

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
23 October 2008 (23.10.2008)

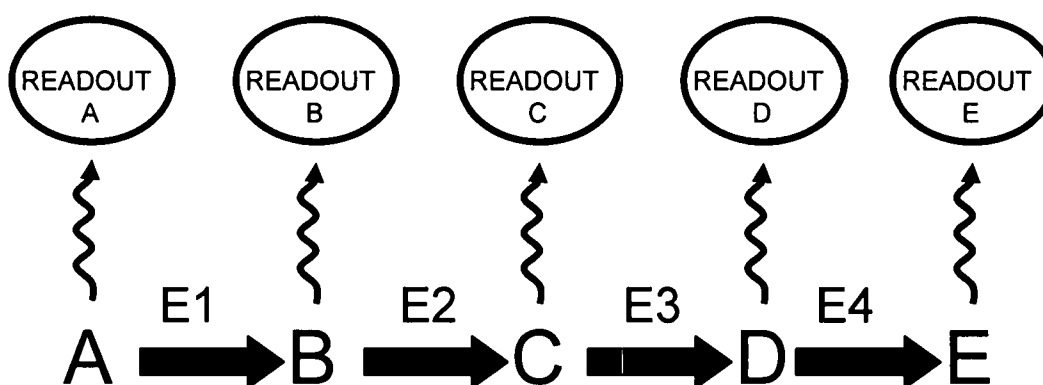
PCT

(10) International Publication Number  
**WO 2008/127283 A2**

- (51) **International Patent Classification:** Not classified
- (21) **International Application Number:**  
PCT/US2007/021473
- (22) **International Filing Date:** 6 October 2007 (06.10.2007)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**  
60/850,017 6 October 2006 (06.10.2006) US
- (71) **Applicant (for all designated States except US):** CODON DEVICES, INC. [US/US]; One Kedall Square, Building 300, Cambridge, MA 02139 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants (for US only):** JACOBSON, Joseph, M. [US/US]; 223 Grant Avenue, Newton, MA 02459 (US). CHURCH, George [US/US]; 218 Kent Street, Brookline, MA 02446 (US). BAYNES, Brian, M. [US/US]; 19 Orrin Street, Cambridge, MA 02138 (US).
- (74) **Agent:** WALLER, Patrick, R.h.; Wolf, Greenfield & Sacks, P.x., Federal Reserve Plaza, 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, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— without international search report and to be republished upon receipt of that report

(54) **Title:** ENGINEERED METABOLIC PATHWAYS

Figure 1



(57) **Abstract:** Certain aspects of the present invention provide methods for designing and engineering metabolic pathways. Aspects of the invention also provide metabolic pathway components and cells containing engineered metabolic pathways. Certain aspects of the invention provide medical, pharmaceutical, industrial, agricultural, environmental, and other Nuses for engineered metabolic pathways of the invention.

## ENGINEERED METABOLIC PATHWAYS

## RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. § 119(e) from U.S. provisional  
5 application serial number 60/850,017, filed October 6, 2006, the entire contents of which are  
herein incorporated by reference.

## FIELD OF THE INVENTION

Aspects of the invention relate to engineered biological pathways.

10

## BACKGROUND OF THE INVENTION

Naturally-occurring metabolic pathways have been extensively studied. Naturally-  
occurring catabolic and anabolic pathways have been identified. In addition, aspects of  
naturally-occurring regulatory mechanisms have been elucidated for certain natural metabolic  
15 pathways.

## SUMMARY OF THE INVENTION

Aspects of the invention relate to engineered biological pathways that are not found in  
nature. Certain aspects of the invention relate to pathways that can perform novel functions,  
pathways that include novel readout features, pathways that include novel regulatory loops,  
20 or combinations thereof. In one aspect, the invention provides methods for designing and/or  
developing engineered biological pathways. In another aspect, the invention provides  
methods for designing and/or developing one or more components (e.g., functional  
components, regulatory components, and/or readout components) of a biological pathway.  
The invention also provides, engineered pathways, engineered pathway components, and  
25 engineered organisms (unicellular and/or multicellular) adapted for and/or containing one or  
more engineered pathways and/or engineered pathway components.

Aspects of the invention are based, at least in part, on the development of novel  
pathways that can be carefully monitored and/or controlled by providing novel readout  
components and/or regulatory components, and/or configurations thereof. In some  
30 embodiments, pathways include one or more readout and/or regulatory components for each  
of one or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) metabolic steps. In some  
embodiments, the readout and/or regulatory component(s) may be different for each of a  
plurality of steps.

In one aspect, embodiments of the invention provide one or more readout components that are useful to obtain feedback information on the status (e.g., expression, level, activity, etc., or any combination thereof) of one or more components, metabolites, or steps in a biological pathway (e.g., in a natural or synthetic metabolic pathway). Aspects of the invention may be useful to provide feedback information on the status of one or more different steps of a metabolic pathway within a functional system (e.g., with a cell, an organism, an *in vitro* preparation, or other system). In some embodiments, a reporter molecule may be used to provide feedback information on the level of one or more metabolites (e.g., substrates, intermediates, products, etc., or any combination of two or more thereof) in a pathway. A pathway may be engineered to include one or more different reporter molecules to provide feedback on the level of each metabolite or a subset of the metabolites associated with the pathway. A reporter molecule can provide direct feedback and/or indirect feedback on the level of a metabolite in a system. Direct feedback can be provided by a reporter molecule that generates a signal in response to the presence or absence of a metabolite. For example, a reporter molecule may generate a signal when it interacts with the metabolite (e.g., when it binds to the metabolite). Indirect feedback can be provided when a reporter molecule interacts with one or more intermediate molecules to generate a signal in response to the presence or absence of a metabolite. A signal may be a detectable signal (e.g., a fluorescent signal), a phenotypic signal (e.g., a change in growth rate, cell death, etc.). In some embodiments, a signal may provide quantitative information about level of a metabolite. For example, the signal intensity may be related to the level of the metabolite. However, in some embodiments, a signal may provide qualitative information relating to the presence or absence of a threshold level of a metabolite. In some embodiments, a reporter molecule may be a nucleic acid or a polypeptide. For example, a reporter may be an aptamer (e.g., a DNA or RNA aptamer), a ribozyme, an antibody, a nucleic acid or polypeptide ligand, or other ligand. It should be appreciated that readout components also may be used to provide direct feedback on the activity of one or more functional and/or regulatory components of a biological pathway. Accordingly, a reporter molecule may interact with a metabolite, a functional component, a regulatory component, or any combination thereof, associated with a pathway system.

According to aspects of the invention, one or more readout components may be useful to identify, understand, design, monitor, influence, and/or provide other feedback information about a biological pathway. Accordingly, a pathway may be engineered by providing one or more readout components to a naturally existing pathway. In other embodiments, one or

more readout components may be provided to an engineered pathway that includes a novel combination of one or more functional and/or regulatory components.

Aspects of the invention may be based, at least in part, on the design and assembly of large nucleic acid molecules and libraries that can be used to make, test, select and screen for engineered pathways and/or pathway components.

Engineered pathways of the invention may include anabolic components, catabolic components, modifying components, or combinations thereof. Aspects of the invention may be used to generate new products, modified levels (e.g., lower or higher) of existing products, new product combinations, and products produced under engineered regulatory control(s).

Aspects of the invention may be used to monitor biological pathways.

Aspects of the invention may be used for industrial (e.g., pharmaceutical, chemical synthesis, manufacturing, etc.), agricultural, mining, environmental, research, and other applications.

## BRIEF DESCRIPTION OF THE FIGURES

FIG. 1 illustrates an embodiment of a metabolic pathway showing readout components for each metabolite;

FIG. 2 illustrates an embodiment of a metabolic pathway showing examples of a feedback loop (10), a feedforward loop (20), and external regulatory pathway (30);

FIG. 3 illustrates an embodiment of two metabolic pathways showing a cross regulatory pathway (40);

FIG. 4 illustrates an embodiment of a metabolic pathway showing readout components for each metabolite and examples of a feedback loop (10), a feedforward loop (20), and external regulatory pathway (30); and

FIG. 5 illustrates an embodiment of two metabolic pathways showing a cross regulatory pathway (40) and readout components for each metabolite in each pathway.

## DETAILED DESCRIPTION OF THE INVENTION

Aspects of the invention relate to engineered metabolic pathways. In some embodiments, the invention provides methods for designing and/or developing novel metabolic pathways. In some embodiments, the invention involves engineering one or more existing metabolic pathways to remove, modify, and/or add one or more functionalities. In some embodiments, the invention involves providing one or more readout components to provide feedback on the level of one or more steps in the pathway. In some embodiments,

the invention involves engineering novel combinations of functional, regulatory, and/or readout components. In some embodiments, the invention involves engineering one or more novel functional and/or regulatory components (or combinations thereof). In some embodiments, a novel biological pathway may include one or more novel and/or existing (e.g., natural) functional and/or regulatory components and/or combinations thereof.

In one aspect, a biological pathway may be engineered by providing one or more readout components. FIG. 1 illustrates a non-limiting example of a linear metabolic pathway (or a linear portion of a metabolic pathway) with functional components (E1 through E4) and metabolites A through E. In one embodiment, A is the starting metabolite (e.g., substrate), B, C and D are intermediate metabolites, and E is the product metabolite (e.g., product).

However, it should be appreciated that reference to starting and product metabolites are made within the context of this pathway or pathway component, and that A and E may be intermediate metabolites within a larger pathway or network of pathways. Accordingly, a pathway may include any number of steps (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more) in a linear, branched, and/or looped configuration. Each step may involve a separate functional component. FIG. 1 also illustrates a separate readout component for each metabolite.

However, it should be appreciated that a pathway may be engineered to include readout components for any number of different metabolites (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) in a pathway. Accordingly, a pathway may be engineered to include readout components for

only a subset of the metabolites in a pathway. In some embodiments, an engineered pathway may include two or more readout components for each metabolite. It also should be appreciated that a pathway may include one or more readout components for the functional and/or regulatory components of the pathway. In some embodiments, a pathway may be defined by one or more user specified (e.g., predetermined) substrates and/or products and include the necessary functional, regulatory and/or readout components. For example, a pathway may be designed and/or engineered to generate at least one specified product from at least one specified substrate (optionally in combination with one or more intermediate metabolites). It should be appreciated that the specified substrates, products, and/or intermediates independently may be foreign to (e.g., not naturally-occurring in) a biological system (e.g., a host cell) that is engineered to metabolize them. Accordingly, in some embodiments, one or more of the engineered pathway components may be foreign to the biological system. However, in some embodiments, all or a subset of the pathway components may be modified components (e.g., recombinant components) based on one or more naturally-occurring components within the biological system. The modified

component(s) may have a *de novo* function, a modified activity level, a modified regulatory response, or a combination thereof.

In one aspect, a pathway may be engineered to include one or more predetermined readout, functional, and/or regulatory components. In one aspect, a biological pathway may be engineered by providing a method for generating novel genetic combinations and/or novel genetic functions and performing assays to identify functions of interest. In some embodiments, a biological system may be used to generate a library comprising a plurality of different combinations of genetic elements. This library can be used to identify one or more genetic combinations that encode a biological (e.g., metabolic) pathway of interest. In some embodiments, a biological system may be engineered to contain a library of different genetic elements (e.g., aptamers) that have different functions. This library may be used to identify one or more genetic elements having functions that can be included in a biological pathway of interest. In some embodiments, a biological system can be designed to include one or more known genetic elements that are predicted to provide useful regulatory and/or functional components for a biological pathway of interest. In some embodiments, two or more alternative biological systems may be designed to include different genetic elements (or different combinations of genetic elements that are predicted to provide equivalent or similar functional and/or regulatory properties for a biological pathway of interest. Pathways may be designed using computer-implemented design techniques. In some embodiments, one or more steps or series of steps in a pathway may be taken from known (e.g., natural or engineered) pathways. In some embodiments, combinatorial pathways may be designed to include different genetic components from different sources (e.g., from different organisms). It should be appreciated that one or more design steps may be automated. Databases of pathways and/or pathway components may be used as resources for engineered pathways of interest. In some embodiments, genetic components encoding one or more designed pathways may be assembled and tested. In some embodiments, a plurality of candidate systems may be designed, assembled and tested. Pathways may be engineered to use two or more different metabolites (e.g., substrates). Pathways may be engineered to produce two or more different metabolites (e.g., products). Pathways may be engineered to include one or more alternative branches (e.g., to produce one or more alternative metabolites) that are regulated by intrinsic signals, extrinsic signals, or a combination thereof.

According to the invention, a metabolic pathway involves one or more steps to convert a substrate metabolite to a product metabolite. A pathway may involve a single step to convert a product metabolite to a substrate metabolite. However, typical pathways involve

a plurality (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 10-20, 20-50, 50-100, or more) of intermediate steps via a plurality of intermediate metabolites. Each step may involve one or more functional components (e.g., catalytic proteins, catalytic nucleic acids, binding proteins, binding nucleic acids, etc., or any combination thereof) that promote the conversion of a first metabolite to a

5 second metabolite. The second metabolite then may be used in a subsequent step of the pathway. The steps in a pathway may be arranged in one or more linear patterns, cyclical patterns, branched patterns (e.g., with one or more converging or diverging branches), or a combination thereof. It should be appreciated that *in vivo* pathways may interact with each other in complex patterns, for example, with similar or identical metabolites being used at  
10 different stages in different pathways, with overlapping regulatory connections, etc., or any combination thereof. However, as described herein, a pathway may be defined by selecting a start point (e.g., a substrate metabolite), an end point (a product metabolite), and by identifying intermediate steps, metabolites, functional components, and regulatory components.

15 Certain aspects of the invention relate to engineering metabolic pathways to act on one or more predetermined substrate or intermediate metabolites. Certain aspects of the invention relate to engineering metabolic pathways to generate one or more intermediate or product metabolites of interest. Certain aspects of the invention relate to engineering metabolic pathways to provide one or more regulatory and/or monitoring functions.

20 According to the invention, an engineered metabolic pathway may be an existing metabolic pathway (e.g., a natural metabolic pathway) that has been changed or a novel metabolic that has been developed or a combination thereof. A change to an existing metabolic pathway may involve removing, modifying, or adding one or more steps (e.g., by removing, adding, or modifying one or more functional components and/or regulatory components). A novel  
25 metabolic pathway may include a novel combination of functional components, novel functional components, or a combination thereof. A novel metabolic pathway also may be developed to include one or more regulatory components (e.g., feedback or feedforward loops, extrinsic regulatory pathways, etc.). A novel metabolic pathway also may be developed to include one or more regulatory connections with other metabolic pathways  
30 (e.g., other natural and/or engineered pathways).

Aspects of the invention may involve providing one more regulatory loops (e.g., one or more feedback or feedforward loops) or extrinsic regulatory connections to an existing or novel metabolic pathway. FIG. 2 illustrates a non-limiting example of a linear metabolic pathway (or a linear portion of a metabolic pathway) with functional components (E1 through

E4) and metabolites A through E. In one embodiment, A is the starting metabolite, B, C and D are intermediate metabolites, and E is the product metabolite. However, it should be appreciated that reference to starting and product metabolites are made within the context of this pathway or pathway component, and that A and E may be intermediate metabolites within a larger pathway or network of pathways. FIG. 2 also shows a feedback loop (10), a feedforward loop (20), and an extrinsic regulatory connection (30). The pathway and regulatory connections shown in FIG. 2 is not limiting.

Aspects of the invention may involve providing one or more regulatory connections between two or more pathways (e.g., between existing pathways, between novel pathways, or a combination of two or more thereof). FIG. 3 illustrates a non-limiting example of two metabolic pathways with a regulatory connection (40) from one pathway to the other. The pathways and regulatory connections shown in FIG. 3 are not limiting.

Aspects of the invention may involve providing one or more readout components for each metabolite (or a subset thereof, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more thereof). FIG. 4 illustrates a non-limiting embodiment of a metabolic pathway showing readout components for each metabolite and examples of a feedback loop (10), a feedforward loop (20), and a regulatory component that is responsive to an external signal or ligand (30). It should be appreciated that the configuration of FIG. 4 is not limiting and that the different components may be independently included in engineered pathways of the invention.

Aspects of the invention may involve providing one or more cross-regulatory connections between two or more different pathways. FIG. 5 illustrates a non-limiting embodiment of two metabolic pathways showing a cross-regulatory pathway (40) and readout components for each metabolite in each pathway. It should be appreciated that the configuration of FIG. 5 is not limiting and that the different components may be independently included in engineered pathways of the invention. One or more of the pathways may be engineered pathways of the invention. However, in some embodiments, the cross-regulatory pathway(s) may be provided by one or more engineered regulatory components of the invention. It should be appreciated that in any of the embodiments described herein, a readout pathway may be provided by one or more readout components (e.g., engineered readout components) as described herein. Similarly, a regulatory pathway may be provided by one or more regulatory components (e.g., engineered regulatory components) as described herein.

In some embodiments, a metabolic pathway is developed or engineered to generate a metabolite (e.g., a product) of interest. In some embodiments, a novel pathway is developed



to act on (e.g., remove, process, modify) a metabolite (e.g., a substrate) of interest. In some embodiments, a pathway is developed or engineered to provide one or more regulatory connections (e.g., so that it may be made responsive to one or more intracellular or extracellular signals). In some embodiments, a pathway may be developed or engineered to provide one or more monitoring functions that provide a detectable readout indicative of the status (e.g., activity or level) of one or more functional components, regulatory components, or metabolites of the pathway.

Pathway Components:

According to aspects of the invention, one or more functional, regulatory, and/or readout components, or combinations of two or more thereof, may be used in an engineered metabolic pathway. A pathway may be designed or identified to process metabolites *in vivo* and/or *in vitro* according to one or more predetermined and/or identified steps. According to aspects of the invention, a metabolite may be a starting metabolite, intermediate metabolite or end product metabolite. In some embodiments, a metabolite may be unique to a metabolic pathway or may be present in one or more metabolic pathways. In certain embodiments, one or more intermediate metabolites may be present in one or more metabolic pathways. Aspects of the invention may be used to synthesize higher levels of one or more predetermined metabolites, synthesize one or more new metabolites, synthesize altered combinations of metabolites, provide internal regulatory connections (e.g., in the form of feedback loops), provide external regulatory connections (e.g., for response to environmental factors, human factors, etc.), provide signals that can be used to monitor one or more intermediate processes or metabolites, etc., or any combination thereof.

Readout Components:

According to aspects of the invention, a readout component may be a reporter molecule that provides information about the status of one or more steps and/or metabolites in an engineered pathway. A reporter molecule may be a nucleic acid or a polypeptide. For example, a reporter molecule may be an enzyme, an enzyme complex, a binding factor, a ligand, or any other molecule that can provide information about the status of one or more steps or metabolites in a pathway. A reporter may be a DNA or RNA aptamer, a ribozyme, or any other DNA or RNA oligonucleotide or molecule that includes a readout function. One or more reporter molecules may be encoded by nucleic acid that is included in an engineered cell or organism (e.g., on one or more plasmids or other vectors and/or integrated into the

genome of the engineered cell). Examples of methods for identifying metabolite-specific reporter molecules are described in more detail herein. Reporter components may be used to monitor one or more steps and/or metabolites in an engineered pathway. Information from reporter components may be used to identify, understand, and/or interfere with an engineered metabolic system (e.g., by modifying levels and/or activities of one or more metabolites, functional components, and/or regulatory components in an engineered metabolic system).

*Functional Components:*

According to aspects of the invention, a functional component may be an enzyme, an enzyme complex, a binding factor, a ligand, or any other molecule that can act in a metabolic pathway to convert a first metabolite to a second metabolite. For example, a functional component may be a protein, RNA, or any other small molecule that can be functional in a metabolic pathway. In some aspects of the invention, an engineered pathway comprises altered combinations of functional components. In some embodiments, an engineered pathway comprises one or more altered functional components. One or more functional components may be encoded by nucleic acid that is included in an engineered cell or organism (e.g., on one or more plasmids or other vectors and/or integrated into the genome of the engineered cell).

*Regulatory Components:*

Aspects of the invention relate to engineering regulatory components of a metabolic pathway. In some embodiments, a regulatory component may respond to an internal or an external signal. In certain embodiments, a regulatory component may involve a feedback or feedforward loop if the regulatory component is responsive to a metabolite in the pathway and acts on one of the functional components (or one of the other metabolites) in the pathway.

According to aspects of the invention, regulatory feedback or feedforward loops may be bipolar. For example, in some embodiments, feedback or feedforward loops may be negative and in other embodiments feedback or feedforward loops may be positive. In certain embodiments, negative feedback or feedforward loops may cause a reduction in a particular process of a metabolic pathway. In some embodiments, positive feedback or feedforward loops may cause an increase in a particular process of a metabolic pathway. In certain embodiments, any combination of feedback and feedforward loops may occur. In some embodiments, a negative feedback and a negative feedforward may occur in a

metabolic pathway. In other embodiments, a positive feedback and a positive feedforward may occur in a metabolic pathway. In certain embodiments, a negative feedback and a positive feedforward may occur in a metabolic pathway. In some embodiments, a positive feedback and a negative feedforward may occur in a metabolic pathway. In other

5       embodiments, one or more negative feedback loops may occur in a metabolic pathway. In some embodiments, one or more negative feedforward loops may occur in a metabolic pathway. In certain embodiments, one or more positive feedback loops may occur in a metabolic pathway. In some embodiments, one or more positive feedforward loops may occur in a metabolic pathway. One or more feedback or feedforward loops may occur

10       simultaneously, consecutively or sequentially. FIGS. 2 and 3 illustrate non-limiting embodiments of feedback and feedforward loops. It should be appreciated that a regulatory component of a regulatory loop may be sensitive to any one or more metabolites in the pathway and may inhibit or activate any one or more functional components of the pathway. Feedback and feedforward loops may be used to provide a form of auto-regulatory control for

15       a pathway so that the level of final product is controlled. The final level of product expression may be tunable as a function of the regulatory component that is used.

According to aspects of the invention, a negative feedback or feedforward loop may be responsive to a signal generated by one or more components or intermediates in a pathway. In some embodiments, an external signal mediated by a regulatory component

20       (e.g., external to the cell that harbors the engineered metabolic pathway) may alter (e.g., up-regulate or down-regulate) one or more steps in a metabolic pathway. In some embodiments, an external signal mediated by a regulatory component may alter the level of feedback or feedforward in a regulatory loop.

In aspects of the invention, a negative feedback loop may be one in which a functional

25       component, metabolite or regulatory component upstream in a metabolic pathway may be reduced or inhibited. In some embodiments, inhibition of a functional component, metabolite or process in a metabolic pathway may be a partial or total reduction of the level of a functional component, metabolite or process, or may be a partial or total inhibition of some activity of a functional component, metabolite or process. In some embodiments, a

30       metabolite may be degraded. In certain embodiments, the level of a metabolite may be indirectly reduced (e.g., via a negative regulatory loop affecting a functional component involved in producing the metabolite). In certain embodiments, a functional component may be partially or totally inhibited (e.g., the expression levels of the functional component may be reduced, and/or the activity of the functional component may be reduced).

In aspects of the invention, a positive feedback loop may be one in which a functional component, metabolite or process upstream in a metabolic pathway may be increased or enhanced. In some embodiments, a functional component, metabolite or process may be affected such that its production or activity is increased or enhanced. In certain

5       embodiments, a functional component of a metabolic pathway may be enhanced resulting in an increase in its activity. In certain embodiments, a metabolite may be increased either as a direct or indirect effect of a positive feedback loop. In some embodiments, an increase in a functional component of a metabolic pathway results in an indirect increase in the production of a metabolite. In certain embodiments, a metabolite in a metabolic pathway may be  
10       directly increased. In certain embodiments, a functional component may be partially or totally stimulated (e.g., the expression levels of the functional component may be increased, and/or the activity of the functional component may be increased).

          In aspects of the invention, a negative feedforward loop may be one in which a functional component, metabolite or process downstream in a metabolic pathway may be  
15       reduced or inhibited. In some embodiments, inhibition of a functional component, metabolite or process in a metabolic pathway may be partial or total reduction of a functional component, metabolite or process, or may be partial or total inhibition of some activity of a functional component, metabolite or process. In some embodiments, a metabolite of a metabolic pathway may be reduced or inhibited. In certain embodiments, a metabolite of a  
20       metabolic pathway may be reduced or inhibited directly or indirectly. In some embodiments, a metabolite may be indirectly reduced or inhibited as a result of a negative feedforward loop affecting a functional component involved in producing the metabolite. In certain  
embodiments, a functional component in a metabolic pathway may be reduced by partial or total inhibition of the functional component or a functional component's activity. Partial  
25       inhibition of a functional component or functional component activity in a metabolic pathway may be sufficient to create the desired effect of a negative feedforward loop. In some  
embodiments, total inhibition of a functional component in a metabolic pathway may be required to create the desired effect of a negative feedforward loop. In certain embodiments,  
partial or total inhibition of a functional component may result in reduction or inhibition of  
30       the production of a metabolite in a metabolic pathway. In certain embodiments, a functional component may be partially or totally inhibited (e.g., the expression levels of the functional component may be reduced, and/or the activity of the functional component may be reduced).

          In aspects of the invention, a positive feedforward loop may be one in which a functional component, metabolite or process downstream in a metabolic pathway may be

increased or enhanced. In some embodiments, a functional component, metabolite or process may be affected such that its production or activity is increased or enhanced. In certain embodiments, a functional component of a metabolic pathway may be enhanced resulting in an increase in its activity. In certain embodiments, a metabolite may be increased either as a direct or indirect effect of a positive feedforward loop. In some embodiments, an increase in a functional component of a metabolic pathway results in an indirect increase in the production of a metabolite. In certain embodiments, a metabolite in a metabolic pathway may be directly increased. In certain embodiments, a functional component may be partially or totally stimulated (e.g., the expression levels of the functional component may be increased, and/or the activity of the functional component may be increased).

In some embodiments, an engineered pathway may include regulatory components that provide feedback and feedforward control based on the level of one or more metabolites. For example, two or more feedback loops that are responsive to the level (e.g., the intracellular level) of a metabolite (e.g., an intermediate or a product) may provide feedback control on one or more upstream functional elements in a pathway. In some embodiments, a first feedback loop may increase the activity of one or more upstream functional elements in response to low levels of the metabolite (e.g., when the metabolite level falls below a first threshold level). In contrast, a second feedback loop may decrease the activity of one or more upstream functional elements in response to high levels of the metabolite (e.g., when the metabolite level rises above a second threshold level). In some embodiments, a plurality of different feedback loops may act on a plurality of upstream functional elements. Similarly, two or more feedforward loops that are responsive to the level (e.g., the intracellular level) of a metabolite (e.g., a substrate or an intermediate) may provide feedforward control on one or more downstream functional elements in a pathway. In some embodiments, a first feedforward loop may decrease the activity of one or more downstream functional elements in response to low levels of the metabolite (e.g., when the metabolite level falls below a first threshold level). In contrast, a second feedforward loop may increase the activity of one or more downstream functional elements in response to high levels of the metabolite (e.g., when the metabolite level rises above a second threshold level). In some embodiments, a plurality of different feedforward loops may act on a plurality of downstream functional elements. In some embodiments, a pathway may comprise one or more feedback and feedforward loops for two or more metabolites in the pathway. In certain embodiments, at least one feedback and at least one feedforward loop are provided for each intermediate in the pathway, along with at least one optional feedforward loop for the substrate and/or at least one optional

feedback loop for the product. Accordingly, an engineered pathway may be designed to include a plurality of feedback and feedforward regulatory loops that maintain a relatively stable metabolite level (e.g., similar molar amounts of each metabolite, or relative molar amounts of different metabolites that are optimized for efficient metabolic processing, or other suitable stable metabolite levels). It should be appreciated that an engineered pathway that maintains relatively stable and desirable (e.g., appropriately balanced) levels of different metabolites may be useful to provide an efficient metabolic process that is not negatively impacted by the inappropriate accumulation of one or more intermediates (e.g., that could otherwise cause the metabolic pathway or one or more steps thereof to be slowed, or divert metabolites into other pathways thereby wasting metabolites, or result in one or more toxic metabolites to accumulate, or have some other negative impact on an engineered pathway or a cell containing an engineered pathway). In some embodiments, the feedback and/or feedforward loops may involve regulatory components that are directly responsive to levels of metabolites at different steps as described herein. In some embodiments, readout components may be used to monitor the level of metabolites in all or a subset of steps in an engineered pathway. It should be appreciated that in some embodiments, one or more (e.g., all or a subset thereof) of the feedback and/or feedforward loops may involve a readout component that provides information about the level of one or more metabolites and a regulatory component that is not directly responsive to the metabolites, but that can be modified by changing one or more external conditions (e.g., addition of one or more regulatory ligands, change of a cellular growth condition such as pH, temperature, salt, etc., or any combination thereof). In order to maintain appropriate levels (e.g., predetermined or specified levels) of the different metabolites, external conditions may be changed as required in response to the readout functions. The readout may be automatically coupled to the changes in the external conditions (e.g., via an automated detector and controller that implements appropriate condition changes in response to different readout changes). However, in some embodiments, an operator (e.g., a human operator) may review or monitor one or more readout changes and implement appropriate condition changes to maintain desired (e.g., specified) levels of one or more different metabolites. It should be appreciated that in some embodiments, two or more different metabolites (e.g., each metabolite in an engineered pathway or a subset thereof) may have different specified levels and different readouts (e.g., different reporter molecules that generate different signals). However, different combinations of similar or different specified levels, readouts, and/or regulatory components may be used as aspects of the invention are not limited in this respect.

It should be appreciated that a regulatory element may exert a negative or positive control on the expression and/or activity of a functional element. For example, a negative or positive control may be exerted indirectly by decreasing or increasing transcription, mRNA stability, and/or translation of the functional element (e.g., an enzyme). In some

5     embodiments, a negative or positive control may be exerted directly on a functional element via binding to, or modification of, the functional element. For example, the functional element may be phosphorylated, dephosphorylated, methylated, demethylated, or otherwise modified to decrease or increase its activity. It should be appreciated that the regulatory

10    loops may be finely tuned to provide appropriate responses (e.g., appropriate levels of activation or inhibition) in response to changes in metabolite levels. For example, if regulation involves promoter activation or inactivation, the promoter strength may be tuned to be appropriately responsive. In some embodiments, the promoter activity may be tuned to provide a dynamic response over a range of metabolite levels that are expected or experimentally observed for the pathway. For example, a promoter activity may be

15    engineered to decrease by between about 5% and 95% (e.g., by about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90%) in response to varying levels of a metabolite. However, higher or lower levels of inactivation may be engineered. Similarly, a promoter activity may be engineered to increase by between about 5% and 95% (e.g., by about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90%) in response to varying levels of a metabolite. However,

20    higher or lower levels of activation may be engineered. However, in other embodiments a promoter activity may be engineered to be responsive to threshold levels of metabolite. For example, a promoter activity may remain substantially constant until a metabolite level goes above or below one or more threshold levels, at which point the promoter activity may change substantially (e.g., increase or decrease by at least 25%, at least 50% , at least 75%, at

25    least 90% or more). It should be appreciated that other forms of control (e.g., stability, expression, and/or activity) may be engineered to be responsive in a similar fashion (e.g., in a dynamic or discrete fashion as described for the promoters above.

According to aspects of the invention, one or more regulatory components may be proteins, RNA's, ribozymes, riboregulators, ligand-controlled riboregulators (Bayer T.S. et

30    a., 2005, Nat. Biotechnol., 23(3):337-43), zinc fingers, small-molecule-dependent switches (Buskirk, A.R. et al., 2005 Chem. Biol., 12(2):151-61), ligand-dependent RNA transcriptional activators (Buskirk, A.R. et al., 2004 Chem. Biol., 11(8):1 157-63), small-molecule activated protein splicers (Buskirk, A.R. et al., 2004 PNAS, 101(29):10505-10), RNA-based transcriptional activators (Buskirk, A.R. et al., 2003 Chem. Biol., 10(6):533-40),

RNA sequences that activate transcriptional regions (Saha S. et al., 2003, Nuc. Acids Res., 31(5): 1565-70) or any other intracellular components that can be responsive to a signal. Each of the aforementioned references are incorporated herein in their entirety by reference.

In some embodiments, a regulatory component (e.g., an aptamer) may be designed or isolated to interact with a signal (e.g., a metabolite) with appropriate kinetics (e.g., appropriate on and off rates) to provide real-time feedback or feedforward to a pathway. For example, a regulatory component may bind reversibly to a metabolite in order to provide feedback or feedforward control that is responsive to the levels of metabolite in the cell.

In some embodiments, a regulatory component (e.g., a single regulatory molecule) may be responsive to two or more different signals (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, or more). The signals may be related (e.g., similar compounds) or unrelated (e.g., distinct compounds). The signals may interact competitively with the regulatory component (e.g., they may bind to the same binding site). However, the signals may interact independently with the regulatory component (e.g., they may bind to different binding sites). The signals may be substrates, intermediates, or products of an engineered pathway that work through one or more regulatory components to provide an intrinsic positive or negative, feedback or feedforward, regulatory loop. In some embodiments, the signals may be compounds that are not produced or consumed by the pathway, but that interact with one or more regulatory components to provide an extrinsic positive or negative regulatory control over one or more steps in an engineered pathway. In some embodiments, a regulatory component (e.g., a single regulatory molecule) may be responsive to both intrinsic and extrinsic signals. In some embodiments, different regulatory components may respond to one or more identical signals.

One or more regulatory components may be encoded by nucleic acid that is included in an engineered cell or organism (e.g., on one or more plasmids or other vectors and/or integrated into the genome of the engineered cell).

#### *Vectors, Host Cells and Organisms:*

Any suitable vector (e.g., plasmid, BAC, YAC, viral vector, etc.) or combination of two or more vectors may be used to harbor one or more genes encoding one or more components of an engineered pathway. In some embodiments, one or more components (or all components) may be encoded on the genome of a host cell or organism. The genes encoding engineered pathway components may be clustered within one or a few (e.g., 2, 3, 4, or 5) genetic regions (e.g., plasmid, genomic regions, chromosomes, etc.), organized on one



or a few (e.g., 2, 3, 4, or 5) operons, or distributed across many genetic regions or operons (e.g., 6-10 or more).

Any suitable host cell or organism may be used or modified to harbor an engineered biological pathway. A host cell may be a unicellular organism (e.g., a bacterial or yeast cell or other prokaryotic or eukaryotic unicellular organism). Non-limiting examples of host cells include *E. coli*, *B. subtilis*, *S. cerevisiae*, and *P. pastoris*. A host cell may be a cell obtained from a multicellular organism but grown in culture (e.g., a mammalian cell grown in culture). A host organism may be a multicellular organism. Examples of multicellular organisms include animals and plants, e.g., mammals, insects, reptiles, fish, birds, land plants, aquatic plants, agricultural plants, monocotyledonous and/or dicotyledonous plants, etc. The type of host chosen may depend on the application. In some embodiments, an engineered pathway may contain one or more components (e.g., functional, readout, and/or regulatory) that are from a different cell type or a different species. In some embodiments, all of the components of an engineered pathway may be derived from a different cell or species than the host cell (for example, prokaryotic components may be used in eukaryotic hosts cells or *vice versa*, components from one species or genus of prokaryotic or eukaryotic organisms may be used in host cells from a different species or genus or prokaryotic or eukaryotic organisms, respectively). However, in some embodiments, a subset of the components (e.g., all or a subset of one or more of the functional, readout, and/or regulatory components independently). However, in some embodiments, all or a subset of the components (e.g., all or a subset of one or more of the functional, readout, and/or regulatory components independently) may be a modified version of a naturally occurring component or may be a *de novo* engineered component as described herein.

In some embodiments, a host cell may be engineered to have a modified genome that is suited to the one or more engineered pathways. For example, a host cell may be engineered to have a reduced genome size (e.g., a genome that is smaller by 10%, 20%, 30%, 40%, 50%, or more). Such a host cell may be adapted to accommodate genetic elements encoding one or more biological pathways of interest. A host cell may be engineered to encode one or more functions for importing (e.g., substrates), synthesizing, or exporting (e.g., products) metabolites, proteins, or other molecules. For example, a host cell may be engineered to encode one or more membrane-bound transporters (e.g., pumps). A host cell may also be engineered to improve growth rate and/or viability in unnatural environments, to detect the presence of a molecule in its environment, to communicate with other cells, to self-organize into patterns, to propagate or die under defined conditions, to act as a scaffold for

extracellular synthesis of materials, or to degrade substances in its environment such as environmental contaminants or pathogens.

Applications:

5 Aspects of the invention may be used to synthesize higher levels of one or more predetermined metabolites, synthesize one or more new metabolites, synthesize altered combinations of metabolites, provide internal regulatory connections (e.g., in the form of feedback loops), provide external regulatory connections (e.g., for response to environmental factors, human factors, etc.), provide intracellular regulatory connection (e.g., between two or  
10 more metabolic pathways) provide signals that can be used to monitor one or more intermediate processes or metabolites, etc., or any combination thereof.

Aspects of the invention may be used for pharmaceutical applications (e.g., to provide engineered pathways that may be useful to gene therapy).

Aspects of the invention may be used for industrial applications (e.g., to provide  
15 engineered pathways that may be useful to increase the synthesis of a product of interest or to provide additional internal or external regulatory connections to regulate the synthesis of a product in response to different factors). Industrial products of interest may include industrial enzymes, metabolites that are useful as feedstocks for industrial syntheses, and other organic or biological products. Industrial products such as propanediol, octane, diesel fuel, ethanol,  
20 butanol, lactic acid, polymers, amino acids, polyhydroxybutyrate, alkaloids, terpenes, polyketides may also be of interest.

Aspects of the invention may be used for agricultural applications (e.g., to provide engineered pathways that may be useful to engineer crops to express one or more products of interest and/or to provide additional internal or external regulatory connections to regulate the  
25 synthesis of a product in response to different factors). In some embodiments, pathways may be engineered to increase photosynthetic yields of agricultural products (e.g., *in vivo* in plants). Pathways may also be engineered to increase aesthetic, odor, or other consumer appeal, or to ingest and/or digest environmental toxins. Products may include fruits, vegetables, grains, flowers, trees, shrubbery, canes, and reeds.

30 In some embodiments, pathways may be adapted to increased levels or scales of production of one or more metabolites (e.g., for agricultural, industrial, pharmaceutical, or other purposes). For example, additional regulatory components may be added (e.g., feedback or feedforward loops, regulatory components that are responsive to external stimuli,

for example to induce a pathway at a desired time during production or at an appropriate time during an agricultural season, etc.).

Aspects of the invention also may be used to develop engineered pathways for environmental applications (e.g., for remediation by providing mixtures of functional components or engineered organisms that can metabolize one or more environmental contaminants to either sequester the contaminants and/or process the contaminants to form one or more environmentally acceptable compounds (e.g., less toxic). In some embodiments, pathways of the invention may be used for scavenging environmental contaminants and/or toxic compounds (e.g., as part of an environmental cleanup or remediation effort). In some  
5  
10  
15  
20  
25  
30  
embodiments, engineered pathways may be used to waste water treatment. In some embodiments, pathways and/or organisms may be engineered to increase absorption or incorporation of environmental toxins or pollutants (e.g., compounds dissolved in water, ground contaminants, air contaminants, carbon dioxide, carbon monoxide, sulfur, etc.).

Aspects of the invention also may be used for energy generation. In some  
15  
20  
25  
30  
embodiments, pathways may be developed to increase the production of a fuel or of a substrate for a industrial fuel processing technique. For example, unicellular or multicellular plants (e.g., algae, crop plants, grasses, trees, etc.) may be developed with engineered pathways to increase the yield of certain compounds or compound substrates. For example, pathways may be engineered to increase the yields of alcohols (e.g., methanol, ethanol, etc.), sugars, animal fats, vegetable oils, hydrocarbons such as isooctane or cetane, other combustible compounds, etc., or any combination thereof. In some embodiments, pathways may be engineered to increase photosynthetic yields of fuel substrates or products (e.g., *in vivo* in plants).

Aspects of the invention also may be used to provide one or more markers of pathway activity. A marker may be responsive to the level or status of a metabolite, a functional component, and/or a regulatory component. A marker may be for example, a binding moiety (e.g., a protein or a nucleic acid, for example, an aptamer) that is responsive (e.g., generates a color) to one or more indicators of pathway activity. The color for example may be generated by expression or activation of an engineered GFP or other protein reporter system.

Aspects of the invention also relate to providing cells that are engineered to include one or more different pathways. For example, a cell may be engineered to include several (e.g., 2, 3, 4, 5, or more) independent pathways or interdependent pathways that are connected via a regulatory network. In some embodiments, the level of one or more metabolites produced in a first pathway may provide positive or negative signal to one or

more functional elements in a second pathway. In some embodiments, two or more pathways may be regulated by the same extrinsic signal(s). Different pathways may be alternative pathways for generating the same product(s). Different pathways may be alternative pathways for metabolizing the same substrate(s). However, different pathways may provide  
5 unrelated synthetic and/or catabolic functions. Accordingly, a multipurpose cell may be engineered that is responsive to a plurality of different signals and/or metabolites. In processes in which cells ferment mixtures such as natural sugars, it may be desirable to utilize a multipurpose cell that can convert all distinct sugar molecules present to a target end product. In some embodiments, the individual sugar molecules may be converted to product  
10 or utilized with different efficiencies, and it may be optimal to adjust the rate of consumption of substrates individually. In another embodiment, a multipurpose cell may be utilized to detect more than one molecule in its environment. The cell may respond in the same manner for each input molecule thus allowing it to be determined that at least one of a set of molecules is present, or it may respond in a different manner for each, thus allowing the  
15 specific molecules present to be identified individually. These may be responsive to different toxins or pollutants and either may process them to reduce their toxicity and/or provide a signal indicating their presence.

The invention therefore provides methods and compositions for generating cells having modified and in some instances novel function. These functions are essentially  
20 unlimited. In some embodiments, such functions arise from the synthesis of a new nucleic acid that imparts a particular biological function as a result of the order of its genetic elements. For example, a particular biochemical pathway in a cell may be altered as a result of a difference in the ratios of enzymes and substrates involved in the pathway. As another example, a particular signaling pathway in a cell may be altered as a result of a difference in  
25 the ratios of kinases, phosphatases, adaptors, and downstream transcription factors. The target nucleic acid (e.g., the final recombined product) can be isolated from the chassis cell and introduced into another cell that is for example amenable to the particular desired function. The target nucleic acid may be integrated into the host cell genome or it may exist as an extragenomic plasmid or vector.

30 Cells comprising these new pathways therefore find wide application including environmental applications such as petroleum metabolism, degradation and/or conversion, pollutant metabolism, degradation and/or conversion, toxic waste metabolism, degradation and/or conversion, greenhouse gas metabolism, degradation and/or conversion, ethanol production, ethanol conversion, synthesis of novel compounds including biologics, altered

enzymes, and the like; agricultural applications such as manure metabolism, degradation and/or conversion, methane metabolism, degradation, conversion and/or capture, corn degradation and conversion (e.g., into ethanol), generation of microbe resistant plants or crops, generation of faster growing or faster maturing plants or crops, generation of plants or crops with particular phenotypes including altered color, smell, taste and the like; food industry application such as generation of faster fermenting yeast for the bread industry, generation of more stable bacteria for the cheese and milk industry; biotechnology applications including increased synthesis of biochemical products such as nucleotides, amino acids, proteins, enzymes, and the like; generation of altered protein complexes such as proteosomes, inflammasomes, transcriptional machinery and complexes, and the like.

Pathway Design or Development:

Aspects of the invention provide methods for designing and making engineered pathways. In one aspect, alternative pathways for making one or more products of interest from one or more available substrates may be made and tested in one or more host cells or organisms of interest. Efficient nucleic acid synthesis methods enable larger numbers of different pathways to be tested. Accordingly, alternative combinations of different pathway components may be designed based on known functional or regulatory properties.

Computer-implemented design techniques may be used to generate alternative pathways for metabolizing one or more substrates of interest and/or generating one or more products of interest. In some embodiments, databases that contain information on genomes and their link to biological systems may be utilized for designing metabolic pathways. The Kyoto Encyclopedia of Genes and Genomes (KEGG) resource is an example of a database that provides a reference knowledge base for linking genomes to biological systems and wiring diagrams of interaction networks and reaction networks. Other examples of database resources are LIGAND (a composite database that provides information about metabolites and other chemical compounds, substrate-product relations representing metabolic and other reactions and information about enzyme molecules), MetaCyc (a database of metabolic pathways and enzymes), the metabolic pathways database (MPW, a database of pathway structures) and the University of Minnesota biocatalysis/biodegradation database (a database of microbial biocatalytic reactions of and biodegradation pathways for organic chemical compounds). A database of pathway components may also contain components of predicted, putative, or unknown functions. It may also contain pseudo-components of defined function that may have an undefined composition. In some embodiments, a program may design

combinations of regulatory and/or functional elements that are in the public domain (e.g., that are not covered by patent rights and/or are not subject to a licensing fee). Databases of freely available genetic elements may be generated and/or used as a source of nucleic acid sequences that can be combined to produce alternative pathways. Alternative pathways  
5 containing different combinations of known functional and/or regulatory elements (e.g., from different species) may be designed, assembled, and/or tested. Libraries including variations in enzymatic element regions may be used to ascertain the relative effects of different types of enzymes or of different variants of the same enzyme. Libraries including variations in regulatory element regions may be used to ascertain the optimal expression level or  
10 regulatory control among a set of genes. In some embodiments, two or more alternative pathways may be provided in a single cell.

Nucleic acids encoding the different pathways may be assembled. In some embodiments, the functional properties of different engineered pathways may be tested *in vivo* by transforming host cells or organisms with the appropriate assembled nucleic acids,  
15 and assaying the properties of the engineered organisms. In some embodiments, the functional properties of different engineered pathways may be tested *in vitro* by isolating components expressed from assembled nucleic acids and testing the appropriate combinations of components in an *in vitro* system.

For example, a plurality of different theoretical metabolic pathways may be  
20 contemplated to obtain one or more moieties of interest (e.g., 1, 2, 3, 4, 5, 5-10, 10-20, 20-50, 50-100 or more). A moiety of interest may be an industrial chemical, an agricultural product (e.g., a fuel such as ethanol, biodiesel, etc.). In some embodiments, different theoretical metabolic pathways may be designed based on a plurality (e.g., 1, 2, 3, 4, 5, 5-10, 10-20, 20-50, 50-100 or more) of different feedstocks that are available. It should be  
25 appreciated that metabolic pathways may be designed to function *in vitro* or *in vivo*. Depending on whether the intended application is *in vitro* or *in vivo*, different factors may be considered and different pathway components may be included.

These different theoretical pathways may be made as described herein and then tested to determine which one or more are the most effective. Alternatively, different candidate  
30 components (e.g., functional and/or regulatory) may be provided as starting components to generate a plurality of different pathways via mixing or recombination (e.g., *in vivo* as described in Example 1). The different pathways then may be tested to determine which one or more are the most effective. Accordingly, engineered combinations of functional and/or regulatory pathway components may be selected or screened for as discussed in Example 1.

In some embodiments, engineered RNA aptamers, proteins, or other molecules that are responsive to one or more metabolites and/or other ligands may be selected or screened for as discussed in Example 2. One or more of these aptamers, proteins, or other molecules may be used as regulatory components of a metabolic pathway. One or more of these  
5 aptamers, proteins, or other molecules may be used to provide a detectable and/or quantifiable readout indicative of the level of one or more intermediates in the pathway.

Aspects of the invention provide sets of aptamers, proteins, or other molecules that can detect the presence of one or more different ligands or effector molecules. In some embodiments, an aptamer, protein, or other molecule set may be provided and transcribed in  
10 a host cell (e.g., from a transcription template that is in a vector or that is integrated into the genome of the host cell). In some embodiments, any additional RNAs and/or proteins that may be required for the different readouts may be transcribed in the host cell.

Aptamers, aptamer sets, proteins, or other molecules of the invention may be used to detect the presence of any type of ligand, including for example, different analytes, metabolic  
15 intermediates and products, toxins, environmental contaminants and pollutants, and any other type of ligand and or effector molecule.

Aptamers, aptamer sets, proteins, or other molecules of the invention may be used as regulatory components of a metabolic pathway. For example, one or more aptamers may provide a positive or negative regulatory feedback or feedforward loop within a pathway. An  
20 aptamer may be designed or isolated to bind to one or more metabolites of interest and, upon binding, upregulate or downregulate (e.g., increase or decrease the expression) of one or more upstream or downstream functional components (e.g., enzymes) within the pathway.

Accordingly, aptamer, protein, or other molecule containing cells (or isolated preparations of aptamer sets) may be used in medicine, biotechnology, industry, agriculture,  
25 environmental studies and remediation, mining, and any other application where one or more ligands may need to be detected. In aspects of the invention, an environmental pollutant may be a water, air, or soil pollutant. Water pollutants may be compounds such as organic and inorganic chemicals, for example, heavy metals, petrochemicals, chloroform, and different types of bacteria. Water pollution also may occur in the form of thermal pollution and  
30 dissolved oxygen depletion. Air pollutants may be compounds such as carbon monoxide, sulfur dioxide, chlorofluorocarbons (CFCs), and nitrogen oxides. Soil pollutants may be compounds such as hydrocarbons, heavy metals, methyl tert-butyl ether (MTBE), herbicides, pesticides and chlorinated hydrocarbons, and others. Such detection methods may be important for detecting changes in pollutants after natural disasters such as hurricanes or

flooding. It should be appreciated that readout and/or regulatory components (and/or a combination thereof) of the invention can be designed or modified to be sensitive to any ligand regardless of whether it is a metabolite of an engineered pathway or a pollutant or other environmental, agricultural, industrial, or mineral molecule. A readout component that is sensitive to a ligand may bind to that ligand and provide detectable quantitative and/or qualitative feedback information about the level of the ligand. Similarly, a regulatory component that is sensitive to a ligand may bind to that ligand and promote a regulatory response to the amount and/or presence or absence of the ligand. It should be appreciated that the readout and/or regulatory components described herein may be used to monitor and/or control a first engineered pathway of the invention as a function of i) one or more metabolite levels (e.g., substrates, intermediates, and/or products) of the first pathway, ii) one or more metabolites (or levels thereof) from at least one second pathway (e.g., a second engineered pathway or a naturally-occurring pathway) that the first engineered pathway is designed to respond to, iii) one or more other external ligands that the first engineered pathway is designed to respond to, or any combination thereof.

Compositions and methods of the invention also may be useful to identify the presence of one or more metabolic intermediates and/or products. In some embodiments, detection may be performed in the natural cellular environment in a live cell rather than in a cellular extract. In some embodiments, metabolic pathways may be studied and individual steps may be identified by providing, *in vivo*, a plurality of different aptamers that are responsive to different intermediate compounds. By determining which aptamers give a positive readout, the nature of the intermediate compounds can be determined and a metabolic pathway may be inferred. In some embodiments, an aptamer set containing different aptamers that are responsive to different substrates, metabolic intermediates, and/or desired end products may be used as a reporter system (e.g., either on a plasmid or integrated into the genome of a host cell) in techniques designed to evolve or select novel biosynthetic pathways. An aptamer set that is selected may include one or more copies of aptamers that are selective for intermediates of analytes that are expected to be produced in a novel biosynthetic pathway of interest. In some embodiments, an appropriate readout from an aptamer set may be used to indicate that a particular combination of enzymes and/or enzyme variants may have a metabolic effect that is desired.

It should be appreciated that a nucleic acid construct encoding an aptamer set of interest may be transcribed *in vitro*. Similarly, a set of RNA aptamers that are responsive to different ligands of interest may be assembled *in vitro*. Sets of aptamers that bind specifically



to a plurality of different ligands also may be used *in vitro*. In some embodiments, the aptamers may be used in an *in vitro* assay to detect any one or more of a plurality of different ligands (e.g., metabolic intermediates, toxins, environmental pollutants, contaminants, pathogens, analytes, etc.). In some embodiments, one or more stabilizing residues (e.g., one or more 2'-O-methyl ribonucleotides or other stabilizing ribonucleotides) may be incorporated into aptamers that are synthesized *in vitro* and/or *in vivo*.

Aspects of the invention may involve one or more nucleic acid assembly reactions in order to make the sets of DNA molecules, RNA encoding fragments, aptamer constructs, modified host cells, and/or other nucleic acids that may be used to isolate and/or use RNA molecules having one or more functions of interest. Aspects of the invention may be used in conjunction with *in vitro* and/or *in vivo* nucleic acid assembly procedures. Non-limiting examples of extension-based and ligation-based assembly reactions are described herein and known in the art (see for example, Published US Patent Applications 20070231805, published October 4, 2007, and 20070122817, published May 31, 2007, the disclosures of which are incorporated herein by reference).

## EXAMPLES

### Example 1. Screening or Selecting for Configurations of Pathway Components

Aspects of the invention relate to nucleic acid libraries and host cells that can be used to generate a variety of different functional nucleic acid configurations *in vivo*. Certain aspects of the invention involve identifying genetic configurations that provide one or more biological functions of interest. In some embodiments, new or alternative regulatory or metabolic pathways may be identified. In some embodiments, methods of producing one or more metabolic products or intermediates may be identified.

Aspects of the invention take advantage of nucleic acid assembly technology that supports the production of any nucleic acid fragments (including large nucleic acid fragments) having a predetermined sequence of interest. Technology described herein allows nucleic acid and cellular libraries of the invention to be designed and assembled to include many different genetic elements of interest. This assembly technology also allows the production of nucleic acids that can be used to modify host organisms as described herein.

Thus, in one aspect, the invention provides a method of altering a cell function comprising introducing into a cell a nucleic acid comprising a set of genetic elements having

recombination sites situated there between, rearranging the genetic elements by recombination at the recombination sites, and screening the cell for an altered cell function.

In some embodiments, the cell has been modified to delete genomic recombination sites. The genomic recombination sites may be reduced by 10-20%, 20-30%, 30-40%, 40-50%, 50-60%, 60-70%, 70-80%, 80-90% or 90-100%. In some embodiments, the genomic recombination sites are reduced by 50% or more. In some embodiments, the genomic recombination sites are reduced by 90% or more.

In some embodiments, the cell is a bacterial cell such as but not limited to an *E. coli* cell. In some embodiments, the cell is a eukaryotic cell such as but not limited a yeast cell, an insect cell, or a mammalian cell.

In some embodiments, the genetic elements are coding sequences. In some embodiments, the genetic elements are regulatory sequences. In some embodiments, the genetic elements are regulatory sequences and coding sequences. In some embodiments, the genetic elements are introns, in others they are exons, and in still others they are introns and exons. In some embodiments, the method further comprises isolating the cell having an altered cell function. In some embodiments, the nucleic acid is a vector. In some embodiments, the vector comprises a selection sequence. In some embodiments, the nucleic acid is integrated into the genome of the cell. In some embodiments, the recombination sites are identical. In other embodiments, the recombination sites comprise at least two different types of recombination sites. In some embodiments, the recombination sites are restriction enzyme sites. In some embodiments, the recombination sites are homologous recombination sites. In some embodiments, the recombination sites are susceptible to single or double stranded cuts.

In another aspect, the invention provides a method of producing a cell having an altered cell function comprising introducing into a cell a nucleic acid comprising a set of genetic elements having recombination sites situated there between, rearranging the genetic elements by allowing recombination between recombination sites, and isolating a cell having an altered cell function. In some embodiments, the method further comprises propagating the cell having an altered function. In another aspect, the invention provides a method for producing a recombined nucleic acid molecule comprising producing a cell according to the method described above, and harvesting from the cell a recombined nucleic acid.

In some embodiments, the target nucleic acid (e.g., the recombined nucleic acid) may be amplified, sequenced or cloned after it is made. In some embodiments, a host cell may be transformed with the assembled target nucleic acid. The target nucleic acid may be

integrated into the genome of the host cell. In some embodiments, the target nucleic acid may encode one or more polypeptides. The polypeptide may be expressed (e.g., under the control of an inducible promoter). The polypeptide may be isolated or purified. A cell transformed with an assembled nucleic acid may be stored, shipped, and/or propagated (e.g.,  
5 grown in culture).

In another aspect, the invention provides methods of obtaining target nucleic acids by sending sequence information and delivery information to a remote site. The sequence may be analyzed at the remote site. The starting nucleic acids may be designed and/or produced at the remote site. The starting nucleic acids may be assembled in a reaction involving a  
10 combination of ligation and extension techniques at the remote site. In some embodiments, the starting nucleic acids, an intermediate product in the assembly reaction, and/or the assembled target nucleic acid may be shipped to the delivery address that was provided.

Other aspects of the invention provide systems for designing starting nucleic acids and/or for assembling the starting nucleic acids to make a target nucleic acid. Other aspects  
15 of the invention relate to methods and devices for automating a multiplex oligonucleotide assembly reaction that involves a combination of ligation and extension assembly techniques. Yet further aspects of the invention relate to business methods of marketing one or more methods, systems, and/or automated procedures that involve a combination of ligation and extension multiplex nucleic acid assembly reactions.

Accordingly, aspects of the invention relate to methods and compositions for  
20 generating functional diversity and for identifying novel biological functions. In some aspects, the invention provides a set of genetic elements associated with recombination sites in an initial configuration (e.g., a vector comprising a linear array of genetic elements alternating with recombination sites). The recombination sites can promote rearrangement of  
25 the genetic elements thereby generating a plurality of different new configurations. Genetic elements may be genes, gene fragments, operons, subsets of genes from an operon, exons, introns, regulatory sequences, or other genetic elements that can confer a functional property (e.g., alone or in combination with one or more additional genetic elements). Accordingly, rearrangement of the genetic elements provides novel genetic configurations that may have  
30 new functional properties.

In some aspects, the invention provides methods for generating functional diversity *in vivo* by providing a population of cells containing an initial configuration of genetic elements associated with recombination sites and allowing or promoting recombination to generate a plurality of rearranged configurations of the genetic elements. Different rearranged

configurations will be present in different cells. Appropriate selection and/or screening techniques may be used to identify cells that have a novel biological function of interest. The rearranged configuration of genetic elements that is associated with a novel biological function may be identified and/or isolated.

5 In some aspects, a cell line may be modified to remove one or more recombination sites (e.g., by deletion or alteration) from its genome. Such a modified cell line may be used as a chassis that can host different initial sets of genetic elements that are configured with the one or more recombination sites that were removed from the host genome. A lack of recombination sites on the host genome reduces the frequency of recombination between the  
10 set of genetic elements and the genome, thereby limiting recombination to rearrangements between the genetic elements of interest.

In one aspect, the invention may be used to generate and identify novel biological pathways, including, for example, novel regulatory pathways, metabolic pathways (e.g., catabolic or anabolic), or other novel biological pathways. In another aspect, proteins or  
15 RNAs with novel or modified functions may be generated and identified. In yet another aspect, methods of the invention may be used to modify existing biological pathways (e.g., to increase or decrease certain functions, to increase or decrease the accumulation of one or more intermediates or products, etc.).

Further aspects of the invention provide modified host cells that are designed to  
20 harbor libraries of the invention and allow for rearrangement of the genetic elements within the library without involving any rearrangement of the host genome. In some embodiments, a host genome may be genetically modified to remove one or more sequences in its genome that are identical or similar to the recombination sites in the library. For example, a host genome may be modified to remove one or more restriction sites that are used to promote  
25 recombination between different genetic elements within a library. Accordingly, a modified host cell of the invention can serve as a chassis for generating functional diversity from an appropriate library of initial nucleic acids that is introduced into the cell.

In aspects of the invention, recombination may result from the actions of endogenous host agents (e.g., nucleic acids, proteins, combinations thereof, and the like). In other  
30 embodiments, a host cell may be modified to express one or more agents that promote recombination between recombination sites. These agents are referred to herein as recombination inducing agents. Examples include recombination enzymes, restriction enzymes, topoisomerases, repair enzymes, and the like. In one illustrative embodiment, a host cell may be modified to express a restriction enzyme that acts on a recombination site.

In another illustrative embodiment, a host cell may be modified to express one or a set of recombination enzymes that act on repeated sequences that are included in the initial nucleic acid library and/or that are introduced into the genome of the cell.

It should be appreciated that genes encoding recombination promoting agents should be expressed at suitable levels. Such levels promote a sufficient rate of genetic rearrangement (e.g., sufficient to provide a large pool of candidate configurations that can be screened or selected for new functions of interest). However, the rate of rearrangement should not be so high that the configurations are too unstable to be screened, selected, or maintained for subsequent analysis and/or propagation.

In some embodiments, genes encoding recombination promoting agents may be inducible thereby temporally limiting rearrangement to times when the genes are induced. In other embodiments, these genes may be constitutively expressed thereby promoting continuous rearrangement during cell growth.

Accordingly, aspects of the invention provide new methods for manipulating genetic elements (e.g., operons, genes, gene fragments, promoters, exons, introns, etc.) thereby opening up new opportunities to modify structure, function and temporal or spatial expression of proteins, protein function, metabolic pathways, and other cellular functions. Assembly methods of the invention can be used to generate any predetermined linked set of genetic elements and recombination sites in any initial configuration of interest. These initial configurations may be incorporated into vectors and/or introduced directly into host cells.

#### *Genetic elements:*

As described herein, a genetic element may be any nucleic acid sequence that confers a biological property of interest (e.g., a biological property that may be altered through rearrangement with other genetic elements to obtain a new or modified biological property of interest). A genetic element may be a coding or a non-coding sequence.

In some embodiments, a genetic element is a nucleic acid that codes for an amino acid, a peptide or a protein. Genetic elements can be as short as a one or a few codons (e.g., a start codon). A genetic element may consist of an entire open reading frame of a protein, or it may consist of the entire open reading frame and one or more (or all) regulatory sequences associated with that open reading frame. Regulatory sequences include but are not limited to promoters, enhancers, silencers, transcriptional attenuation sequences, and the like. Genetic elements may be exons, introns, or nucleic acid sequences comprising both exons and introns. A genetic element can comprise a plurality of coding sequences and/or regulatory sequences.

In some embodiments, a genetic element may be one or more regulatory and/or one or more coding sequences from a naturally-occurring operon (e.g., those found in bacterial sequences).

In some embodiments, nucleic acids that can adopt a particular secondary structure may be genetic elements. An example of such a nucleic acid is a poly-G sequence. As another example, a genetic element may be a nucleic acid having a sequence that induces polymerase slippage.

The genetic elements are linked together, preferably with recombination sites there between. As used herein, linked refers to a covalent bond between genetic elements and recombination sites. The covalent bond in its simplest form is a phosphodiester backbone of the nucleic acid molecule which comprises the genetic elements and recombination sites. Other linkages are also possible provided they do not interfere with the recombination of genetic elements and ultimately the transcription of the recombined nucleic acid.

The nucleic acids may further comprise mRNA stability and/or stabilization sequences. The location of these sequences may similarly be rearranged and thus they too may be genetic elements.

#### *Recombination sites:*

As used herein, a recombination site is a nucleotide sequence that induces or facilitates recombination in vitro or in vivo. In many instances the site is recognized, bound by, and/or acted upon by a recombination promoting agent such as a protein.

In some embodiments, a recombination site is a restriction enzyme site (i.e., a site recognized by and/or cleaved by a restriction enzyme). After cleavage by a restriction enzyme, a restriction site can promote recombination. Restriction sites may be of any length (e.g., 4- 20 base pairs). The longer the restriction site, the less frequently it will normally occur in a genome. Enzymes that cut these longer sequences are sometimes referred to as "rare cutters". Suitable restriction enzyme sites may be found, for example, in a commercial catalog (e.g., New England Biolabs). Most restriction enzymes will induce a double strand break. However, the action of certain restriction enzymes will result in a single strand nick only. A single strand nick also may promote recombination because the processing of this nick by a replication fork or DNA repair enzymes can induce a recombination event. It should be appreciated that for a restriction site to act as a recombination site *in vivo*, the appropriate restriction enzyme must also be present in the cell. The enzyme may be

endogenous to the cell or may be ectopically expressed or introduced into the cell directly as a protein.

In certain embodiments, a recombination site is a sequence-specific recombination site (e.g., a *lox* P site) that is recognized by a recombinase (e.g., the Cre enzyme). It also should be appreciated that for a sequence-specific recombination site to act as a recombination site *in vivo*, the appropriate recombinase enzyme must also be present in the cell. The enzyme may be endogenous to the cell or may be ectopically expressed or introduced into the cell directly as a protein.

In some embodiments, any repeated nucleic acid sequence can be a recombination site. For example, any nucleotide sequence can be a recombination site if there are two or more identical or homologous nucleotide sequences interspersed between genetic elements. Since recombination is promoted by homology, a greater homology (e.g., either in length or percentage) promotes a higher recombination frequency. Preferably, these types of recombination sites share 100% identity (i.e., their nucleotide sequences are identical).

However, homologous recombination can also occur between sequences that are not identical yet still share a high degree of homology. Thus these sequences may share greater than 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homology. The entire nucleotide sequence located between consecutive genetic elements may be a recombination sequence, or only a fragment thereof may be. The nucleotide sequences located between genetic elements may determine their propensity to participate in desired recombination events. For example, a particular recombination site can be designed to recombine specifically with only one other recombination site. This can be accomplished if the two sites have sequences that are rare and highly homologous if not identical. In some embodiments, recombination sites can be designed to recombine with many other locations by using sequences that are identical or highly homologous to sequences that occur frequently.

Examples of recombination enzymes include but are not limited to tyrosine recombinases, serine recombinases, FIp, RecA, Pre (plasmid recombination enzyme) and ERCC1.

In some embodiments, recombination can be induced by certain nucleotide modifications or processes. For example, DNA strand breaks (e.g., double strand breaks and/or single strand breaks) can promote recombination. Damaged or modified bases, or abasic sites also can induce recombination. Any nucleotide modification that results in the stalling of a replication fork also can induce recombination. Accordingly, modified or

damaged nucleotides can be recombination sites, as can sites acted upon by enzymes that modify and/or damage nucleic acids in this manner.

In certain embodiments, a recombination site is any stretch of nucleotides that can induce recombination through a triggering event. For example, bases that are susceptible to modification may be recombination sites. Such bases, when modified, can be removed by repair enzymes or through a physical action (e.g., exposure to heat or light). Removal of damaged bases produces abasic sites that can induce recombination. In some embodiments, if multiple damaged sites are located opposite from each other, removal of damaged bases can lead to DNA double strand breaks that also promote recombination.

The linked set of genetic elements having recombination sites situated there between may utilize a single type of recombination site, such that recombination between any and all genetic elements may occur with an approximately equal probability. In other instances, the linked set of genetic elements may utilize two or more types of recombination sites. In these latter instances, there should be at least two copies of each recombination site so that each site has at least one recombination partner. In these embodiments, the initial nucleic acid is designed to increase the recombination frequency between particular genetic elements while almost precluding these recombination with other genetic elements.

#### *Vectors:*

Initial configurations of genetic elements and recombination sites may be provided in the form of a single or double-stranded linear or circular nucleic acid molecule with or without vector sequence. These initial configurations are referred to herein as initial nucleic acids. In some embodiments, an initial configuration of genetic elements and recombination sites may be cloned into a vector. A vector may be any suitable vector. For example, a vector may be a plasmid, a cosmid, a phagemid, a BAC, a YAC, an F factor, or any other suitable prokaryotic, eukaryotic or viral vector. A vector may include an origin of replication and/or one or more selectable markers (e.g., antibiotic resistant markers, etc.) and/or detectable markers (e.g., fluorescent markers, etc.). In some embodiments, a vector may be a shuttle vector that is functional in two or more different types (e.g., species) of host cells.

It should be appreciated that a vector may be selected or modified to remove recombination sites that could interfere with the desired recombination events involving the recombination sites that are being used to promote rearrangements of the genetic elements of interest. Vectors may therefore be modified to reduce the number of one or more



recombination sites by at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% or by 100%.

Vectors can be introduced into a host cell through a variety of mechanisms. They can be transformed, transfected or introduced by physical techniques like microinjection or electroporation. In other embodiments, vectors may be introduced through biological means, for example using phages or viruses. Many methods of introducing oligonucleotides into cells are known to persons of ordinary skill in the art and are incorporated herein by reference.

Upon introduction of an initial nucleic acid (whether or not in the context of a vector) into a host cell, the genetic elements can undergo recombination. Recombination can be initiated by replication-associated events, or through other triggering events such as the initiation of DNA strand breaks in the recombination sites through the action of restriction enzymes or the creation of DNA strand breaks through other means.

In some embodiments, a low copy number vector (e.g., plasmid) may be used to maintain the initial linked set of genetic elements and recombination sites and avoid a potential loss of elements due to toxicity or other issues that may be associated with high copy number vectors.

In some embodiments, an initial set of genetic elements and associated recombination sites may be integrated into the genome of the host cell. This may involve integrating a vector into the genome of a host cell. Accordingly, a host cell and/or a plasmid may be modified to introduce a homologous sequence that could promote integration of the plasmid into the genome. The plasmid may be replication defective in the host that is being used to generate the rearranged configurations of genetic elements (e.g., the target nucleic acids). By incorporating the genetic elements and recombination sites into the host genome, multiple recombination events can occur without losing the vector or needing to select for the vector. Also, in some embodiments the set of genetic elements and recombination sites may be more stable if they are integrated as a single copy into the genome of the host.

#### *Host cells:*

Any cell type may be suitable as a host cell provided it can perform the recombination functions required for rearrangement of the genetic elements. The cell may be inherently or endogenously capable of such recombination or it may be manipulated to be so. In some embodiments, a host cell expresses one or more restriction enzymes and/or one or more recombinase enzymes that can act on one or more of the recombination sites being used to

generate rearrangements. The enzyme may be encoded in a vector or in the genome of the host cell (e.g., the gene encoding the enzyme may be integrated into the genome of the host cell). In some embodiments, expression of the enzyme can be controlled (e.g., inducible). For example, the gene encoding the enzyme can be placed under the control of a specific promoter. This can be used to control the timing and duration of recombination by turning enzyme expression on or off when appropriate. Accordingly, the extent of recombination to be controlled. By switching off enzyme expression, a pool of rearranged configurations (i.e., target nucleic acids) can be maintained in a stable form and exposed to appropriate selection and or screens for a biological function of interest.

In some embodiments, a host cell may be modified to provide a platform or chassis that can be used for multiple screens or selections starting with different sets and/or configurations of genetic elements. The genome of a host cell may be modified to remove sequences that can induce unwanted recombination.

In some embodiments, to accommodate multiple rounds of recombination events between sites on one or more vectors that are introduced into a host cell, the genome of the cell may be modified to remove recombination sites that potentially may interfere with intra- or inter-vector recombination. In certain embodiments, the chassis-cell will be engineered to have no recombination sites at all. In other embodiments, the chassis will have a subset of its recombination sites removed (e.g., about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more).

Any restriction site may be used as a recombination site and may therefore be removed (completely or some fraction thereof) from the genome of a host cell. In some embodiments, rarer restriction sites may be selected (e.g., ones that recognize a unique long site). In some embodiments, one or more 4 base cutters or 6 base cutters recognition sequences may be removed. In some embodiments, the recognition sequences of one or more of I-SceI, I-CeuI, PI-PspI, PI-SceI, and NotI restriction sites may be removed. In some embodiments, CTAG sites may be removed from the genome of the host and used as part of a recombination site in association with the set of genetic elements.

It should be appreciated that certain or all recombination sites can be removed from the genome without a penalty if the site is essential to the genome (e.g., if it is a non-transcribed sequence). If one or more recombination sites is in for example an actively transcribed part of the genome, and cannot be removed without compromising the viability of the cell, its ability to act as a recombination site may be reduced or eliminated by mutation. For example, if the site is a homologous recombination site then the site may be mutated by

reducing the level of identity or homology to the point where it would no longer recombine with the recombination sites in between the genetic elements of the initial nucleic acid. If the site is in a coding region, it may be mutated by using alternate codons, and thereby not affecting the protein sequence. If the recombination sites of the vector(s) are based on  
5 restriction sites that need to be activated, genomic restriction sites having the same sequence can be removed or inactivated to avoid or reduce the frequency of vector-genome recombination.

In some embodiments, a "chassis" cell may be modified to remove all (or a subset, including for example at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% or more) of  
10 two or more (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) different recombination sites. For example, two or more different restriction sites may be removed. These cells also may be modified to express two or more different restriction enzymes that recognize these sites. These enzymes may be independently inducible. These cells may be used to promote recombination of different sets of genetic elements that are associated with different  
15 restriction sites. In some embodiments, an initial configuration of a set of genetic elements may include two or more different restriction sites (e.g., distributed in the same configuration or in different configurations). A chassis that expresses the corresponding restriction enzymes under different regulatory controls can be used to promote independent rearrangement of different components of the initial set of genetic elements by expressing  
20 different restriction enzymes.

In some embodiments, the genome of the host may be modified to introduce a sequence that is homologous to a sequence on a vector or other nucleic acid containing the set of genetic elements and recombination sites in order to help integrate them into the genome of the host cell. In some embodiments, the genome of a host cell may be modified to provide  
25 recombination sequences to allow genomic integration of two or more different sets of genetic elements.

In some embodiments, the genome of a host cell may be reduced (e.g., by 5%, 10%, 15%, or more) in order to accommodate the sets of genetic elements being integrated.

A host cell may be prokaryotic (e.g., bacterial such as *E. coli* or *B. subtilis*) or  
30 eukaryotic (e.g., a yeast, mammal or insect cell). It should be appreciated that when integrating a nucleic acid into a eukaryotic genome (e.g., a mammalian genome) care should be taken to select sites that will allow sufficient expression (e.g., silenced regions of the genome should be avoided, whereas a site comprising an enhancer may be appropriate).

In some embodiments, a host cell may be selected for its recombination properties. In some embodiments, a host cell may be selected for its metabolic properties. For example, if a selection or screen is related to a particular metabolic pathway, it may be helpful to use a host cell that has a related pathway. Such a host cell may have certain physiological adaptations  
5 that allow it to process or import or export one or more intermediates or products of the pathway. However, in other embodiments, a host cell that expresses no enzymes associated with a particular pathway of interest may be selected in order to be able to identify all of the components required for that pathway using appropriate sets of genetic elements and not relying of the host cell to provide one or more missing steps.

10 Examples of organisms that may have useful phenotypes include, but are not limited to *Deinococcus radiodurans* and *Vibrio furnissii*. *Deinococcus radiodurans* has evolved a strong capability for recombination and introduced vectors can be recombined at high frequency. *V. furnissii* can produce n-alkenes from products found in waste-water and is commercially interesting. (Park et al., 2005, J. Appl. MicroB. 98, 324). The bacterium  
15 already has a pathway in place for w-alkene synthesis.

The genome of a host organism may be modified through the re-synthesis of large parts of the genome and replacing the original genome (or a portion thereof) with a new optimized genome (or a portion thereof) through recombination. In some embodiments, assembly methods described herein may be used to generate these large genome parts.

20 In some aspects of the invention, cells may be modified to add recombination elements between naturally occurring genomic genetic elements (e.g., between predetermined genomic elements of interest). Recombination within such cells also generates functional diversity that can be used to screen or select for one or more novel functions of interest. This approach may be particularly useful if the host cell genome encodes an operon or other  
25 cluster of genes selected for analysis.

*Configuration of sets of genetic elements and recombination sites:*

Aspects of the invention may involve any combination of any appropriate number of genetic elements and recombination sites. For example, 2-5, 5-10, 10-20, 20-50, 50-100 or  
30 more different genetic elements may be included. Each genetic element may be flanked by two recombination sites resulting in a configuration of alternating genetic elements and recombination sites. However, other configurations may be used. For example, several genetic elements may be grouped together and not separated by recombination sites (e.g., if they perform a core function to the desired biological function being screened or selected

for). In some embodiments, the genetic elements are genes (including sequences required for transcription and translation). In some embodiments, the genetic elements are part of a natural operon and are under transcriptional control of a single promoter. In some embodiments a plurality of different genetic elements may be separated by restriction sites but artificially brought under the control of a single promoter in an artificial operon. In some embodiments, the identity of the genetic elements that are included in the initial set may be determined by the type of biological function that is being selected or screened for. For example, if an improved or altered enzyme function is desired, multiple copies of a gene encoding the enzyme may be used, each copy having one or more sequence variations. The recombination sites may be designed to allow rearrangement of different regions of the gene so that different sequence combinations can be sampled. In contrast, if a new or modified metabolic pathway is desired, a plurality of different enzymes that have functions related to the desired pathway may be used along with different promoter and other regulatory sequences. Recombination sites may be placed between these different genetic elements so that different combinations of genes expressed at different levels may be sampled. It should be appreciated that combinations of these strategies may be implemented. It also should be appreciated that combinations of genetic elements from different organisms also may be grouped together in an initial set.

As discussed above, the recombination sites that flank the genetic elements can be the same or different. In another embodiment, multiple copies of the recombination site are inserted in the vector thereby increasing the likelihood of a recombination event. In other embodiments, genetic elements are flanked by different recombination sites. Having different recombination sites has the advantage that more than one recombination event can be triggered independently. Any combination of recombination sites (e.g., restriction sites, homologous sequences, etc.) can be used when assembling these different recombination sites.

#### *Screens and selections for biological functions:*

Under appropriate conditions, rearrangement of genetic elements is promoted *in vivo* inside a host cell due to the presence of the recombination sites. Rearrangement of the genetic elements provides novel genetic configurations that may have new functional properties. In some aspects, the invention provides methods for selecting or screening for novel functions. Host cells harboring the libraries of target nucleic acids (i.e., recombined nucleic acids) may be exposed to appropriate conditions to identify one or more novel

functions of interest. Novel functions may include altered activities of existing enzymes, novel regulatory responses (e.g., altered patterns of response to a signal, response to a novel signal, etc., or combinations thereof), novel combinations of enzymes that result in novel pathways (e.g., novel metabolic pathways), other novel functions, or combinations thereof.

5 In some embodiments, selection or screening may be performed on the host cell in which genetic rearrangement occurred. In other embodiments sets of genetic elements are allowed to undergo recombination in chassis cells and are subsequently extracted from the chassis cells. The rearranged vectors can then be screened *in vitro* or can be introduced in an alternative cell line, which does not have to be a chassis cell, to be analyzed *in vivo*.

10 Aspects of the invention may be used for pathway engineering *in vivo*. The evolution of entire metabolic pathways rather than just one enzyme may be particularly useful because compounds are often produced or metabolized in a process involving multiple steps of a pathway rather than by one enzyme. Multi-enzyme pathways can also be engineered through the manipulation of certain key enzymes in the pathway.

15 Once a pool of rearranged genetic elements has been generated, the candidates can be used to screen or select for biological properties of interest. Candidates can be screened while recombination is still proceeding. In some embodiments, candidates can be screened after a certain number of recombination events have taken place.

Candidates can be screened for by selective pressure (e.g., whether the organism  
20 survives when a toxin is added to the growth environment or when an essential nutrient is removed, etc.). Further non-limiting examples of screening or selection techniques may include growing organisms at high temperatures or in organic solvents. If a specific enzyme is targeted for optimization, an enzyme-specific selection process can be used.

In some embodiments, metabolic pathways can be screened for in functional screens.  
25 Screening for metabolic pathway can involve screening for the occurrence of a desired final product or one or more intermediate products by using a reporter assay for the one or more products. Other non-limiting techniques that can be used may include monitoring decreases in precursor amounts, monitoring metabolism on a related fluorescent compound, and others.

In one embodiment, binding assays can be used to detect the synthesis of a desired  
30 end product. In one embodiment, the desired product is detected using a binding partner such as an aptamer. Aptamers can consist of DNA and/or RNA sequences. Aptamers that bind to metabolites, intermediates or a variety of other compounds can be used. An aptamer that binds a metabolite or intermediate of interest can be developed and used. Binding of the aptamer to the metabolite or intermediate can be assayed for with a reporter that can be an

integral part of the aptamer. The reporter can also be a molecule that detects the difference between bound and unbound aptamer. It should be appreciated that a plurality of different aptamers (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10 or more) with different readouts may be used to monitor the levels of different metabolites or intermediates. It should be appreciated that aptamers may be DNA, RNA, or other nucleic acid molecules. In some embodiments, aptamers such as those disclosed by Smolke *et al.* (Nucleic Acids Research, 2006, 34(19):5670-5682; Published US Application No. 20060088864, published April 27, 2006, the entire contents of which are incorporated herein by reference) or modified forms thereof may be used. In some embodiments, aptamers obtained using methods described herein (e.g., in Example 2 below) may be used.

#### Example 2. Screening or Selecting for Aptamers

Aspects of the invention relate to nucleic acid libraries and host cells that can be used to screen many different nucleic acids *in vivo* and identify rare nucleic acids that have predetermined structural or functional properties of interest. Certain aspects of the invention involve identifying RNA aptamers using *in vivo* selections or screens. In some embodiments, recombinant cells may include several different *in vivo* aptamers associated with different reporter readouts. Aptamers may be used as reporter molecules, regulatory components, and/or functional components of an engineered biological pathway.

Aspects of the invention take advantage of nucleic acid assembly technology that supports the production of any nucleic acid fragments (including large nucleic acid fragments) having a predetermined sequence of interest. Technology described herein allows libraries of the invention to be designed and assembled to include many different predetermined sequences of interest. This assembly technology also allows the production of nucleic acids that can be used to modify host organisms as described herein.

Aspects of the invention relate to RNA libraries that can be used to screen or select for RNA molecules with functional or structural properties *in vivo* (e.g., RNA aptamers). Other aspects of the invention relate to libraries of RNA molecules having predetermined structural and/or functional properties. Aspects of the invention provide compositions and methods for expressing RNA libraries *in vivo*. Further aspects of the invention provide modified host cells that are adapted to express RNA libraries of interest. For example, a host cell may express a specific polymerase for transcribing the RNA, a ribonuclease that can specifically cleave long RNA transcripts, an RNA polymerase that can incorporate modified nucleotides, or any combination thereof.

Aspects of the invention relate to nucleic acid libraries and methods and compositions for preparing libraries containing very high numbers of nucleic acid regions. Aspects of the invention involve preparing a library comprising a plurality of cells, each transformed with one or more separate nucleic acid molecules, wherein each nucleic acid molecule comprises a plurality of nucleic acid regions, and wherein each nucleic acid region can be assayed to evaluate one or more structural and/or functional properties. Accordingly, aspects of the invention can be used to assay a large number of nucleic acid regions for the presence of one or more regions having structural and/or functional properties of interest (e.g., one or more nucleic acid aptamers having selective ligand-binding properties).

In some embodiments, each nucleic acid fragment can transcribe an RNA molecule. The RNA molecules can be assayed (e.g., *in vivo* or *in vitro*) to determine whether any of them have a structure or function of interest. Accordingly, in one aspect the invention provides *in vivo* libraries of transcribed RNA molecules that can be evaluated *in vivo* for the presence of one or more RNAs having structural and/or functional properties of interest (e.g., one or more RNA aptamers having selective ligand-binding properties under biological conditions). In some embodiments, the complexity of a library that comprises a plurality of different vectors wherein each vector encodes a plurality of different RNA molecules may be calculated as the number of transformants multiplied by the number of different RNA-encoding regions on each vector. By using vectors that encode a large number of different RNA molecules (e.g., 10-100 or more), a library of the invention provides a large number of different RNA variants. Accordingly, methods of the invention can be useful to sample a large number of potential nucleic acid sequence variants. By providing a platform for *in vivo* selection or screening, methods of the invention can be useful for identifying one or more nucleic acids (e.g., RNAs) that have structural and/or functional properties of interest under biological conditions. In contrast, aptamers that are identified through *in vitro* aptamer screening and selection technology may not maintain their selective ligand-binding properties under biological conditions.

In some aspects, the invention provides different cell lines, each comprising a plurality of different aptamers that each recognizes a different ligand and provides a different readout (e.g., signal) when its ligand is present *in vivo*. These cell lines, and the sets of aptamers that they contain, can be used in medicine, agriculture, industry, mining, or for other applications where the ability to detect and distinguish between different ligands can be very important. In some embodiments, a cell containing a plurality of different aptamers that can selectively bind to, and signal the presence of, different metabolic intermediates (e.g.,



intracellular metabolic intermediates) can be used to dissect and/or monitor metabolic pathways. Such cells, and the sets of aptamers that they contain, also can be used as markers to select and/or screen for enzymes, enzyme variants, or combinations thereof, that can form novel or modified metabolic pathways. Accordingly, aspects of the invention may be used to

5 develop novel or modified metabolic pathways that may catalyze the conversion of a first compound to a second compound, that may degrade or modify certain compounds, that may synthesize certain compounds, or any combination thereof. For example, methods of the invention may be useful to develop pathways for degrading or modifying environmental contaminants to reduce their toxicity. In some embodiments, metabolic pathways for

10 generating commercially useful compounds may be useful (e.g., ethanol, and other commercially useful compounds).

In some embodiments, methods of the invention relate to *in vivo* aptamer identification and production. A library of RNA molecules may be transcribed and individual RNA molecules with functional and or structural properties of interest may be identified.

15 In aspects of the invention, nucleic acid regions encoding different RNA molecules may be of any length. In some embodiments, a nucleic acid region and the encoded RNA may be at least 50 to at least 200 nucleotide bases long. In certain embodiments, a transcribed RNA may be at least 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200

20 nucleotide bases long. However, certain RNAs may be shorter than 50 bases long (e.g., between about 10 and about 50 bases long).

In aspects of the invention, each vector may encode one or more separate RNA molecules. In certain embodiments, a single vector encodes about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, or more RNA molecules. In some

25 embodiments, the RNA sequences are all different. However, in some embodiments several identical copies of one or more RNA sequences may be transcribed from a single vector. The sequences encoding the separate RNA molecules may be arranged in a linear array.

In some embodiments, transcription of one or more RNA molecules may be under the control of the same promoter. In certain embodiments, transcription of one or more RNA

30 molecules may be under the control of separate promoters. In some embodiments, each RNA is transcribed from its own separate promoter. The separate promoters may be separate copies of the same promoter or different promoters. In some embodiments, one or more promoters may be inducible. In some embodiments, RNA transcription may involve transcription enzymes of the host cell.

In some embodiments, nucleic acid regions encoding separate RNA molecules may be transcribed as a single RNA transcript. A single RNA transcript may include 2 or more RNA molecules. In some embodiments, a single RNA transcript may include 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more RNA molecules. The single RNA transcript may include one or more cleavage sites that can be acted on to release one or more individual RNAs from the RNA transcript. In some embodiments, one or more enzymes may cut the cleavage sites to release individual RNAs. In some embodiments, the cleavage sites may be autocatalytic RNA cleavage sites. In other embodiments, RNAs may be transcribed as individual transcripts. In certain embodiments, a plurality of RNAs may be transcribed in a combination of individual RNA transcripts and RNA transcripts that include two or more RNAs.

In aspects of the invention, a nucleic acid sequence encoding an RNA molecule and one or more regulatory sequences may be "operably" joined. The nucleic acid sequence and one or more regulatory sequences may be covalently linked in such a way as to place the transcription of the coding nucleic acid sequence under the influence or control of the regulatory sequences. A promoter region is operably joined to a coding nucleic acid sequence if the promoter region is capable of promoting transcription of that nucleic acid sequence such that the resulting transcript may be an RNA molecule of the invention.

The precise nature of the regulatory sequences needed for transcription may vary between species or cell types. In some embodiments, a 5' non-transcribed regulatory sequences may be used that includes a promoter region having a promoter sequence for transcriptional control of the operably joined nucleic acid sequence. Regulatory sequences may also include enhancer sequences or upstream activator sequences as desired.

Transcription vectors containing all the necessary elements for transcription are commercially available and known to those skilled in the art. See, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, 1989. Systems and promoters for nucleic acid transcription in mammalian cells are known to those of ordinary skill in the art and available commercially.

In some embodiments, one or more transcribed RNA sequences may be identical. However, in order to maximize the number of different RNA sequences that may be sampled, each vector may encode a plurality of unique RNA sequences. The vector inserts that encode the unique RNA sequences may be made in a nucleic acid assembly procedure that is designed to generate a linear array of unique sequences. In addition, the nucleic acid assembly may be designed to produce a large number of different vector inserts each

encoding a plurality of unique RNA sequences that are not repeated in any of the other different vector inserts. However, it should be appreciated that multiple copies of each different vector insert may be produced in order to clone the inserts into the vectors and/or in order to transform the host cells. The number of different vector inserts that are designed and assembled may be a function of the expected number of transformants. For example, if a host system can generate up to  $10^{10}$ ,  $10^{12}$ ,  $10^{14}$  or more different transformants, the number of different unique vector inserts should be similar or higher. It should be appreciated that if each insert encodes 100 unique RNA sequences, then a library will encode a number of different RNA molecules that is 100 times the number of transformants.

The distribution of different RNA sequences across the library may be random or systematic depending on the design. In some embodiments, the RNAs expressed on one vector may differ from each other by 1-5 nucleotide substitutions. However, in some embodiments, RNAs encoded on one DNA insert may have sequences that differ from each other by about 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190 or 200 more nucleotide substitutions. A library may not sample all different sequence variants that are possible for an RNA of a predetermined length. The sequence variants that are assembled may be determined at the design stage based on one or more factors that could include design and assembly considerations and/or any information that may suggest that certain sequence variants are more likely to result in structural or functional properties of interest.

In some embodiments, a library may be assembled to include a plurality of identical or similar RNA sequences, and additional sequence variation may be introduced using mutagenesis, error-prone PCR, or other suitable methods. However, such methods introduce sequence variations randomly and are unlikely to generate as much sequence variation as a procedure that involves a design stage at which each unique RNA sequence may be predetermined.

In aspects of the invention, nucleic acids encoding RNA molecules may be cloned into vectors. A vector may be any suitable vector. For example, a vector may be a plasmid, a cosmid, a phagemid, a BAC, a YAC, an F factor, or any other suitable prokaryotic, eukaryotic or viral vector. A vector may include an origin of replication and/or one or more selectable markers (e.g., antibiotic resistant markers, etc.) and/or detectable markers (e.g., fluorescent markers, etc.). In some embodiments, a vector may be a shuttle vector that is functional in two or more different types (e.g., species) of host cells.

In aspects of the invention, vectors or expression systems may be transfected or transformed into a cell or other system capable of transcribing the RNA molecules of the invention. A host cell may be prokaryotic (e.g., bacterial such as *E. coli* or *B. subtilis*) or eukaryotic (for example a yeast, mammal, insect, or other eukaryotic cell).

5 In aspects of the invention, a modified RNA polymerase that incorporates one or more modified ribonucleotides (e.g., 2'-O-methyl ribonucleotides) that may stabilize RNA molecules could be expressed in the host cell.

In certain embodiments, a population of cells may be grown under conditions suitable for the expression of the RNA molecules of the invention. Such conditions may involve  
10 providing a suitable nutrient medium to allow growth and proliferation of the cells. The nutrient medium may contain any of the following in an appropriate combination: isotonic saline, buffer, amino acids, serum or serum replacement, and other exogenously added factors. In some embodiments, the nutrient medium may contain one or more drugs, such as antibiotics, used for selection of a cell having a particular characteristic. In some  
15 embodiments the nutrient medium is serum free. Nutrient medium is commercially available from sources such as Life Technologies (Gaithersburg, MD).

In certain embodiments, a nucleic acid encoding different RNA molecules may be integrated into the host cell genome.

In some embodiments, a population of transformed host cells can produce many  
20 different unique RNA molecules. In some embodiments, at least  $10^8$ ,  $10^{10}$ ,  $10^{12}$ ,  $10^{13}$ ,  $10^{14}$ ,  $10^{15}$ ,  $10^{16}$ ,  $10^{17}$ ,  $10^{18}$ ,  $10^{19}$ , or  $10^{20}$  or more different unique RNA molecules may be transcribed.

#### *Selection and screening:*

25 In aspects of the invention, a library of transcribed RNA molecules may be subjected to a screen or selection to identify one or more RNA molecules having a structural and/or functional property of interest. The presence of an RNA of interest in an intracellular library of transcribed RNA molecules may be determined directly or indirectly. In some  
embodiments, the presence of an RNA of interest may be detected directly if the desired  
30 function can be directly screened or selected for. For example, if an enzymatic function is desired, a screen or selection may be based on the presence or absence of the enzymatic properties of interest. Such an assay may be an *in vivo* assay. However, in some  
embodiments, an *in vitro* assay may be performed on cell extracts. In some embodiments, the presence of an RNA that binds to a ligand with high affinity and/or specificity may be

detected directly if the binding to the ligand results in a detectable signal (e.g., an increase or decrease in fluorescence intensity). For example, an RNA aptamer bound to malachite green may fluoresce whereas the dye alone does not fluoresce. In other embodiments, a fluorescent ligand or effector may be used and the assay to detect an RNA aptamer that binds to the  
5 ligand or effector may involve detecting quenching of the fluorescent signal associated with aptamer binding. In some embodiments, the ligand or effector may be toxic and RNA aptamer binding may lower the toxicity. In certain embodiments, an RNA that cleaves or modifies an effector molecule may be detected if cleavage or modification alters a detectable or selectable property of the ligand or effector.

10 In some aspects, RNA (e.g., an RNA aptamer) binding to a ligand may not be readily detectable using a direct detection technique. In some embodiments, RNA binding to a ligand may be detected indirectly if the candidate RNA is fused to a predetermined reporter RNA domain and binding of the candidate RNA to a ligand affects the structure and  
15 properties of the reporter domain to an extent that can be detected using one or more different readouts. A reporter domain may be a riboregulator or switch domain that changes conformation to either expose or sequester an antisense sequence when a ligand binds to the candidate domain. Accordingly, if the candidate RNA is an aptamer that specifically binds a ligand, the readout could be any detectable or selectable phenotype that can be regulated by antisense technology. According to the invention, any detectable or selectable phenotype  
20 may be used. For example, a readout may be drug resistance or susceptibility (e.g., antibiotic resistance or susceptibility), one or more detectable cell surface properties, a change in fluorescence intensity, auxotrophy, or one or more anabolic or catabolic phenotypes. It should be appreciated that a reporter domain may be fused to each candidate RNA transcribed in a library. In some embodiments, a DNA encoding the reporter RNA may be  
25 fused to each of the DNAs encoding the different RNA candidates in the library so that each candidate is transcribed along with a reporter domain. A DNA encoding a reporter RNA domain may be fused at the 3' end or the 5' end of each DNA encoding a candidate RNA, and accordingly transcribed candidate RNAs may have a reporter RNA at either their 3' or 5' end. In some embodiments, a reporter RNA may be fused at both the 3' and 5' ends. The  
30 reporter domains fused at the 3' and 5' ends may control different readouts. In some embodiments, different groups of candidate RNAs in the transcribed RNA library may be fused to different reporter RNAs. In some embodiments, a reporter RNA domain may be an enzyme that can be disrupted by ligand binding to an adjacent aptamer domain. In some

embodiments, a reporter RNA domain may be a protein binding domain that can be disrupted by ligand binding to an adjacent aptamer domain.

Accordingly, in some embodiments each nucleic acid sequence expressing an RNA molecule has a different reporter system. In certain embodiments, two or more nucleic acid sequences have the same reporter system. In some embodiments, the reporter system is the system disclosed by Smolke et al. (2005, *Nature Biotechnology*, 23(3):337-343), the entire contents of which are incorporated herein by reference. For example, a ligand responsive riboregulator may be used to regulate the expression of any target transcript in response to any ligand. An example of such a construct may be a riboregulator having an antisense domain that controls gene expression and an aptamer domain that recognizes specific effector ligands. Ligand binding induces a conformational change in the molecule that allows the antisense domain to interact with a target mRNA and inhibit or reduce translation. As an example, the aptamer may bind a xanthine derivative, theophylline, causing a conformational change allowing the antisense domain to interact with the mRNA encoding green fluorescent protein (GFP).

In certain embodiments, the reporter system may be a yeast three-hybrid system such as that disclosed by SenGupta D. J. et al. (1996, *Proc. Natl. Acad. Sci. USA*, 93:8496-8501), the entire contents of which are incorporated herein by reference. For example, a hybrid protein containing a DNA-binding domain (for example LexA) with RNA-binding domain 1 localizes to the promoter of an appropriate reporter gene. A second hybrid protein containing a transcriptional activation domain with RNA binding domain 2 activates transcription of the reporter gene when in close proximity to the gene's upstream regulatory sequences. A hybrid RNA containing sites recognized by the two RNA-binding proteins links the two hybrid proteins to one another and the complex results in detectable expression of the reporter gene.

Accordingly, a reporter domain may be any domain that is sensitive to (e.g., can be disrupted by) a ligand binding to an aptamer sequence that is fused to the reporter domain. As a result of ligand binding, the readout of mediated by the reporter domain may involve any detectable or selectable direct or indirect phenotype. The reporter may act via one or more protein, RNA, DNA, and/or other domains to produce a readout. Accordingly, an RNA reporter domain may be a ribozyme, an RNA switch, an antisense RNA, an allosteric effector RNA, an RNA that regulates the expression or activity of another RNA molecule, or an RNA that binds to a detectable compound. Therefore, the reporter domain also may be an aptamer domain.

*RNA identification:*

In some embodiments, if each cell contains only one type of RNA candidate molecule, the isolation of a cell that has a selected or screened for phenotype provides the identify of the RNA having a desired structure or function (e.g., enzymatic activity, binding affinity, etc.). The nucleic acid encoding the transcribed RNA may be isolated and sequenced. However, in embodiments where each cell contains a plurality of different RNA candidates, the isolation of a cell having a selected or screened for phenotype only narrows the identity of the targeted RNA down to one of the different RNAs that are transcribed in that cell. In some embodiments, the RNA with the desired structural and/or functional properties may be identified by independently testing each of the different RNAs that are transcribed in the cell. The RNAs may be tested by cloning each one and transcribing them and assaying them individually *in vivo*. In some embodiments, individual RNAs may be synthesized or assembled and tested *in vivo* or *in vitro*. It should be appreciated that other techniques may be used to identify the RNA of interest. In some embodiments, a cell that is isolated as having a desired phenotype may contain a set of RNA coding sequences that is enriched for one or a few variants. In some embodiments, regardless of the number of different transcribed RNAs in the isolated cell, further rounds of selection and or screening may be performed to enrich host cells for RNAs that have the desired properties. Repeated selection and/or screening may favor cells that have more copies of the RNA of interest relative to other transcribed RNA variants (e.g., due to gene conversion or other process that results in the RNA of interest spreading across the set of transcribed RNAs).

*Aptamer expressing cells:*

Aspects of the invention relate to cells capable of transcribing a plurality of different aptamers. In some embodiments, a plurality of aptamers may be preselected by their ability to bind to one or more different molecules of interest (e.g., one or more different ligands or effector molecules).

In some embodiments, a plurality of different aptamers may be transcribed by a single cell line. In certain embodiments, each cell expresses at least 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more aptamers. In some embodiments, the transcribed aptamers are all different. In other embodiments, the transcribed aptamers may include one or more copies of the same aptamer.

In some embodiments, the transcription of one or more aptamers may be under the control of the same promoter. In certain embodiments, transcription of one or more aptamers

molecules may be under the control of separate promoters. The separate promoters may be separate copies of the same promoter or different promoters. In some embodiments, one or more promoters may be inducible. In some embodiments, aptamer transcription may involve transcription enzymes of the host cell.

5 In some embodiments, transcribed aptamers may be of different lengths. In some embodiments, an aptamer may be at least 50 to at least 200 nucleotide bases long. In certain embodiments, an aptamer may be at least 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200 nucleotide bases long or longer. However, certain aptamers may be shorter than 50 bases  
10 long (e.g., between about 10 and about 50 bases long). In some embodiments, each transcribed aptamer may be of a different length.

In some embodiments, certain aptamers may be transcribed as a single RNA chain. A single transcribed RNA may include two or more aptamers. In some embodiments, a single transcribed RNA may include 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80,  
15 85, 90, 95, or 100 aptamers. The single RNA transcript may include one or more cleavage sites that can be acted on to release one or more different aptamers from the RNA transcript. In some embodiments, one or more enzymes may cut the cleavage sites to release individual aptamers. In some embodiments, the cleavage sites may be autocatalytic RNA cleavage sites. In other embodiments, aptamers may be transcribed as individual transcripts. In certain  
20 embodiments, a plurality of aptamers may be transcribed in a combination of individual aptamer transcripts and RNA transcripts that include two or more aptamers.

In certain embodiments, one or more aptamer coding sequences may be integrated into the genome of a host cell.

In aspects of the invention, an aptamer may be transcribed fused to a reporter RNA.  
25 The reporter RNA may produce a signal (either directly or indirectly) if the aptamer binds to its ligand. In aspects of the invention, an aptamer readout using a reporter RNA could be drug resistance or susceptibility, a cell surface property, a change in fluorescence intensity, auxotrophy, or other anabolic or catabolic phenotypes.

Methods described herein may be used to obtain readout and/or regulatory aptamers  
30 for any ligand (e.g., one or more metabolites or other molecules, including, for example, environmental contaminants, toxins, minerals, ores, etc., such as those described herein, or any combination of two or more thereof). It should be appreciated that an aptamer that provides a transcriptional or translational switch in response to a ligand as described herein may be used for readout or regulatory (e.g., via controlling the expression of a gene)



applications described herein. In some embodiments, an aptamer region that is responsive to a ligand (e.g., by changing configuration as described herein) may be used for both readout and regulatory purposes.

5           Example 3. Multiplex Nucleic Acid Assembly

Aspects of the invention may involve one or more nucleic acid assembly reactions in order to make the sets of genetic elements and recombination sites, the modified host cells, the aptamers, and/or other nucleic acids that may be used to generate biological diversity and screen or select for one or more functions of interest.

10           Aspects of the invention involve assembling nucleic acids that contain one or more components of a metabolic pathway. Aspects of the invention involve assembling nucleic acids that can be used to modify the genome of a host cell. For example, the genome of a host cell may be reduced in size (e.g., by 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, or more) in order to accommodate nucleic acids that encode components of an

15           engineered metabolic pathway. Nucleic acids of the invention may be assembled using any suitable method including a combination of one or more ligation, recombination, or extension reactions. Multiplex nucleic acid assembly reactions may be used to assemble one or more nucleic acid components. Multiplex nucleic acid assembly relates to the assembly of a plurality of nucleic acids to generate a longer nucleic acid product. In one aspect, multiplex

20           oligonucleotide assembly relates to the assembly of a plurality of oligonucleotides to generate a longer nucleic acid molecule. However, it should be appreciated that other nucleic acids (e.g., single or double-stranded nucleic acid degradation products, restriction fragments, amplification products, naturally occurring small nucleic acids, other polynucleotides, etc.) may be assembled or included in a multiplex assembly reaction (e.g., along with one or more

25           oligonucleotides) in order to generate an assembled nucleic acid molecule that is longer than any of the single starting nucleic acids (e.g., oligonucleotides) that were added to the assembly reaction. In certain embodiments, one or more nucleic acid fragments that each were assembled in separate multiplex assembly reactions (e.g., separate multiplex oligonucleotide assembly reactions) may be combined and assembled to form a further

30           nucleic acid that is longer than any of the input nucleic acid fragments. In certain embodiments, one or more nucleic acid fragments that each were assembled in separate multiplex assembly reactions (e.g., separate multiplex oligonucleotide assembly reactions) may be combined with one or more additional nucleic acids (e.g., single or double-stranded nucleic acid degradation products, restriction fragments, amplification products, naturally

occurring small nucleic acids, other polynucleotides, etc.) and assembled to form a further nucleic acid that is longer than any of the input nucleic acids.

In aspects of the invention, one or more multiplex assembly reactions may be used to generate target nucleic acids having predetermined sequences. In one aspect, a target nucleic acid may have a sequence of a naturally occurring gene and/or other naturally occurring nucleic acid (e.g., a naturally occurring coding sequence, regulatory sequence, non-coding sequence, chromosomal structural sequence such as a telomere or centromere sequence, etc., any fragment thereof or any combination of two or more thereof). In another aspect, a target nucleic acid may have a sequence that is not naturally-occurring. In one embodiment, a target nucleic acid may be designed to have a sequence that differs from a natural sequence at one or more positions. In other embodiments, a target nucleic acid may be designed to have an entirely novel sequence. However, it should be appreciated that target nucleic acids may include one or more naturally occurring sequences, non-naturally occurring sequences, or combinations thereof.

In one aspect of the invention, multiplex assembly may be used to generate libraries of nucleic acids having different sequences. In some embodiments, a library may contain nucleic acids having random sequences. In certain embodiments, a predetermined target nucleic acid may be designed and assembled to include one or more random sequences at one or more predetermined positions.

In certain embodiments, a target nucleic acid may include a functional sequence (e.g., a protein binding sequence, a regulatory sequence, a sequence encoding a functional protein, etc., or any combination thereof). However, some embodiments of a target nucleic acid may lack a specific functional sequence (e.g., a target nucleic acid may include only non-functional fragments or variants of a protein binding sequence, regulatory sequence, or protein encoding sequence, or any other non-functional naturally-occurring or synthetic sequence, or any non-functional combination thereof). Certain target nucleic acids may include both functional and non-functional sequences. These and other aspects of target nucleic acids and their uses are described in more detail herein.

A target nucleic acid may be assembled in a single multiplex assembly reaction (e.g., a single oligonucleotide assembly reaction). However, a target nucleic acid also may be assembled from a plurality of nucleic acid fragments, each of which may have been generated in a separate multiplex oligonucleotide assembly reaction. It should be appreciated that one or more nucleic acid fragments generated via multiplex oligonucleotide assembly also may be combined with one or more nucleic acid molecules obtained from another source (e.g., a

restriction fragment, a nucleic acid amplification product, etc.) to form a target nucleic acid. In some embodiments, a target nucleic acid that is assembled in a first reaction may be used as an input nucleic acid fragment for a subsequent assembly reaction to produce a larger target nucleic acid.

5           Accordingly, different strategies may be used to produce a target nucleic acid having a predetermined sequence. For example, different starting nucleic acids (e.g., different sets of predetermined nucleic acids) may be assembled to produce the same predetermined target nucleic acid sequence. Also, predetermined nucleic acid fragments may be assembled using one or more different *in vitro* and/or *in vivo* techniques. For example, nucleic acids (e.g.,  
10   overlapping nucleic acid fragments) may be assembled in an *in vitro* reaction using an enzyme (e.g., a ligase and/or a polymerase) or a chemical reaction (e.g., a chemical ligation) or *in vivo* (e.g., assembled in a host cell after transfection into the host cell), or a combination thereof. Similarly, each nucleic acid fragment that is used to make a target nucleic acid may be assembled from different sets of oligonucleotides. Also, a nucleic acid fragment may be  
15   assembled using an *in vitro* or an *in vivo* technique (e.g., an *in vitro* or *in vivo* polymerase, recombinase, and/or ligase based assembly process). In addition, different *in vitro* assembly reactions may be used to produce a nucleic acid fragment. For example, an *in vitro* oligonucleotide assembly reaction may involve one or more polymerases, ligases, other suitable enzymes, chemical reactions, or any combination thereof.

20

#### Example 4. Business Applications

Aspects of the invention may be useful to generate engineered metabolic pathways, components thereof, related engineered cells, nucleic acid libraries that represent very large numbers of nucleic acid sequence variants (e.g., RNA candidates for an aptamer screen)  
25   nucleic acid assembly reactions, etc., or combinations thereof. Accordingly, aspects of the invention relate to marketing methods, compositions, kits, devices, and systems for generating engineered metabolic pathways, components thereof, related engineered cells, nucleic acid libraries that represent very large numbers of nucleic acid sequence variants, methods and compositions for *in vivo* aptamer screening and selection, methods and  
30   compositions for identifying, monitoring, and generating metabolic pathways, and methods for designing and assembling libraries as described herein.

Aspects of the invention may be useful for reducing the time and/or cost of production, commercialization, and/or development of engineered metabolic pathways, related synthetic nucleic acids, and/or related compositions. Accordingly, aspects of the

invention relate to business methods that involve collaboratively (e.g., with a partner) or independently marketing one or more methods, kits, compositions, devices, or systems for analyzing and/or assembling engineered metabolic pathways, obtaining related libraries, and identifying aptamers *in vivo* as described herein. For example, certain embodiments of the invention may involve marketing a procedure and/or associated devices or systems involving techniques and assays described herein. In some embodiments, synthetic nucleic acids, libraries of synthetic nucleic acids, host cells containing synthetic nucleic acids, expressed polypeptides or proteins, etc., also may be marketed.

Marketing may involve providing information and/or samples relating to methods, kits, compositions, devices, and/or systems described herein. Potential customers or partners may be, for example, companies in the pharmaceutical, biotechnology and agricultural industries, as well as academic centers and government research organizations or institutes. Business applications also may involve generating revenue through sales and/or licenses of methods, kits, compositions, devices, and/or systems of the invention.

#### EQUIVALENTS

The present invention provides among other things methods for assembling large polynucleotide constructs and organisms having increased genomic stability. While specific embodiments of the subject invention have been discussed, the above specification is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

#### INCORPORATION BY REFERENCE

All publications, patents and sequence database entries mentioned herein, including those items listed below, are hereby incorporated by reference in their entirety as if each individual publication or patent was specifically and individually indicated to be incorporated by reference. In the event of a conflict, the disclosure and description of the present invention shall control.

We claim:

## CLAIMS

1. An engineered biological pathway in an engineered biological system comprising:  
a plurality of functional components that promote/catalyze a plurality of sequential  
5 reaction steps converting a first substrate to a first product, and  
at least one engineered readout component that provides feedback on the status of at  
least one of the plurality of sequential reaction steps, wherein the engineered readout  
component does not exist in combination with the plurality of functional components in a  
natural biological system.  
10
2. The engineered pathway of claim 1, wherein the engineered biological system is a  
recombinant organism.
3. The engineered pathway of claim 1, wherein the engineered biological system is a  
15 recombinant cell.
4. The engineered pathway of claim 3, wherein the recombinant cell is a plant cell, a  
bacterial cell, an insect cell, or a mammalian cell.
- 20 5. The engineered pathway of claim 1, wherein the plurality of functional components  
are components of a natural metabolic pathway.
6. The engineered pathway of claim 5, wherein one or more of the functional  
components are enzymes.  
25
7. The engineered pathway of claim 5, wherein one or more of the functional  
components are proteins.
8. The engineered pathway of claim 1, wherein the plurality of functional components  
30 does not exist in a natural biological system.
9. The engineered pathway of claim 8, wherein one or more of the functional  
components are recombinant molecules that do not exist in a natural biological system.

10. The engineered pathway of claim 8, wherein the plurality of functional components do not exist together in a natural biological system.

11. The engineered pathway of claim 1, further comprising one or more regulatory components.

12. The engineered pathway of claim 11, wherein one or more regulatory components are proteins.

13. The engineered pathway of claim 11, wherein one or more regulatory components are nucleic acids.

14. The engineered pathway of claim 1, wherein the engineered readout component provides a detectable signal that is indicative of the status of at least one of the plurality of sequential reaction steps.

15. The engineered pathway of claim 14, wherein the detectable signal is a fluorescent signal.

16. The engineered pathway of claim 1, wherein the engineered readout component provides a physiological response that is indicative of the status of at least one of the plurality of sequential reaction steps.

17. The engineered pathway of claim 16, wherein the physiological response is a cell growth rate change or cell death.

18. The engineered pathway of claim 1, wherein the engineered readout component is responsive to a level of a metabolite.

19. The engineered pathway of claim 18, wherein the metabolite is the substrate.

20. The engineered pathway of claim 18, wherein the metabolite is the product.

21. The engineered pathway of claim 18, wherein the metabolite is an intermediate metabolite produced in one of the sequential reaction steps.

22. The engineered pathway of claim 1, wherein the engineered biological system  
5 comprises two or more engineered readout components.

23. The engineered pathway of claim 22, wherein the engineered biological system comprises engineered readout components for two or more intermediates produced in the sequential reaction steps.

10

24. The engineered pathway of claim 22, wherein the engineered biological system comprises engineered readout components for the substrate and the product.

15

25. The engineered pathway of claim 1 or 22, wherein the engineered readout component comprises a nucleic acid or a polypeptide.

26. The engineered pathway of claim 25, wherein the nucleic acid comprises a DNA aptamer, an RNA aptamer, an RNA molecule comprising a regulatory domain and a reporter domain, or an RNA molecule comprising a binding domain and a reporter domain.

20

27. The engineered pathway of claim 25, wherein the nucleic acid comprises a ribozyme.

28. The engineered pathway of claim 25, wherein the polypeptide is an antibody.

25

29. The engineered pathway of claim 1, comprising at least two engineered readout components.

30. The engineered pathway of claim 1, comprising 2-5 engineered readout components.

30

31. The engineered pathway of claim 1, comprising 5-10 engineered readout components.

32. The engineered pathway of claim 1, comprising 10-15 engineered readout components.

33. The engineered pathway of claim 1, wherein the plurality of reaction steps synthesize a product that incorporates two or more substrates.

34. The engineered pathway of claim 1, wherein the plurality of reaction steps modify a substrate.

35. The engineered pathway of claim 1, wherein the substrate is degraded.

36. The engineered pathway of claim 1, wherein the plurality of reaction steps reduce the toxicity of a toxic substrate.

37. The engineered pathway of claim 1, wherein the product is an amino acid, a nucleotide, a nucleoside, a polypeptide, a nucleic acid, an alcohol, a carbohydrate, or other complex organic compound.

38. The engineered pathway of claim 1, wherein the product is ethanol, cellulose, or lignin.

39. The engineered pathway of claim 1, wherein the product is a photosynthetic product.

40. The engineered pathway of claim 1, wherein the substrate is toxic.

41. An engineered host cell comprising one or more engineered readout components that provide feedback on the level of one or more target molecules inside the cell or in the growth environment of the cell.

42. The engineered host cell of claim 41, wherein the one or more target molecules are toxins.

43. The engineered host cell of claim 41, wherein the one or more target molecules are environmental contaminants.



44. The engineered host cell of claim 41, further comprising a metabolic pathway that converts the one or more target molecules into one or more products that are detected by the one or more engineered readout components.

5 45. A nucleic acid preparation that encodes:  
a plurality of functional components that promote/catalyze a plurality of sequential  
reaction steps converting a first substrate to a first product, and  
at least one engineered readout component that provides feedback on the status of at  
least one of the plurality of sequential reaction steps, wherein the engineered readout  
10 component does not exist in combination with the plurality of functional components in a  
natural biological system.

46. The nucleic acid preparation of claim 45, wherein the plurality of functional  
components and the engineered readout component are encoded on a single nucleic acid  
15 molecule.

47. The nucleic acid preparation of claim 45, wherein the plurality of functional  
components and the engineered readout component are encoded on two or more nucleic acid  
molecules.

20 48. The nucleic acid preparation of claim 45, wherein the plurality of functional  
components and the engineered readout component are encoded on 2-10 nucleic acid  
molecules.

25 49. The nucleic acid preparation of any one of claims 45-48, wherein the nucleic acid  
molecules are plasmid or vector molecules.

50. The nucleic acid preparation of any one of claims 45-48, wherein one or more of the  
nucleic acid molecules are engineered nucleic acids derived from nucleic acids assembled in  
30 one or more multiplex assembly reactions, codon optimized nucleic acids, non-natural  
nucleic acids, or have less than 90%, less than 80%, or less than 70% identity with a natural  
nucleic acid.

51. An engineered biological pathway in an engineered biological system comprising:

a plurality of engineered functional components that promote/catalyze a plurality of sequential reaction steps converting a first substrate to a first product, wherein the plurality of functional components does not exist in a natural biological system, and wherein the plurality of functional components comprises at least five/ten different functional components.

5

52. The engineered pathway of claim 51, wherein one or more of the functional components are recombinant molecules that do not exist in a natural biological system.

53. The engineered pathway of claim 51, wherein the plurality of functional components  
10 do not exist together in a natural biological system.

54. The engineered pathway of claim 53, wherein the plurality of functional components comprises two or more different functional components derived from different species.

15 55. The engineered pathway of claim 51, wherein the engineered biological system is a recombinant organism.

56. The engineered pathway of claim 51, wherein the engineered biological system is a recombinant cell.  
20

57. The engineered pathway of claim 56, wherein the recombinant cell is a plant cell, a bacterial cell, an insect cell, or a mammalian cell.

58. The engineered pathway of claim 56, wherein the plurality of engineered functional  
25 components are components of a natural metabolic pathway, and wherein each of the plurality of functional components are derived from a species that is different from the recombinant cell.

59. The engineered pathway of claim 51, wherein each of the engineered functional  
30 components is expressed from a recombinant nucleic acid.

60. The engineered pathway of claim 59, wherein each of the engineered functional components is expressed from the same recombinant nucleic acid.

61. The engineered pathway of claim 59, wherein the recombinant nucleic acid is on a plasmid or other vector.

62. The engineered pathway of claim 59, wherein the recombinant nucleic acid is  
5 integrated into the genome of the host cell.

63. The engineered pathway of claim 51, wherein one or more of the engineered functional components are enzymes.

10 64. The engineered pathway of claim 51, wherein one or more of the engineered functional components are proteins.

65. The engineered pathway of claim 51, further comprising one or more engineered regulatory components.

15 66. The engineered pathway of claim 51, wherein one or more engineered regulatory components are proteins.

67. The engineered pathway of claim 51, wherein one or more engineered regulatory  
20 components are nucleic acids.

68. The engineered pathway of claim 51 or claim 65, further comprising an engineered readout component.

25 69. The engineered pathway of claim 68, wherein the engineered readout component provides a detectable signal that is indicative of the status of at least one of the plurality of sequential reaction steps.

70. The engineered pathway of claim 69, wherein the detectable signal is a fluorescent  
30 signal.

71. The engineered pathway of claim 68, wherein the engineered readout component provides a physiological response that is indicative of the status of at least one of the plurality of sequential reaction steps.

72. The engineered pathway of claim 71, wherein the physiological response is a cell growth rate change or cell death.

5 73. The engineered pathway of claim 51, comprising 10-15 engineered functional components.

74. The engineered pathway of claim 51, comprising 15-50 engineered functional components.

10

75. The engineered pathway of claim 65, comprising 2-5 engineered regulatory components.

15

76. The engineered pathway of claim 65, comprising 5-10 engineered regulatory components.

77. The engineered pathway of claim 65, comprising 10-20 engineered regulatory components.

20

78. The engineered pathway of claim 65, comprising 20-50 engineered regulatory components.

79. The engineered pathway of claim 68, comprising 2-5 engineered readout components.

25

80. The engineered pathway of claim 68, comprising 5-10 engineered readout components.

81. The engineered pathway of claim 68, comprising 10-20 engineered readout components.

30

82. The engineered pathway of claim 68, comprising 20-50 engineered readout components.

83. The engineered pathway of claim 65, wherein one or more engineered regulatory components are expressed from a recombinant nucleic acid.

84. The engineered pathway of claim 83, wherein the one or more engineered regulatory components are expressed from the same recombinant nucleic acid.

85. The engineered pathway of claim 83, wherein the recombinant nucleic acid is on a plasmid or other vector.

86. The engineered pathway of claim 83, wherein the recombinant nucleic acid is integrated into the genome of the host cell.

87. The engineered pathway of claim 68, wherein one or more engineered readout components are expressed from a recombinant nucleic acid.

88. The engineered pathway of claim 87, wherein the one or more engineered readout components are expressed from the same recombinant nucleic acid.

89. The engineered pathway of claim 87, wherein the recombinant nucleic acid is on a plasmid or other vector.

90. The engineered pathway of claim 87, wherein the recombinant nucleic acid is integrated into the genome of the host cell.

91. The engineered pathway of claim 51, wherein the engineered functional components are expressed from a nucleic acid derived from a nucleic acid assembled in a multiplex assembly reaction.

92. The engineered pathway of claim 65, wherein one or more engineered regulatory components are expressed from a nucleic acid derived from a nucleic acid assembled in a multiplex assembly reaction.

93. The engineered pathway of claim 68, wherein one or more engineered readout components are expressed from a nucleic acid derived from a nucleic acid assembled in a multiplex assembly reaction.

5 94. The engineered pathway of claim 51, comprising wherein the plurality of reaction steps synthesize a product that incorporates two or more substrates.

95. The engineered pathway of claim 51, wherein the plurality of reaction steps modify a substrate.

10

96. The engineered pathway of claim 51, wherein the substrate is degraded.

97. The engineered pathway of claim 51, wherein the plurality of reaction steps reduce the toxicity of a toxic substrate.

15

98. The engineered pathway of claim 51, wherein the product is an amino acid, a nucleotide, a nucleoside, a polypeptide, a nucleic acid, an alcohol, a carbohydrate, or other complex organic compound.

20 99. The engineered pathway of claim 51, wherein the product is ethanol, cellulose, or lignin.

100. The engineered pathway of claim 51, wherein the product is a photosynthetic product.

25 101. The engineered pathway of claim 51, wherein the substrate is toxic.

102. An engineered host cell comprising:

a plurality of engineered functional components that promote/catalyze a plurality of sequential reaction steps converting a first substrate to a first product, wherein the plurality of functional components does not exist in a natural biological system, and wherein the plurality of functional components comprises at least five/ten different functional components.

30

103. A nucleic acid preparation that encodes:

a plurality of engineered functional components that promote/catalyze a plurality of sequential reaction steps converting a first substrate to a first product, wherein the plurality of functional components does not exist in a natural biological system, and wherein the plurality of functional components comprises at least five/ten different functional components.

5

104. The nucleic acid preparation of claim 103, wherein the plurality of functional components are encoded on a single nucleic acid molecule.

105. The nucleic acid preparation of claim 103, wherein the plurality of functional components are encoded on two or more nucleic acid molecules.

10

106. The nucleic acid preparation of claim 103, wherein the plurality of functional components are encoded on 2-10 nucleic acid molecules.

107. The nucleic acid preparation of any one of claims 103-106, wherein the nucleic acid molecules are plasmid or vector molecules.

15

108. The nucleic acid preparation of any one of claims 103-106, wherein the nucleic acid molecules are engineered nucleic acids derived from nucleic acids assembled in one or more multiplex assembly reactions, codon optimized nucleic acids, non-natural nucleic acids, or have less than 90%, less than 80%, or less than 70% identity with a natural nucleic acid.

20

109. The engineered pathway of claim 65, comprising 2-5, at least 2, at least 3, at least 4, or at least 5 engineered regulatory components.

25

110. A host cell comprising two or more engineered biological pathways.

111. The host cell of claim 110, further comprising one or more cross-regulatory controls between components of the two pathways.

30

112. An engineered nucleic acid preparation comprising one or more nucleic acids that encode components of two or more engineered biological pathways.

113. The engineered pathway of claim 112, wherein the one or more nucleic acids encode one, two, three, four, five, or more engineered functional components.

114. The engineered pathway of claim 112, wherein the one or more nucleic acids encode one, two, three, four, five, or more engineered regulatory components.

115. The engineered pathway of claim 112, wherein the one or more nucleic acids encode one, two, three, four, five, or more engineered readout components.

116. An *in silico* method for designing an engineered biological pathway, the method comprising computer-implemented acts of:

analyzing a plurality of feedstocks;

analyzing a plurality of target products;

comparing a plurality of alternative combinations of different pathway components;

and,

identifying one or more combinations of pathway components that are predicted to convert one or more feedstocks into one or more products.

117. A method for identifying an engineered biological pathway, the method comprising:

providing a plurality of systems having alternative combinations of different pathway components;

assaying the plurality of systems for the production of one or more products of interest using one or more members of a plurality of feedstocks; and,

identifying one or more systems having a combination of pathway components that convert one or more feedstocks into one or more predetermined products with at least a threshold level of efficiency.

118. The method of claim 116, wherein information about one or more of the pluralities of feedstocks, target products, or pathway components is stored in an electronically accessible database.

119. The method of claim 117, wherein the plurality of systems comprise a library of nucleic acids that encode the alternative combinations of different pathway components.



120. The method of claim 117, wherein levels of one or more predetermined products are assayed using one or more readout components.

5 121. The method of claim 120, wherein one or more of the readout components are a nucleic acid, a polypeptide, a DNA aptamer, an RNA aptamer, an RNA molecule comprising a regulatory domain and a reporter domain, or an RNA molecule comprising a binding domain and a reporter domain.

10 122. A method of engineering a biological pathway, the method comprising,  
designing a first pathway to synthesize a target product,  
assembling a plurality of nucleic acids, each nucleic acid encoding a functional or a regulatory component of the pathway, and  
combining products encoded by the nucleic acids, thereby producing an engineered biological pathway.

15 123. A method of engineering a biological pathway, the method comprising,  
designing a first pathway to metabolize a target substrate,  
assembling a plurality of nucleic acids, each nucleic acid encoding a functional or a regulatory component of the pathway, and  
20 combining products encoded by the nucleic acids, thereby producing an engineered biological pathway.

Figure 1

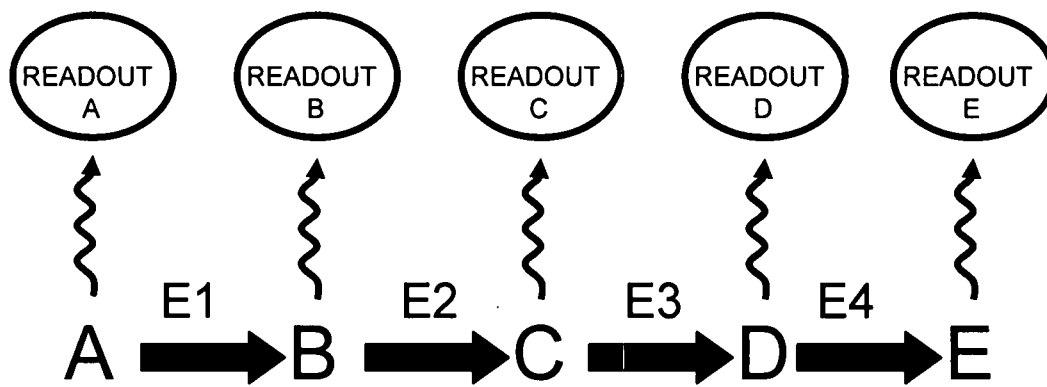


Figure 2

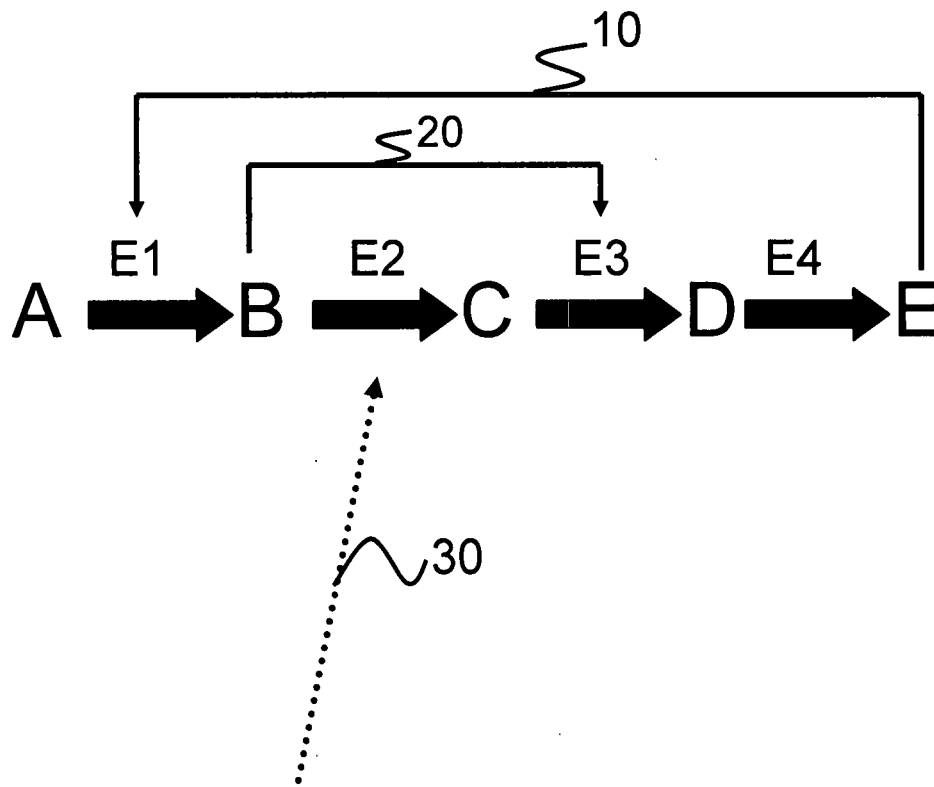


Figure 3

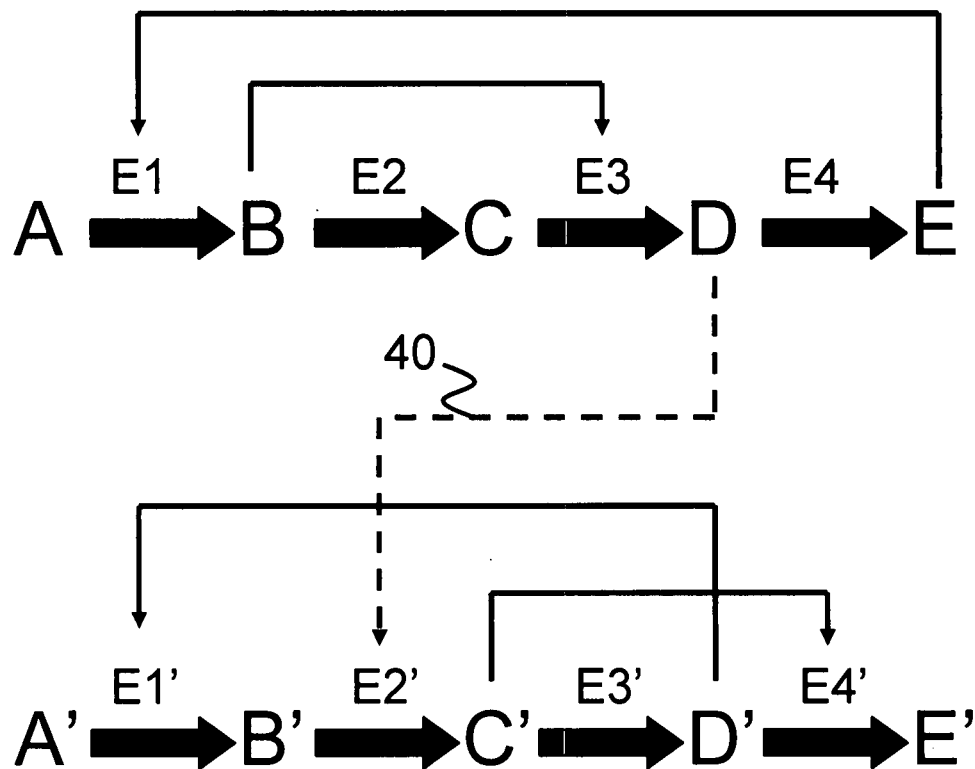


Figure 4

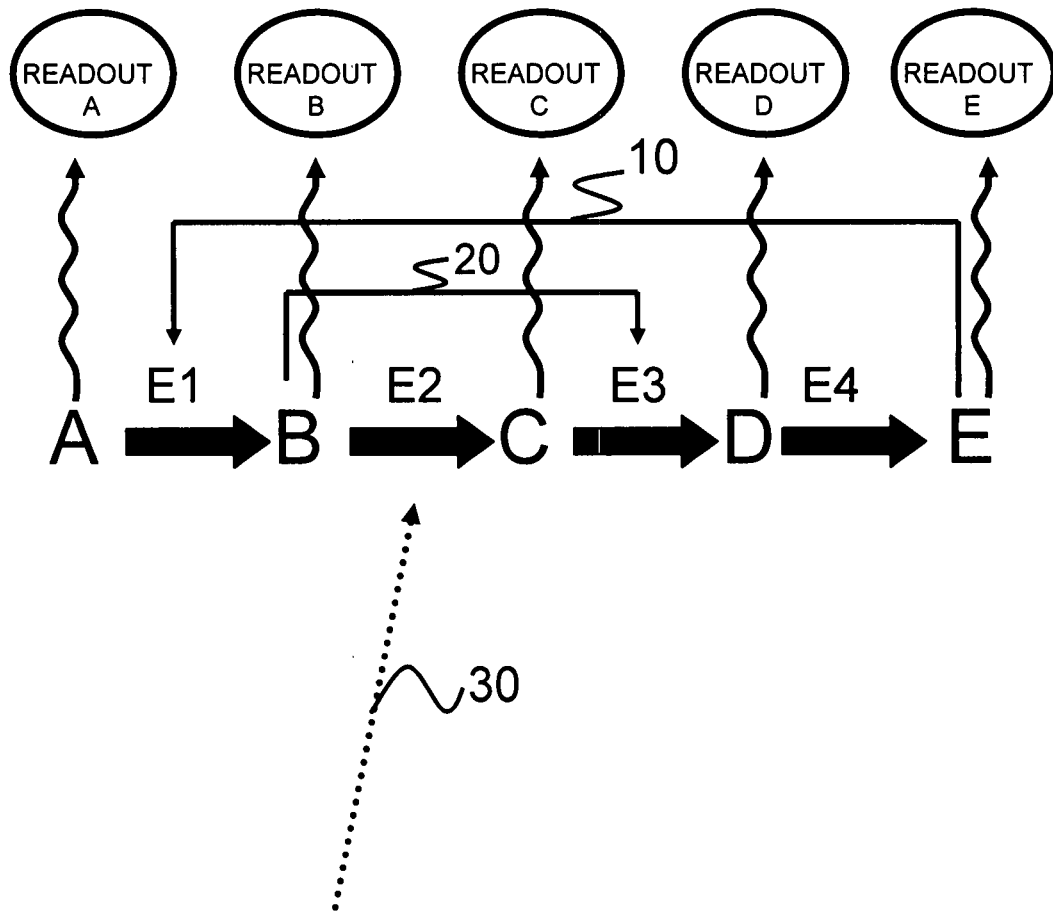


Figure 5

