METHODS OF DESIGNING PROGRAMMABLE INDUCIBLE PROMOTERS

Described herein is a method for identifying synthetic inducible promoters that have specified induction and/or repression for DNA binding proteins such as an allosteric transcription factor and an inducer molecule. The method includes an in vitro selection from an unselected polynucleotide library comprising a plurality of random degeneracies, and an in vivo selection to produce an induced promoter library. Produced is an induction table, which allows the selection of a promoter with specific induction and/or repression properties. Also included are biosensors containing the synthetic inducible promoters.
In vitro selection

Unselected DNA sequences

Selected DNA sequences

In vivo selection

RNAP

-35

-10

GFP

fluorescence binning and quantitative next-generation sequencing

LBD = Ligand Binding Domain

RNAP = RNA Polymerase

DBD = DNA Binding Domain

= Ligand

= High GFP fluorescence

= Low GFP fluorescence

FIG. 1
In vitro selection

FIG. 2
In vivo selection

FIG. 3
CS1 Control Library-Bsal (SEQ ID NO: 2):
TGCGACGGTCTCAGGAGCGCGCCTTGACATCGCATCTTTTTGCTACCAAAAGAT-
TCATGATGAGAAATTCAATAGGAGAAAGGT
CS1 16N-Bsal (SEQ ID NO: 3):
TGCGACGGTCTCAGGAGCGCGCCTTGACNNNNNNNNNNNNNNNNNNNTAAATTAGAT-
TCATGATGAGAAATTCAATAGGAGAAAGGT
CS1 17N-Bsal (SEQ ID NO: 4):
TGCGACGGTCTCAGGAGCGCGCCTTGACNNNNNNNNNNNNNNNNNNNTAAATTAGAT-
TCATGATGAGAAATTCAATAGGAGAAAGGT
CS1 18N-Bsal (SEQ ID NO: 5):
TGCGACGGTCTCAGGAGCGCGCCTTGACNNNNNNNNNNNNNNNNNNNTAAATTAGAT-
TCATGATGAGAAATTCAATAGGAGAAAGGT
CS1 19N-Bsal (SEQ ID NO: 6):
TGCGACGGTCTCAGGAGCGCGCCTTGACNNNNNNNNNNNNNNNNNNNTAAATTAGAT-
TCATGATGAGAAATTCAATAGGAGAAAGGT

Color scheme:
Bsal recognition site
Bsal cut site
-35 and -10 sites
RBS
Control spacer
Degenerated spacer library

FIG. 4
FIG. 5
METHODS OF DESIGNING PROGRAMMABLE INDUCIBLE PROMOTERS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Application 62/481,426 filed on Apr. 4, 2017, which is incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH & DEVELOPMENT

This invention was made with government support under DE-FC02-07ER64494 awarded by the US Department of Energy. The government has certain rights in the invention.

FIELD OF THE DISCLOSURE

The present disclosure is related to de novo methods for designing synthetic inducible promoters and the use of the promoters in, for example, biosensor applications.

BACKGROUND

Biosensors to detect small molecule ligands (e.g., metabolites) have applications in synthetic biology, medical diagnosis, environmental monitoring, bioremediation, and bioenergy. Protein-based biosensors are autonomous, self-powered, miniaturizable, and programmable macromolecules that function in both in vivo and ex vivo environments. Allosteric transcription factors (aTFs), a family of regulatory proteins found in all kingdoms of life, are widely used as biosensors in synthetic biology. Bacterial aTFs, such as LacI or TetR, are composed of allosterically-linked ligand- and DNA-binding domains. When an aTF binds to a ligand, the protein undergoes a conformational change, causing a change in affinity for DNA. In the case of transcription repressors, the loss of DNA affinity allows the RNA polymerase to access the promoter and initiate transcription of the downstream gene. Therefore, the concentration of small molecule can be measured in terms of a reporter gene expression in a dose-dependent manner.

Natural aTF-promoter pairs, however, may be unsuitable for a wide-range of biosensor applications. What is needed are methods of designing synthetic promoters for aTFs that can provide desired induction properties for aTF-based biosensors.

BRIEF SUMMARY

In an aspect, a method of identifying a synthetic inducible promoter, comprises

in vitro selecting a first population of polynucleotides that bind to a DNA binding protein to produce an enriched polynucleotide library by
providing an unselected polynucleotide library comprising a plurality of random degeneracies over one or more regions of a 12 to 40 base pair polynucleotide sequence, and
selecting from the unselected polynucleotide library a plurality of polynucleotide sequences that bind the DNA binding protein to provide the enriched polynucleotide library;
in vivo selecting from the enriched polynucleotide library a second population of polynucleotides that repress or induce production of a reporter protein to produce an induced promoter library by operably linking the enriched polynucleotide library to
a ribosome binding site and a reporter gene to provide a plurality of reporter vectors,
transforming the plurality of reporter vectors into a host strain which co-expresses the DNA binding protein, and growing the host strain provide a culture, and dividing the culture into two split cultures and adding an inducer molecule for the DNA binding protein into one of the two split cultures to provide a non-induced culture and an induced culture, and sorting a control culture transformed with the plurality of reporter vectors with no DNA binding protein expression, the non-induced culture, and the induced culture by reporter protein intensity, to provide a sorted control culture, a sorted non-induced culture, and a sorted induced culture, and binning the sorted control culture, the sorted non-induced culture and the sorted induced culture to produce one or more control gates, one or more non-induced gates and one or more induced gates, wherein the one or more control gates comprises a control promoter library, the one or more non-induced gates comprises a non-induced promoter library, and the one or more induced gates comprises the induced promoter library, wherein a gate is a culture portion comprising a plurality of promoters of specified reporter intensities;
sequencing and analyzing the control promoter library, the non-induced promoter library and the induced promoter library by
culturing and then amplifying the control promoter library, the non-induced promoter library and the induced promoter library, to provide an amplified control promoter library, an amplified non-induced promoter library and an amplified induced promoter library,
quantitatively next generation sequencing the amplified control promoter library, the amplified non-induced promoter library, and the amplified induced promoter library to provide a plurality of sequenced promoters,
providing a control metric, a non-induced metric, and/or an induced metric for at least a portion of the plurality of sequenced promoters based upon identification of each of the at least a portion of the plurality of sequenced promoters in the control promoter library, the non-induced promoter library and/or the induced promoter library,
determining, from the control metric, the non-induced metric and the induced metric for each of the at least a portion of the plurality of sequenced promoters an induction/repression property, and providing an induction table including each of the at least a portion of the plurality of sequenced promoters comprising the promoter sequence and the control metric, non-induced metric, and induced metric, wherein the control metric, non-induced metric, and induced metric provide an induction/repression property of the promoter sequence for the DNA binding protein and the inducer molecule;
and
selecting, based upon the induction table, a synthetic promoter having a specified induction and/or repression for the DNA binding protein and the inducer molecule.
In another aspect, also included herein is a biosensor comprising a synthetic promoter identified by the foregoing process, wherein the biosensor is responsive to the concentration of the inducer molecule.

BRIEF DESCRIPTION OF THE DRAWINGS

The present application recognized that known aTF-promoter pairs suffer from several limitations. An aTF from one host may not work in a heterologous host due to, for example, promoter incompatibility and the presence of cryptic internal regulatory sequences. The dynamic range of a natural promoter may be unsuitable for biosensor applications. In addition, the DNA binding sequence of many natural aTFs is unknown. These shortcomings restrict the wide applicability and portability of aTF biosensors.

The inventors have developed a method for de novo design of synthetic inducible promoters for transcription factors and other DNA binding proteins such as aTFs with tunable dynamic range behavior, and compatibility with virtually any host organism. The method can include selecting inducible promoters, for example, by converting a constitutive promoter of an organism into an inducible promoter by introducing binding sites near the RNA polymerase binding site. By controlling the access of a transcription factor and the RNA polymerase to the promoter, the dynamic range of the system can be controlled. Briefly, the method includes three steps: first, in vitro selection is used to identify tens of thousands of DNA sequences with varying affinities to the transcription factor or other DNA binding protein; second, in vivo transcription of a fluorescent reporter is driven by synthetic promoters containing these in vitro selected transcription factor or other DNA binding protein DNA binding sequences to provide a population of polynucleotides that repress or induce production of the reporter protein; and third, the induction properties of the promoters (tens of thousands) are simultaneously characterized by fluorescence binning and quantitative next-generation sequencing to enable selection of synthetic promoters having a specified induction and/or repression for the transcription factor or other DNA binding protein and the inducer molecule. An embodiment of the method is provided in FIG. 1.

In an embodiment, the method can provide novel promoters of known transcription factors and other DNA binding proteins such as aTFs with tunable transcription factor-promoter and DNA binding protein-promoter pairs for biosensor applications. An embodiment of the method comprises in vitro selecting a first population of polynucleotides that bind to a DNA binding protein to produce an enriched polynucleotide library by providing an unselected polynucleotide library comprising a plurality of random degeneracies over one or more regions of a 12 to 40 base pair polynucleotide sequence, and selecting from the unselected polynucleotide library a plurality of polynucleotide sequences that bind the DNA binding protein to provide the enriched polynucleotide library; in vivo selecting from the enriched polynucleotide library a second population of polynucleotides that repress or induce production of a reporter protein to produce an induced polynucleotide library by operably linking the enriched polynucleotide library to a ribosome binding site and a reporter gene to provide a plurality of reporter vectors, transforming the plurality of reporter vectors into a host strain which co-expresses the DNA binding protein, and growing the host strain to provide a culture, and dividing the culture into two split cultures and adding an inducer molecule for the DNA binding protein into one of the two split cultures to provide a non-induced culture and an induced culture, and sorting a control culture transformed with the plurality of reporter vectors with no DNA binding protein expression, the non-induced culture, and the induced culture by reporter protein intensity, to provide a sorted control culture, a sorted non-induced culture, and a sorted induced culture, and binning the sorted control culture, the sorted non-induced culture and the sorted induced culture to produce one or more control gates, one or more non-induced gates and one or more induced gates, wherein the one or more control gates comprises a control promoter library, the one or more non-induced gates comprises a non-induced promoter library, and the one or more induced gates comprises the induced promoter library, wherein a gate is a culture portion comprising a plurality of promoters of specified reporter intensities; sequencing and analyzing the control promoter library, the non-induced promoter library and the induced promoter library by culturing and then amplifying the control promoter library, the non-induced promoter library and the induced promoter library, to provide an amplified control promoter library, an amplified non-induced promoter library and an amplified induced promoter library,
quantitatively next generation sequencing the amplified control promoter library, the amplified non-induced promoter library, and the amplified induced promoter library to provide a plurality of sequenced promoters, providing a control metric, a non-induced metric, and/or an induced metric for at least a portion of the plurality of sequenced promoters based upon identification of each of the at least a portion of the plurality of sequenced promoters in the control promoter library, the non-induced promoter library and/or the induced promoter library, determining, from the control metric, the non-induced metric and the induced metric for each of the at least a portion of the plurality of sequenced promoters an induction/repression property, and providing an induction table including each of the at least a portion of the plurality of sequenced promoters comprising the promoter sequence and the control metric, non-induced metric, and induced metric, wherein the control metric, non-induced metric, and induced metric provide an induction/repression property of the promoter sequence for the DNA binding protein and the inducer molecule; and selecting, based upon the induction table, a synthetic promoter having a specified induction and/or repression for the DNA binding protein and the inducer molecule.

In an embodiment, the DNA binding protein is an allostERIC transcription factor such as a bacterial transcription factor. In other embodiments, the DNA binding protein is an allosteric activator, a two-component signaling protein, or a eukaryotic nuclear receptor.

In an embodiment, the steps in the method can be performed in a different order. For example, the sequencing step can be performed after the in vivo selection. In this embodiment, one can identify the highly enriched sequences and clone them for the in vivo step. Alternatively, one can start in the in vitro step to identify promoters with high affinity to the DNA binding protein and the inducer molecule. One can then follow up with binning and sorting to identify genotypes and phenotypes.

Exemplary DNA binding proteins include bacterial transcription factor (repressor): TetR, LacI, TglG, MphR; bacterial transcription factor (activator): ArsC, LysR; two-component signaling proteins: PhoP/PhoQ, EnvZ/OmpR, KdpE/KdpD; eukaryotic nuclear receptor: glucocorticoid receptor, mineralocorticoid receptor, estrogen receptor, and others. Table 1 provides exemplary transcription factors.

### Table 1

<table>
<thead>
<tr>
<th>System</th>
<th>Transcription factor family</th>
<th>Example transcription factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArsC/NylS</td>
<td>ArsR, XylR, RhaR, RhaS, UreR, Rob, Rns, MetR</td>
<td></td>
</tr>
<tr>
<td>DnrR</td>
<td>DnrR, SloR, MssR, MetR</td>
<td></td>
</tr>
</tbody>
</table>

### Exemplary transcription factors

<table>
<thead>
<tr>
<th>System</th>
<th>Transcription factor family</th>
<th>Example transcription factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fur</td>
<td>Fur, Zif, Mur, Nur</td>
<td></td>
</tr>
<tr>
<td>LuxR</td>
<td>LuxR, Atyb, QscR, HapR</td>
<td></td>
</tr>
<tr>
<td>Lrp/AmtC</td>
<td>Lrp, AmtC, LrpA, Lpc</td>
<td></td>
</tr>
<tr>
<td>Cpt/Fur</td>
<td>Cpt, FmR, Vfr</td>
<td></td>
</tr>
<tr>
<td>IclR</td>
<td>IclR, ResZ, SwfR</td>
<td></td>
</tr>
<tr>
<td>LacI/CylR</td>
<td>LacI, CylA, GnlR, PucR, CryR, Cya</td>
<td></td>
</tr>
<tr>
<td>Prokaryotes-two component</td>
<td>PhoP/PhoQ, EnvZ/OmpR, KdpE/KdpD, ComA/ComP</td>
<td></td>
</tr>
<tr>
<td>Eukaryotes</td>
<td>Nuclear receptor</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Types of molecules</th>
<th>Example molecules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sugar molecules</td>
<td>lactose, IPTG, L-arabinose, maltose, trehalose, glucose-6P, glycerol-3P, galactose, L-ascorbate, deoxyribonucleic acid, inositol, fructose</td>
</tr>
<tr>
<td>Metallic ions</td>
<td>Hg(II), Cu(II), Ag(I), Au(I), Zn(II), Pb(II), Cd(II)</td>
</tr>
<tr>
<td>Antimicrobial agents</td>
<td>anthrohydroxyacin, chlorophanolid, resorcinol, preflavine, rifamycin, actinorhodin, simocycline D8, trichosan, ethanol, rhodamineg, tetraphenylphosphonium</td>
</tr>
<tr>
<td>Flavonoids</td>
<td>quercetin, fisetin, galangin, phloretin, naringenin, or quercetin</td>
</tr>
<tr>
<td>Steroid hormones</td>
<td>glucocorticoid, estrogen mineralocorticoid</td>
</tr>
</tbody>
</table>

moter. Alternatively, the unselected polynucleotide library comprises any set of random degeneracies, for example, to identify operator sequences for a DNA binding protein of unknown specificity.

If the DNA binding protein binds as a dimer, the length of the polynucleotide can be varied to select for one dimer or multiple sequential dimers with user-defined spacing.

A plurality of polynucleotide sequences that bind the DNA binding protein are then selected from the unselected polynucleotide library to provide the enriched promoter library. The enriched polynucleotide library is enriched for sequences that bind the DNA binding protein factor with a range of affinities that can be determined by the experimenter, for example, by varying the buffer conditions used for binding, such as salt concentration and pH. For example, using low salt conditions may allow for a wide range of binding affinities, while high salt conditions will generally select higher affinity sequences. The enriched polynucleotide library thus includes a plurality of operator sequences that bind to the DNA binding protein with a range of affinities from sub-nanomolar to micromolar.

The in vitro selection can be performed by any suitable method such as a pull-down method, an electrophoretic mobility shift assay (EMSA) method, protein-binding microarrays, or other similar methods. In general, the DNA binding protein, the unselected polynucleotide library, or both may comprise a label such as an affinity tag, a radioactive tag, or a fluorescent tag.

In an example of a pull-down method, the DNA binding protein is labeled with an affinity tag. Exemplary affinity tags include green fluorescent protein (GFP), glutathione-S-transferase (GST) and the FLAG®-peptide tag consisting of eight amino acids (AspTyrLysAspAspAspAspLys) including an enterokinase-cleavage site. The unselected polynucleotide library is incubated with the labeled DNA binding protein under selected buffer conditions that allow for binding of the allosteric transcription factor to selected polynucleotide sequences. Exemplary buffer conditions (PBS (Phosphate-buffered saline) pH 7.5, TBS(Tris-buffered saline) pH 7.5, HBS(HEPES-buffered saline) pH 7.5, or buffers optimized for individual proteins. Additional components for the buffer mix include poly dI-dC, BSA (bovine serum albumin), non-ionic detergent, reducing agent, a protease inhibitor cocktail, and the like. Magnetic beads, for example, can be used to pull down affinity labeled proteins-DNA complexes, separating them from the unbound DNA sequences. The bound DNA sequences are amplified using PCR, for example, so that another round of selection may be performed. In an aspect, 3-5 rounds of in vitro selection are performed.

In another aspect, an EMSA method is used to select the enriched polynucleotide library. Either purified DNA binding protein or a cell lysate containing the DNA binding protein may be employed. Optionally, the binding specificity of the DNA binding protein factor is confirmed using the known genomic binding sequence (e.g., the same length DNA as the library) and an inducer for the DNA binding protein. The unselected polynucleotide library can be labeled with a radioactive or fluorescent label. The unselected polynucleotide library and the allosteric transcription factor are incubated under selected buffer conditions that allow for binding of the DNA binding protein to selected polynucleotide sequences. The binding mixture and a control are loaded on the same gel. The polynucleotide library sequences are cut from the gel at the same position as the bound genomic sequence. The DNA sequences are eluted from the gel, and magnetic beads, for example, are used to collect the DNA molecules. After amplification of the bound DNA sequences, one or more additional rounds of selection may be performed.

For a given DNA binding protein, the number of unique sequences found post in vitro enrichment depends on several factors including the existence of potential operator sequences in the random pool, expression of the DNA binding protein in the host cell, idiosyncratic DNA binding domain conformation, and potential toxicity of DNA binding protein expression.

The enriched polynucleotide library is a library of operator sequences for the DNA binding protein that bind the DNA binding protein with a range of affinities. The enriched library size can range from a few hundred to tens of thousands of operator sequences depending on the stringency of the selection. The higher the round of enrichment is, the lower the effective promoter diversity will be, however, the benefit is to obtain sequences with higher level of binding affinity.

Once the enriched polynucleotide library that is enriched for sequences that bind the DNA binding protein has been produced, the gene expression activity of the enriched polynucleotide library is measured in vivo to simultaneously compute the induction level of the members of the enriched polynucleotide library. The method thus includes in vivo selecting from the enriched polynucleotide library a second population of polynucleotides that repress or induce production of a reporter protein to produce an induced polynucleotide library.

The in vivo selection is carried out by first operably linking the enriched polynucleotide library to a constant ribosome binding site and a reporter gene to provide a plurality of reporter vectors. “Operably linked” means that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the reporter gene. The enriched polynucleotide library can be operably linked in any location, or multiplicity, so long as it operably linked to allow for expression of the reporter gene.

Exemplary reporter proteins include GFP and FbFP (an aerobic condition).

In an aspect, the DNAs from the first enriched polynucleotide library are inserted into a constitutive promoter either between or flanking the RNA polymerase binding sites (e.g., containing the consensus –35 (TTGACA) and –10 sites (TATAAT) in the case of E. coli). This unit is operably linked to a ribosome binding site and a reporter gene (e.g., GFP), for example in an appropriate vector such as a plasmid, to produce a reporter vector. The enriched polynucleotide library can be operably linked to a transcription regulatory site, such as A TATA box. For example, the enriched polynucleotide library can be in the general vicinity of transcription regulatory sites such as TATA box in eukaryotes.

In an aspect, the polynucleotide library can be inserted at different locations on a promoter, as for example downstream of the –10 site, upstream of –35 site or overlapping these sites.

In an aspect, multiple polynucleotide libraries can be inserted along a promoter, that is, multiple polynucleotide libraries can be operably linked to the ribosome binding site and the reporter gene. The plurality of reporter vectors is then transformed into a host strain which co-expresses the allosteric transcription factor, and the host strain is grown provide a culture.

Exemplary host can include both prokaryotes and eukaryotes. In an embodiment, the host strain comprises organisms of the genus Escherichia, Bacillus, Staphylococcus, Cau-
vated-cell-sorting (FACS), for example. FACS allows the separation of the target cell population from the rest of the population distribution graph in the corresponding program. Then the cell sorter executes the sorting based on the gates drawn.

Once the host strain culture has been grown, the culture is divided into two split cultures, and an inducer molecule for the DNA binding protein is added into one of the two split cultures to provide a non-induced culture and an induced culture. Exemplary aTF-inducer pairs are provided in Table 4.

<table>
<thead>
<tr>
<th>System</th>
<th>Organism genus and example species</th>
<th>Example species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prokaryotes</td>
<td>E. coli</td>
<td>Bacillus subtilis, B. thermophilus, B. anthracis</td>
</tr>
<tr>
<td></td>
<td>S. aureus, S. epidermidis</td>
<td>Staphylococcus</td>
</tr>
<tr>
<td></td>
<td>C. Crecentis</td>
<td>Caulobacter</td>
</tr>
<tr>
<td></td>
<td>S. pyogenes, S. pneumoniae</td>
<td>Streptococcus</td>
</tr>
<tr>
<td></td>
<td>T. aquaticus, T. thermophilus</td>
<td>Thermus</td>
</tr>
<tr>
<td></td>
<td>S. coelicolor, S. antibioticus, S. avermiformis</td>
<td>Streptomyces</td>
</tr>
<tr>
<td></td>
<td>M. Pneumoniae, M. genitalium</td>
<td>Mycoplasma</td>
</tr>
<tr>
<td></td>
<td>A. fischeri</td>
<td>Aliivibrio</td>
</tr>
<tr>
<td></td>
<td>A. vinelandii</td>
<td>Azotobacter</td>
</tr>
<tr>
<td></td>
<td>P. aeruginosa, P. putida, P. syringae</td>
<td>Pseudomonas</td>
</tr>
<tr>
<td></td>
<td>A. tumefaciens</td>
<td>Agrobacterium</td>
</tr>
<tr>
<td></td>
<td>Z. mobilis</td>
<td>Zymomonas</td>
</tr>
<tr>
<td></td>
<td>S. cerevisiae</td>
<td>Saccharomyces</td>
</tr>
<tr>
<td></td>
<td>Y. lipolytica</td>
<td>Yarrowia</td>
</tr>
<tr>
<td></td>
<td>P. pastoris</td>
<td>Pichia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eukaryotes</th>
<th>S. cerevisiae</th>
<th>S. cerevisiae</th>
</tr>
</thead>
</table>

<p>| TABLE 3 |
| Exemplary aTF-inducer pairs |</p>
<table>
<thead>
<tr>
<th>Inducer</th>
<th>Native host</th>
</tr>
</thead>
<tbody>
<tr>
<td>CreR</td>
<td>Taurocholate, chelate, taurocholate, saucylate</td>
</tr>
<tr>
<td>CymR</td>
<td>p-cumarate, p-cymene</td>
</tr>
<tr>
<td>DecI</td>
<td>acarate</td>
</tr>
<tr>
<td>HimA</td>
<td>flavonoids, flavonoids</td>
</tr>
<tr>
<td>NaIC</td>
<td>clarinated phenols</td>
</tr>
<tr>
<td>PheK</td>
<td>flavonoids</td>
</tr>
<tr>
<td>QueR</td>
<td>berberine (alkaloid)</td>
</tr>
<tr>
<td>RelR</td>
<td>resorcinol</td>
</tr>
<tr>
<td>SmiT</td>
<td>triacetoxy-protoplast</td>
</tr>
<tr>
<td>TetR</td>
<td>tetracycline</td>
</tr>
<tr>
<td>TglR</td>
<td>phloretin, naringenin, quercetin</td>
</tr>
</tbody>
</table>

A control culture is produced that is a culture transformed with the plurality of reporter vectors with no DNA binding protein expression.

The control culture, the non-induced culture and the induced culture are sorted by reporter protein intensity (e.g., fluorescence intensity), to provide a sorted control culture, a sorted non-induced culture, and a sorted induced culture. Sorting can be done, for example, using fluorescent-activated-cell-sorting (FACS), for example. FACS allows the separation of the target cell population from the rest of the population based on a given parameter (e.g., fluorescence). To separate the desired population, users can directly circle out the desired population (referred as “drawing gates” or “gating”) on the population distribution graph in the corresponding program. Then the cell sorter executes the sorting based on the gates drawn.

Binning is then done to provide culture portions comprising a plurality of promoters that provide specified reporter intensities, referred to as gates. The gates provide a convenient means by which to group the control promoter sequences, non-induced promoter sequences and induced promoter sequences by activity level. The method thus includes binning the sorted control culture, the sorted non-induced culture and the sorted induced culture to produce one or more control gates, one or more non-induced gates and one or more induced gates, wherein the one or more control gates comprises a control promoter library, the one or more non-induced gates comprises a non-induced promoter library, and the one or more induced gates comprises the induced promoter library.

For example, the activity of each promoter sequence (enriched polynucleotide library member operably linked to a ribosome binding site) is measured in terms of reporter level, e.g., GFP level. To simultaneously measure the activity of all the promoter sequences, we classify them into bins based on GFP fluorescence. The number of bins is adjusted depending on the range of promoter activities exhibited. For instance, promoter activities over a 100 fold range are captured in 4-5 bins. The width of each bin is limited by the accuracy of cell sorting. A “gate” is applied in the cell sorter to collect cells within a specified fluorescence range. The gates provide a convenient means by which to group the control sequences, non-induced sequences and induced sequences by activity level. A gate is a portion of a culture comprising a plurality of promoters of specified reporter intensities. Cells containing the enriched polynucleotide library, enriched polynucleotide library+DNA binding protein, enriched polynucleotide library+DNA binding protein+inducer are thus sorted based on their activity.

The control promoter library, the non-induced promoter library, and the induced promoter library are then sequenced and analyzed. The control promoter library, the non-induced promoter library and the induced promoter library are cultured and amplified to provide an amplified control promoter library, an amplified non-induced promoter library and an amplified induced promoter library. Culturing is done, for example, overnight. Amplification can be done with a sequencing primer and a barcode primer, for example, using PCR techniques. The barcode primer is an identifier sequence that allows the sequence reads to be separated into the appropriate library. Thus, the barcode primer allows for demultiplexing after sequencing.

Then each of the amplified control promoter library, the amplified non-induced promoter library and the amplified induced promoter library are sequenced using quantitative next generation sequencing to provide a plurality of sequenced promoters. Each of the plurality of sequenced promoters can be found in one or more of the control promoter library, the non-induced promoter library and the induced promoter library.

In an aspect, sequencing comprises next generation sequencing (NGS), also called high throughput, massively parallel, or deep sequencing. While there are different platforms that allow for NGS, all NGS platforms provide for the parallel sequencing of millions of DNA sequences. Next generation sequencing allows for the quantitation of the proportion of a sequence in each gate, and for normalization across the total of sequence number.

For each promoter sequence, the fluorescence can be ascertained from the gate in which it was found. If the gate width is too narrow, a promoter sequence may be found in multiple gates because fluorescence follows a Gaussian distribution. In such a case, the effective fluorescence of that
A promoter sequence is measured as a weighted sum across multiple, where the weights represent normalized abundance within each gate.

The fluorescence change for different conditions can then be determined for each sequenced promoter under the different conditions, specifically, control with no DNA binding protein, non-induced and induced.

Metrics are then provided for the sequenced promoters. At least five metrics can be measured for each promoter sequence: fluorescence level of constitutive promoter, promoter repressed by DNA binding protein, and promoter induced by adding inducer; fold repression (ratio of fluorescence of constitutive promoter to promoter repressed by DNA binding protein), and fold induction or an association constant (ka) dynamic range (ratio of fluorescence of promoter+DNA binding protein+inducer to promoter+DNA binding protein). The method thus includes providing a control metric, a non-induced metric, and/or an induced metric for at least a portion of the plurality of sequenced promoters based upon identification of each of the at least a portion of the plurality of sequenced promoters in the control promoter library, the non-induced promoter library, and/or the induced promoter library. The metric(s) can be obtained for each unique promoter sequence. Depending on the size of the promoter library, we can quantify the “phenotype” of thousands to hundreds of thousands of promoter sequences simultaneously. The control metric, non-induced metric, and/or induced metric are then used to determine the induction/repression for the at least a portion of the plurality of sequenced promoters.

From the control metric, the non-induced metric and the induced metric for each of the at least a portion of the plurality of sequenced promoters, an induction/repression property is then determined.

An induction table is then provided including each of the at least a portion of the plurality of sequenced promoters comprising the promoter sequence and the control metric, non-induced metric, and induced metric, wherein the control metric, non-induced metric, and induced metric provide an induction/repression property of the promoter sequence for the DNA binding protein and the inducer molecule.

Kinetic profiles can also be extracted. For example, a dose-response experiment, carrying out induction at different inducer concentrations, will allow us to measure additional performance characteristics such as operational range (ligand responsiveness upper and lower bound) and transfer function (through Hill coefficient curve fitting) of induction response.

Finally, based upon the induction table, a synthetic promoter having a specified induction and/or repression for the DNA binding protein and the inducer molecule is selected.

Further included herein are biosensors comprising a synthetic promoter identified by the process described herein, wherein the biosensor is responsive to the concentration of the inducer molecule. In an aspect, the synthetic promoter does not comprise the sequence bound by the DNA binding protein in nature.

The invention is further illustrated by the following non-limiting examples.

## EXAMPLES

### Example 1: Brief Description of the Method of Designing Synthetic Promoters

#### Step 1: In vitro Selection for Non-Native Promoters that Bind to an Allosteric Transcription Factor

Starting from random sequences, a first population of DNA sequences that bind to an allosteric transcription factor (aTf) is selected to produce an enriched library of operator sites. An embodiment of the in vitro selection is provided in FIG. 2.

First, an unselected polynucleotide library is provided that includes a plurality of random degeneracies over one or more regions of a known constitutive promoter. Exemplary libraries are provided in Example 2. An exemplary unselected polynucleotide is 12 to 40 base pairs, specifically 16 to 19 base pairs in length. In an alternative aspect, the unselected polynucleotide library comprises any set of random degeneracies, for example, to identify operator sequences for an aTf of unknown specificity.

Second, the polynucleotide sequences that bind the allosteric transcription factor are selected from the unselected polynucleotide library to provide the enriched polynucleotide library. An example of the in vitro selection is provided in sequence: fluorescence level of constitutive promoter, promoter repressed by 15% to a ribosome binding site and a reporter gene (e.g., GFP), control metric, non-induced metric, and/or induced metric are then used to determine the induction level of the members of the enriched polynucleotide library. The control metric, non-induced metric, and/or induced metric are then used to determine the induction level of the members of the enriched polynucleotide library.

### Example 2: In vivo Selection of the Enriched Library

Once the enriched polynucleotide library that is enriched for sequences that bind the allosteric transcription factor has been produced, the gene expression activity of the enriched polynucleotide library is measured in vivo to simultaneously compute the induction level of the members of the enriched polynucleotide library. This is carried out by binning and sequencing three populations: enriched polynucleotide library alone, enriched polynucleotide library+aTf, enriched polynucleotide+aTf+inducer. FIG. 3 shows an embodiment of the in vivo activity measurement.

In an aspect, the DNAs from the first enriched polynucleotide library are inserted into a constitutive promoter either between or flanking the RNA polymerase binding sites (e.g., containing the consensus −35 (TTGACA) and −10 sites (TATAAT) in the case of E. coli). This promoter is operably linked to a ribosome binding site and a reporter gene (e.g., GFP), for example in an appropriate vector such as a plasmid, to produce a reporter vector, which is then transformed into a host strain such as E. coli. The host strain co-expresses the allosteric transcription factor used for the in vitro selection.

The host strain is then grown to provide a culture. The host strain is then grown to provide a culture. The control culture, the non-induced culture and the induced culture are sorted by dose-response experiment, carrying out induction at different concentrations, specifically, control with no DNA binding protein, non-induced and induced.

The in vitro selection can be performed by any suitable method such as a pull-down method or an electrophoretic mobility shift assay (EMSA) method. In either method, the allosteric transcription factor, the unselected polynucleotide library, or both may comprise a label such as an affinity tag, a radioactive tag or a fluorescent tag.

### Step 2: In vivo Selection of the Enriched Library

Once the enriched polynucleotide library that is enriched for sequences that bind the allosteric transcription factor has been produced, the gene expression activity of the enriched polynucleotide library is measured in vivo to simultaneously compute the induction level of the members of the enriched polynucleotide library. This is carried out by binning and sequencing three populations: enriched polynucleotide library alone, enriched polynucleotide library+aTf, enriched polynucleotide+aTf+inducer. FIG. 3 shows an embodiment of the in vivo activity measurement.

In an aspect, the DNAs from the first enriched polynucleotide library are inserted into a constitutive promoter either between or flanking the RNA polymerase binding sites (e.g., containing the consensus −35 (TTGACA) and −10 sites (TATAAT) in the case of E. coli). This promoter is operably linked to a ribosome binding site and a reporter gene (e.g., GFP), for example in an appropriate vector such as a plasmid, to produce a reporter vector, which is then transformed into a host strain such as E. coli. The host strain co-expresses the allosteric transcription factor used for the in vitro selection.

The host strain is then grown to provide a culture. The culture is divided into two split cultures and an inducer molecule for the allosteric transcription factor is added into one of the split cultures to provide a non-induced culture and an induced culture. A control culture transformed with the plurality of reporter vectors with no allosteric transcription factor expression is also prepared. The control culture, the non-induced culture and the induced culture are sorted by

### Step 3: In vivo Selection of the Enriched Library

Once the enriched polynucleotide library that is enriched for sequences that bind the allosteric transcription factor has been produced, the gene expression activity of the enriched polynucleotide library is measured in vivo to simultaneously compute the induction level of the members of the enriched polynucleotide library. This is carried out by binning and sequencing three populations: enriched polynucleotide library alone, enriched polynucleotide library+aTf, enriched polynucleotide+aTf+inducer. FIG. 3 shows an embodiment of the in vivo activity measurement.

In an aspect, the DNAs from the first enriched polynucleotide library are inserted into a constitutive promoter either between or flanking the RNA polymerase binding sites (e.g., containing the consensus −35 (TTGACA) and −10 sites (TATAAT) in the case of E. coli). This promoter is operably linked to a ribosome binding site and a reporter gene (e.g., GFP), for example in an appropriate vector such as a plasmid, to produce a reporter vector, which is then transformed into a host strain such as E. coli. The host strain co-expresses the allosteric transcription factor used for the in vitro selection.

The host strain is then grown to provide a culture. The culture is divided into two split cultures and an inducer molecule for the allosteric transcription factor is added into one of the split cultures to provide a non-induced culture and an induced culture. A control culture transformed with the plurality of reporter vectors with no allosteric transcription factor expression is also prepared. The control culture, the non-induced culture and the induced culture are sorted by
reporter protein intensity (e.g., fluorescence intensity) to provide a sorted control culture, a sorted non-induced culture, and a sorted induced culture. Sorting can be done, for example, using FACS.

Binning of the sorted control culture, the sorted non-induced culture and the sorted induced culture produces one or more control gates, one or more non-induced gates and one or more induced gates. The gates provide a convenient means by which to group the control promoter sequences, non-induced promoter sequences and induced promoter sequences by activity level. A gate is a portion of a culture comprising a plurality of polynucleotides (e.g., promoters) of specified reporter intensities. The one or more control gates comprises a control promoter library, the one or more non-induced gates comprises a non-induced promoter library, and the one or more induced gates comprises the induced promoter library.

In a specific aspect, the activity of each promoter sequence is measured in terms of reporter GFP level. To simultaneously measure the activity of all the promoters, we classify them into bins based on GFP fluorescence. The number of bins is adjusted depending on the range of promoter activities exhibited. The width of each bin is limited by the accuracy of cell sorting. A “gate” is applied in the cell sorter to collect cells within a fluorescence range. The gates provide a convenient means by which to group the control sequences and induced sequences by activity level. A gate is a portion of a culture comprising a plurality of promoters of specified reporter intensities. Cells containing the enriched polynucleotide library, enriched polynucleotide library+aTF, enriched polynucleotide library+aTF+inducer are sorted based on their activity.

Step 3: Sequencing and Analyzing the Control Promoter Library, the Non-Induced Promoter Library and the Induced Promoter Library

The control promoter library, the non-induced promoter library, and the induced promoter library are then sequenced and analyzed.

The control promoter library, the non-induced promoter library and the induced promoter library are cultured, for example, overnight, prior to sequencing. The control promoter library, the non-induced promoter library, and the induced promoter library are amplified, for example, using PCR, prior to sequencing. Amplification can be done with a sequencing primer and a barcode primer, for example.

In a specific example, cells from each bin are grown till they reach optical density (OD600) of 0.1-0.2. Their plasmids are harvested and promoter sequence is amplified by 10-15 cycles of PCR amplification. Each bin is assigned a unique barcode primer for demultiplexing after next-generation sequencing.

In an aspect, sequencing comprises next generation sequencing (NGS), also called high throughput, massively parallel, or deep sequencing. While there are different platforms that allow for NGS, all NGS platforms provide for the parallel sequencing of millions of DNA sequences. Next generation sequencing allows for the quantification of the proportion of a sequence in each gate, and for normalization across the total of sequence number.

For each promoter sequence, the fluorescence can be ascertained from the bin in which it was found. If the bin width is too narrow, a promoter sequence may be found in multiple bins because fluorescence follows a Gaussian distribution. In such a case, the effective fluorescence of that promoter sequence is measured as a weighted sum across multiple, where the weights represent normalized abundance within each bin.

The fluorescence change for different conditions can then be determined for each sequenced promoter under the different conditions, specifically, control with no allosteric transcription factor, non-induced and induced.

At least five metrics can be measured for each promoter sequence: fluorescence level of constitutive promoter, promoter repressed by aTF, and promoter induced by adding inducer, fold repression (ratio of fluorescence of constitutive promoter to promoter repressed by aTF), and fold induction or an association constant (ka) dynamic range (ratio of fluorescence of promoter+aTF+inducer to promoter+aTF). The metric(s) can be obtained for each unique promoter sequence. Depending on the size of the promoter library, we can quantify the “phenotype” of thousands to hundreds of thousands of promoter sequences simultaneously.

Thus, the quantitative NGS allows for providing a control metric, a non-induced metric, and/or an induced metric for at least a portion of the plurality of sequenced promoter based upon identification of the at least a portion of the plurality of sequenced promoters in the control promoter library, the non-induced promoter library and/or the induced promoter library. The control metric, non-induced metric, and/or induced metric are then used to determine the induction/repression for the at least a portion of the plurality of sequenced promoters.

The induction/repression properties of the at least a portion of the plurality of sequenced promoters can be expressed as a table of promoters that states the sequence of the promoter, metric (e.g., expression level as reported by fluorescence) under all three conditions (promoter, promoter+aTF, promoter+aTF+inducer), fold repression and fold induction (or dynamic range).

Kinetic profiles can also be extracted. For example, a dose-response experiment, carrying out induction at different inducer concentrations, will allow us to measure additional performance characteristics such as operational range (ligand responsiveness upper and lower bound) and transfer function (through Hill coefficient curve fitting) of induction response.

Using the induction table, a synthetic promoter having a specified induction and/or repression for the allosteric transcription factor and the inducer molecule is selected. In an aspect, the selected synthetic promoter may be used to prepare a vector comprising the synthetic promoter, e.g., operator operably linked to a ribosome binding site. The allosteric transcription factor may be expressed from the same vector, or may be expressed from a different vector.

Example 2: Exemplary Unselected Promoter Library

The following exemplary unselected promoter libraries are based on the TetR family transcription repressors. The library is designed to include the E. coli RNA polymerase holoenzyme interacting 35 and 10 regions. The library also includes a constant RBS site. Other accessory sequences are for the promoter architecture or cloning purposes. The library is illustrated in FIG. 4.

The components of the library sequences include:

Ribosome binding site: (SEQ ID NO: 1)
GAAATTCATTTAAAGGAGAGGAGGAGG
-35 site: TTGACA
Control Sequence:

\[
\begin{align*}
\text{BsaI recognition sequence: GGTCTC} \\
\text{BsaI recognition cut site: CTGA}
\end{align*}
\]

Libraries:

\[
\begin{align*}
\text{CSI 16N-BsaI: } & \quad \text{(SEQ ID NO: 2)} \\
& \quad \text{TTGACGTTCAGCTGAGGCGGCTTTGACATCCCTCTCTCTTCTT} \\
& \quad \text{ATAAGATTAGTCAATGAGAATTTCAGTATAAGGAGAAGAAGG} \\
\text{CSI 17N-BsaI: } & \quad \text{(SEQ ID NO: 3)} \\
& \quad \text{TTGACGTTCAGCTGAGGCGGCTTTGACATCCCTCTCTCTTCTT} \\
& \quad \text{ATAAGATTAGTCAATGAGAATTTCAGTATAAGGAGAAGAAGG} \\
\text{CSI 18N-BsaI: } & \quad \text{(SEQ ID NO: 4)} \\
& \quad \text{TTGACGTTCAGCTGAGGCGGCTTTGACATCCCTCTCTCTTCTT} \\
& \quad \text{ATAAGATTAGTCAATGAGAATTTCAGTATAAGGAGAAGAAGG} \\
\text{CSI 19N-BsaI: } & \quad \text{(SEQ ID NO: 5)} \\
& \quad \text{TTGACGTTCAGCTGAGGCGGCTTTGACATCCCTCTCTCTTCTT} \\
& \quad \text{ATAAGATTAGTCAATGAGAATTTCAGTATAAGGAGAAGAAGG}
\end{align*}
\]

Example 3: Results for PmeR-a Flavonoid-Responsive Allosteric Transcription Factor

In this example, the allosteric transcription factor was PmeR (repressor) and the inducer molecule was naringenin.

The results are shown in FIG. 5. The post in vitro enrichment of polynucleotides (with PmeR protein) was cloned into a vector to drive the expression of a green fluorescent protein using the Golden Gate assembly method. Following the transformation and overnight growth in E. coli DH10B, the fluorescent intensity distribution was profiled using a flow cytometer with 488 nm laser and 510 (25) nm filter. The delivery of the transcription repressor was through a second vector containing constitutively expressing PmeR protein. The fluorescent intensity distribution shift was captured with a flow cytometer after overnight growth. A decrease of overall fluorescent intensity was observed. The induction assay was conducted with the addition of 300 µM naringenin to the previous E. coli population with decreased fluorescence. Following the overnight growth, an overall increase in fluorescent intensity was captured with a flow cytometer measurement.

Example 4: Results for DesT-a Stearate-Responsive Allosteric Transcription Factor

In this example, the allosteric transcription factor was DesT (repressor) and the inducer molecule was stearate (sodium salt). The results are shown in FIG. 6. The post in vitro enrichment of polynucleotides (with DesT protein) was cloned into a vector to drive the expression of a green fluorescent protein using the Golden Gate assembly method. Following the transformation and overnight growth in E. coli DH10B, the fluorescent intensity distribution was profiled using a flow cytometer with 488 nm laser and 510 (25) nm filter. The delivery of the transcription repressor was through a second vector containing constitutively expressing DesT protein. The fluorescent intensity distribution shift was captured with a flow cytometer after overnight growth. A decrease of overall fluorescent intensity was observed. The induction assay was conducted with the addition of 250 µM stearate to the previous E. coli population with decreased fluorescence. Following the overnight growth, an overall increase in fluorescent intensity was captured with a flow cytometer measurement.

Example 5: Results for bacterial repressor TtgR, results are given in Table 5:

### TABLE 5

<table>
<thead>
<tr>
<th>Promoter sequence (including -35 and -10 regions)</th>
<th>Induced with 400 µM naringenin (au)</th>
<th>Fold induction</th>
<th>Induced with 400 µM phloretin (au)</th>
<th>Fold induction</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTGACATACTACCCGTGTTGATATTATAAT</td>
<td>6512</td>
<td>20241</td>
<td>3.1</td>
<td>33643</td>
</tr>
<tr>
<td>TTGACATACATACCGGTTGATATTATAAT</td>
<td>26598</td>
<td>50162</td>
<td>1.9</td>
<td>77021</td>
</tr>
<tr>
<td>TTGACATACATACCGATGGATGATATTATAAT</td>
<td>5135</td>
<td>16300</td>
<td>2.0</td>
<td>16383</td>
</tr>
<tr>
<td>TTGACATACATACCGATGGATGATATTATAAT</td>
<td>2939</td>
<td>18533</td>
<td>6.3</td>
<td>40363</td>
</tr>
</tbody>
</table>

au - arbitrary fluorescence units
Example 5: Systematic Evolution and Characterization of Inducible Promoters for Each of the 6 TetR Family aTFs

Besides the sequences provided in examples 3 and 4, we also systematically evolved and characterized thousands of inducible promoters for each of the 6 TetR family aTFs. The results are shown in FIG. 7.

In FIG. 7, each dot presents a promoter with unique fold induction and repression profiles. We also tested the native promoters for 6 TetR family aTFs by operably linking each native promoter to a RBS and a reporter protein (the same RBS and reporter protein used for the promoter libraries). Native promoters' fold induction and repression profiles are included in the figure and marked as a single black dot in each panel of the figure. Native promoters are the promoters repressed by aTFs in their native host organisms.

The fold induction results are also summarized Table 6. Sequences of native promoters are shown in Table 7.

### TABLE 6

<table>
<thead>
<tr>
<th>Transcription factor</th>
<th>Native promoter fold induction</th>
<th>Designed promoters' fold induction range</th>
<th>Presence of designed promoters (Yes/No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PmeR</td>
<td>6</td>
<td>1-50</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50-100</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;200</td>
<td>Yes</td>
</tr>
<tr>
<td>TtgR</td>
<td>6</td>
<td>1-50</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50-100</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;200</td>
<td>Yes</td>
</tr>
<tr>
<td>NaIc</td>
<td>12</td>
<td>1-50</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50-100</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;200</td>
<td>Yes</td>
</tr>
<tr>
<td>CmeR</td>
<td>1</td>
<td>1-10</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10-20</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20-30</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30-40</td>
<td>No</td>
</tr>
<tr>
<td>SmeT</td>
<td>1</td>
<td>1-10</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10-20</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>20-30</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30-40</td>
<td>Yes</td>
</tr>
<tr>
<td>DesT</td>
<td>1</td>
<td>1-50</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>50-100</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>&gt;200</td>
<td>No</td>
</tr>
</tbody>
</table>

### TABLE 7

<table>
<thead>
<tr>
<th>Transcription factors and native promoter sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcription factor</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>PmeR</td>
</tr>
<tr>
<td>TtgR</td>
</tr>
<tr>
<td>NaIc</td>
</tr>
<tr>
<td>CmeR</td>
</tr>
<tr>
<td>SmeT</td>
</tr>
<tr>
<td>DesT</td>
</tr>
<tr>
<td>Seq No</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

- **SEQ ID NO 1**: Length 24, Type DNA, Organism Artificial Sequence, Feature: Ribosome binding site.
- **SEQ ID NO 2**: Length 91, Type DNA, Organism Artificial, Feature: Control Sequence.
- **SEQ ID NO 3**: Length 90, Type DNA, Organism Artificial, Feature: CSI 16N−BsaI.
- **SEQ ID NO 4**: Length 91, Type DNA, Organism Artificial, Feature: CSI 17N−BsaI.
- **SEQ ID NO 5**: Length 92, Type DNA, Organism Artificial, Feature: CSI 18N−BsaI.
catgatgaga attcattaaa gaggagaag gt

<210> SEQ ID NO 6
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: CSI 19N-BsaI
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (32) .. (50)
<223> OTHER INFORMATION: n is a, c, g, or t
<400> SEQUENCE: 6
tgcgacggtc tcactgaggc gcgccttgac annnnnnnn nnnnnnnn tataatagat 60
tcatgatgag aattcattaa agaggagaaa ggt 93

<210> SEQ ID NO 7
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: promoter sequence
<400> SEQUENCE: 7
ttgacataca tagcgtgtg tatgtataat 30

ttgacataca atacggttgt tatgtataat 30

ttgacataca tgcgtgaatg tatgttataa t 31

ttgacataca tgcgtgaatg tatgttataa t 31

ttgacataca tgcgtgaatg tatgttataa t 31

ttgacataca tgcgtgaatg tatgttataa t 31

ttgacataca tgcgtgaatg tatgttataa t 31

ttgacataca tgcgtgaatg tatgttataa t 31

ttgacataca tgcgtgaatg tatgttataa t 31

ttgacataca tgcgtgaatg tatgttataa t 31

<210> SEQ ID NO 11
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: promoter sequence
<400> SEQUENCE: 11
The invention claimed is:

1. A method of identifying a synthetic inducible promoter, comprising

   in vitro selecting a first population of polynucleotides that bind to a DNA binding protein to produce an enriched polynucleotide library by

   providing an unselected polynucleotide library comprising a plurality of random degeneracies over one or more regions of a 12 to 40 base pair polynucleotide sequence, and

   selecting from the unselected polynucleotide library a plurality of polynucleotide sequences that bind the DNA binding protein to provide the enriched polynucleotide library;

   in vivo selecting from the enriched polynucleotide library a second population of polynucleotides that repress or induce production of a reporter protein to produce an induced promoter library by

   operably linking the enriched polynucleotide library to a ribosome binding site and a reporter gene to provide a plurality of reporter vectors,

   transforming the plurality of reporter vectors into a host strain which co-expresses the DNA binding protein, and

   growing the host strain provide a culture, and

   dividing the culture into two split cultures and adding an inducer molecule for the DNA binding protein into one of the two split cultures to provide a non-induced culture and an induced culture, and
The method of claim 1, wherein the DNA binding protein is an allosteric transcription factor that is from an LTR family, an AraC/XylS family, a DsrR family, a DesR family, a Fur family, a GntR family, an Lrp/AsnC family, a Crp/Fnr family, an IclR family, a MerR family, a MarR family, a LacI/GalR, a TetR family, a transcription factor associated with amino acid metabolism, a prokaryotic two-component transcription factor, or an eukaryotic nuclear receptor transcription factor.

The method of claim 1, wherein the DNA binding protein is an allosteric transcription factor that is CmeR, CymR, DesT, LmeA, NalC, PmeR, QscR, RolR, SmeT, TetR, or TlgR.

The method of claim 1, wherein the enriched polynucleotide library is operably linked to a transcription regulatory site.

The method of claim 9, wherein the transcription regulatory site is a TATA box.

The method of claim 1, wherein multiple polynucleotide libraries are operably linked to the ribosome binding site and the reporter gene.

The method of claim 1, wherein the inducer is a sugar molecule, a metallic ion, an antimicrobial agent, a dye, a flavonoid, or a combination comprising at least one of the foregoing.

The method of claim 1, wherein the inducer is lactose, IPTG, L-arabinose, maltose, trehalose, glucose-6P, glyceroP, glucitol, L-ascorbate, deoxyribonucleoside, inositol, fructose, Hg(II), Cu(II), Ag(I), Au(I), Zn(II), Pb(II), Cd(II), anhydrotetracyclin, chloramphenicol, resorcinol, proflavine, rifamycin, actinorhodin, simocyclinone D8, triclosan, ethidium, rhodamine6G, tetraphenylphosponim, quercetin, fisetin, galangin, chlorophorin, naringin, catechin, coumestrol, stearate, Oleate, c-AMP, cholate, salicylate, or a glucocorticoid, or an estrogen mineralocorticoid.

The method of claim 1, wherein the gate comprising a plurality of promoters of specified reporter intensities is selected by florescent-activated-cell-sorting.

The method of claim 1, wherein the host strain comprises organisms of the genus Escherichia, Bacillus, Staphylococcus, Caulobacter, Streptococcus, Thermus, Streptomyces, Mycoblasma, Allilbrio, Synochocysis, Azotobacter, Pseudomonas, Agrobacterium, Zymomonas, Saccharomyces, Yarrowia, Pichia, or a combination comprising at least one of the foregoing.

A biosensor comprising a synthetic promoter identified by the process of claim 1, wherein the biosensor is responsive to the concentration of the inducer molecule.

The biosensor of claim 16, wherein the synthetic promoter does not comprise the sequence bound by the DNA binding protein in nature.