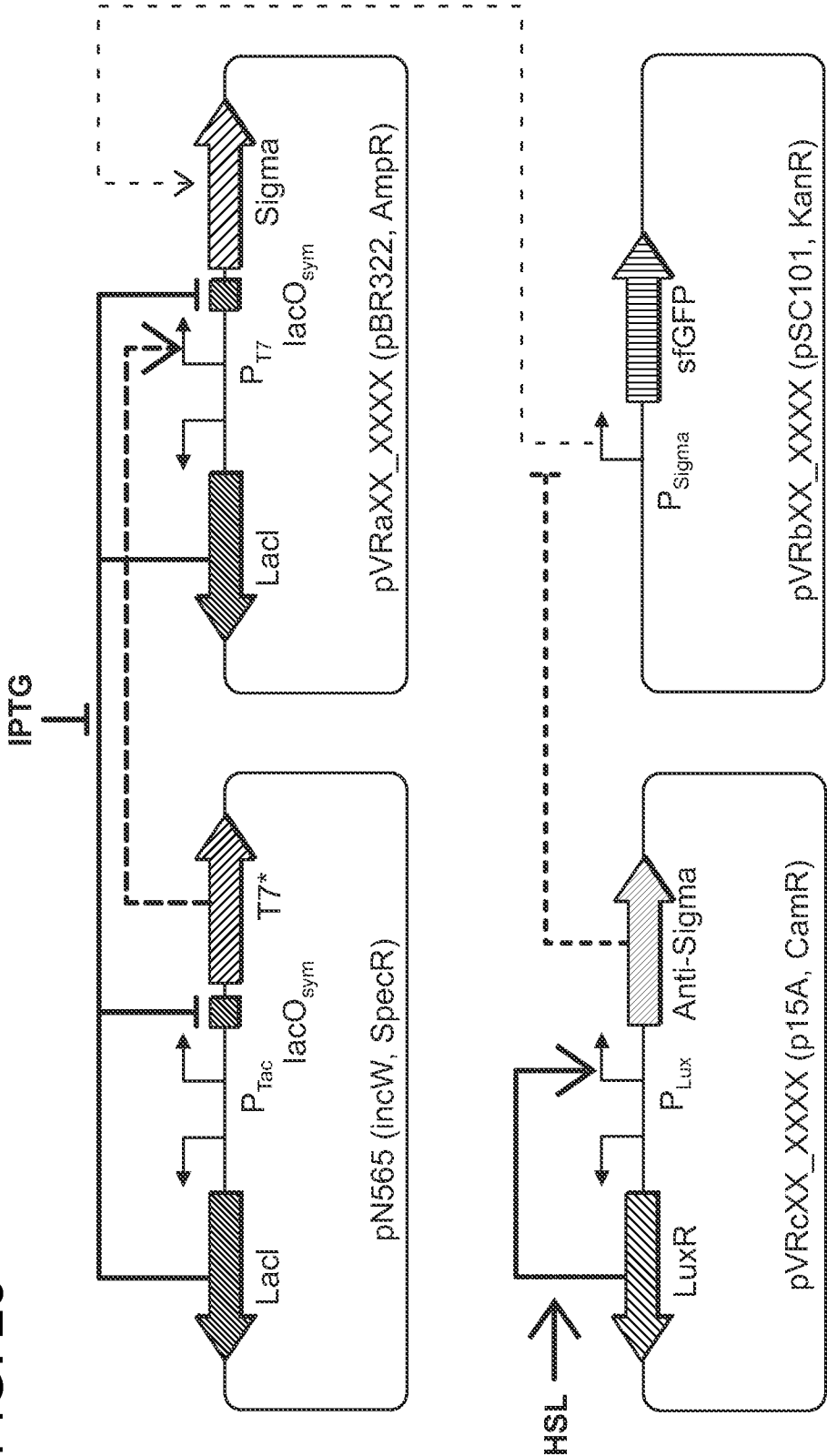
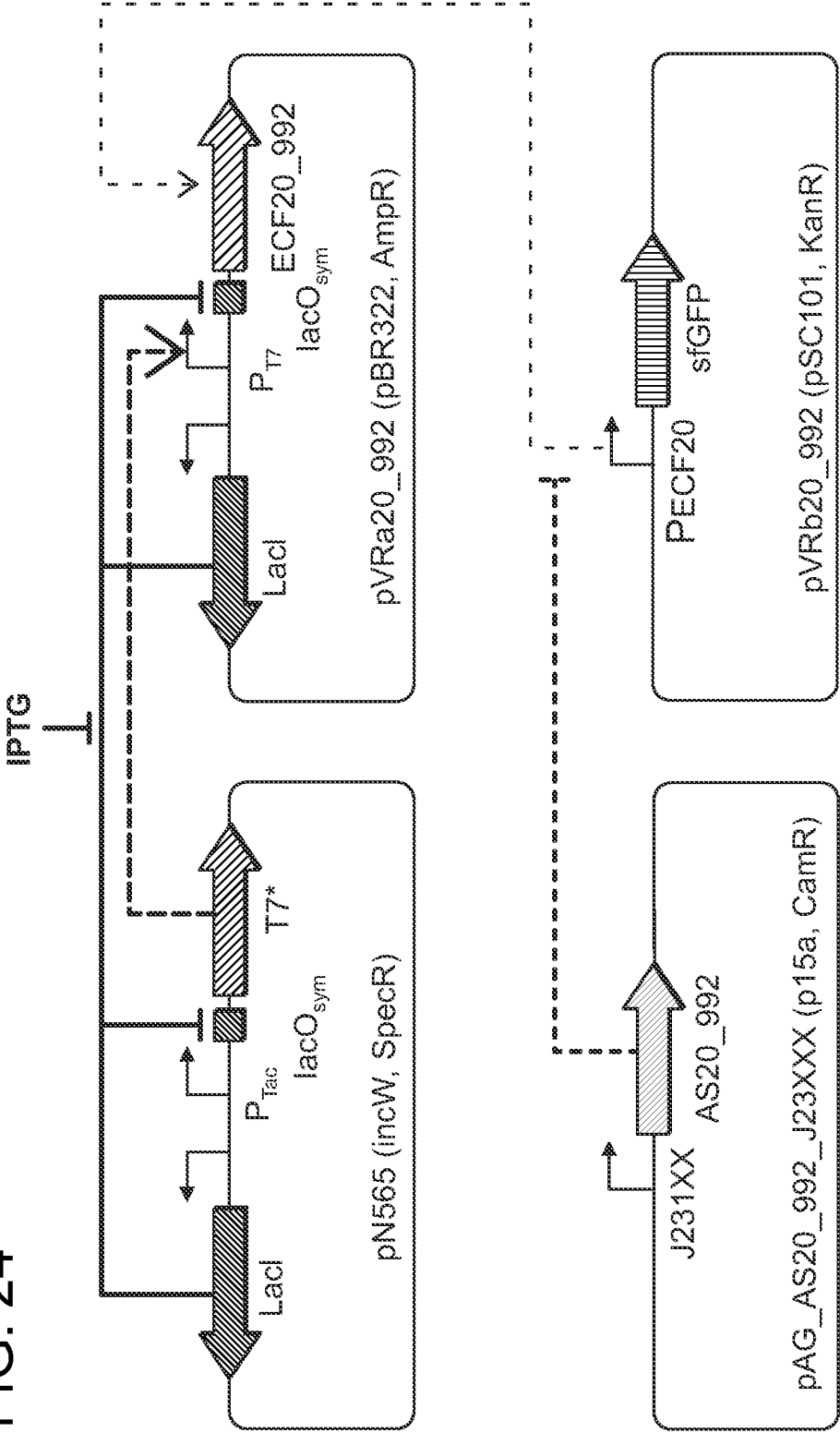
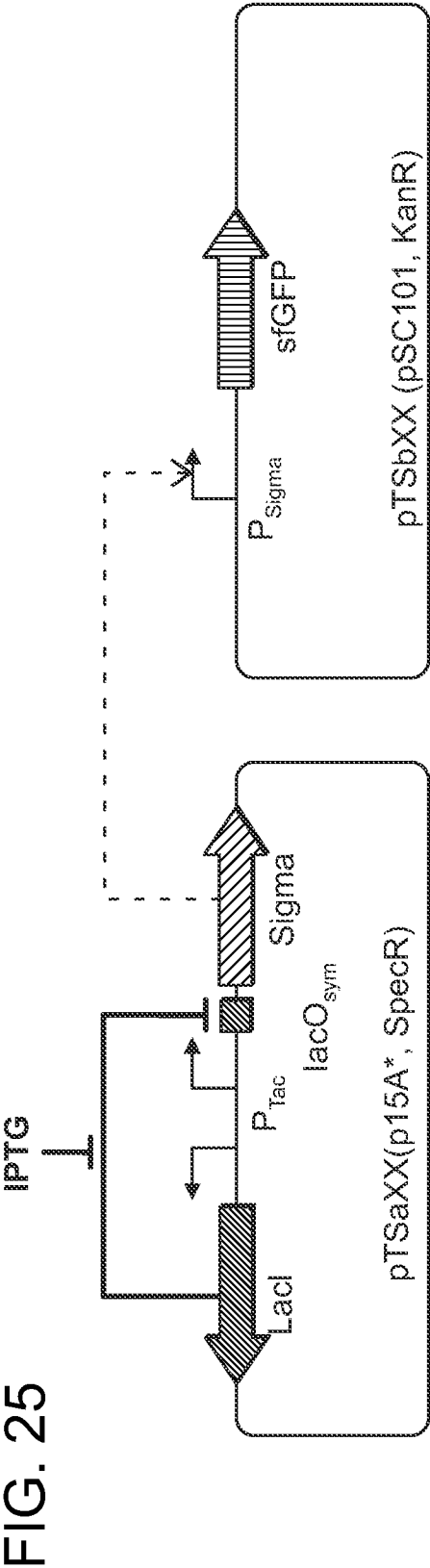


FIG. 24





INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/032145

A. CLASSIFICATION OF SUBJECT MATTER

INV. C07K14/195 C12N15/63 C40B40/10 G06N3/12
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N C07K C40B G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	WO 2012/170436 A1 (UNIV CALIFORNIA [US]; LOU CHUNBO [US]; MOON TAE SEOK [US]; RHODIUS VIR) 13 December 2012 (2012-12-13) cited in the application paragraph [0132] - paragraph [0230]; claims 1-94; examples 1-6	1-35
X,P	DAVID CHEN ET AL: "Sequestration-based bistability enables tuning of the switching boundaries and design of a latch", MOLECULAR SYSTEMS BIOLOGY, vol. 8, 23 October 2012 (2012-10-23), XP055065402, ISSN: 1744-4292, DOI: 10.1038/msb.2012.52 cited in the application the whole document	1,2,4,9, 11-14, 17-23
	----- -/-	



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

17 June 2013

Date of mailing of the international search report

25/06/2013

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Westphal-Daniel, K

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2013/032145

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	JINGFANG LIU ET AL: "Identification and Characterization of the Cognate Anti-Sigma Factor and Specific Promoter Elements of a <i>T. tengcongensis</i> ECF Sigma Factor", PLOS ONE, vol. 7, no. 7, 16 July 2012 (2012-07-16), page e40885, XP055066883, DOI: 10.1371/journal.pone.0040885 the whole document	1,2,4, 17,18,20
X	J. C. W. LOCKE ET AL: "Stochastic Pulse Regulation in Bacterial Stress Response", SCIENCE, vol. 334, no. 6054, 21 October 2011 (2011-10-21), pages 366-369, XP055065400, ISSN: 0036-8075, DOI: 10.1126/science.1208144 cited in the application	1,2, 4-14,17, 18,20-23
Y	the whole document	3,15,16, 19,34,35
X	ELIZABETH A. CAMPBELL ET AL: "A Conserved Structural Module Regulates Transcriptional Responses to Diverse Stress Signals in Bacteria", MOLECULAR CELL, vol. 27, no. 5, 1 September 2007 (2007-09-01), pages 793-805, XP055065518, ISSN: 1097-2765, DOI: 10.1016/j.molcel.2007.07.009 cited in the application	1,2,4, 17,18,20
X	ANNA STARON; ET AL: "The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family", MOLECULAR MICROBIOLOGY, vol. 74, no. 3, 1 November 2009 (2009-11-01), pages 557-581, XP055065323, ISSN: 0950-382X, DOI: 10.1111/j.1365-2958.2009.06870.x cited in the application	24-27, 29-33
Y	the whole document	28

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/032145

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	I. L. GRIGOROVA: "Fine-tuning of the Escherichia coli E envelope stress response relies on multiple mechanisms to inhibit signal-independent proteolysis of the transmembrane anti-sigma factor, RseA", GENES & DEVELOPMENT, vol. 18, no. 21, 1 November 2004 (2004-11-01), pages 2686-2697, XP055065354, ISSN: 0890-9369, DOI: 10.1101/gad.1238604 the whole document	1,4,17, 20
Y	----- ASHOK KUMAR ET AL: "A hybrid sigma subunit directs RNA polymerase to a hybrid promoter in Escherichia coli", JOURNAL OF MOLECULAR BIOLOGY, vol. 246, no. 5, 1 January 1995 (1995-01-01), pages 563-571, XP055065512, ISSN: 0022-2836, DOI: 10.1016/S0022-2836(05)80105-6 the whole document	3,19,28
Y	----- W. AN ET AL: "Synthesis of orthogonal transcription-translation networks", PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, vol. 106, no. 21, 26 May 2009 (2009-05-26) , pages 8477-8482, XP055065319, ISSN: 0027-8424, DOI: 10.1073/pnas.0900267106 the whole document	15,16,19
Y	----- WEISS R ET AL: "Genetic circuit building blocks for cellular computation, communications, and signal processing", NATURAL COMPUTING, KLUWER, NL, vol. 2, no. 1, 1 January 2003 (2003-01-01) , pages 47-84, XP002662632, ISSN: 1567-7818, DOI: 10.1023/A:1023307812034 the whole document	34,35
Y	----- JONGHYEON SHIN AND VINCENT NOIREAUX: "An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells", ACS SYNTHETIC BIOLOGY, AMERICAN CHEMICAL SOCIETY, USA, vol. 1, no. 1, 27 December 2011 (2011-12-27), pages 29-41, XP009170159, ISSN: 2161-5063, DOI: 10.1021/SB200016S the whole document	15,16
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/032145

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ALVIN TAMSIR ET AL: "Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'", NATURE, vol. 469, no. 7329, 13 January 2011 (2011-01-13), pages 212-215, XP055065330, ISSN: 0028-0836, DOI: 10.1038/nature09565 the whole document	15,16
A	----- ABHINAV TIWARI ET AL: "Bistable responses in bacterial genetic networks: Designs and dynamical consequences", MATHEMATICAL BIOSCIENCES, vol. 231, no. 1, 9 March 2011 (2011-03-09) , pages 76-89, XP028193359, ISSN: 0025-5564, DOI: 10.1016/J.MBS.2011.03.004 [retrieved on 2011-03-06] the whole document	1-35
A	----- R. K. SHULTZABERGER ET AL: "Anatomy of Escherichia coli 70 promoters", NUCLEIC ACIDS RESEARCH, vol. 35, no. 3, 26 January 2007 (2007-01-26), pages 771-788, XP055065517, ISSN: 0305-1048, DOI: 10.1093/nar/gkl956 the whole document	1-35
A	----- VIRGIL A. RHODIUS ET AL: "Conserved and Variable Functions of the [sigma]E Stress Response in Related Genomes", GENES & DEVELOPMENT, vol. 7, no. 1, 1 January 2006 (2006-01-01) , page 2618, XP055065396, ISSN: 0890-9369, DOI: 10.1371/journal.pbio.0040002 cited in the application the whole document -----	1-35

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2013/032145

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2012170436 A1	13-12-2012	US 2013005590 A1 WO 2012170436 A1	03-01-2013 13-12-2012
