(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 August 2017 (10.08.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/134602 Al

(51) International Patent Classification:
G06F 19/12 (201 1.01)

(21) International Application Number:
PCT/IB2017/050576

(22) International Filing Date:
2 February 2017 (02.02.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/291,308 4 February 2016 (04.02.2016) US

(71) Applicant: **KING ABDULLAH UNIVERSITY OF SCI¬ ENCE AND TECHNOLOGY** [SA/SA]; 4700 King Abdullah University of Science and Technology, Thuwal, 23955-6900 (SA).

(72) Inventors: **GAO, Xin;** c/o King Abdullah University of Science and Technology, 4700 King Abdullah University of Science and Technology, Thuwal, 23955-6900 (SA). **KUWAHARA, Hiroyuki;** c/o King Abdullah University of Science and Technology, 4700 King Abdullah University of Science and Technology, Thuwal, 23955-6900 (SA). **ALAZMI, Meshari Saud;** c/o King Abdullah University of Science and Technology, 4700 King Abdullah University of Science and Technology, Thuwal, 23955-6900 (SA). **CUI, Xuefeng;** c/o King Abdullah University of Science and Technology, 4700 King Abdullah University of Science and Technology, Thuwal, 23955-6900 (SA).

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

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

Declarations under Rule 4.17 :

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.1 7(H))*

[Continued on next page]

(54) Title: METHOD FOR DETERMINING HETEROLOGOUS BIOSYNTHESIS PATHWAYS

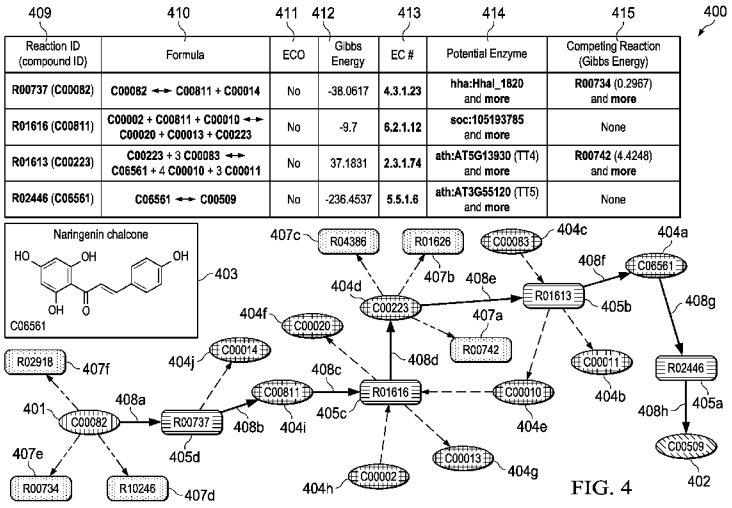


FIG. 4

(57) Abstract: The present invention relates to a method and system for dynamically analyzing, determining, predicting and displaying ranked suitable heterologous biosynthesis pathways for a specified host. The present invention addresses the problem of finding suitable pathways for the endogenous metabolism of a host organism because the efficacy of heterologous biosynthesis is affected by competing endogenous pathways. The present invention is called MRE (Metabolic Route Explorer), and it was conceived and developed to systematically and dynamically search for, determine, analyze, and display promising heterologous pathways while considering competing endogenous reactions in a given host organism.

WO 2017/134602 A1

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(Hi))*

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

METHOD FOR DETERMINING HETEROLOGOUS BIOSYNTHESIS PATHWAYS

RELATED APPLICATION DATA

[0001] This application claims the benefit of U.S. Provisional Patent Application Serial No. 62/291,308 filed February 4, 2016.

TECHNICAL FIELD

[0002] The present invention relates to a method and system for analyzing, determining, predicting and displaying ranked suitable heterologous biosynthesis pathways for a specified host.

BACKGROUND OF THE INVENTION

[0003] High-value natural products can be constructed through biosynthesis using recent advances in genome editing and metabolic engineering. Known methods and systems for graphically displaying biosynthesis pathways for natural product construction, for the most part, simply provide for the display of a selection of certain data on a graphical user interface. These prior art graphical systems fail to account for essential analytical functionality of host parameters that is needed to accurately calculate biosynthesis pathways in a high speed system with accuracy and enhanced usability.

[0004] Many design decisions must be made when analyzing possible biosynthesis pathways for a natural product, but prior art graphical display programs do not adequately account for several key decisions, such as the problems associated with foreign gene introduction into a host organism and the suitability of pathways for the endogenous metabolism of a host organism. Specifically, one design decision that may be made for engineering of heterologous biosynthesis systems concerns the decision of which foreign metabolic genes to introduce into a given host organism. The introduction of foreign metabolic genes into the biosynthesis analysis is a decision that must be made based on multifaceted factors, such as the suitability of pathways for the endogenous metabolism of a host organism, in part because the efficacy of heterologous biosynthesis is affected by competing endogenous pathways.

[0005] Known graphical user display systems do not accurately calculate biosynthesis pathways considering this suitability of pathways for the endogenous metabolism of a host

organism to maximize speed of the system with accuracy and enhanced usability, which means known systems are not as accurate as possible concerning the design decision of introduction of foreign metabolic genes into a given host organism.

[0006] For instance, several known graphical display systems do not allow the user to specify a host organism in the determination of pathways of construction for a natural product using biosynthesis, such as the graphical systems known as BNICE, PredPath and Metabolic tinker, which were developed to explore pathways irrespective of the consideration for host organisms.

[0007] Table la. Graphical Display Systems.

Tool	Access	Chassis	Chemical transformation	Thermodynamic consideration	Ranking score	Information given for each pathway	(Ref.)
BNICE	Closed access	No host	Predicted reactions	No	No pathway ranking	3-level EC number for each predicted chemical transformation	7
PathPred	Open access web server	No host	Predicted reactions	No	Chemical similarity	Final compound of biodegradation, predicted intermediates and reactions, confidence for each predicted reaction	5
Metabolic tinker	Open access web server	No host	RHEA reactions	Directionality, favorability	Net favorability	Possible reactions for each chemical transformation step and net favorability	4

[0008] These graphical systems cannot assess the suitability of pathways in a specific context without appropriately considering the introduction foreign metabolic genes into a given host organism or considering the endogenous metabolic system of a host organism. Therefore, these prior art graphical display systems fail to account for essential functionality that is needed to accurately calculate suitability of biosynthesis pathways in a high speed system with accuracy and enhanced usability.

[0009] Several other known graphical display systems do not adequately analyze the basis for the chemical transformation of intermediate precursors that form metabolic routes in a pathway display, such as the FMM, DESHARKY and Metabolic tinker display systems (in Table la, above), which specify chemical transformation using metabolic reaction sets from databases. Through the use of metabolic reaction sets from databases, these display systems that do not

adequately consider the basis for chemical transformation of intermediate precursors that forms metabolic routes.

[0010] Table lb. Graphical Display Systems.

Tool	Access	Chassis	Chemical transformation	Thermodynamic consideration	Ranking score	Information given for each pathway	(Ref.)
FMM	Open access web server	Many choices	KEGG reactions	No	Number of reaction steps	EC numbers for enzymes, availability of each enzyme in various host organisms, suggestion for foreign enzymes	1
DESHARKY	Free download	*E. coli*	KEGG reactions	No	Growth rate	Source or target compound, EC numbers for enzymes, genes for some foreign enzymes, growth rate reduction measures	6

[0011] The above display systems do not adequately consider the basis for chemical transformation of intermediate precursors that form metabolic routes, but instead consider only reaction sets from databases. For that reason, these graphical display systems cannot adequately assess the suitability of pathways in a specific context without appropriately considering the introduction foreign metabolic genes into a given host organism or considering the endogenous metabolic system of a host organism. Therefore, these prior art graphical display systems fail to account for essential functionality that is needed to accurately calculate suitability of biosynthesis pathways in a high speed system with accuracy and enhanced usability.

[0012] Other known graphical display systems do not adequately analyze the basis for the chemical transformation of intermediate precursors that form metabolic routes in a pathway display, such as graphical display systems that include BNICE (in Table la, above), PredPath (in Table la, above) and XTMS, which merely predict some generalized chemical transformation rules using such curated reaction sets and apply those generalized rules to expand potentially feasible metabolic routes.

[0013] Table lc. Graphical Display Systems

Tool	Access	Chassis	Chemical transformation	Thermodynamic consideration	Ranking score	Information given for each pathway	(Ref.)
XTMS	Open access web server	*E. coli*	Predicted reactions	Favorability	Gene scores, reaction steps, toxicity, yield, Gibbs energy	Source compound for the retrosynthesis path, predicted reactions with EC numbers, genes for foreign enzymes, toxicity, production yield	3

[0014] Because these graphical systems only consider curated reaction sets and use generalized rules to expand on possible routes, these graphical display systems cannot adequately assess the suitability of pathways in a specific context without appropriately considering the introduction foreign metabolic genes into a given host organism or considering the endogenous metabolic system of a host organism. Therefore, these prior art graphical display systems fail to account for essential functionality that is needed to accurately calculate suitability of biosynthesis pathways in a high speed system with accuracy and enhanced usability.

[0015] Other known graphical display systems do not adequately analyze the basis for the chemical transformation of intermediate precursors that form metabolic routes in a pathway display, such as Metabolic tinker (in Table la, above) and XTMS (in Table lc, above), which use thermodynamic data to constrain the reaction directionality or to rank pathways based on their net favorability. These systems do not adequately consider competing endogenous reactions; and, therefore, these graphical display systems cannot adequately assess the suitability of pathways in a specific context without appropriately considering the introduction of foreign metabolic genes into a given host organism or considering the endogenous metabolic system of a host organism. Therefore, these prior art graphical display systems fail to account for essential functionality that is needed to accurately calculate suitability of biosynthesis pathways in a high speed system with accuracy and enhanced usability.

[0016] Some graphical display systems allow for the consideration of one specific host organism in the analysis, such as the display systems that restrict the user to consider *Escherichia coli* as a host organism. Graphical display systems that restrict the user to consider *Escherichia coli* as a host organism are based on flux balance analysis (FBA), such as XTMS (in Table la, above), DESHARKY (in Table la, above), OptStrain and GEM-Path, are specific

to the *Escherichia coli* chassis. While FBA-based tools tend to offer certain information to evaluate *de novo* pathways, these systems demand detailed knowledge of a given metabolic system with tight reaction-flux boundaries in order to identify meaningful steady-state flux distributions among a large number of candidate solutions.

[0017] Such detailed data are only available for well-studied organisms, and this may be a major reason why FBA-based tools focus exclusively on the pathway design in *E. coli.* Because these graphical display systems are restricted in the type of host organism to be evaluated, these graphical display systems cannot adequately assess the suitability of pathways in a specific context without appropriately considering the introduction of foreign metabolic genes into a given host organism or considering the endogenous metabolic system of a host organism. Therefore, these prior art graphical display systems fail to account for essential functionality that is needed to accurately calculate suitability of biosynthesis pathways in a high speed system with accuracy and enhanced usability.

[0018] Some other graphical display systems, such as FMM and PHT, allow the user to select a host organism from a large set of choices, but these graphical display systems do not use the chassis information to rank suitable biosynthesis pathways for a given endogenous metabolic system. Instead, the PHT display system just reports and displays which enzymes are not natively available in the host, and the FMM display system suggests the introduction of foreign enzymes for certain reactions in heterologous pathways.

[0019] Table Id. Graphical Display Systems

Tool	Access	Chassis	Chemical transformation	Thermodynamic consideration	Ranking score	Information given for each pathway	(Ref.)
FMM	Open access web server	Many choices	KEGG reactions	No	Number of reaction steps	EC numbers for enzymes, availability of each enzyme in various host organisms, suggestion for foreign enzymes	1
PHT	Open access web server	Many choices	KEGG reactions	No	Number of reaction steps	EC numbers for enzymes, local and global compound similarities for each reaction step	2

[0020] Because these systems do not use the chassis information to rank suitable biosynthesis pathways for a given endogenous metabolic system, these graphical display systems cannot adequately assess the suitability of pathways in a specific context without appropriately considering the introduction of foreign metabolic genes into a given host organism or

considering the endogenous metabolic system of a host organism. Therefore, these prior art graphical display systems fail to account for essential functionality that is needed to accurately calculate suitability of biosynthesis pathways in a high speed system with accuracy and enhanced usability.

[0021] Overall, known methods and systems for displaying biosynthesis pathways for natural products, for the most part, simply select and display data for disclosure on a graphical user interface, but these known systems do not accurately or adequately analyze pathways for biosynthesis by properly considering introduction of foreign metabolic genes into a given host organism or the endogenous metabolic system of a host organism. These known prior art display systems: (1) do not specify host organisms at all, or (2) do not analyze the basis for the chemical transformation of intermediate precursors that form metabolic routes in a pathway display, or (3) predict some generalized chemical transformation rules using such curated reaction sets and apply them to expand potentially feasible metabolic routes, or (4) restrict the user to use one specific host organism, use thermodynamic data to constrain the reaction directionality or to rank pathways based on their net favorability, which does not consider competing endogenous reactions, or (5) do not use chasis information to rank suitable biosynthesis pathways for a given endogenous metabolic system.

[0022] All of these known graphical display systems cannot adequately assess the suitability of pathways in a specific context without appropriately considering the introduction of foreign metabolic genes into a given host organism or considering the endogenous metabolic system of a host organism. For the above reasons, these prior art graphical display systems fail to account for essential functionality that is needed to accurately calculate suitability of biosynthesis pathways in a high speed system with accuracy and enhanced usability.

SUMMARY OF THE INVENTION

[0023] The present invention relates to a method and system for analyzing, determining, predicting and displaying ranked suitable heterologous biosynthesis pathways for a specified host. The present invention addresses the problem of finding suitable pathways for the endogenous metabolism of a host organism because the efficacy of heterologous biosynthesis is affected by competing endogenous pathways. The present invention is called MRE (Metabolic Route Explorer), and it was conceived and developed to systematically and dynamically search for, determine, analyze, and display promising heterologous pathways while considering competing endogenous reactions in a given host organism.

Table le. Feature summary of present invention.

Tool	Access	Chassis	Chemical transformation	Thermodynamic consideration	Ranking score	Information given for each pathway
MRE	Open access web server	Many choices	Verified KEGG reactions	Boltzmann factor	Fraction of conversions via normalized Boltzmann weights	Required metabolites, EC numbers for enzymes, genes for foreign enzymes, reaction free energy, competing native reactions

[0024] Unlike known prior art display systems, the present invention Metabolic Route Explorer (MRE) disclosed herein focuses on the suggestion of foreign enzymes with well-characterized activities for promising heterologous pathways by taking into account the effects of the existing, endogenous metabolic infrastructure of a host organism. To find promising biosynthesis routes from a large number of potential candidates, thermodynamic data offer useful information. Unlike some other existing pathway display systems, such as Metabolic tinker and XTMS (which use thermodynamic data to constrain the reaction directionality or to rank pathways based on their net favorability, which does not consider competing endogenous reactions), the present invention MRE system uses thermodynamic data to rank pathways in a host-dependent manner from the perspective of the integration of new reactions into the endogenous metabolic system.

[0025] In order to suggest actual foreign enzymes for the design of heterologous biosynthesis pathways, the present invention MRE only considers verified reactions as metabolic parts. For each foreign reaction in a suggested heterologous pathway, present invention MRE generates information about endogenous reactions competing for metabolites. Since one effective approach to increase the productivity is to attenuate or eliminate competing reactions, MRE also offers useful insights into how to debottleneck and optimize heterologous pathways.

[0026] To rationally design a productive heterologous biosynthesis system, it is essential to consider the suitability of foreign reactions for the specific endogenous metabolic infrastructure of a host. The present invention MRE has been developed, which, for a given pair of starting and desired compounds in a given chassis organism, and dynamically ranks biosynthesis routes from the perspective of the integration of new reactions into the endogenous metabolic system.

[0027] The present invention is more than a mere "a mathematical algorithm," "a fundamental economic or longstanding commercial practice," or "a challenge in business." The present

invention is a method and system that more accurately, more comprehensively, more systematically and dynamically searches for, determines, analyzes, and displays promising heterologous pathways in the field of natural product construction while considering competing endogenous reactions in a given host organism. The claimed invention has a specific, structured graphical user interface paired with the above prescribed functionality that directly relates to the graphical user interface's structure, which resolves identified problems in the prior art display systems.

[0028] For instance, the present invention pairs its graphical user interface with its analysis programming to reduce the time for searching, analysis, and dynamic determination and display of suitable biosynthesis pathways over known prior art display systems, and the present invention achieves more accurate predictions of suitable biosynthesis pathways by adequately assessing the suitability of pathways in a specific context, appropriately considering the introduction foreign metabolic genes into a given host organism, and appropriately considering the endogenous metabolic system of a host organism. The combination of these attributes in the present invention allows researchers to more efficiently and accurately search for, determine, analyze, and display promising heterologous pathways while considering competing endogenous reactions in a given host organism.

[0029] The use of an endogenous pathway score (calculated based on one or more of the reaction weights in a given pathway for the reaction, the route from the source compound to the target product, the number of reactions that are native and foreign to the host organism and whether the reactions are endogenous to the host), specific context factors, host organism factors, and endogenous metabolic system factors are inventive concepts in the context of the present system, which allows the present invention to decrease the design cycle time-periods over known display systems by eliminating erroneous, flawed or unsuitable pathways from the display and consideration in the biosynthesis efforts. For the above reasons, the present invention is a graphical display system that properly accounts for essential factors in the biosynthesis analysis to more accurately calculate suitability of biosynthesis pathways in a high speed system with greater accuracy, enhanced usability, and dynamic displays.

BRIEF DESCRIPTION OF THE DRAWINGS

[0030] The above, and other objects and advantages of the present invention will be understood upon consideration of the following detailed description taken in conjunction with

the accompanying drawings, in which like reference characters refer to like parts throughout, and in which:

[0031] FIG. 1A is the present invention Metabolic Route Explorer (MRE) user-interface query input page showing "auto-completion" field of product ;

[0032] FIG. IB is the present invention Metabolic Route Explorer (MRE) user-interface query input page displaying fields available with "advanced options" checked;

[0033] FIG. 2 is the present invention Metabolic Route Explorer (MRE) user-interface summary page for the top-ranked biosynthesis routes;

[0034] FIG. 3 is the present invention Metabolic Route Explorer (MRE) user-interface page showing a graph of the top-ranked biosynthesis routes;

[0035] FIG. 4 is the present invention Metabolic Route Explorer (MRE) user-interface page showing a graph of the pathway-level information for a specified biosynthesis route;

[0036] FIG. 5 is the present invention Metabolic Route Explorer (MRE) user-interface page showing competing reactions for a specified biosynthesis route;

[0037] FIG. 6 is the present invention workflow diagram of the Metabolic Route Explorer (MRE);

[0038] FIG. 7 is the present invention display of an illustration of a thermodynamic favorability-based weighting scheme;

[0039] FIG. 8 is the present invention display of an illustration of a competition-based weighting scheme for the same reaction in FIG. 7;

[0040] FIG. 9 is the present invention display of a simplified metabolic network showing biosynthesis routes from the CI starting compound to the C6 product;

[0041] FIG. 10 is the present invention display of an illustration of the difference in ranking outcomes for a thermodynamic favorability-based approach and a competition-based approach;

[0042] FIG. 11 is the present invention display of a graph demonstrating the computational performance of MRE for various settings;

[0043] FIG. 12 is the present invention display of the structure of an experimentally derived heterologous biosynthetic pathway for producing naringenin from L-tyrosine in an *E. coli* host;

[0044] FIG. 13 is the present invention Metabolic Route Explorer (MRE) user-interface page showing a chart of the reactions for the top-ranked biosynthesis pathway shown in FIG. 12;

[0045] FIG. 14 is the present invention display of a pathway-level graph generated in MRE of the top-ranked pathway for the production of 1,3-PDO from glycerol in *E. coli;*

[0046] FIG. 15 is the present invention display of a pathway-level graph generated in MRE of the top-ranked pathway for the production of 1,3-PDO from glycerol in yeast;

[0047] FIG. 16 is the present invention display of a pathway-level graph generated in MRE of a known route for the production of artemisinic acid from acetyl-CoA in yeast; and

[0048] FIG. 17 is the present invention display of a pathway-level graph generated in MRE of the top-ranked pathway for the production of artemisinic acid from acetyl-CoA in yeast.

[0049] While the invention is susceptible to various modifications and alternative forms, specific embodiments thereof have been shown by way of example in the drawings and are herein described in detail. It should be understood that the description herein of specific embodiments is not intended to limit the invention to the particular forms disclosed, but on the contrary, the intention is meant to cover all modifications, equivalents, and alternatives falling within the spirit and scope of the invention as defined by the appended claims.

DETAILED DESCRIPTION

[0050] The present invention is a method and system of determining heterologous biosynthesis pathways in a specified host, which not only takes into account thermodynamic criteria for the desired reaction, but also considers the effect of competing endogenous reactions and suggests heterologous enzymes that may increase the favorability of the reaction route. Put another way, the present invention relates to a method and system for analyzing, determining, predicting and displaying ranked suitable heterologous biosynthesis pathways for a specified host.

[0051] The present invention addresses the problem of finding suitable pathways for the endogenous metabolism of a host organism because the efficacy of heterologous biosynthesis is affected by competing endogenous pathways. The present invention is called MRE

(Metabolic Route Explorer), and it was conceived and developed to systematically and dynamically search for, determine, analyze, and display promising heterologous pathways while considering competing endogenous reactions in a given host organism and to suggest possible foreign enzymes that may be suitable for use in the reactions. To rationally design a productive heterologous biosynthesis system, it is essential to consider the suitability of foreign reactions for the specific endogenous metabolic infrastructure of a host. The present invention MRE has been developed, which, for a given pair of starting and desired compounds in a given chassis organism, and dynamically ranks biosynthesis routes from the perspective of the integration of new reactions into the endogenous metabolic system.

[0052] To explore biosynthesis routes with MRE, the user specifies a host organism and a pair of the starting and target compounds. To increase its usability and to help the user specify organisms and compounds, MRE comes with an auto-completion feature. With advanced options, the user can override the default setting for the metabolic route search. These options include the maximum number of reaction steps (denoted by *n),* the number of top-ranked pathways to generate (denoted by *K),* and a list of compounds that are not considered as primary metabolic precursors in the search, called the *exclusion list.* By default, *n* and *K* are set to 8 and 50, respectively, while the exclusion list has 10 1 compounds that have high degrees of connectivity in its metabolic network graph, for example, water, ATP and ADP. This exclusion list can also be customized to have other compounds (e.g., CO2). In addition, MRE allows the user to constrain the chemical transformation of precursors based on RPAIR types (e.g., main, cofac and trans). These filtering schemes to constrain possible chemical transformations were reported to increase the relevance of the *de novo* biosynthesis route suggestion. By default, MRE considers chemical transformations based on main, cofac and trans RPAIR types.

[0053] For each promising heterologous biosynthesis pathway, the present invention MRE suggests actual enzymes for foreign metabolic reactions and dynamically generates information on competing endogenous reactions for the consumption of metabolites. These unique, chassis-centered features distinguish the present invention MRE from existing display systems and allow synthetic biologists to dynamically evaluate the design of their biosynthesis systems from a different perspective. As disclosed herein, the present invention MRE (Metabolic Route Explorer) was developed that systematically searches for promising heterologous pathways by considering competing endogenous reactions in a given host organism. The present invention supports biosynthesis of a range of high-value natural

products as a case study, and the present inventions MRE has been shown to be an effective tool to guide the design and optimization of heterologous biosynthesis pathways.

[0054] The present invention is a novel method and system for determining heterologous biosynthesis pathways to achieve a desired product from a specified host organism considering the suitability of foreign reactions for the specific endogenous metabolic infrastructure of the specified host organism and suggestions of foreign enzymes needed for the reactions using a competition-based weighting approach to determine the top-ranked biosynthesis routes. The present invention has a host-independent metabolic network constructed from databases containing verified metabolic reactions. Weights are assigned to the reactions in host dependent fashion by classifying which enzymatic reactions are native and foreign in the given host organism, by using thermodynamic data, and by identifying competing endogenous reactions. The host-independent metabolic network and weight data are used to construct a metabolic network with host-dependent weights.

[0055] The present invention is dynamic and versatile in that it allows user input to select a host organism, source and target compounds, and search options. The present invention Metabolic Route Explorer exhaustively explores and ranks biosynthesis routes for the selected criteria. The present invention MRE generates ranked biosynthesis routes, genes for foreign enzymes and competing native reactions. The results are displayed in summary tables and reaction pathway graphs with links to more detailed graphs and tables with more specific reaction details including reaction formulas, native and foreign enzymes and competing native reactions.

[0056] The present invention is a computer program based method and system for determining heterologous biosynthesis pathways to produce a target product in a host organism from a selected starting material. The method and system are characterized as follows:

a. First, a user inputs data, including a host organism, a starting compound, and a desired product. A user may also select other criteria such as the number of reactions per route or the number of routes, KEGG RPAIR constraints, or additional compounds to exclude in the search.

b. Second, a summary of pathways is generated by the Metabolic Route Explorer that ranks the pathways by score, and displays the pathway score summing all of the reaction weights in a given pathway for the reaction, the route from the source

compound to the target product, the number of reactions that are native and foreign to the host organism and whether the reactions are endogenous to the host.

c. Next, a graph is generated consisting of the top ten or top thirty routes from the starting compound to the target product, in which vertices represent metabolites and edges represent chemical transformations via verified metabolic reactions.

   i. Color coding of vertices and edges indicate starting and ending compounds, and which reactions are native or foreign to the host organism, and the width of the arrows (edges) indicate the value of the Gibbs energy, or strength, of the reaction path.

   ii. Hovering a cursor over nodes or edges will dynamically display compound names or the reaction Gibbs energy, respectively.

d. From the summary table, a specific route may also be selected and a graph is generated showing the specified route which indicates the pathway from starting compound to target product.

   i. The graph shows the reactions steps and metabolites, as well as competing endogenous reactions.

   ii. With this graph a table of the specific reaction steps for the route is displayed with the reaction identification, reaction formula, whether the reaction is native to the host, the Gibbs energy of the reaction step, the native enzymes, potential foreign enzymes and data for the competing endogenous reactions.

   iii. From this page, a competing reaction can be selected and specific pathway details will be displayed. Selecting an enzymes from the table will generate a display of enzyme data.

[0057] The present invention is more than a mere "a mathematical algorithm," "a fundamental economic or longstanding commercial practice," or "a challenge in business." Utilizing the data supplied in the tables and graphs allows a user to select the best path for a specified chemical transformation which is inclusive of thermodynamic criteria for the foreign reaction in view of competing native reactions, and MRE suggests foreign enzymes that may be used to catalyze the desired foreign reactions to increase the desired end product.

[0058] The present invention is a method and system that more accurately, more comprehensively, more systematically and dynamically searches for, determines, analyzes, and displays promising heterologous pathways in the field of natural product construction while considering competing endogenous reactions in a given host organism. The claimed invention has a specific, structured graphical user interface paired with the above prescribed functionality that directly related to the graphical user interface's structure, which resolves identified problems in the prior art display systems.

[0059] For instance, the present invention pairs its graphical user interface with its analysis programming to reduce the time for searching, analysis, and dynamic determination and display of suitable biosynthesis pathways over known prior are display systems, and the present invention achieves more accurate predictions of suitable biosynthesis pathways by adequately assessing the suitability of pathways in a specific context, appropriately considering the introduction of foreign metabolic genes into a given host organism, and appropriately considering the endogenous metabolic system of a host organism. The combination of these attributes in the present invention allows researchers to more efficiently and accurately search for, determine, analyze, and display promising heterologous pathways while considering competing endogenous reactions in a given host organism.

[0060] The use of endogenous pathway score (calculated based on one or more of the reaction weights in a given pathway for the reaction, the route from the source compound to the target product, the number of reactions that are native and foreign to the host organism and whether the reactions are endogenous to the host), specific context factors, host organism factors, and endogenous metabolic system factors are inventive concepts in the context of the present system, which allows the present invention to decrease the design cycle time periods over known display systems by eliminating erroneous, flawed or unsuitable pathways from the display and consideration in the biosynthesis efforts. For the above reasons, the present invention is a graphical display system that properly accounts for essential factors in the biosynthesis analysis to more accurately calculate suitability of biosynthesis pathways in a high speed system with greater accuracy, enhanced usability, and dynamic displays.

[0061] FIGS. 1-5 are typical user interface pages of MRE. FIG. 1A is a query input page (100) where a user inputs a host organism, a starting material and a desired product. In the host organism input field (101), a host organism is entered, either by name (e.g. *Escherichia coli* K-12 MG 1655) or by KEGG organism code (e.g. ECO). A starting material is entered in the starting material input field (102) by either KEGG compound ID (e.g. C00031) or name (e.g.

D-glucose or grape sugar). The desired product is entered in the desired product input field (103) by either KEGG compound ID (e.g. C00022) or by name (e.g. pyruvate or 2-oxopropanoate). MRE also provides an auto-completion feature. For example, inputting the desired product name generates a drop down list (104) of possible product matches from which the desired product may be selected. When this information is entered, MRE generates a summary page for the top-ranked biosynthesis routes.

[0062] FIG. IB is the query input page (100) with the advanced options feature (105) selected. A user may customize pathway details that will be displayed by specifying ranges in the option fields. The maximum number of reactions in each route (105) can be selected. The default setting for number of reactions is $n = 8$, however, a user can specify up 20 reactions for a route. The maximum number of biosynthesis routes (106), can be selected. The default setting for number of reactions is $K = 50$, however, a user can specify up 500 routes. The KEGG RPAIR (reaction-pair) constraints (108) can be selected. The MRE default setting includes Main, Cofac and Trans, however, a user can remove any of these options or add the Leave or Ligase options. Additional compounds, such as CO2, can be added to the exclusion list in the excluded compounds field (109).

[0063] Based on the input query for biosynthesis requirements in FIG. 1, MRE generates the *top-K* metabolic routes, and the main result page (FIG. 2) summarizes these routes. For each metabolic route, MRE highlights whether it is endogenous or heterologous to the host organism. For each foreign reaction in a heterologous biosynthesis route, MRE predicts which metabolites may not be available in the host, and it lists exogenous genes for the corresponding enzymatic activity and suggests a list of foreign genes based on a taxonomic similarity measure whose cDNA sequences can be downloaded in the FASTA format. It also shows a list of native reactions competing for a metabolic precursor with each foreign enzymatic reaction. MRE provides a means to visualize a specific pathway with competing endogenous reactions as well as a graph aggregating top-ranked routes.

[0064] Unlike known prior art display systems, the present invention Metabolic Route Explorer (MRE) disclosed herein focuses on the suggestion of foreign enzymes with well-characterized activities for promising heterologous pathways by taking into account the effects of the existing, endogenous metabolic infrastructure of a host organism. To find promising biosynthesis routes from a large number of potential candidates, thermodynamic data offer useful information. Unlike some other existing pathway display systems, such as Metabolic tinker and XTMS (which use thermodynamic data to constrain the reaction directionality or to

rank pathways based on their net favorability, which does not consider competing endogenous reactions), the present invention MRE system uses thermodynamic data to rank pathways in a host-dependent manner from the perspective of the integration of new reactions into the endogenous metabolic system. In order to suggest actual foreign enzymes for the design of heterologous biosynthesis pathways, the present invention MRE only considers verified reactions as metabolic parts. For each foreign reaction in a suggested heterologous pathway, present invention MRE generates information about endogenous reactions competing for metabolites. Since one effective approach to increase the productivity is to attenuate or eliminate competing reactions, MRE also offers useful insights into how to debottleneck and optimize heterologous pathways.

[0065] FIG. 2 is an example of a summary page (200) showing a table representation of the ten top-ranked biosynthesis routes for production of naringenin (KEGG compound ID C00509) from L-tyrosine (KEGG compound ID C00082) in an *E. coli* host organism. The summary page will typically display 50 routes (the default setting for *K*) unless a different number of routes have been selected in the user-input page. In the example table (200), only 10 of the routes are shown. The No. column (201) in the table shows the ranking of the pathway. The Score column (202) shows the score of the pathway wherein the higher the score, the better the route. The score is based on a predetermined calculation, such as, for example, the sum of all reaction weights in a given pathway taking into consideration the thermodynamic criteria of the reactions in the route as well as the effect of endogenous competing reactions.

[0066] The Route column (203) shows the steps in the specified metabolic route from the starting material to the target compounds by KEGG compound ID. Alternatively, a user can choose to view the compounds by name instead of ID numbers. The Reactions column (204) shows the number of reactions in the route indicating a ratio of how many of the steps in the pathway are reactions native to the host organism (first number) and how many are foreign reactions for the host organism (second number). Column 205 shows whether the reaction pathway is natively present in the user specified host. The ECO column heading in the example table specifies the host is *E. coli.* "Yes" indicates that the all the reactions exist in the user-specified host and "No" indicates that they do not. In the example table, the reactions listed are not native to the host organism.

[0067] FIG. 3 is an example of a graphical representation page (300) of the top-ten ranked biosynthesis routes seen in the FIG. 2 summary page showing the pathway steps for production of naringenin (KEGG compound ID C00509) from L-tyrosine (KEGG compound ID C00082)

in an *E. coli* host organism. Alternatively, a user can choose to view a graph of the top 30 routes. The displayed graphs are scalable by the user to allow visualization of graph details.

[0068] The example graph shows the L-tyrosine starting compound (301) as an oval with the KEGG compound ID and shows the naringenin desired product (302) as an oval with the KEGG compound ID. Alternatively, a user can choose to view the graph with compound names instead of ID numbers. By hovering over a selected compound, a pop-up box (303) displaying the selected compound by common name, chemical structure and KEGG compound ID can be viewed. In the example, naringenin is displayed in the pop-up box (303). Metabolites are shown by KEGG compound ID in ovals designated 304a-304q along the reaction pathways.

[0069] Reactions in the pathway are shown as arrows or edges in the graph wherein the arrows indicate the direction of the reaction (i.e., the reactants and the products). The width of the arrow indicates the value of the Gibbs energy for the reaction, wherein the stronger the reaction, the wider the arrow will be. In the example, the foreign reaction designated 305e has a thicker arrow than the foreign reaction designated 305d, indicating that 305e is the stronger reaction. Hovering the cursor over a reaction pathway will display the reaction compounds and the reaction's Gibbs energy. Foreign reactions are shown by KEGG reaction IDs along arrows 305a-305y in the example graph. Native reactions are shown by KEGG reaction IDs along arrows 306a-306d in the example graph.

[0070] A user viewing the page would see the compounds and reaction paths in color, for example, the starting compound (301), the desired product (302) and the metabolites (304a-304q) would be seen in red, green and yellow, respectively, and the foreign (305a-305y) and native reaction (306a-306d) arrows would be seen in cyan and purple, respectively. This allows a user to quickly identify the reaction pathways and whether the pathways are native or foreign to the host organism. For instance, in the example graph, cyan colored arrows (305a, 305b and 305d) would indicate that all three of the reactions beginning with the starting compound (301) are foreign reactions to the *E. coli* host organism.

[0071] FIG. 4 is an example of a pathway-level page (400) showing the graph for only the top-rated biosynthesis route shown on the FIG. 2 summary page for the production of naringenin (KEGG compound ID C00509) from L-tyrosine (KEGG compound ID C00082) in an *E. coli* host organism. This page shows detailed information about each reaction step of a given biosynthesis pathway. This pathway information is displayed in a table and on a graph in this

page. This page can be viewed in terms of KEGG IDs or names, and this display choice can be changed with a click on the link shown on top of the page. For each reaction, the table shows the reactants, the products, whether the reaction is endogenous, the standard reaction Gibbs energy, and the corresponding EC number (if it's an enzymatic reaction). If the reaction is based on a heterologous enzyme, a list of potential enzymes and a list of native competing reactions are also shown (with links to more detailed pages) in the table. Moreover, all potential cDNA gene sequences for the pathway are available for download in FASTA format.

[0072] The selected reaction pathway from the starting compound (401) to the desired product (402) proceeds along the reaction pathway arrows (408a-408h) and includes the KEGG reaction IDs for each reaction (405a-405d and 407a-407f) and the KEGG compound IDs for compounds that are utilized or produced by the reactions (404a-404j). The pop-up box (403), which can be viewed by hovering the cursor over a compound, shows the common name, chemical structure and KEGG compound ID for an intermediate compound in the reaction pathway. For this pathway, the reactions seen along route 408a-408h are foreign reactions (405a-405d). An important additional piece in information on the detailed graph in FIG. 4 is the inclusion of competing reactions (407a-407f), which can impede the progress of the reaction.

[0073] At the top of FIG.4, the pathway details for the route are broken down in table format. The Reaction ID column (409) displays the KEGG reaction ID for the step as well as the KEGG compound ID for the primary reactant in that step. The Formula column (410) displays the primary reactants and products in the step by KEGG compound ID. Column 411 shows whether the reaction pathway is natively present in the user specified host. The ECO column heading in the example table specifies the host is *E. coli.* "Yes" indicates that the reaction exists in the user-specified host and "No" indicates that it does not. In the example table, the reactions listed are not native to the host organism. The Gibbs energy column (412) displays the energy associated with the reaction which indicates the favorability of the reaction based on thermodynamic criteria. The EC# column (413) displays the Enzyme Commission number for the enzyme that catalyzes the step and the Potential Enzyme column (414) suggests foreign enzymes that may be utilized in the reaction. The Competing Reaction column (415) displays competing reactions by KEGG reaction ID and Gibbs Energy associated with the competing reaction.

[0074] A user viewing the page would see the compounds and reaction paths in color. For example, the desired path is shown with blue arrows (408a-408h). The starting compound

(401), the desired product (402) and the metabolites (404a-404j) would be seen in red, green and yellow, respectively, and the foreign (405a-405d) reaction boxes would be seen in cyan. Competing endogenous reactions (407a-407f) are shown as gray boxes. This allows a user to quickly identify the reaction pathways and whether the pathways are native, foreign or competing reaction for the host organism. For instance, in the example graph, gray boxes 407a, 407f would indicate that there are six competing reactions on this route.

[0075] Unlike known prior art display systems, the present invention Metabolic Route Explorer (MRE) disclosed herein focuses on the suggestion of foreign enzymes with well-characterized activities for promising heterologous pathways by taking into account the effects of the existing, endogenous metabolic infrastructure of a host organism. To find promising biosynthesis routes from a large number of potential candidates, thermodynamic data offer useful information. Unlike some other existing pathway display systems, such as Metabolic tinker and XTMS (which use thermodynamic data to constrain the reaction directionality or to rank pathways based on their net favorability, which does not consider competing endogenous reactions), the present invention MRE system uses thermodynamic data to rank pathways in a host-dependent manner from the perspective of the integration of new reactions into the endogenous metabolic system. In order to suggest actual foreign enzymes for the design of heterologous biosynthesis pathways, the present invention MRE only considers verified reactions as metabolic parts. For each foreign reaction in a suggested heterologous pathway, present invention MRE generates information about endogenous reactions competing for metabolites. Since one effective approach to increase the productivity is to attenuate or eliminate competing reactions, MRE also offers useful insights into how to debottleneck and optimize heterologous pathways.

[0076] FIG. 5 is an example of a competing reaction information page (500) displaying the details for the competing endogenous reactions associated with compound 404d in FIG. 4. Compound 404d is a metabolite in the pathway from reaction 405c to reaction 405b and three endogenous reactions in the host organism (407a, 407b, 407c) are competing with the desired pathway for that compound. The reaction column (501) displays the specifics of the three competing reactions. The Gibbs energy column (502) displays the standard reaction Gibbs energy associated with that competing reaction and the Enzyme gene ID column (503) displays the Enzyme Gene ID for the enzyme in the host organism associated with the competing reaction.

[0077] As the display pages represented in FIGS. 1-5 progress, more specific reaction data is provided. FIGS. 1A and IB are the user input and FIG. 2 shows the top ten routes resulting from that input. FIG. 3 is a graphical representation of the top ten routes displayed in FIG.2. FIG. 4 is a graphical representation of one selected route from FIG. 3 and provides specific details of the reaction mechanism for that pathway, including any known competing endogenous reactions for each step. The FIG. 5 table displays expanded details of the competing reactions for a given step. MRE also provides self-explanatory pages showing detailed information on compounds, reactions and EC numbers. These pages can be accessed by clicking the internal links on the compound IDs, the reaction IDs and the EC numbers. Moreover, external links to other databases are also provided on these pages as alternatives.

[0078] FIG. 6 depicts the workflow of MRE (600). Metabolic reaction data (601) from several data sources, including but not limited to, the Kyoto Encyclopedia of Genes and Genomes (KEGG), ExPASy ENZYME database, and eQuilibrator dataset are complied. KEGG lists around 4000 organisms, which MRE uses for the selection of a host organism. The KEGG COMPOUND database is used to identify metabolites, while the KEGG REACTION database and the ExPASy ENZYME database are used to find metabolic reactions with verified activities. The eQuilibrator dataset is used to obtain the reaction Gibbs energy in the standard 1M concentration setting. The KEGG RPAIR database is used to restrict search space based on the relation between reactants and products. The KEGG GENES database is used for DNA sequence data for enzymatic genes, and the KEGG taxonomy mapping dataset is used to calculate taxonomic distances.

[0079] As seen in FIG. 6, MRE first constructs a directed graph representing a host-independent metabolic network with verified reactions (603). This graph (603) comprises all metabolic reactions with verified activities found in the data source, and it is built regardless of the choice of a host organism for a biosynthesis system. It next assigns weights to the edges in the graph in a host-dependent fashion by classifying which enzymatic reactions are native and foreign in the given host organism and by using the thermodynamic data.

[0080] From the metabolic reaction data compiled from the data sources (601), MRE constructs a host-independent metabolic network with verified reactions (602) by first identifying reactions with verified activities. Enzymatic reactions are categorized based on Enzyme Commission numbers (EC numbers). Each EC reaction (i.e., a reaction class corresponding to each EC number) denotes a class of catalytic reactions with the same chemical transformation. To retrieve verified metabolic reactions with known enzymes,

reaction classes with partially qualified EC numbers are filtered out as these partial EC reactions are unverified and can lead to misinterpretation of enzymatic activities. EC reactions that do not contain any enzymes are also removed. With this filtering process, 5389 complete EC reactions and 76 spontaneous reactions with verified activities were identified.

[0081] Next, standard reaction Gibbs energy $\Delta rG'°$ is estimated for each of these verified reactions using eQuilibrator with absolute temperature set to 298.15K. Each verified EC reaction is then split into two reactions: the forward reaction with the reaction Gibbs energy $\Delta rG'°$ and the backward reaction with the reaction Gibbs energy $-ArG'°$. Those EC reactions whose $\Delta rG'°$ could not be estimated were assigned the largest of the estimated values for both directions. This conservative approach is used to avoid the suggestion of biosynthesis routes containing reactions with no thermodynamic information as much as possible.

[0082] Using these reactions, a directed graph of the host-directed metabolic network (603) is built that models the transformation of metabolites where its vertices represent metabolites and its edges represent chemical transformations via verified metabolic reactions. Since this directed graph unifies all metabolic reactions with verified activities in the reaction databases, its structure is independent of the endogenous metabolic system of any host organism.

[0083] User input (604), including host organism, source and target compounds and advanced search options, are used in conjunction with the host independent metabolic network (603) to assign weights to edges of the directed graph in a host-dependent fashion (605). To assign the weight of each outgoing edge from a given compound node, the assumption was made that this reaction was in the host organism and computed the probability of converting the precursor via this reaction over the competing native reactions.

[0084] By representing the competition for a metabolic precursor with endogenous reactions by a statistical mechanical model, the probability of each reaction with $\Delta_r G'°$ through the Boltzmann distribution was computed. The logarithm of this computed probability was assigned as the weight of this outgoing edge. The data from the user input (604), the host independent metabolic network (603) and the assigned weights leads to a metabolic network with host-dependent weights (606). Use of this type of statistical mechanics modeling in the context of the biosynthesis system design is novel. Given the metabolic network graph with host-dependent weights (606), MRE will explore and rank biosynthesis routes (607) by exhaustively searching for biosynthesis paths from the given starting material to the given product and generating results (608) of the *top-K* metabolic routes, each of which has at most *n*

reaction steps. The results (608) include the ranked biosynthesis routes, genes for foreign enzymes and competing native reactions.

[0085] In this search to explore and rank biosynthesis routes (607), the compounds in the exclusion list are not considered as intermediate precursors of the product. To rank routes, MRE computes their scores by summing all reaction weights in each route and keeps K routes with the highest scores. MRE transforms the metabolic route search problem into a classical computer science problem known as K-shortest loopless path problem and uses an efficient algorithm to solve it. The core part of the search was implemented in C++.

[0086] The weighting scheme used to assign weights to edges in a host-dependent fashion (605) depends on a host organism and models the competition for metabolic precursors with the endogenous reactions. Importantly, this competition-based weighting scheme can capture the effects of competing endogenous reactions on heterologous reactions, while a thermodynamic favorability-based weighting scheme cannot. This can make their weight assignments widely different from each other, as illustrated in FIGS. 7 and 8.

[0087] To derive a mathematical description of the weighting scheme, a scenario is used to generate weights for edges in the reactions transforming precursor C. Here, RNRN represents a set of native reactions that can transform C in a given host organism. For each reaction r that can transform C, $e\text{-}ArG'\circ/RTe\text{-}\Delta rG'\circ/RT$ was set as its Boltzmann factor. Thenar), the normalized Boltzmann factor for r, is defined as follows:

$$f(r) = \frac{e^{-\Delta_r G'_\circ/RT}}{1 + e^{-\Delta_r G'_\circ/RT} + \sum_{r' \in \mathbf{R}_N \setminus \{r\}} e^{-\Delta_{r'} G'_\circ/RT}, f(r)^{e-\Delta rG''/RT}}$$

(1) where R is the gas constant and T is the absolute temperature. Those reactions that are not in the host organism do not affect the calculation of the Boltzmann distribution. If rSRNrSRN, thenar) is simply based on the Boltzmann distribution of the native reaction system transforming compound C. On the other hand, if rgRNrgRN , thenar) is based on the Boltzmann distribution of the reaction system that contains all native reactions transforming C and foreign reaction r. With this scheme, every edge in the graph that transforms C in reaction r has the weight $\log f(r)$.

[0088] FIGS. 7 and 8 demonstrate the differences between a thermodynamic-favorability-based weighting scheme (700) illustrated in FIG. 7 and a competition-based weighting scheme (800) illustrated in in FIG. 8.

[0089] In the thermodynamic-favorability-based weighting scheme (700) illustrated in FIG. 7, node 701 represents a starting metabolite and nodes 704, 707 and 710 are metabolites from the metabolic conversions via reactions represented by the edges (703, 706 and 709). Edge (arrow) 709 represents a reaction that is native to the host organism and edges 703 and 706 represent foreign reactions. The value within the ovals (702, 706, 708) for each edge represents the weight $\Delta_r G'^{\circ}/RT$ where R is the gas constant and T is the absolute temperature.

[0090] In the competition-based weighting scheme (800) illustrated in FIG. 8, node 801 represents a starting metabolite and nodes 804, 807 and 810 are metabolites from the metabolic conversions via reactions represented by the edges (803, 806 and 809). Edge 809 represents a reaction that is native to the host organism and edges 803 and 806 represent foreign reactions. The value within the ovals (802, 806, 808) for each edge represents its weight. With this scheme, edges (or arrows) with the same $\Delta_r G'^{\circ}$ value can have different weights in a host-dependent fashion. For example, the weight of CI $\rightarrow$ C3 is In [eV(l + e^{1} + e^{1})], while that of C1 $\rightarrow$ C4 is In [e^{1}l{ l + e^{1})].

[0091] The competition-based weighting scheme illustrated in FIG. 8 is based on the logarithm of normalized Boltzmann weights. Unlike thermodynamic favorability-based measure in FIG. 7, the competition-based weighting scheme estimates a fraction of a given precursor that is converted into next intermediate metabolites. Thus, a pathway score based on the sum of all reaction weights in a given pathway can characterize the lower bound of a fraction of starting material that is converted into the product through this pathway, and the competition-based score can capture the productivity of each pathway more appropriately.

[0092] FIGS. 9 and 10 are an example set showing differences in ranking outcomes between the thermodynamic favorability based approach and utilization of the competition-based approach. FIG. 9 is a simplified metabolic network (900). The nodes (901, 904, 907, 912, 915, 920, 923) are metabolites and the edges (903, 905, 909, 911, 913, 917, 919, 921 and 925) are metabolic conversions. Edges 909, 917, 919 and 921 indicate native reactions, while edges 903, 905, 911, 913 and 925 indicate foreign reactions. The value within the oval for each edge (902, 906, 908, 910, 914, 916, 918, and 922) indicates $A_r G'^{\circ}/RT$ where R is the gas constant and T is the absolute temperature. In this example, compound CI (901) is the starting metabolite, and compound C6 (915) is the target product. Three possible routes are shown for the conversion of CI to C6.

[0093] FIG. 10 shows the ranking of the three biosynthesis routes (1000) with the thermodynamic favorability approach (1001, 1002, 1003), wherein the lower the score, the better or more favorable the route, and the competition-based approach (1004, 1005, 1006), wherein the higher the score, the better or more favorable the route. For example, the score of $CI \rightarrow C4 \rightarrow C6$ (1003) is - 1 + 2 = 1 with the thermodynamic favorability approach, indicating that the 1003 route is the least favorable of the three routes shown. However, for the competition-based approach the score of $CI \rightarrow C4 \rightarrow C6$ (1004) is $\ln [eV(1 + e^{1})] + \ln [e^{-2}/(1 + e^{-2} + e^{-10})] = -2.44$ indicating that the 1004 route is the most favorable.

[0094] Biosynthesis pathways of interest are often those that transform a higher fraction of a starting material to a target product. One heuristic to rank pathways based on this productivity criterion is the net favorability of pathways. At a first glance, the net thermodynamic favorability (as illustrated in FIGS. 9 and 10) can be seen as a good measure to rank pathways based on this criterion. However, this measure can only quantify the ratio of the target concentration to the source concentration at equilibrium, which may not correspond well with the true picture of the titer of the target product, especially when a given pathway has strong competing reactions and the equilibrium concentration of the starting material is substantially lowered.

[0095] FIG. 11 is graph showing computational performance of MRE for various settings. On the graph (1100), n (1101) is the maximum number of reaction steps in biosynthesis routes, and K (1102) is the number of top routes that MRE generates. For this analysis, six different pairs of source and target compounds, four settings for the maximum number of reaction steps (n= 5, 10, 15 and 20), and five settings for the number of top metabolic routes (K=100, 200, 500, 1000 and 2000) were included. Each point represents the average computational time in seconds (1103) of the six source-target pairs with a given setting for n and K. In this performance test, the default setting of the exclusion list was used. The computation was performed on an Intel Xeon E5-2680 workstation with 256 GB of memory. The computational time increased as the value of n and K increases, but the magnitude of an increase was found to be reasonable. Even with a very computationally demanding setting (i.e. «=20 and £=2000), MRE was able to process the queries within 25 seconds on average. The data also indicated that, with the default setting (i.e. «=8 and K=50), the processing time is expected to be less than 5 seconds. Also, since MRE caps the value of n and K that the user can set at 20 and 500, respectively, the maximum computational time is expected to be around 10 seconds. These

show that the exhaustive search employed in MRE does not compromise the user experience based on its processing time.

[0096] To further evaluate the computational performance of MRE, the processing time in the runtime environment was measured. 1000 reachable pairs of source and target compounds were randomly selected. With the setting of the largest reaction step size and the largest number of top-ranked pathways (i.e., $n = 20$ and $K = 500$), it took less than 10 seconds for MRE to exhaustively explore routes and process queries on average. In 95% of the samples, the processing time was less than 20 seconds, and even in the worst case, it was just less than 30 seconds. With the default setting (i.e., $n = 8$ and $K = 50$), the processing time was at most 1.36 seconds. The exhaustive pathway search employed in MRE should not compromise the user experience based on its processing time.

Case Study

[0097] As a case study, MRE was applied to search for pathways for various biosynthesis specifications using either *E. coli* K-12 MG1655 or *Saccharomyces cerevisiae* as the host organism. Table 2 summarizes the top-ranked heterologous pathways that MRE discovered. This shows that, in biosynthesis of a range of high-value natural products, MRE was able to identify pathways that are known to be productive. The MRE results were also analyzed by comparing them with results from four open-access web servers that can design heterologous biosynthesis pathways, namely, FMM, Metabolic tinker, PHT and XTMS. To explore biosynthesis pathways with these tools, default configurations were used.

Table 2. Top-ranked pathways identified by MRE for various biosynthesis specifications

Biosynthesis specification			Results of top-ranked pathway identified by MRE			Comparison with existing tools	
Source	Target	Host	Steps	Necessary foreign enzymes	Remark	Found a path [a]	Match with MRE [b]
L-tyrosine (C00082)	Naringenin (C00509)	*E. coli* (ECO)	4	4.3.1.23, 6.2.1.12, 2.3.1.74, 5.5.1.6	Recovered a known route as the top route [1]	FMM, XTMS	FMM
glycerol (C00116)	1,3-PDO (C02457)	*E. coli* (ECO)	2	4.2.1.30, 1.1.1.202	Recovered a known route as the top route [2]	FMM, PHT	FMM, PHT
glycerol (C00116)	R-1.2-PDO (C02912)	Yeast (SCE)	5	4.2.3.3, 1.1.1.79, 1.1.1.77	Recovered a known route as the top route [3]	MT	MT
acetyl-CoA (C00024)	artemisinic acid (C20309)	Yeast (SCE)	10	2.5.1.92, 4.2.3.50, 4.2.3.82, 4.2.3.24, 1.14.13.158	Recovered a known route , and predicted better ones [4]	none	none
L-tyrosine (C00082)	resveratrol (C03582)	*E. coli* (ECO)	3	4.3.1.23, 6.2.1.12, 2.3.1.95	Recovered a known route as the top route [5]	FMM	FMM
D-xylose (C00181)	xylitol (C00379)	*E. coli* (ECO)	2	1.1.1.21, 1.1.1.307	Recovered two known routes as the top routes [6]	FMM, PHT	FMM, PHT
PRPP (C00119)	histidine (C00135)	*E. coli* (ECO)	8	2.6.1.27	Predicted better and shorter routes than a known native route [7]	FMM, MT	none
chorismate (C00251)	tryptophan (C00078)	yeast (SCE)	5	none	Predicted the native route as the best, and found shorter routes [8]	FMM, MT, PHT	FMM

For each biosynthesis specification, the source and target compounds are specified in KEGG ID, and the host organism is in KEGG organism code. For each pathway, the number of

reaction steps and the necessary foreign enzymes (in EC number) are specified. Comparison with FMM, Metabolic tinker (MT), PHT and XTMS is also shown. For each tool, its default setting was used, except for the configuration of a pathway length, which was set to accommodate known pathways. In the Table 2, the notation "a" denotes tools that have identified at least one path for a given biosynthesis specification, and the notation "b" denotes tools whose top-ranked pathway is the same as the top-ranked one from MRE.

Biosynthesis of naringenin

[0098] Naringenin is a plant secondary metabolite, which is reported to have various health benefits, including high antioxidant capacities and significant antiviral effects on the hepatitis C virus. Hollman P.C., Katan M.B. Bioavailability and health effects of dietary flavonols in man. Arch. Toxicol. Suppl. 1998;20:237-248. Owing to inefficiencies in the production of naringenin from natural plant sources, metabolic engineering to have an efficient microbial synthesis of this high-value natural product is thought to be a commercially viable alternative.

[0099] FIGS. 12 and 13 show a heterologous biosynthesis pathway to produce naringenin from L-tyrosine in an *E. coli* host. In the analysis seen in FIGS. 12 and 13, L-tyrosine (KEGG compound ID: C00082), an aromatic non-essential amino acid, was selected as the starting material since a state-of-the-art heterologous naringenin production from L-tyrosine in an *E. coli* strain is known (see FIG. 12). This heterologous biosynthesis route comprises four foreign enzymatic reactions. To analyze the performance of MRE in comparison with other tools, two open-access biosynthesis pathway web servers were applied, Metabolic tinker and XTMS. Since these two recently developed tools also rely on reaction thermodynamic data for their pathway ranking, it was possible to also analyze the effects on the competition-based ranking scheme.

[00100] Given this biosynthesis requirement, Metabolic tinker and PHT were not able to find any pathways, while XTMS generated a predicted pathway with hypothetical reactions as its top-ranked candidate. In contrast, the top-ranked route from MRE and FMM was identical to the state of the art. The pathway information given by MRE indicates that the third reaction in the pathway, which transforms p-coumaroyl-CoA into naringenin chalcone, is a bottleneck and competes for the availability of cofactor malonyl-CoA with a more favorable native reaction involved in the fatty acid biosynthesis in the *E. coli* host (Figure 13). This suggests that an increase in the concentration of malonyl-CoA or the inhibition of the fatty acid biosynthesis could enhance the productivity of this naringenin biosynthesis pathway. Indeed, previous studies demonstrated that both an increase in the availability of malonyl-CoA in the host and a

decrease in the activities in the fatty acid pathway can increase the naringenin titers. While FMM was also able to identify the heterologous naringenin biosynthesis pathway that MRE found, the pathway information given by FMM was not helpful to find an optimization target as FMM does not have a feature to quantify the effects of competing reactions in the host.

[00101] FIG. 12 shows the structure of an experimentally derived biosynthesis pathway (1200) from the L-tyrosine starting compound (1201) through the intermediate metabolites (1205, 1209, 1213) to the desired naringenin product (1217). The KEGG compound ID of each metabolite appears in an oval (1202, 1206, 1210, 1214, 1218) below each structure. The required enzyme for each step is noted above the arrows (1203, 1207, 1211, 1215). The abbreviations for the enzymes are: tyrosine ammonia lyase (TAL); 4-coumarate:CoA ligase (4CL); chalcone synthase (CHS) and chalcone isomerase (CHI). The EC numbers (1204, 1208, 1212, 1216) for each reaction are indicated below each arrow.

[00102] FIG. 13 displays the information of the top-ranked biosynthesis pathway in MRE for the L-tyrosine to naringenin conversion (which is the same pathway associated with FIG. 4). Column 1301 displays the KEGG reaction and compound IDs. Column 1302 displays the reaction step by KEGG compounds ID. The ECO column (1303) heading in the example table specifies the host is *E. coli.* "Yes" indicates that the reaction exists in the user-specified host and "No" indicates that it does not. In the example table, the reactions listed are not native to the host organism. Column 1304 displays Gibbs energy. Column 1305 displays EC number for the enzyme if it is an enzymatic reaction. Column 1306 displays potential enzymes that may be used for the reaction and Column 1307 displays competing reactions by KEGG ID and Gibbs energy for the competing reaction.

Production of value-added chemicals from glycerol

[00103] Glycerol is a readily available and relatively inexpensive chemical compound that can be generated in large amounts as a byproduct of biodiesel and bioethanol production processes. Because of its economic viability and long-term sustainability, fermentative production of high-value materials from glycerol has gained much attention recently. Using glycerol as the starting material, pathways were searched for the production of two value-added chemicals, 1,3-propanediol (1,3-PDO), a commodity chemical mainly used to make polyester fiber, and 1,2-propanediol (1,2-PDO), another high-demand commodity chemical used to make a wide range of products including antifreeze, thermoset plastics and cosmetics.

[00104] FIGS. 14 and 15 are pathway level graphs generated in MRE for production of 1,3-propanediol (1,3-PDO) from glycerol in two different host organisms. FIG. 14 shows the MRE top-ranked pathway for the production of 1,3-PDO from glycerol in *E. coli* and FIG. 15 shows the MRE top-ranked pathway for the production of R-l,2-PDO from glycerol in yeast.

[00105] MRE was first applied to search for pathways for the production of 1,3-PDO in *E. coli* chassis. The top-ranked pathway (Fig. 14) that MRE identified is a known two-step heterologous pathway, which requires the introduction of a glycerol dehydratase gene and a 1,3-propanediol dehydrogenase gene in the host. Since the first glycerol dehydratase reaction competes for the utilization of glycerol against several native reactions including glycerol kinase, MRE predicts that this can be a target for productivity optimization. Metabolic tinker and XTMS were not able to find any pathways for the 1,3-PDO production, whereas FMM and PHT found the same pathway that MRE identified.

[00106] In FIG. 14, the glycerol starting material (1401) is converted to the end product 1,3-PDO (1402) by heterologous reactions 1405a and 1405b, which are foreign to the *E. coli* host, via the reaction route shown byl408a-1408d. Intermediate metabolites produced or used by these reactions are shown by 1404a- 1404e. There are three competing endogenous reactions (1407a, 1407b, 1407c) shown along arrows the arrows (1417a, 1417b, 1417c) which indicate that these reactions, which are native to the *E. coli* host, are competing with reaction 1405a along path 1408a for the glycerol starting product (1401) and foreign reaction pathways are indicated by 1415a-d. In an MRE generated pathway level graph, the display would show starting, ending and intermediate compounds shown in red, green and yellow, respectively. The top-ranked route is shown in blue, while native, foreign and competing reactions and paths are shown in purple, cyan and gray. The color display of pathways and compounds would allow a user to quickly identify the components in each route.

[00107] Next, MRE was applied to search for pathways for the synthesis of R-l,2-PDO in the yeast chassis. The top-ranked pathway (Figure 15) found was a known synthesis pathway for 1,2-PDO. In this pathway, glycerol is first converted to dihydroxyacetone phosphate (DHAP) via two native enzymatic reactions. Methylglyoxal synthase then transforms DHAP into methylglyoxal, which is, in turn, converted into (R)-lactaldehyde. Finally, lactaldehyde reductase is used to produce R-1,2-PDO from (R)-lactaldehyde. FMM and PHT were not able to find any pathways that convert glycerol into R-1,2-PDO, whereas Metabolic tinker identified the same pathway that MRE found as the top-ranked one. Since XTMS focuses on

the *E. coli* chassis, this tool was applied to search for heterologous R-l,2-PDO production pathways in *E. coli;* however, no pathways were found.

[00108] In FIG. 15, the glycerol starting material (1501) is converted to the end product R 1,2-PDO (1502) by native reactions 1506a and 1506b, which are endogenous to the yeast host and by heterologous reactions 1505a, 1505b and 1505c, which are foreign to the yeast host, via the reaction route shown byl508a-1508j. Intermediate metabolites produced or used by these reactions are shown by 1504a- 15041. There are five competing endogenous reactions, three (1507a, 1507b, 1507c) are shown along arrows (1517a, 1517b, 1517c) as competing with the desired route for metabolite 1504f, and two reactions (1507d, 1507e) are shown along arrows (15 17d, 15 17e) as competing with the desired route metabolite 1504h and foreign reaction pathways are indicated by 1515a-f and 1516a-d, respectively. In an MRE generated pathway level graph, starting, ending and intermediate compounds are shown in red, green and yellow, respectively, the top-ranked route is shown in blue, and native, foreign and competing reactions and paths are shown in purple, cyan and gray, which allow a user to quickly identify by color the components in each route.

Production of Artemisinic Acid

[00109] Artemisinic acid is an intermediate precursor for antimalaria drug artemisinin, and its production is often celebrated as one of the early success stories of the combination of metabolic engineering and synthetic biology. This engineered biosynthesis pathway utilizes the endogenous mevalonate pathway in budding yeast to transform acetyl-CoA into farnesyl pyrophosphate (FPP), which is then converted into artemisinic acid with heterologous amorphadiene synthase and three-step oxidation reactions.

[00110] In FIGS. 16 and 17, two routes are shown for the production of artemisinic acid from acetyl-CoA in yeast. FIG. 16 is a known route and FIG. 17 is an MRE top-ranked route. To see if MRE could recover this engineered pathway, MRE was applied to explore pathways for the production of artemisinic acid from acetyl-CoA in yeast. It was found that one of the top-ranked pathways that MRE generated was this known heterologous pathway (Figure 16). Interestingly, the pathway that MRE identified as the top candidate (Figure 17) was slightly different from the previously engineered pathway. The difference lies in how isopentenyl pyrophosphate (IPP) is converted into farnesyl pyrophosphate (FPP). In the top-ranked path, IPP is first converted into (2Z,6Z)-farnesyl diphosphate (Z,Z-FPP). This route is chosen because IPP is a precursor of a thermodynamically highly favorable native reaction, and the

conversion reaction from IPP to Z,Z-FPP is much more favorable than that from IPP to FPP, enabling a higher fraction of IPP to be utilized in the route. By using Z,Z-FPP as the precursor, this route introduces three foreign carbon-oxygen lyases to form FPP. FMM, Metabolic tinker and PHT were not able to find any pathways. XTMS found a partial pathway that converts FPP into artemisinic acid, albeit it is for the *E. coli* chassis.

[00111] In the known route (1600) shown in FIG. 16, the acetyl-CoA starting material (1601) is converted to desired product artemisinic acid (1602) via reaction route 1608a-1608p. The route includes endogenous reactions 1606a-1606f shown along arrows 1616a-16 16p (which indicate paths native to the host) and heterologous reactions 1605a, 1605b shown along arrows 1615a-1615f (which indicate paths foreign to the host). Intermediate metabolites produced or used by these reactions are shown by 1604a-16041 There are three competing endogenous reactions (nodes 1607a, 1607b, 1607c and arrows 1617a, 1617b, 1617c) that compete with desire route (16081-1608m) for the intermediate metabolite trans, trans, farnesyl diphosphate (1604k). In an MRE generated pathway level graph, starting, ending and intermediate compounds are shown in red, green and yellow, respectively, the top-ranked route is seen in dark blue, and native, foreign and competing reactions and paths are shown in purple, cyan and gray respectively, to allow a user to quickly identify the components by color in each route.

[00112] In an MRE top-ranked route (1700) shown in FIG. 17, the acetyl-CoA starting material (1701) is converted to desired product artemisinic acid (1702) via reaction route 1708a-1708t. The route includes endogenous reactions 1706a-1706e shown along arrows 1716a-1716n (which indicate paths native to the host) and heterologous reactions 1705a-1705e shown along arrows 1715a-1715i (which indicate paths foreign to the host). Intermediate metabolites produced or used by these reactions are shown by 1704a-1704v. There are three competing endogenous reactions (nodes 1707a, 1707b 1707c and arrows 1717e, 1717f, 1717g) that compete with desire route (1708m-1708n) for the intermediate metabolite dimethylallyl diphosphate (1704i), and three competing endogenous reactions (nodes 1707b, 1707c, 1707d and arrows 1717a, 1717b, 1717c) that compete with desire route (1708p-1708q) for the intermediate metabolite trans, trans, farnesyl diphosphate (1704k). In an MRE generated pathway level graph, starting, ending and intermediate compounds are shown in red, green and yellow, respectively, the top-ranked route is seen in dark blue, and native, foreign and competing reactions and paths are shown in purple, cyan and gray, respectively, to allow a user to quickly identify the components in each route.

[00113] The present invention, MRE, is an open-access biosynthesis design tool, that searches for promising metabolic routes for a given biosynthesis specification and suggests exogenous enzymes for heterologous biosynthesis pathways based on the infrastructure of an endogenous metabolic system. The present invention relies on the data sources (mainly KEGG) to mine verified metabolic reactions and to search for biosynthesis routes based on them. Indeed, while painstaking effort has resulted in a large collection of annotated metabolic reaction data, among the 9910 reactions in the KEGG REACTION database (Release 76.0), 1272 reactions with no EC numbers were found, 1079 with partial EC numbers were found and 2170 with no annotations for associated genes were found. The number of verified reactions in KEGG is expected to increase over time which would alleviate any issues related to a lack of verified reactions. Other metabolic reaction databases, such as Rhea, may also be integrated.

[00114] Several existing tools took an approach to expand a list of metabolic parts in hand by defining specific transformation rules, albeit such rules can be subjective. To design biosynthesis systems, this approach relies on the prediction of metabolic parts with specific metabolic activities, which may or may not exist. Thus, the design of biosynthesis systems via this top-down approach may require the *de novo* design of unnatural proteins to achieve specific metabolic activities. MRE was developed to suggest actual enzymes for heterologous pathways. Thus, it takes a complementary, bottom-up approach in which a biosynthesis system is designed by using well-characterized metabolic parts. To this end, only verified reactions were used.

[00115] Here, by using the biosynthesis of a range of high-value natural products as a case study, it has shown that MRE can suggest promising heterologous biosynthesis pathways and provide useful information to pinpoint bottlenecks of pathways. With the host-dependent competition-based pathway ranking scheme, along with the suggestion of foreign enzymes with competing endogenous reactions, MRE offers novel insights into the design and optimization of heterologous biosynthesis systems.

[00116] While the invention is susceptible to various modifications and alternative forms, specific embodiments thereof have been shown by way of example in the drawings and are herein described in detail. For example, the user-interface example pages shown in Figures 1-5 do not itemize or describe in detail the dimensions, shapes, sizes, inputs or outputs, or exact specification of the identified items (e.g. user input, summary tables, etc.), which are all understood to exist and be within the scope of the invention as described and claimed.

[00117] Furthermore, size and shapes of display pages, input fields and linked data are not described in detail, but such details are understood to be varied or modifiable while still complying with the scope of the invention set forth herein and covered by the claims. It should be understood, however, that the description herein of specific embodiments is not intended to limit the invention to the particular forms disclosed, but on the contrary, the intention is meant to cover all modifications, equivalents, and alternatives falling within the spirit and scope of the invention as defined by the appended claims.

WHAT IS CLAIMED IS:

1. A method of determining and displaying suitable heterologous biosynthesis pathways to produce a target product in a specified host organism from a selected starting material, comprising the steps of:

selecting and inputting user input data that includes designation of a host organism, a starting compound, and a desired target product into the pathway computer system;

searching for one or more possible heterologous pathways with consideration of competing endogenous reactions in a given host organism;

determining one or more suitable heterologous biosynthesis pathways from the starting compound to achieve the desired target product in light of consideration of competing endogenous reactions in a given host organism;

dynamically ranking suitable heterologous biosynthesis pathways from the starting compound to achieve the desired target product by an endogenous pathway score;

generating a graph of a predetermined number of most suitable heterologous biosynthesis pathways that shows reactions steps and metabolites, as well as competing endogenous reactions,

displaying one or more suitable heterologous biosynthesis pathways from the starting compound to the target product, in which vertices on the graph represent metabolites and arrow edges on the graph represent chemical transformations via verified metabolic reactions; and,

utilizing the data supplied in graphical display to select a preferred heterologous biosynthesis pathway for a specified chemical transformation.

2. The method of Claim 1 wherein said determining step includes consideration of thermodynamic criteria for the foreign reaction in view of competing native reactions.

3. The method of Claim 1 wherein said user input data selected and input may also include selecting the number of reactions per pathway or the number of pathways.

4. The method of Claim 1 wherein said user input data may also include designating the KEGG RPAIR constraints.

5. The method of Claim 1 wherein said user input data may also include selecting one or more compounds to exclude from consideration in the suitable pathway designation.

6. The method of Claim 1 wherein said determining step may include providing the user with possible suggestions on one or more heterologous enzymes that may increase the favorability of the reaction pathway.

7. The method of Claim 1 wherein said determining step may include consideration of thermodynamic criteria involved in one or more reactions.

8. The method of Claim 1 wherein dynamically ranking includes consideration of the integration of new reactions into the endogenous metabolic system

9. The method of Claim 1 further comprises the step of:

identifying and suggesting foreign enzymes that may be used to catalyze the desired foreign reactions to increase the efficiency in achieving the target end product.

10. The method of Claim 1 wherein the endogenous pathway score is the sum all of the reaction weights in a given pathway for the reaction, the pathway from the source compound to the target product, the number of reactions that are native and foreign to the host organism and whether the reactions are endogenous to the host.

11. The method of Claim 1 wherein said predetermined number of suitable heterologous biosynthesis pathways on the generated graph is ten.

12. The method of Claim 1 wherein said predetermined number of suitable heterologous biosynthesis pathways on the generated graph is thirty.

13. The method of Claim 1 wherein the graphical display includes color coding of vertices and edges to indicate starting and ending compounds, and which reactions are native or foreign to the host organism.

14. The method of Claim 1 wherein the graphical display includes a variation of the width of the arrow edges to indicate the value of the Gibbs energy, or strength, of the reaction path.

15. The method of Claim 1 wherein the graphical display includes dynamically changing the display based on the placement of a computer pointer over nodes or edges of compound names or the reaction Gibbs energy.

16. The method of Claim 1 further comprising the step of:

displaying a table of the specific reaction steps for the route with the reaction identification, reaction formula, whether the reaction is native to the host, the Gibbs energy of the reaction step, the native enzymes, potential foreign enzymes and data for the competing endogenous reactions.

17. The method of Claim 15 further comprising the step of:

selecting one or more enzymes from the table to generate and display enzyme data relating to the selected enzyme.

18. A system for determining and displaying suitable heterologous biosynthesis pathways to produce a target product in a specified host organism from a selected starting material, comprising:

an input terminal that receives selected user input data that includes designation of a host organism, a starting compound, and a desired target product into the pathway computer system;

a computer engine that searches for one or more possible heterologous pathways with consideration of competing endogenous reactions in a given host organism, determines one or more suitable heterologous biosynthesis pathways from the starting compound to achieve the desired target product in light of consideration of competing endogenous reactions in a given host organism, dynamically ranking suitable heterologous biosynthesis pathways from the starting compound to achieve the desired target product by an endogenous pathway score; and generates a graph of a predetermined number of most suitable heterologous biosynthesis pathways that shows reactions steps and metabolites, as well as competing endogenous reactions,

a display that displays one or more suitable heterologous biosynthesis pathways from the starting compound to the target product, in which vertices on the graph represent metabolites and arrow edges on the graph represent chemical transformations via verified metabolic reactions; said display of data supplied in graphical display being

utilized by said user to select a preferred heterologous biosynthesis pathway for a specified chemical transformation.

19. The system of Claim 18 wherein said determining step includes consideration of thermodynamic criteria for the foreign reaction in view of competing native reactions.

20. The system of Claim 18 wherein said user input data selected and input may also include selecting the number of reactions per pathway or the number of pathways.

21. The system of Claim 18 wherein said user input data may also include designating the KEGG RPAIR constraints.

22. The system of Claim 18 wherein said user input data may also include selecting one or more compounds to exclude from consideration in the suitable pathway designation.

23. The system of Claim 18 wherein said determining step may include providing the user with possible suggestions on one or more heterologous enzymes that may increase the favorability of the reaction pathway.

24. The system of Claim 18 wherein said determining step may include consideration of thermodynamic criteria involved in one or more reactions.

25. The system of Claim 18 wherein dynamically ranking includes consideration of the integration of new reactions into the endogenous metabolic system

26. The system of Claim 18 wherein said computer engine also identifies and suggests foreign enzymes that may be used to catalyze the desired foreign reactions to increase the efficiency in achieving the target end product.

27. The system of Claim 18 wherein the endogenous pathway score is the sum all of the reaction weights in a given pathway for the reaction, the pathway from the source compound to the target product, the number of reactions that are native and foreign to the host organism and whether the reactions are endogenous to the host.

28. The system of Claim 18 wherein said predetermined number of suitable heterologous biosynthesis pathways on the generated graph is ten.

29. The system of Claim 18 wherein said predetermined number of suitable heterologous biosynthesis pathways on the generated graph is thirty.

30. The system of Claim 18 wherein the graphical display includes color coding of vertices and edges to indicate starting and ending compounds, and which reactions are native or foreign to the host organism.

31. The system of Claim 18 wherein the graphical display includes a variation of the width of the arrow edges to indicate the value of the Gibbs energy, or strength, of the reaction path.

32. The system of Claim 18 wherein the graphical display includes dynamically changing the display based on the placement of a computer pointer over nodes or edges of compound names or the reaction Gibbs energy.

33. The system of Claim 18 wherein the display shows a table of the specific reaction steps for the route with the reaction identification, reaction formula, whether the reaction is native to the host, the Gibbs energy of the reaction step, the native enzymes, potential foreign enzymes and data for the competing endogenous reactions.

34. The system of Claim 33 wherein the computer engine selects one or more enzymes from the table to generate and display enzyme data relating to the selected enzyme.

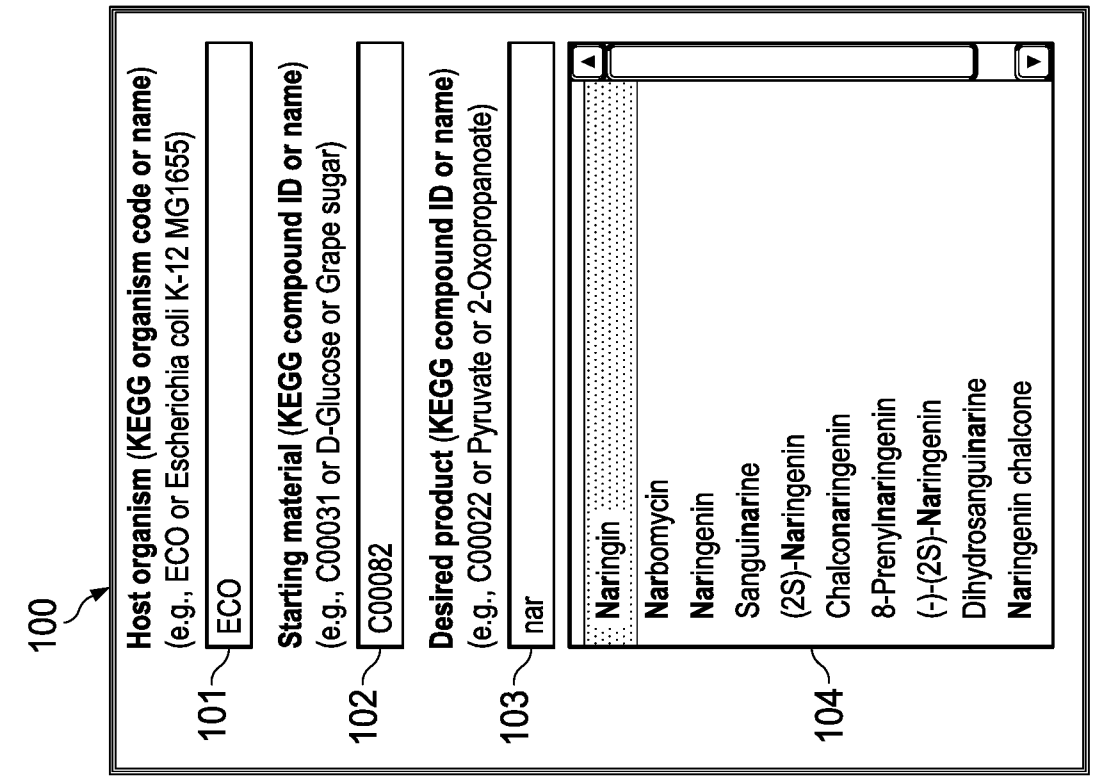

100
Host organism (KEGG organism code or name)
(e.g., ECO or Escherichia coli K-12 MG1655)
101
ECO
Starting material (KEGG compound ID or name)
(e.g., C00031 or D-Glucose or Grape sugar)
102
C00082
Desired product (KEGG compound ID or name)
(e.g., C00022 or Pyruvate or 2-Oxopropanoate)
103
nar
104
Naringin
Narbomycin
Naringenin
Sanguinarine
(2S)-Naringenin
Chalconaringenin
8-Prenylnaringenin
(-)-(2S)-Naringenin
Dihydrosanguinarine
Naringenin chalcone

FIG. 1A

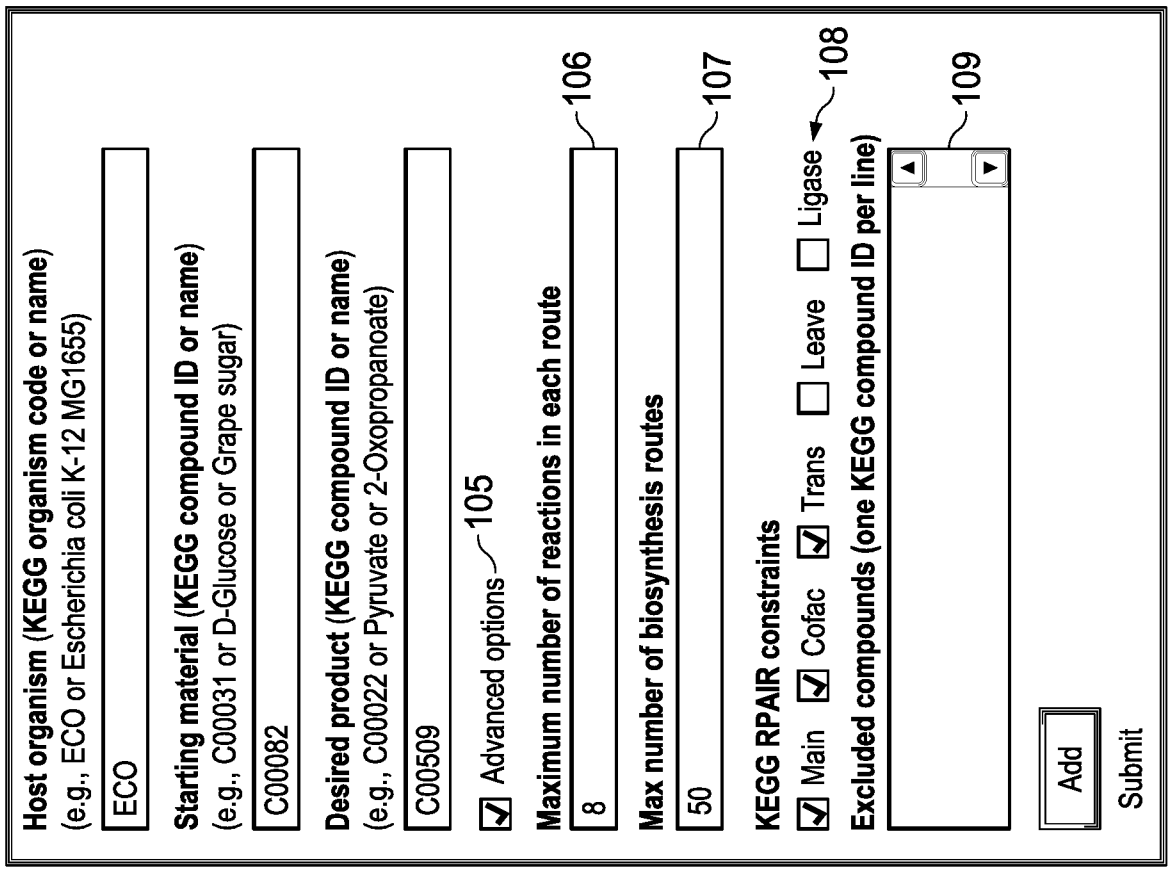
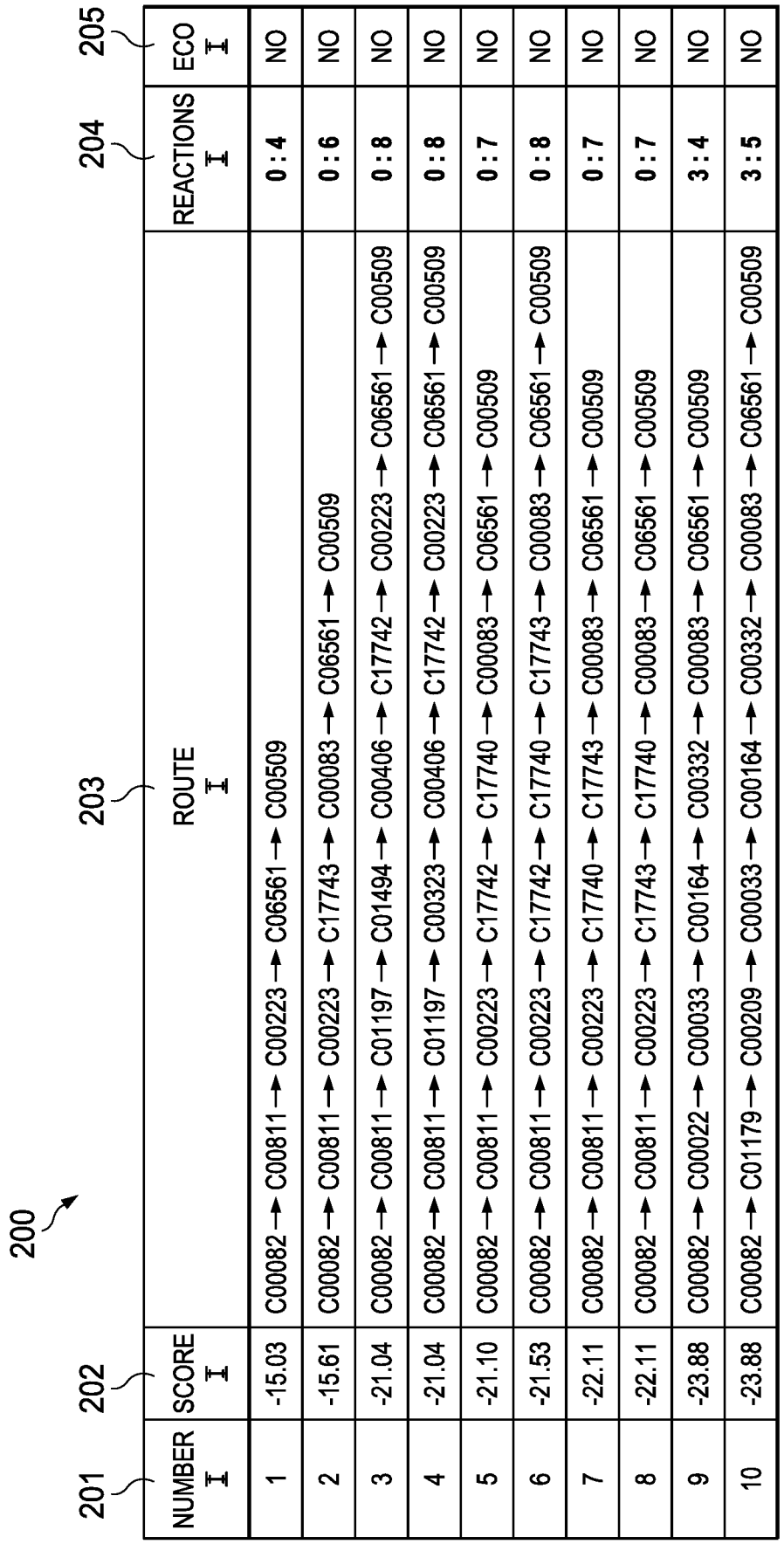

Host organism (KEGG organism code or name)
(e.g., ECO or Escherichia coli K-12 MG1655)
ECO
Starting material (KEGG compound ID or name)
(e.g., C00031 or D-Glucose or Grape sugar)
C00082
Desired product (KEGG compound ID or name)
(e.g., C00022 or Pyruvate or 2-Oxopropanoate)
C00509
Advanced options
105
Maximum number of reactions in each route
106
8
Max number of biosynthesis routes
107
50
KEGG RPAIR constraints
Main
Cofac
Trans
Leave
Ligase
108
Excluded compounds (one KEGG compound ID per line)
109
Add
Submit

FIG. 1B



NUMBER	SCORE	ROUTE	REACTIONS	ECO
1	-15.03	C00082 → C00811 → C00223 → C06561 → C00509	0 : 4	NO
2	-15.61	C00082 → C00811 → C00223 → C17743 → C00083 → C06561 → C00509	0 : 6	NO
3	-21.04	C00082 → C00811 → C01197 → C01494 → C00406 → C17742 → C00223 → C06561 → C00509	0 : 8	NO
4	-21.04	C00082 → C00811 → C01197 → C00323 → C00406 → C17742 → C00223 → C06561 → C00509	0 : 8	NO
5	-21.10	C00082 → C00811 → C00223 → C17742 → C17740 → C00083 → C06561 → C00509	0 : 7	NO
6	-21.53	C00082 → C00811 → C00223 → C17742 → C17740 → C17743 → C00083 → C06561 → C00509	0 : 8	NO
7	-22.11	C00082 → C00811 → C00223 → C17740 → C17743 → C00083 → C06561 → C00509	0 : 7	NO
8	-22.11	C00082 → C00811 → C00223 → C17743 → C17740 → C00083 → C06561 → C00509	0 : 7	NO
9	-23.88	C00082 → C00022 → C00033 → C00164 → C00332 → C00083 → C06561 → C00509	3 : 4	NO
10	-23.88	C00082 → C01179 → C00209 → C00033 → C00164 → C00332 → C00083 → C06561 → C00509	3 : 5	NO

FIG. 2

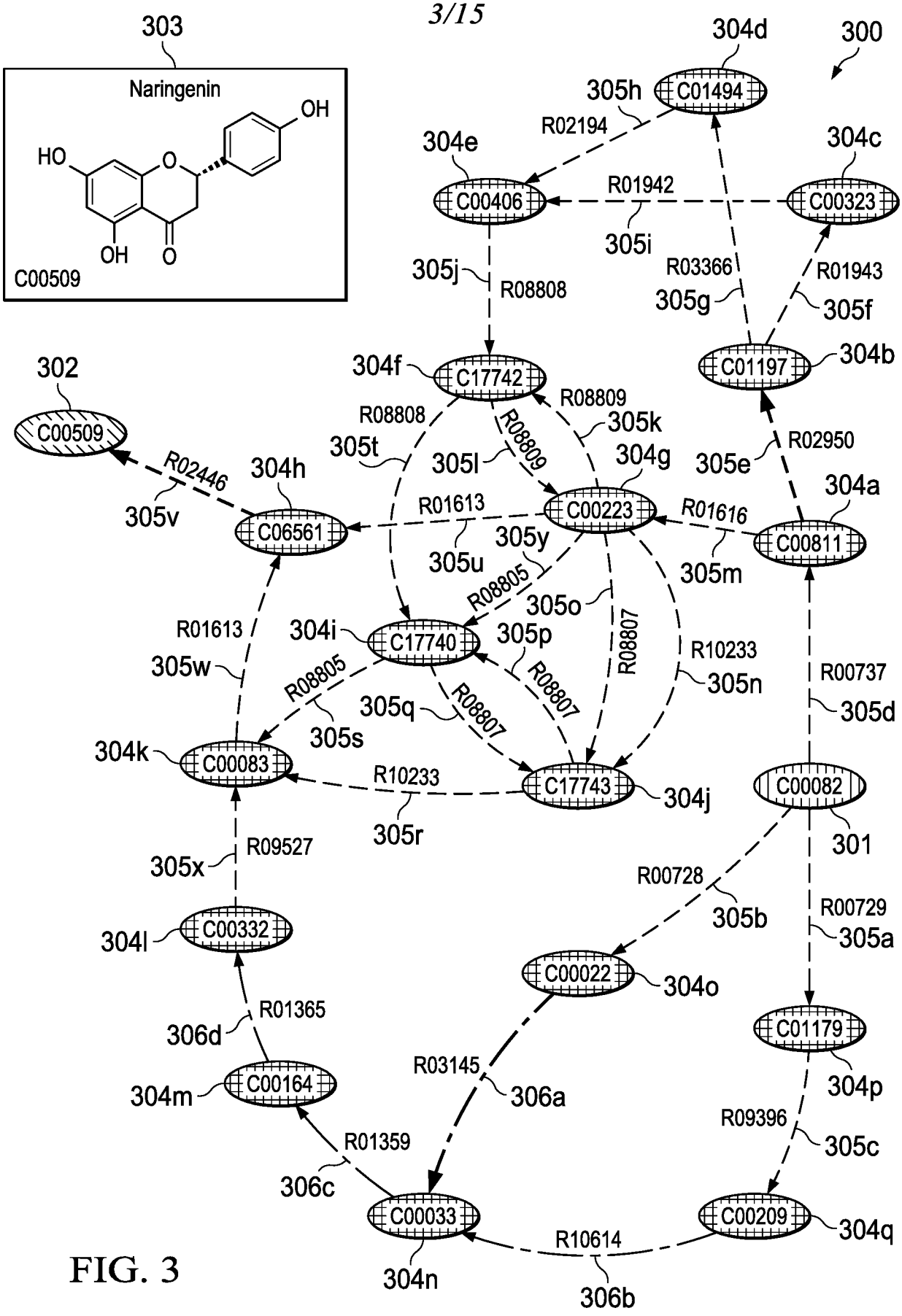

303
Naringenin
C00509
300
304d
C01494
305h
R02194
304e
C00406
304c
C00323
R01942
305i
305j
R08808
R03366
305g
R01943
305f
C01197
304b
302
C00509
304f
C17742
R08809
R08808
305k
305t
305l
304g
305e
R02950
R02446
304h
305v
C06561
R01613
C00223
R01616
C00811
304a
305y
305u
305m
R08805
305o
R01613
304i
C17740
305p
R10233
305w
R08805
R08807
305n
R00737
305s
305q
R08807
305d
304k
C00083
R10233
C17743
304j
C00082
305r
301
R09527
305x
R00728
R00729
305b
305a
304l
C00332
C00022
304o
R01365
C01179
306d
R03145
304p
304m
C00164
306a
R09396
305c
R01359
306c
C00033
C00209
304q
R10614
304n
306b

FIG. 3

409 Reaction ID (compound ID)	410 Formula	411 ECO	412 Gibbs Energy	413 EC #	414 Potential Enzyme	415 Competing Reaction (Gibbs Energy)
R00737 (C00082)	**C00082** ↔ **C00811** + **C00014**	No	-38.0617	**4.3.1.23**	**hha:Hhal_1820** and **more**	**R00734** (0.2967) and **more**
R01616 (C00811)	**C00002** + **C00811** + **C00010** ↔ **C00020** + **C00013** + **C00223**	No	-9.7	**6.2.1.12**	**soc:105193785** and **more**	None
R01613 (C00223)	**C00223** + 3 **C00083** ↔ **C06561** + 4 **C00010** + 3 **C00011**	No	37.1831	**2.3.1.74**	**ath:AT5G13930** (TT4) and **more**	**R00742** (4.4248) and **more**
R02446 (C06561)	**C06561** ↔ **C00509**	No	-236.4537	**5.5.1.6**	**ath:AT3G55120** (TT5) and **more**	None

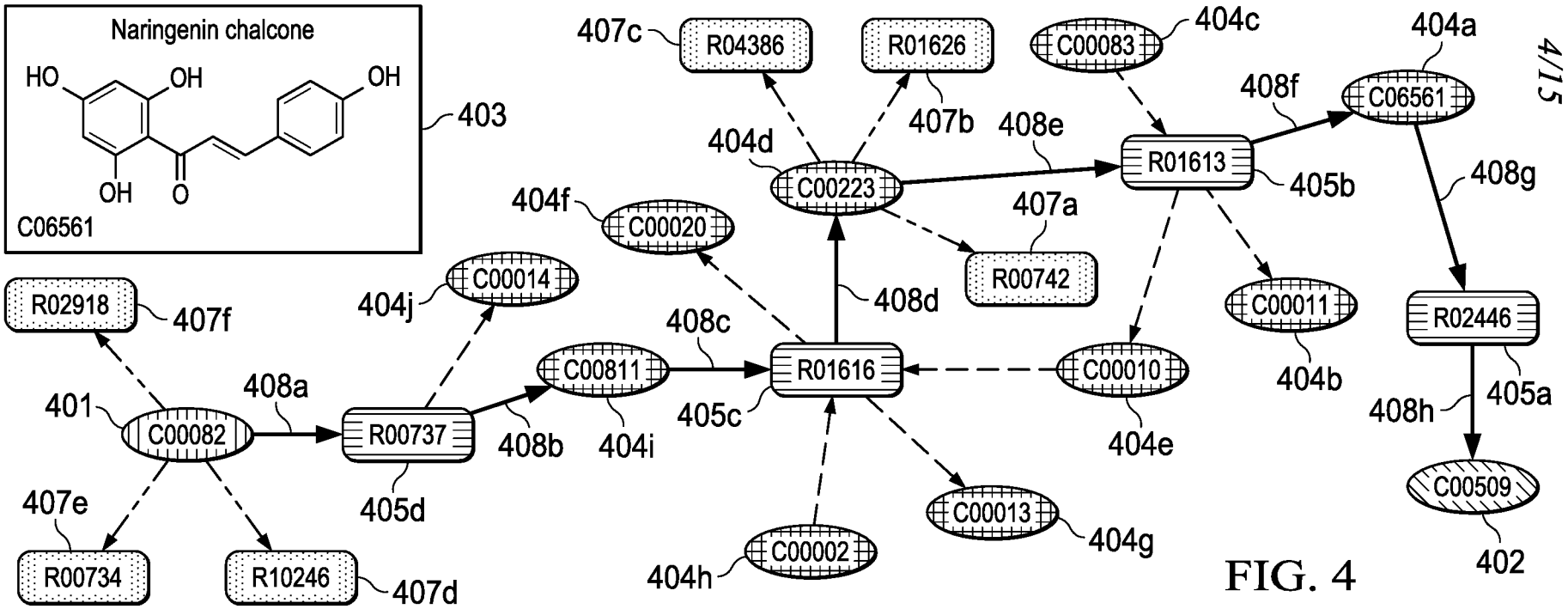


FIG. 4

500

List of the endogenous reactions competing with reaction **Malonyl-CoA:4-Coumaroyl-CoA malonyltransferase(cyclizing)(R01613)** for compound **Malonyl-CoA (C00083)** in **Escherichia coli K-12 MG1655(ECO):**

Reaction	Gibbs Energy	Enzyme Gene ID
acetyl-CoA:carbon dioxide ligase (ADP-forming) (R00742)	4.4248	**b0185**(accA) **b2316**(accD) **b3256**(accC)
carboxybiotin-carboxyl-carrier-protein-CoA carboxytranferase (R04386)	not determined	**b0185**(accA) **b2316**(accD) **b3256**(accC)
Malonyl-CoA:[acyl-carrier-protein] S-malonyltransferase (R01626)	not determined	**b1092**(fabD)

501 502 503

FIG. 5

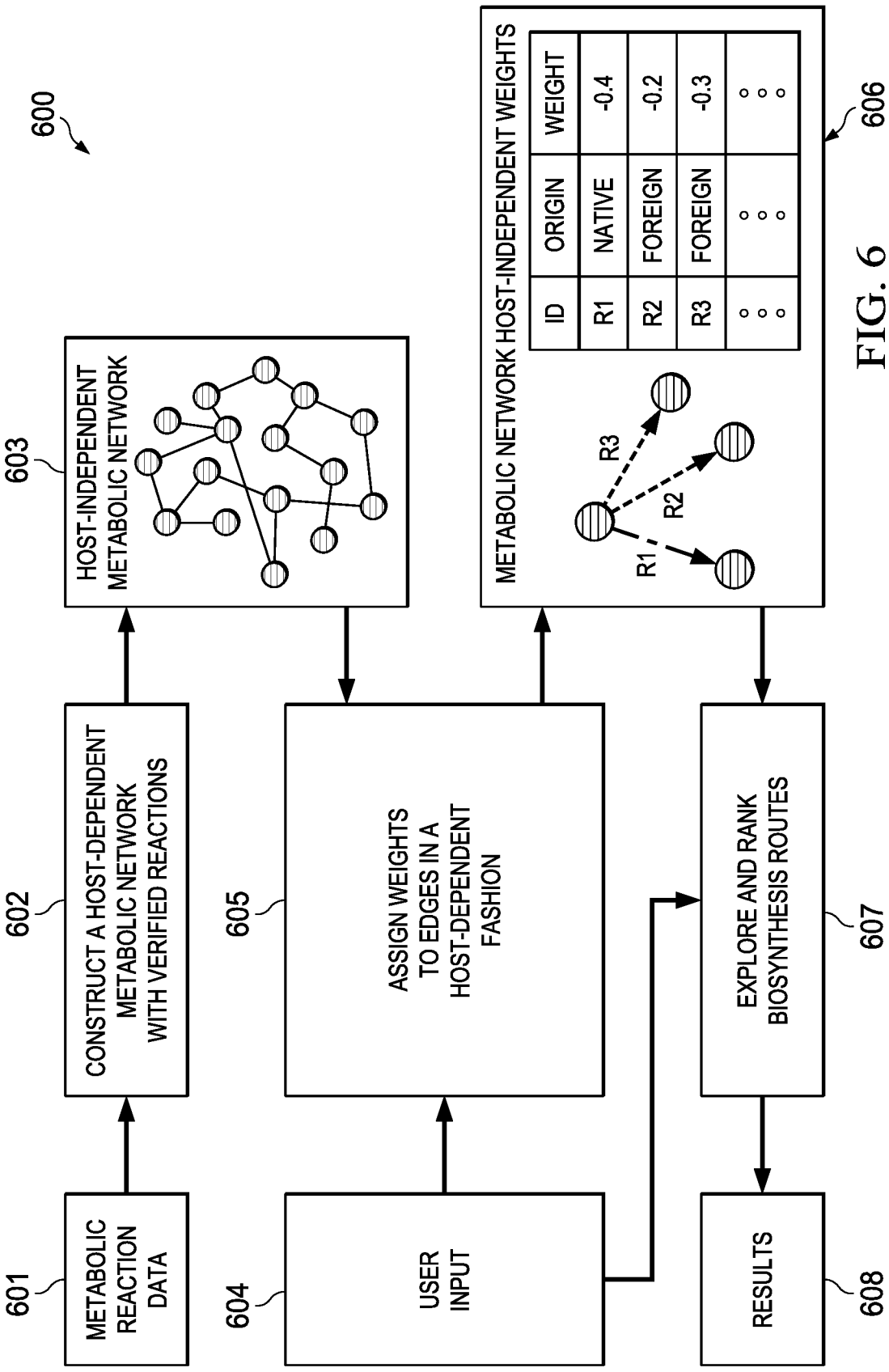

600
601
METABOLIC REACTION DATA
602
CONSTRUCT A HOST-DEPENDENT METABOLIC NETWORK WITH VERIFIED REACTIONS
603
HOST-INDEPENDENT METABOLIC NETWORK
604
USER INPUT
605
ASSIGN WEIGHTS TO EDGES IN A HOST-DEPENDENT FASHION
606
METABOLIC NETWORK HOST-INDEPENDENT WEIGHTS
ID
ORIGIN
WEIGHT
R1
NATIVE
-0.4
R2
FOREIGN
-0.2
R3
FOREIGN
-0.3
R1
R2
R3
607
EXPLORE AND RANK BIOSYNTHESIS ROUTES
608
RESULTS

FIG. 6

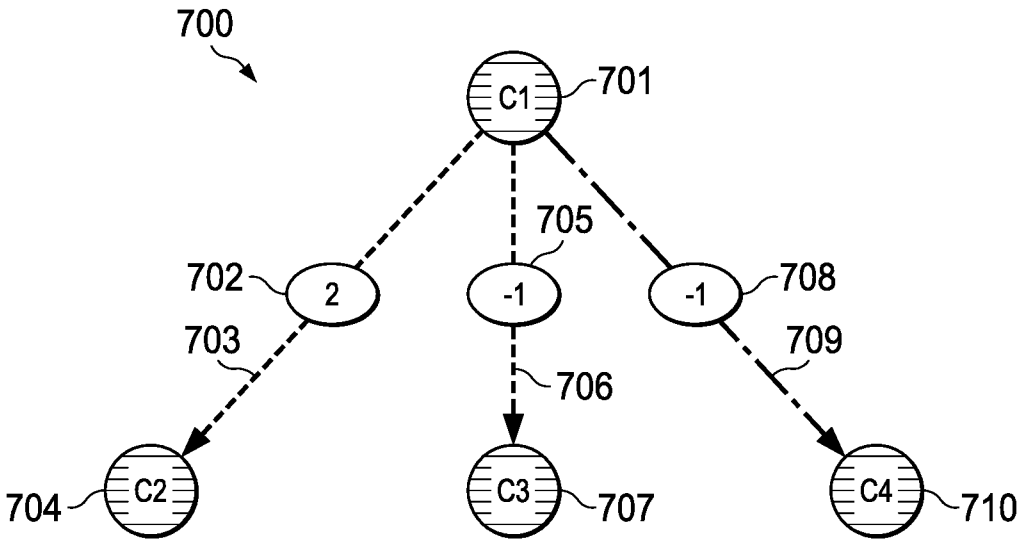

700
C1
701
705
702
2
-1
-1
708
703
709
706
704
C2
C3
707
C4
710

FIG. 7

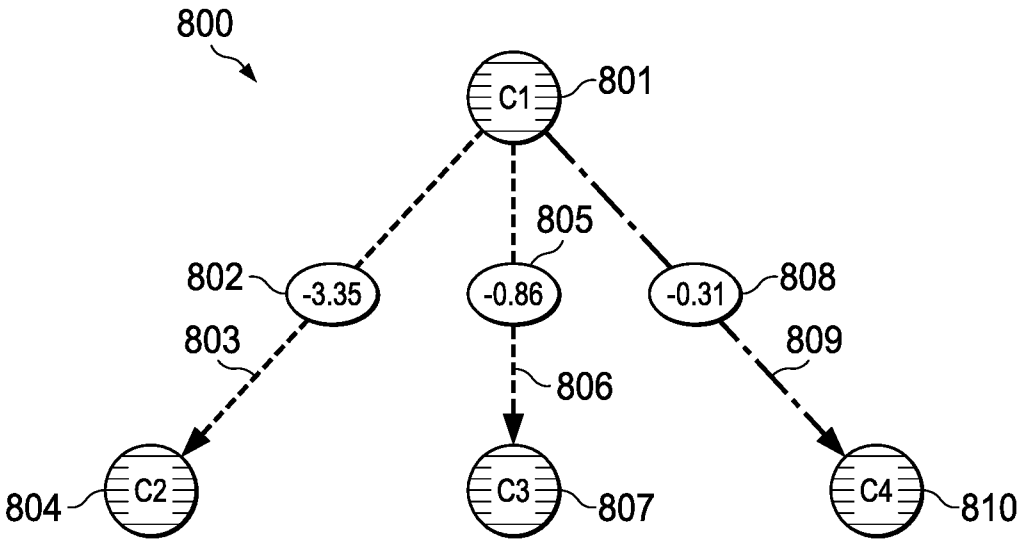

800
C1
801
805
802
-3.35
-0.86
-0.31
808
803
809
806
804
C2
C3
807
C4
810

FIG. 8

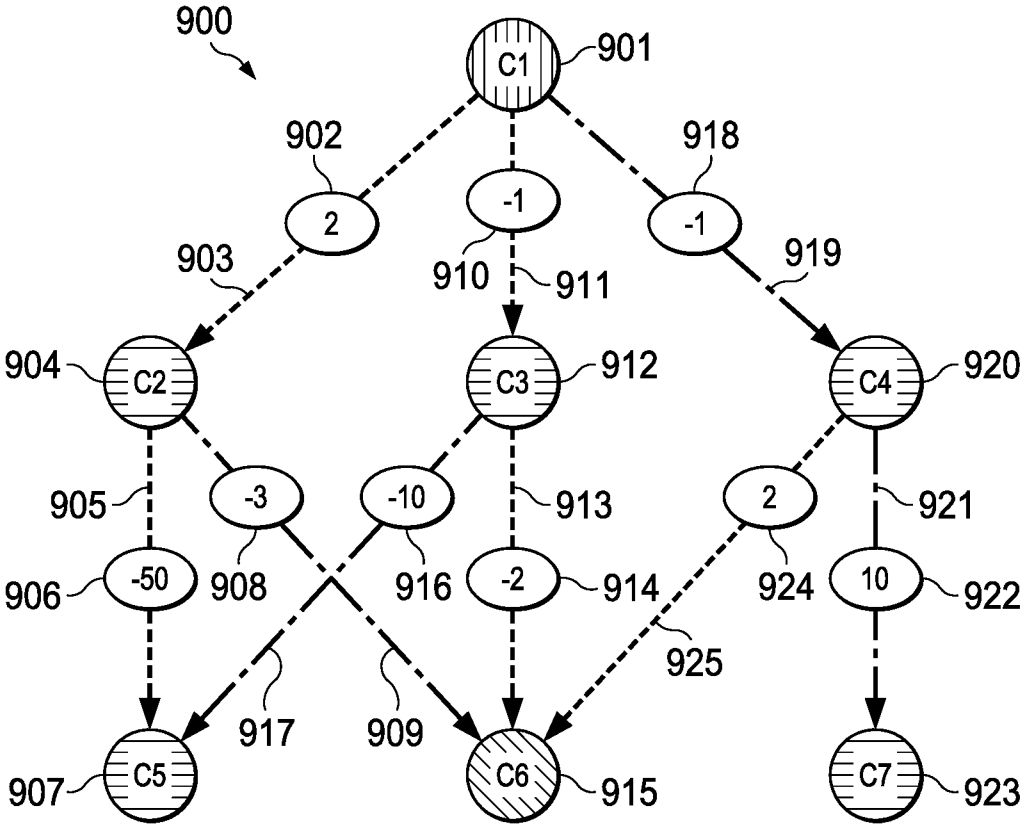

900
C1
901
902
2
903
-1
910
911
918
-1
919
904
C2
C3
912
C4
920
905
-3
-10
913
2
921
906
-50
908
916
-2
914
924
10
922
925
917
909
907
C5
C6
915
C7
923

FIG. 9

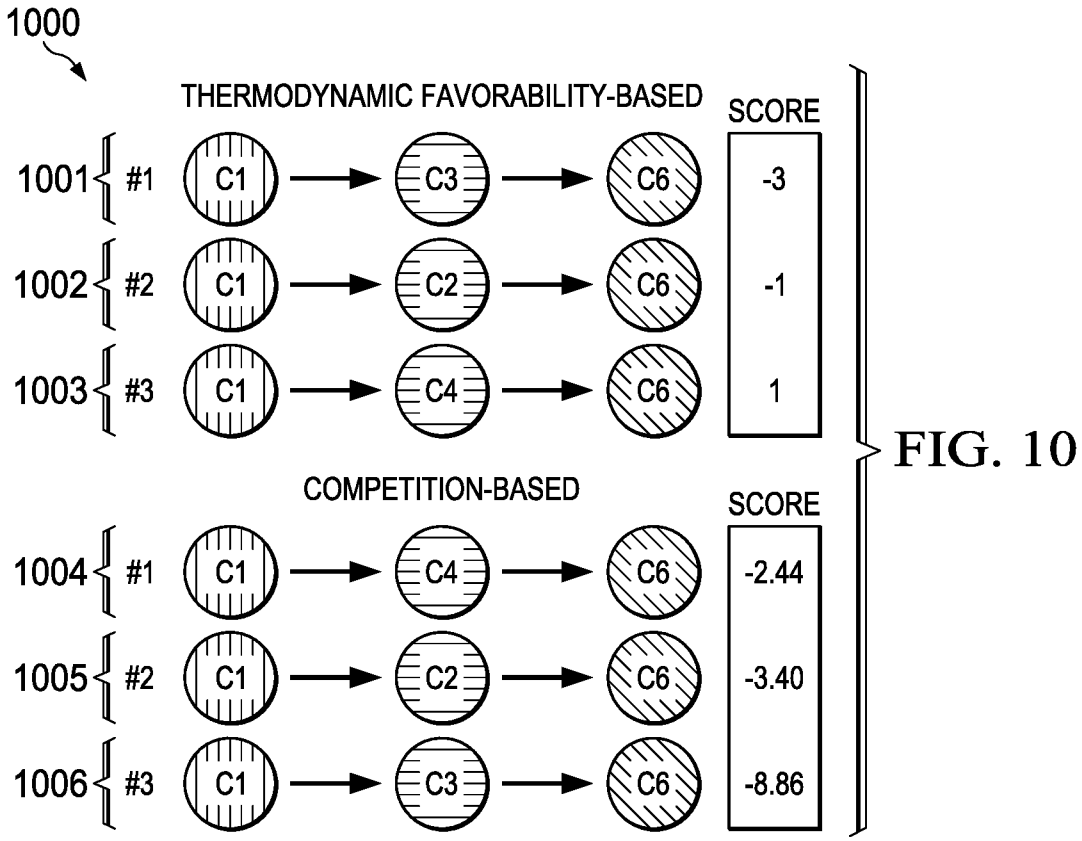
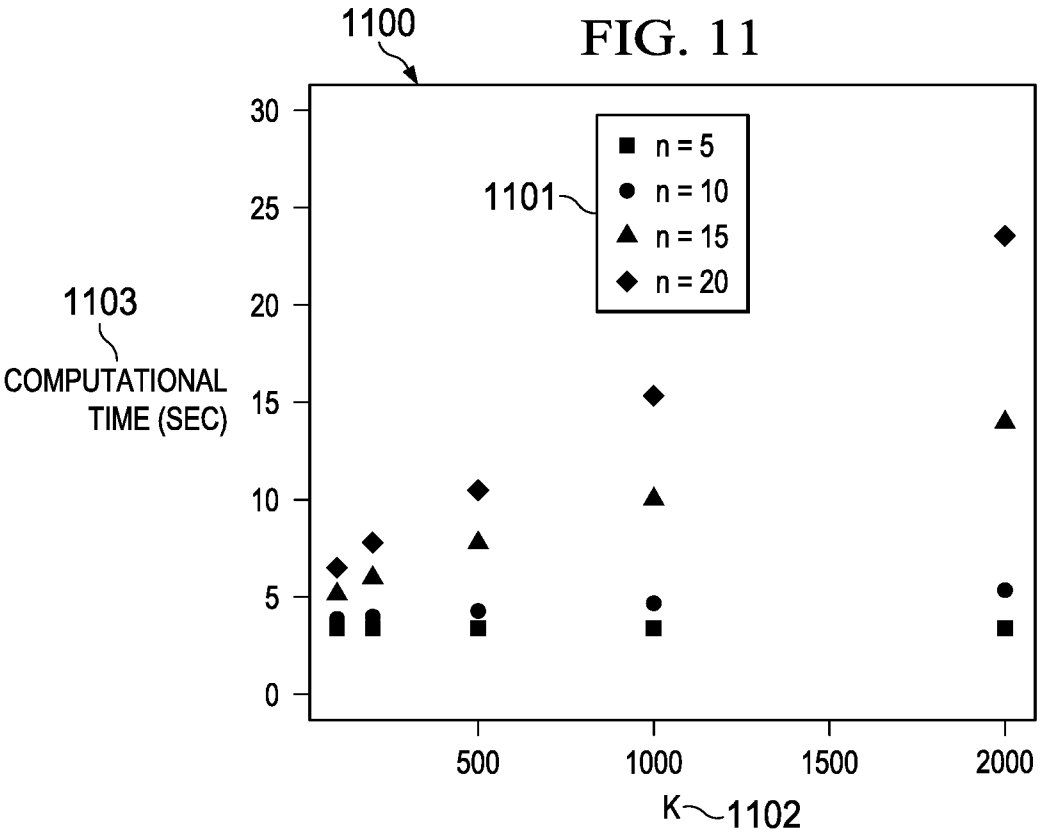

1000
THERMODYNAMIC FAVORABILITY-BASED
SCORE
1001 #1 C1 C3 C6 -3
1002 #2 C1 C2 C6 -1
1003 #3 C1 C4 C6 1
COMPETITION-BASED
SCORE
1004 #1 C1 C4 C6 -2.44
1005 #2 C1 C2 C6 -3.40
1006 #3 C1 C3 C6 -8.86

FIG. 10

FIG. 11


1100
1101
n = 5
n = 10
n = 15
n = 20
1103
COMPUTATIONAL TIME (SEC)
30
25
20
15
10
5
0
500
1000
1500
2000
K
1102

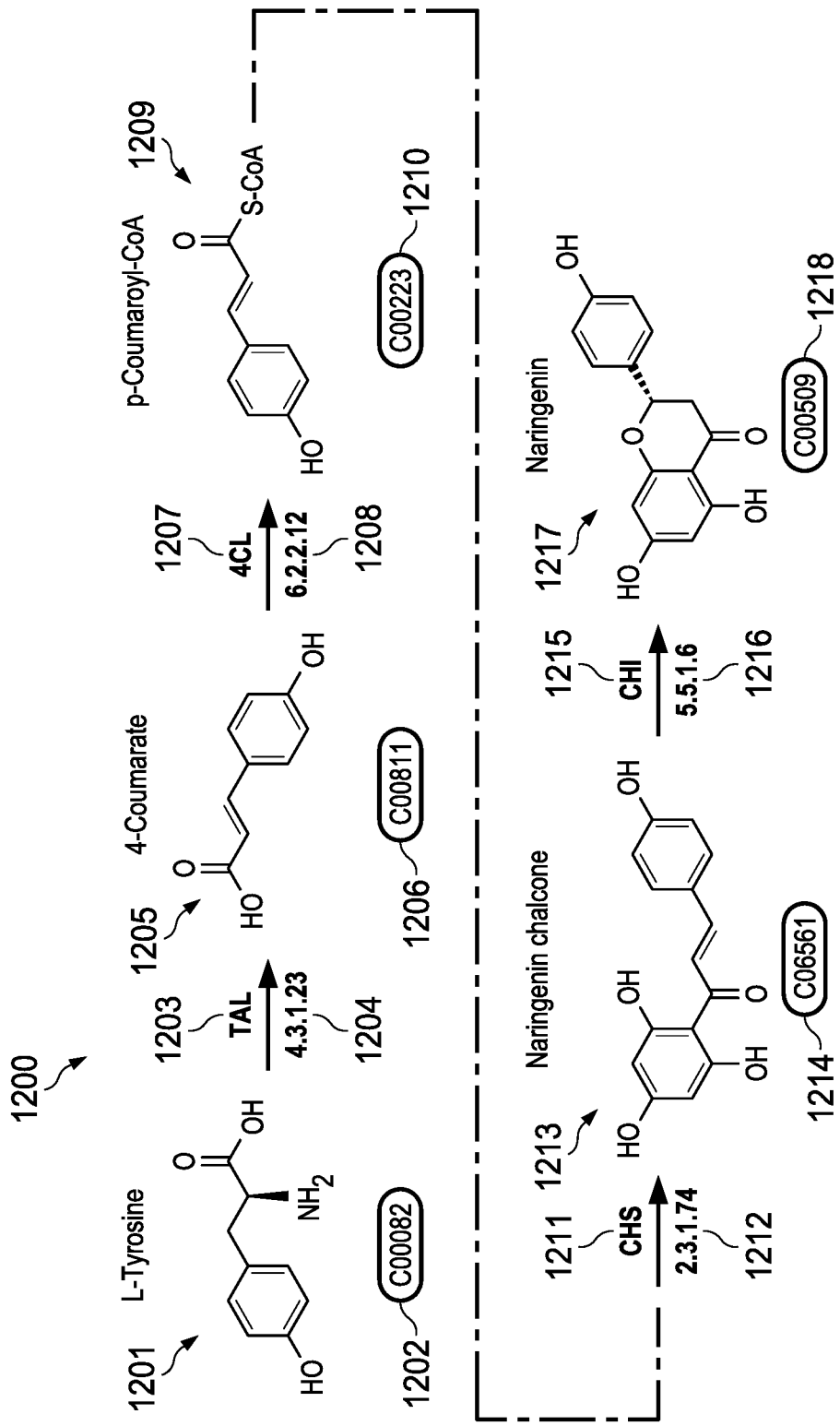

1200
1201
L-Tyrosine
1202
C00082
1203
TAL
4.3.1.23
1204
1205
4-Coumarate
1206
C00811
1207
4CL
6.2.2.12
1208
1209
p-Coumaroyl-CoA
1210
C00223
1211
CHS
2.3.1.74
1212
1213
Naringenin chalcone
1214
C06561
1215
CHI
5.5.1.6
1216
1217
Naringenin
1218
C00509

FIG. 12

1300

Pathway Number 1: C00082 (R00737) C00811 (R01616) C00223 (R01613) C06561 (R02446) C00509
Download all potential cDNA gene sequences in fasta format for this pathway

Reaction ID (Compound ID)	Formula	ECO	Gibbs Energy	EC #	Potential Enzyme	Competing Reaction (Gibbs Energy)
R00737 (C00082)	**C00082 ⟷ C00811 + C00014**	No	-38.0617	**4.3.1.23**	**hha:Hhal_1820** and **more**	**R00734** (0.2967) and **more**
R01616 (C00811)	**C00002 + C00811 + C00010 ⟷ C00020 + C00013 + C00223**	No	-9.7	**6.2.1.12**	**soc:105193785** and **more**	None
R01613 (C00223)	**C00223** + 3 **C00083 ⟷ C06561** + 4 **C00010** + 3 **C00011**	No	37.1831	**2.3.1.74**	**ath:AT5G13930** (TT4) and **more**	**R00742** (4.4248) and **more**
R02446 (C06561)	**C06561 ⟷ C00509**	No	-236.4537	**5.5.1.6**	**ath:AT3G55120** (TT5) and **more**	None

1301 1302 1303 1304 1305 1306 1307

FIG. 13

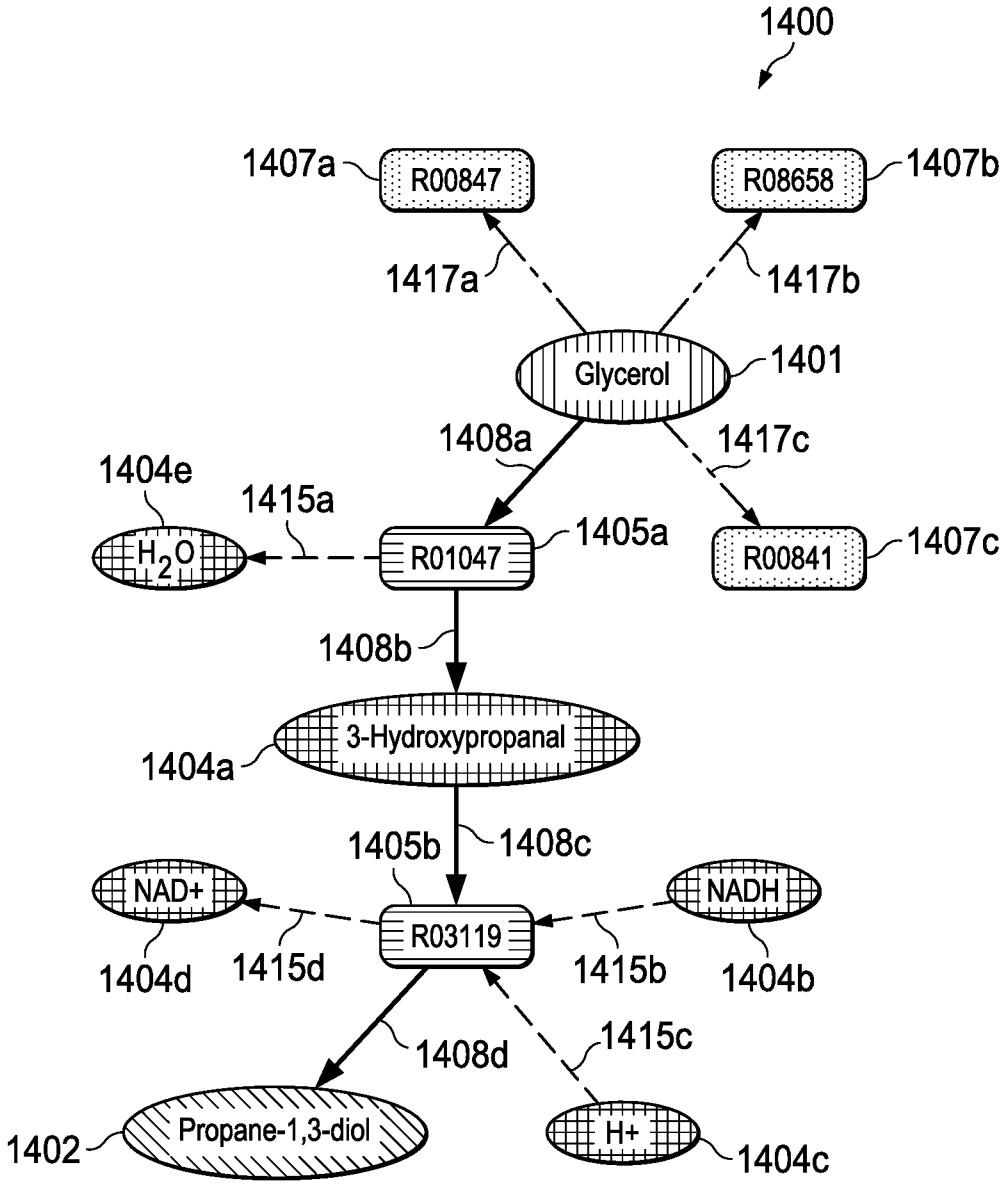

1400
1407a
R00847
R08658
1407b
1417a
1417b
Glycerol
1401
1408a
1417c
1404e
1415a
H_2O
R01047
1405a
R00841
1407c
1408b
3-Hydroxypropanal
1404a
1405b
1408c
NAD+
NADH
R03119
1404d
1415d
1415b
1404b
1408d
1415c
1402
Propane-1,3-diol
H+
1404c

FIG. 14

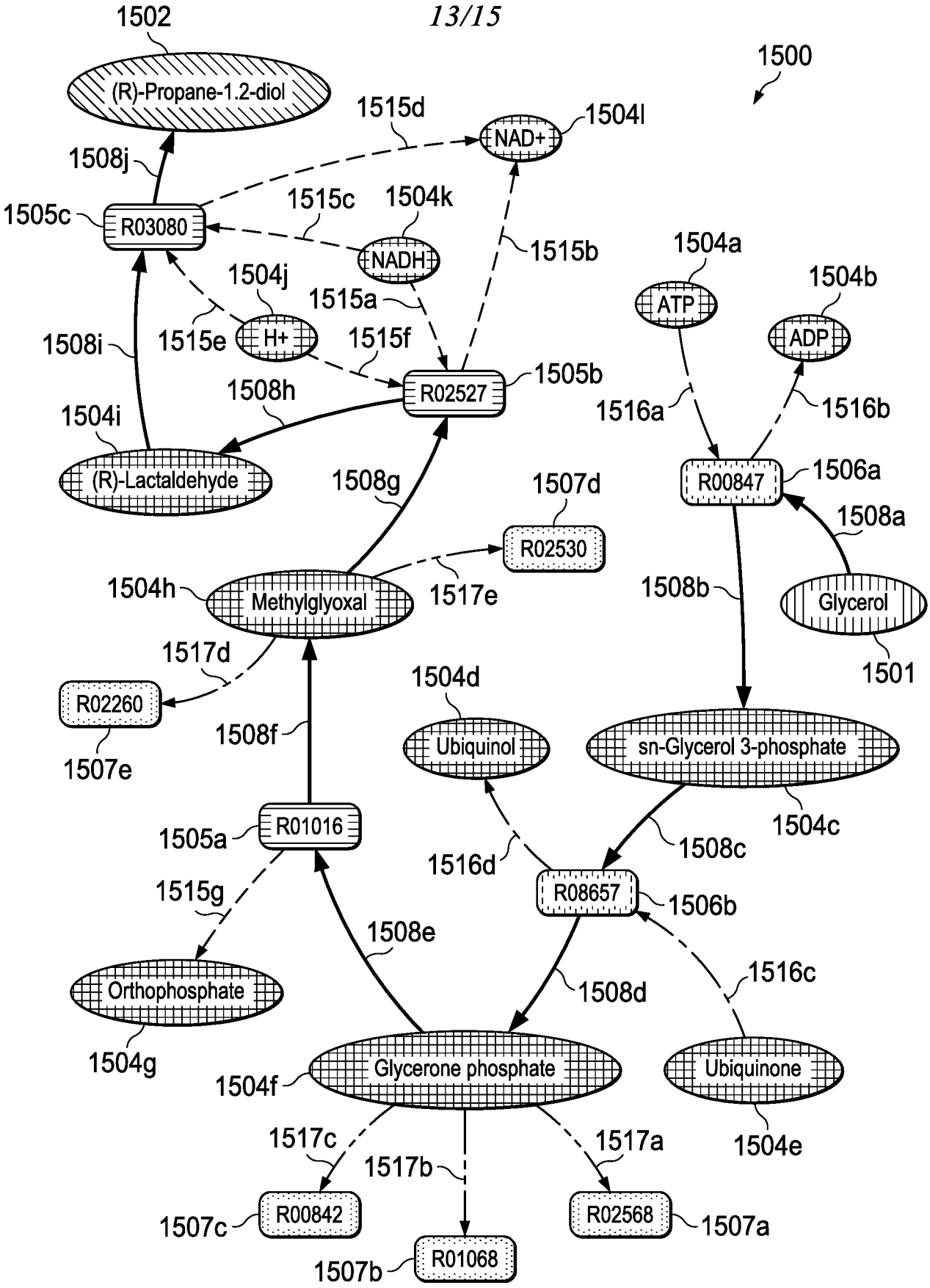

1502
1500
(R)-Propane-1.2-diol
1515d
NAD+
1504l
1508j
1505c
R03080
1515c
1504k
NADH
1515b
1504a
1504b
ATP
ADP
1504j
1515a
1508i
1515e
H+
1515f
R02527
1505b
1516a
1516b
1508h
1504i
(R)-Lactaldehyde
1508g
1507d
R00847
1506a
1508a
R02530
1504h
Methylglyoxal
1517e
1508b
Glycerol
1501
1517d
R02260
1507e
1504d
Ubiquinol
sn-Glycerol 3-phosphate
1504c
1508f
1505a
R01016
1516d
1508c
1515g
R08657
1506b
1508e
1516c
Orthophosphate
1504g
1508d
1504f
Glycerone phosphate
Ubiquinone
1504e
1517c
1517b
1517a
1507c
R00842
R02568
1507a
1507b
R01068

FIG. 15

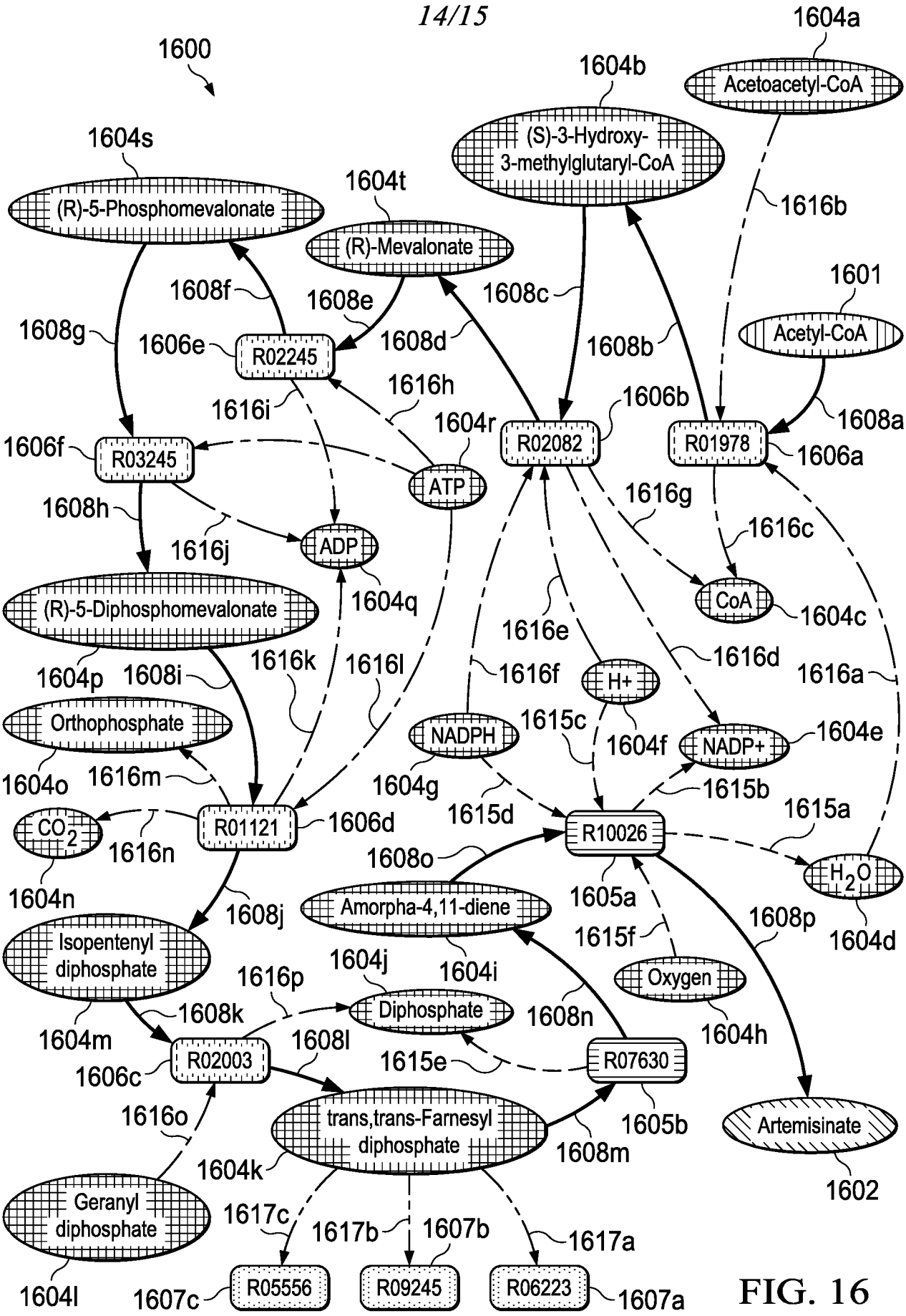

1600
1604a
Acetoacetyl-CoA
1604b
(S)-3-Hydroxy-3-methylglutaryl-CoA
1604s
(R)-5-Phosphomevalonate
1604t
(R)-Mevalonate
1616b
1601
Acetyl-CoA
1608f
1608e
1608c
1608g
1606e
R02245
1608d
1608b
1616h
1616i
1606b
1604r
1608a
1606f
R03245
R02082
R01978
1606a
ATP
1608h
1616g
1616c
1616j
ADP
(R)-5-Diphosphomevalonate
1604q
CoA
1604c
1616e
1604p
1608i
1616k
1616l
1616f
H+
1616d
1616a
Orthophosphate
NADPH
1615c
1604f
NADP+
1604e
1604o
1616m
1604g
1615b
1615d
1615a
CO2
R01121
1606d
R10026
1616n
1608o
H2O
1604n
1605a
1608p
1608j
Amorpha-4,11-diene
1615f
1604d
Isopentenyl diphosphate
1604j
1604i
1616p
Oxygen
1608k
Diphosphate
1608n
1604h
1604m
1608l
1615e
R07630
R02003
1606c
1616o
trans,trans-Farnesyl diphosphate
1605b
Artemisinate
1608m
1604k
1602
Geranyl diphosphate
1617c
1617b
1607b
1617a
1604l
1607c
R05556
R09245
R06223
1607a

FIG. 16

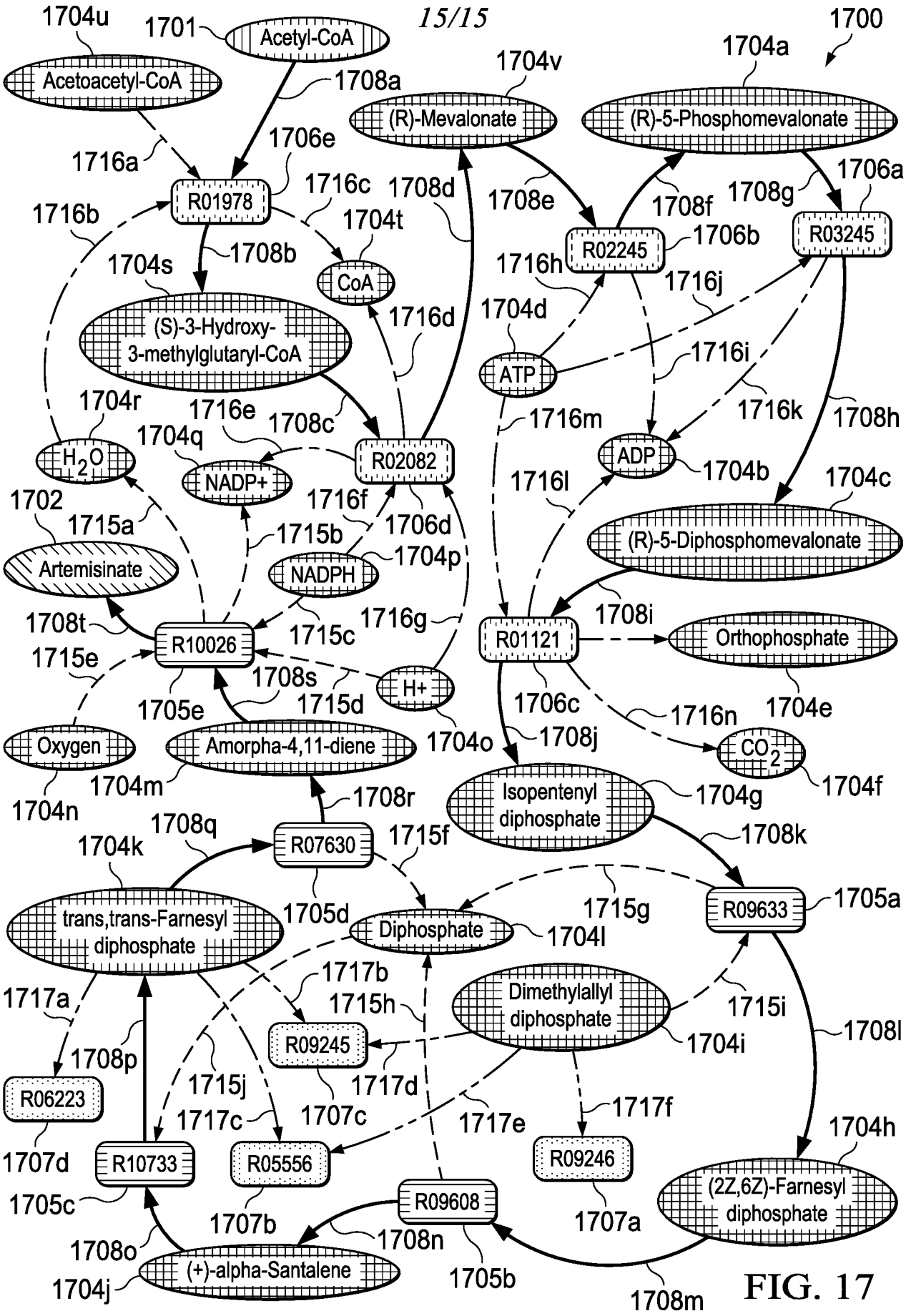

1700
1701
Acetyl-CoA
1702
1704a
(R)-5-Phosphomevalonate
1704b
ADP
1704c
(R)-5-Diphosphomevalonate
1704d
ATP
1704e
Orthophosphate
1704f
CO2
1704g
Isopentenyl diphosphate
1704h
(2Z,6Z)-Farnesyl diphosphate
1704i
Dimethylallyl diphosphate
1704j
(+)-alpha-Santalene
1704k
trans,trans-Farnesyl diphosphate
1704l
Diphosphate
1704m
Oxygen
1704n
1704o
Amorpha-4,11-diene
1704p
NADPH
1704q
NADP+
1704r
H2O
1704s
(S)-3-Hydroxy-3-methylglutaryl-CoA
1704t
CoA
1704u
Acetoacetyl-CoA
1704v
(R)-Mevalonate
1705a
R09633
1705b
1705c
R10733
1705d
1705e
1706a
R03245
1706b
R02245
1706c
1706d
1706e
R01978
1707a
R09246
1707b
R09608
1707c
R05556
1707d
R06223
1708a
1708b
1708c
1708d
1708e
1708f
1708g
1708h
1708i
R01121
1708j
1708k
1708l
1708m
1708n
1708o
1708p
1708q
R07630
1708r
1708s
R10026
1708t
Artemisinate
1715a
1715b
1715c
1715d
1715e
1715f
1715g
1715h
1715i
1715j
R09245
1716a
1716b
1716c
1716d
1716e
1716f
1716g
1716h
1716i
1716j
1716k
1716l
1716m
1716n
1717a
1717b
1717c
1717d
1717e
1717f
H+

FIG. 17

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2017/050576

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F19/12
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal , WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	HIROYUKI KUWAHARA ET AL: "MRE: a web tool to suggest foreign enzymes for the biosynthesis pathway design with competing endogenous reactions in mind", NUCLEIC ACIDS RESEARCH, vol. 44, no. W1, 29 April 2016 (2016-04-29), pages W217-W225, XP055364895, ISSN: 0305-1048, DOI: 10.1093/nar/gkw342 the whole document	1-34

-/- ·

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 April 2017	04/05/2017

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Konak, Eyiip

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2017/050576

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	T. BLUM ET AL: "MetaRoute: fast search for relevant metabolic routes for interactive network navigation and visualization", BIOINFORMATICS., vol. 24, no. 18, 16 July 2008 (2008-07-16), pages 2108-2109, XP055364897, GB ISSN: 1367-4803, D0I: 10.1093/bioinformatics/btn360 the whole document	1-34
A	G. R0DRIG0 ET AL: "DESHARKY: automatic design of metabolic pathways for optimal cell growth", BIOINFORMATICS., vol. 24, no. 21, 1 November 2008 (2008-11-01), pages 2554-2556, XP055364899, GB ISSN: 1367-4803, D0I: 10.1093/bioinformatics/btn471 the whole document	1-34
A	PITKÃ NEN ESA ET AL: "Inferring branching pathways in genome-scale metabolic networks", BMC SYSTEMS BIOLOGY, BIOMED CENTRAL LTD, L0, vol. 3, no. 1, 29 October 2009 (2009-10-29), page 103, XP021064661, ISSN: 1752-0509, D0I: 10.1186/1752-0509-3-103 the whole document	1-34
A	R0DRIG0 LIBERAL ET AL: "PathwayBooster: a tool to support the curation of metabolic pathways", BMC BIOINFORMATICS, BIOMED CENTRAL, LONDON, GB, vol. 16, no. 1, 15 March 2015 (2015-03-15), page 86, XP021217658, ISSN: 1471-2105, D0I: 10.1186/S12859-014-0447-2 the whole document	1-34

1