(12) STANDARD PATENT (11) Application No. **AU 2013317782 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Compositions and methods for producing chemicals and derivatives thereof

(51) International Patent Classification(s)
C12P 7/58 (2006.01) ***C12N 9/04*** (2006.01)

(21) Application No: **2013317782** (22) Date of Filing: **2013.09.20**

(87) WIPO No: **WO14/047510**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/704,408	**2012.09.21**	**US**

(43) Publication Date: **2014.03.27**
(44) Accepted Journal Date: **2016.10.06**

(71) Applicant(s)
Synthetic Genomics, Inc.

(72) Inventor(s)
Kambourakis, Spiros;Griffin, Benjamin M.;Martin, Kevin V.

(74) Agent / Attorney
FB Rice, Level 23 44 Market Street, Sydney, NSW, 2000

(56) Related Art
US2011/0124065 A1

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 March 2014 (27.03.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/047510 A1

(51) International Patent Classification:
C12P 7/58 (2006.01) *C12N 9/04* (2006.01)

(21) International Application Number:
PCT/US2013/061036

(22) International Filing Date:
20 September 2013 (20.09.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/704,408 21 September 2012 (21.09.2012) US

(71) Applicant: **SYNTHETIC GENOMICS, INC.** [US/US]; 11149 North Torrey Pines Road, Suite 100, La Jolla, CA 92037 (US).

(72) Inventors: **KAMBOURAKIS, Spiros**; 4885 Kensington Drive, San Diego, CA 92116 (US). **GRIFFIN, Benjamin M.**; 3488 Voyager Circle, San Diego, CA 92130 (US). **MARTIN, Kevin V.**; 246 Barbara Avenue, Solana Beach, CA 92075 (US).

(74) Agents: **HAILE, Lisa A.** et al.; DLA Piper LLP (US), 4365 Executive Drive, Suite 1100, San Diego, CA 92121-2133 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, 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 4.17(ii))*

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

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2014/047510 A1

(54) Title: COMPOSITIONS AND METHODS FOR PRODUCING CHEMICALS AND DERIVATIVES THEREOF

(57) Abstract: The present invention provides methods for producing a product of one or more enzymatic pathways. The methods include both enzymatic and chemical conversions as steps. Various pathways are provided for converting glucose into 5-dehdyro-4-deoxyglucarate (DDG), and for converting glucose into 2,5-furandicarboxylic acid (FDCA). The methods also involve the use of engineered enzymes that perform reactions with high specificity and efficiency.

COMPOSITIONS AND METHODS FOR PRODUCING CHEMICALS AND DERIVATIVES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. provisional application serial number 61/704,408, filed August 21, 2012, which is hereby incorporated by reference in its entirety, including all tables, figures, and claims.

INCORPORATION OF SEQUENCE LISTING

[0002] The material in the accompanying Sequence Listing is hereby incorporated by reference into this application. The accompanying sequence listing text file, name SGI1660_1WO_PCT_Sequence Listing_ST25, was created on August __, 2013 and is __ KB. The file can be assessed using Microsoft Word on a computer that uses Windows OS.

BACKGROUND OF THE INVENTION

[0003] In recent years, an increasing effort has been devoted to identify new and effective ways to use renewable feedstocks for the production of organic chemicals. Among a plethora of downstream chemical processing technologies, the conversion of biomass-derived sugars to value-added chemicals is considered very important. In particular, six-carboned carbohydrates, i.e. hexoses such as fructose and glucose, are widely recognized the most abundant monosaccharides existing in nature, therefore can be suitably and economically used as the chemical feedstocks.

[0004] The production of furans and furan derivatives from sugars has attracted increasing attention in chemistry and in catalysis studies, and is believed to have the potential to provide one of the major routes to achieving sustainable energy supply and chemicals production. Indeed, dehydration and/or oxidation of the sugars available within biorefineries with integrated biomass conversion processes can lead to a large family of products including a wide range of furans and furan derivatives.

[0005] Among the furans having the most commercial values, furan-2,5-dicarboxylic acid (also known as 2,5-furandicarboxylic acid, hereinafter abbreviated as FDCA) is a valuable intermediate with various uses in several industries including pharmaceuticals, pesticides, antibacterial agents, fragrances, agricultural chemicals, as well as in a wide range of manufacturing applications of polymer materials, *e.g.* bioplastic resins. As such, FDCA is considered a green alternative of terephthalic acid (TPA), a petroleum-based monomer that is one of the largest-volume petrochemicals produced yearly worldwide. In fact, the US Department of Energy has identified FDCA as one of the top 12 priority compounds made from sugars into a value-added chemical for establishing the "green" chemistry of the future, and as such, it has been named one of the "sleeping giants" of the renewable intermediate chemicals (Werpy and Petersen, *Top Value Added Chemicals from Biomass*. US Department of Energy, Biomass, Vol 1, 2004).

[0006] Although various methods have been proposed for commercial scale production of FDCA (for review, see, *e.g.*, Tong et al., *Appl. Catalysis A: General,* 385, 1-13, 2010), the main industrial synthesis of FDCA currently relies on a chemical dehydration of hexoses, such as glucose or fructose, to the intermediate 5-hydroxymethylfurfural (5-HMF), followed by a chemical oxidation to FDCA. However, it has been reported that current FDCA production processes via dehydration are generally nonselective, unless immediately upon their formation, the unstable intermediate products can be transformed to more stable materials. Thus, the primary technical barrier in the production and use of FDCA is the development of an effective and selective dehydration process from biomass-derived sugars.

[0007] It is therefore desirable to develop methods for production of this highly important compound, as well as many other chemicals and metabolites, by alternative means that not only would substitute renewable for petroleum-based feedstocks, but also use less energy and capital-intensive technologies. In particular, the selective control of sugar dehydration could be a very powerful technology, leading to a wide range of additional, inexpensive building blocks.

SUMMARY OF THE INVENTION

[0008] The present invention provides methods for producing a product of one or more enzymatic pathways. The pathways used in the methods of the invention involve one or more conversion steps such as, for example, an enzymatic conversion of guluronic acid into D-glucarate (Step 7); an enzymatic conversion of 5-ketogluconate (5-KGA) into L-Iduronic acid (Step 15); an enzymatic conversion of L-Iduronic acid into Idaric acid Step 7b); and an enzymatic conversion of 5-ketocluconate into 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH) (Step 16). In some embodiments the methods of the invention produce 2,5-furandicarboxylic acid (FDCA) as a product. The methods include both enzymatic and chemical conversions as steps. Various pathways are also provided for converting glucose into 5-dehdyro-4-deoxy-glucarate (DDG), and for converting glucose into FDCA. The methods can also involve the use of engineered enzymes that perform reactions with high specificity and efficiency.

[0009] In a first aspect the invention provides a method for producing a product of an enzymatic or chemical pathway from a starting substrate. The pathway can contain any one or more of the following conversion steps: an enzymatic conversion of guluronic acid into D-glucarate (Step 7); an enzymatic conversion of 5-ketogluconate (5-KGA) into L-Iduronic acid (Step 15); an enzymatic conversion of L-Iduronic acid into Idaric acid (Step 7b); and an enzymatic conversion of 5-ketocluconate into 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH) (Step 16); an enzymatic conversion of 1,5-gluconolactone to gulurono-lactone (Step 19).

[0010] In one embodiment the product of the enzymatic pathway is 5-dehydro-4-deoxy-glucarate (DDG). In various embodiments the substrate of the method can be glucose, and the product can 5-dehydro-4-deoxy-glucarate (DDG). The method can involve the steps of the enzymatic

conversion of D-glucose to 1,5-gluconolactone (Step 1); the enzymatic conversion of 1,5-gluconolactone to gulurono-lactone (Step 19); the enzymatic conversion of gulurono-lactone to guluronic acid (Step 1B); the enzymatic conversion of guluronic acid to D-glucarate (Step 7); and the enzymatic conversion of D-glucarate to 5-dehydro-4-deoxy-glucarate (DDG) (Step 8).

[0011] In another method of the invention the substrate is glucose and the product is DDG, and the method involves the steps of: the conversion of D-glucose to 1,5-gluconolactone (Step 1); the conversion of 1,5-gluconolactone to gluconic acid (Step 1a); the conversion of gluconic acid to 5-ketogluconate (5-KGA) (Step 14); the conversion of 5-ketogluconate (5-KGA) to L-Iduronic acid (Step 15); the conversion of L-Iduronic acid to Idaric acid (Step 7b); and the conversion of Idaric acid to DDG (Step 8a).

[0012] In another method of the invention the substrate is glucose and the product is DDG and the method involves the steps of the conversion of D-glucose to 1,5-gluconolactone (Step 1); the conversion of 1,5-gluconolactone to gluconic acid (Step 1a); the conversion of gluconic acid to 5-ketogluconate (5-KGA) (Step 14); the conversion of 5- ketogluconate (5-KGA) to 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH) (Step 16); the conversion of 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH) to 4-deoxy-5-threo-hexosulose uronate (DTHU) (Step 4); and the conversion of 4-deoxy-5-threo-hexosulose uronate (DTHU) to DDG (Step 5).

[0013] In another method of the invention the substrate is glucose and the product is DDG, and the method involves the steps of: the conversion of D-glucose to 1,5-gluconolactone (Step 1); the conversion of 1,5-gluconolactone to gluconic acid (Step 1a); the conversion of gluconic acid to 5-ketogluconate (5-KGA) (Step 14); the conversion of 5-ketogluconate (5-KGA) to L-Iduronic acid (Step 15); the conversion of L-Iduronic acid to 4-deoxy-5-threo-hexosulose uronate (DTHU) (Step 7B); and the conversion of 4-deoxy-5-threo-hexosulose uronate (DTHU) to DDG (Step 5).

[0014] Any of the methods disclosed herein can further involve the step of converting the DDG to 2,5-furan-dicarboxylic acid (FDCA). Converting the DDG to FDCA in any of the methods can involve contacting DDG with an inorganic acid to convert the DDG to FDCA.

[0015] In another aspect the invention provides a method for synthesizing derivatized (esterified) FDCA. The method involves contacting DDG with an alcohol, an inorganic acid at a temperature in excess of 60 C to form derivatized FDCA. In different embodiments the alcohol is methanol, butanol or ethanol .

[0016] In another aspect the invention provides a method for synthesizing a derivative of FDCA. The method involves contacting DDG with an alcohol, an inorganic acid, and a co-solvent to produce a derivative of DDG; optionally purifying the derivative of DDG; and contacting the derivative of DDG with an inorganic acid to produce a derivative of FDCA. The inorganic acid can be sulfuric acid and the alcohol can be ethanol or butanol. In various embodiments the co-solvent can

be any of THF, acetone, acetonitrile, an ether, butyl acetate, an dioxane, chloroform, methylene chloride, 1,2-dichloroethane, a hexane, toluene, and a xylene.

[0017] In one embodiment in the derivative of DDG is di-ethyl DDG and the derivative of FDCA is di-ethyl FDCA, and in another embodiment the derivative of DDG is di-butyl DDG and the derivative of FDCA is di-butyl FDCA.

[0018] In another aspect the invention provides a method for synthesizing FDCA. The method involves contacting DDG with an inorganic acid in a gas phase.

[0019] In another aspect the invention provides a method for synthesizing FDCA. The method involves contacting DDG with an inorganic acid at a temperature in excess of 120 C.

[0020] In another aspect the invention provides a method for synthesizing FDCA. The method involves contacting DDG with an inorganic acid under anhydrous reaction conditions.

DESCRIPTION OF THE DRAWINGS

[0021] Figure 1 is a electrophoretic gel of crude lysates and purified enzymes of proteins 474, 475, and 476.

[0022] Figures 2a-h is a schematic illustration of the pathways of Routes 1, 2, 2A, 2C, 2D, 2E, 2F, respectively.

[0023] Figures 3a-c present a schematic illustration of the pathways of Routes 3, 4, and 5, respectively.

[0024] Figure 4 is an HPCL-MS analysis of the dehydration of gluconate with gluconate dehydratase to produce DHG by pSGI-359.

[0025] Figure 5 is a graphical illustration of semicarbizide assay plots for measuring the activity of gluconate dehydratases.

[0026] Figure 6a-b provides Lineweaver-Burk plots for the oxidation of glucuronate and iduronate with three enzymes of the invention.

[0027] Figure 7a shows the results of an HPLC analysis of time points for the isomerization of 5KGA and Iduronate using enzymes DTHU isomerases in the EC 5.3.1.17 family. Controls: dead enzyme is a control with heat inactivated enzyme. Med Bl refers to reactions without isomerase add/n. Time points, x axis 1=0.5 h; 2=1; 3=2 h; 4=16h. Figure 7b shows an HPLC analysis of time points for the isomerization of 5KGA and iduronate using enzymes in the EC 5.3.1.17 family. Controls: dead enzyme is a control with heat inactivated enzyme; Med Bl: refers to reactions without isomerase add/n. Time points, X axis: 1=0 h; 2= 1 h ; 3= 2 h ; 4=17 h.

[0028] Figure 8 shows product formation for the isomerization of 5KGA and iduronate with enzymes in the EC 5.3.1.n1 family. The data were obtained from enzymatic assays.

[0029] Figure 9: HPLC analysis of the formation of 2,5-DDH and the reduction of 5KGA concentration over time. Total ion counts for 2,5-DDH are shown.

[0030] Figure 10 is a HPLC-MS chromatogram showing the production of guluronic acid lactone from 1,5-gluconolactone. An overlay of a trace of authentic guluronic acid is shown.

[0031] Figure 11 is a schematic illustration of the Scheme 6 reaction pathway.

[0032] Figures 12a and 12b are LC-MS chromatograms showing 5-KGA and DDG reaction products, respectively.

[0033] Figure 13 is a an LC-MS chromatogram showing FDCA and FDCA dibutyl ester derivative reaction products.

[0034] Figure 14a is a GC-MS analysis of a crude reaction sample of the diethyl-FDCA synthesis from the reaction of DDG with ethanol. Single peak corresponded to diethyl-FDCA. Figure 14b is an MS fragmentation of the major product from the reaction of DDG with ethanol.

[0035] Figure 15a is a GC-MS analysis of a crude reaction sample of the diethyl-FDCA synthesis from the reaction of DDG with ethanol. Single peak corresponded to diethyl-FDCA. Figure 15b is a MS fragmentation of the major product from the reaction of DDG with ethanol.

[0036] Figure 16 is a schematic illustration of the synthesis of FDCA and its derivatives from DTHU.

[0037] Figure 17 is a schematic illustration of Scheme 1. Cell free enzymatic synthesis of DDG from glucose. Enzymes are **ST-1**: glucose oxidase; **ST-1A:** hydrolysis-chemical; **ST-14:** gluconate dehydrogenase (pSGI-504); **ST-15:** 5-dehydro-4-deoxy-D-glucuronate isomerase (DTHU IS, pSGI-434); **ST-7B:** Uronate dehydrogenase (UroDH, pSGI-476)); **ST-8A** Glucarate dehydratase (GlucDH, pSGI-353); **ST-A:** NAD(P)H oxidase (NADH_OX, pSGI-431); **ST-B:** Catalase. Figure 17b shows the concentration of reaction intermediates over the first 3h as analyzed by HPLC. Formation of DDG is shown in both reactions.

DETAILED DESCRIPTION OF THE INVENTION

[0038] The present invention provides methods for producing a product of an enzymatic pathway. The methods can comprise the enzymatic conversion of a substrate into a product. By utilizing the enzymatic and chemical pathways of the invention it is possible to synthesize a wide variety of products in a highly efficient and economical manner. One product that can be produced by the methods and pathways of the invention is 2,5-furanyl dicarboxylic acid (FDCA), which can be produced at commercial scales according to the invention. The methods can comprise one or more enzymatic and/or chemical substrate-to-product conversion steps disclosed herein.

[0039] The pathways of the invention are comprised of one or more steps. It is understood that a step of a pathway of the invention can involve the forward reaction or the reverse reaction, i.e., the substrate A being converted into intermediate B and product C, while in the reverse reaction substrate C is converted into intermediate B and product A. In the methods both the forward and the reverse reactions are described as the step unless otherwise noted.

[0040] The methods involve producing a product of a pathway, which can be an enzymatic pathway. In some embodiments the pathways can include one or more chemical steps. The methods involve one or more enzymatic and/or chemical conversion steps, which convert a substrate to a product. Steps that can be included in the methods include, for example, any one or more of: an enzymatic conversion of guluronic acid into D-glucarate (Step 7); an enzymatic conversion of L-iduronic acid to Idaric acid (7B); an enzymatic conversion of L-Iduronic acid to 4-deoxy-5-threo-hexosulose uronate (DTHU)(7B); an enzymatic conversion of 5-ketogluconate (5-KGA) into L-Iduronic acid (Step 15); an enzymatic conversion of L-Iduronic acid into Idaric acid Step 7B); and an enzymatic conversion of 5-ketocluconate into 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH) (Step 16); an enzymatic conversion of 1,5-gluconolactone to gulurono-lactone (Step 19). Any one or more of the forementioned steps can be included in a method or pathway of the invention. An enzymatic step or pathway is a step or pathway that requires an enzyme as a catalyst in the reaction to make the step proceed. Chemical steps can be performed without an enzyme as a catalyst in the reaction. Any one or more of the steps recited in the methods can be an enzymatic step. In some embodiments every step of the pathway is an enzymatic step, while in other embodiments one or more steps in the pathway is a chemical step.

[0041] In some embodiments any of the methods can include a step involving the addition of the substrate of the reaction to a reaction mix containing the enzyme that performs the conversion. Thus the method of converting guluronic acid into D-glucarate (step 7) can involve the addition of guluronic acid as starting substrate to the reaction mix; the enzymatic conversion of L-iduronic acid to Idaric acid (7B) can involve the addition of L-Iduronic acid as starting substrate to the reaction mix; the enzymatic conversion of L-Iduronic acid to 4-deoxy-5-threo-hexosulose uronate (DTHU) (7B) can involve the addition of DTHU as starting substrate to the reaction mix. Another step that can be included in any of the methods is a step of purifying from the reaction mixture a reaction product. Thus, a step of purifying D-glucarate or L-Iduronic acid, or Idaric acid, or 4,6-dihydroxy 2,5-diketo hexanoate can be included in any of the methods described herein. Any of the methods disclose can include a step of isolating or purifying DDG or FDCA from the reaction mixture.

[0042] The reaction mix used in the methods can be a cell lysate of cells that contain one or more enzymes that perform the enzymatic conversion, but can also be a reaction mixture containing components added by the user to form a reaction mixture, or can contain components purified from a cell lysate, or may be contained in a whole cell biocatalyst.

[0043] In various embodiments the methods of the invention are methods of converting glucose to DDG, or glucose to FDCA, or glucose to DTHU or DEHU, or for converting DDG to FDCA. The methods can involve converting the starting substrate in the method into the product. The starting substrate is the chemical entity considered to begin the method and the product is the chemical entity considered to be the final end product of the method. Intermediates are those

chemical entities that are created in the method (whether transiently or permanently) and that are present between the starting substrate and the product. In various embodiments the methods and pathways of the invention have about four or about five intermediates or 4-5 intermediates, or about 3 intermediates, or 3-5 intermediates, or less than 6 or less than 7 or less than 8 or less than 9 or less than 10 or less than 15 or less than 20 intermediates, meaning these values not counting the starting substrate or the final end product.

[0044] The invention provides methods of producing FDCA and/or DDG, from glucose that have high yields. The theoretical yield is the amount of product that would be formed if the reaction went to completion under ideal conditions. In different embodiments the methods of the invention produce DDG from glucose, fructose, or galactose with a theoretical yield of at least 50% molar, or at least 60% molar or at least 70% molar, or at least 80% molar, at least 90% molar or at least 95% molar or at least 97% molar or at least 98% molar or at least 99% molar, or a theoretical yield of 100% molar. The methods of the invention also can provide product with a carbon conservation of at least 80% or at least 90% or at least 99% or 100%, meaning that the particular carbon atoms present in the initial substrate are present in the end product of the method at the recited percentage. In some embodiments the methods produce DDG and/or FDCA from glucose via dehydration reactions.

Synthesis Routes

[0045] The invention also provides specific pathways for synthesizing and producing a desired product. Any of the following described routes or pathways can begin with glucose and flow towards a desired product. In some embodiments D-glucose is the starting substrate and the direction of the pathway towards any intermediate or final product of the pathway is considered to be in the downstream direction, while the opposite direction towards glucose is considered the upstream direction. It will be realized that routes or pathways can flow in either the downstream or upstream direction. It is also understood that glucose, fructose, galactose, or any intermediate in any of the pathways can be the starting substrate in a method of the invention, and DDG, FDCA, or any intermediate in any of the routes or pathways of the invention can be the final end product of a method of the invention. The disclosed methods therefore include any one or more steps disclosed in any of the routes or pathways of the invention for converting any starting substrate or intermediate into any end product or intermediate in the disclosed routes or pathways using one or more of the steps in the disclosed routes or pathways. Thus, for example the methods can be methods for converting glucose to DDG, or glucose to guluronic acid, or glucose to galactarate, or glucose to DTHU, or glucose to DEHU, or for converting glucose to guluronic acid, or for converting glucose to iduronic acid, or for converting glucose to idaric acid, or for converting glucose to glucaric acid, or for converting galactarate to DDG, or for converting guluronic acid to D-glucarate, or for converting 5-KGA to L-Iduronic acid, or for converting L-Iduronic acid to Idaric acid, or for converting 5-KGA to 2,5-DDH or DTHU, or for converting DHG to DEHU. In these embodiments the methods utilize

the steps disclosed in the methods and pathways of the invention from glucose as starting substrate to the relevant end product.

[0046] Route 1 is illustrated in Figure 2a. Route 1 converts D-glucose (or any intermediate in the pathway) into 5-dehydro-4-deoxy-glucarate (DDG) via an enzymatic pathway via a series of indicated steps. Route 1 converts D-glucose into DDG via a pathway having 1,5-gluconolactone, gluconic acid, 3-dehydro-gluconic acid (DHG), 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH), and 4-deoxy-L-threo-hexosulose uronate (DTHU) as intermediates and DDG as the final end product. For any of the pathways additional intermediates not shown can also be present. The steps are the enzymatic conversion of D-glucose to 1,5-gluconolactone (Step 1); the enzymatic conversion of 1,5-gluconolactone to gluconic acid (Step 1A); the enzymatic conversion of gluconic acid to 3-dehydro-gluconic acid (DHG) (Step 2); the enzymatic conversion of 3-dehydro-gluconic acid (DHG) to 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH) (Step 3); the enzymatic conversion of 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH) to 4-deoxy-L-threo-hexosulose uronate (DTHU) (Step 4); and the enzymatic conversion of 4-deoxy-L-threo-hexosulose uronate (DTHU) to 5-dehydro-4-deoxy glucarate (DDG) (Step 5). Route 1 also comprises sub-routes where the glucose or any intermediate in the pathway is converted into any other downstream intermediate as final product, and each substrate to product sub-route is considered disclosed as if each is set forth herein in full.

[0047] Route 2 is illustrated in Figure 2b and converts D-glucose into DDG. The steps in the Route 2 pathway are the enzymatic conversion of D-glucose into 1,5-gluconolactone (Step 1); the enzymatic conversion of 1,5-gluconolactone to gluconic acid (Step 1A); the enzymatic conversion of gluconic acid to guluronic acid (Step 6); the enzymatic conversion of guluronic acid to D-glucarate (Step 7); the enzymatic conversion of D-glucarate to DDG (Step 8). Route 2 also comprises sub-routes where glucose or any intermediate in the pathway is converted into any other downstream intermediate as final product, and each sub-route is considered disclosed as if each is set forth herein in full.

[0048] Route 2A is illustrated in Figure 2c. The steps in Route 2A are the enzymatic conversion of D-glucose to 1,5-gluconolactone (Step 1); the enzymatic conversion of 1,5-gluconolactone to guluronic acid lactone (Step 19); the enzymatic conversion of guluronic acid lactone to guluronic acid (Step 1B); the enzymatic conversion of guluronic acid to D-glucarate (Step 7); the enzymatic conversion of D-glucarate to 5-dehydro-4-deoxy-glucarate (DDG) (Step 8). Route 2A also comprises sub-routes where glucose or any intermediate in the pathway as starting substrate is converted into any other downstream intermediate as final end product, and each sub-route is considered disclosed as if each is set forth herein in full.

[0049] Route 2B is illustrated in Figure 2d. The steps in Route 2B are the enzymatic conversion of D-glucose into gluconic acid (Steps 1 and 1A); the enzymatic conversion of gluconic acid into 5-ketogluconate (5-KGA) (Step 14); the enzymatic conversion of 5-KGA into L-Iduronic

acid (Step 15); the enzymatic conversion of L-Iduronic acid into Idaric acid (Step 7B); the enzymatic conversion of Idaric acid into DDG (Step 8A). Route 2B also comprises sub-routes where glucose or any intermediate in the pathway as starting substrate is converted into any other downstream intermediate as final end product, and each sub-route is considered disclosed as if each is set forth herein in full.

[0050] Route 2C is illustrated in Figure 2e. The steps in Route 2C are the enzymatic conversion of D-glucose to gluconic acid (Steps 1 and 1A); the enzymatic conversion of gluconic acid to 5-ketogluconate (5-KGA) (Step 14); the enzymatic conversion of 5-KGA to 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH) (Step 16); the enzymatic conversion of 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH) to 4-deoxy-5-threo-hexosulose uronate (DTHU) (Step 4); the enzymatic conversion of DTHU to DDG (Step 5). Route 2C also comprises sub-routes where glucose or any intermediate in the pathway as starting substrate is converted into any other downstream intermediate as final end product, and each sub-route is considered disclosed as if each is set forth herein in full.

[0051] Route 2D is illustrated in Figure 2f. The steps in Route 2D are the enzymatic conversion of D-glucose to gluconic acid (Steps 1 and 1A); the enzymatic conversion of gluconic acid to 5-ketogluconate (5-KGA) (Step 14); the enzymatic conversion of 5-KGA to Iduronic acid (Step 15); the enzymatic conversion of L-Iduronic acid to DTHU (Step 17); the enzymatic conversion of DTHU to DDG (Step 5). Route 2D also comprises sub-routes where glucose or any intermediate in the pathway as starting substrate is converted into any other downstream intermediate as final end product, and each sub-route is considered disclosed as if each is set forth herein in full.

[0052] Route 2E is illustrated in Figure 2g. The steps in Route 2D are the enzymatic conversion of D-glucose to 1,5-gluconolactone (Step 1); the enzymatic conversion of 1,5-gluconolactone to guluronic acid lactone (Step 19); the enzymatic conversion of guluronic acid lactone to guluronic acid (Step 1B); the enzymatic conversion of guluronic acid to 4-deoxy-erythro-hexosulose uronate (DEHU) (Step 17A); the enzymatic conversion of DEHU to 3-deoxy-D-erythro-2-hexulosaric acid (DDH) (Step 7A). Route 2E also comprises sub-routes where glucose or any intermediate in the pathway as starting substrate is converted into any other downstream intermediate as final end product, and each sub-route is considered disclosed as if each is set forth herein in full.

[0053] Route 2F is illustrated in Figure 2h. The steps in Route 2F are the enzymatic conversion of D-glucose to gluconic acid (Steps 1 and 1A); the enzymatic conversion of gluconic acid to guluronic acid (Step 6); the enzymatic conversion of guluronic acid to 4-deoxy-erythro-hexosulose uronate (DEHU) (Step 17); the enzymatic conversion of DEHU to 3-deoxy-D-erythro-2-hexulosaric acid (DDH) (Step 7A). Route 2F also comprises sub-routes where glucose or any intermediate in the pathway as starting substrate is converted into any other downstream intermediate as final end product, and each sub-route is considered disclosed as if each is set forth herein in full.

[0054] Route 3 is illustrated in Figure 3a. The steps in Route 3 are the enzymatic conversion of D-glucose to gluconic acid (Steps 1 and 1A); the enzymatic conversion of gluconic acid to 3-dehydro-gluconic acid (DHG) (Step 2); the enzymatic conversion of DHG to 4-deoxy-erythro-hexosulose uronate (DEHU) (Step 6A); the enzymatic conversion of DEHU to DDG (Step 7A). Route 3 also comprises sub-routes where glucose or any intermediate in the pathway as starting substrate is converted into any other downstream intermediate as final end product, and each sub-route is considered disclosed as if each is set forth herein in full.

[0055] Route 4 is illustrated in Figure 3b. The steps in Route 4 are the enzymatic conversion of D-glucose to a-D-gluco-hexodialdo-1,5-pyranose (Step 9); the enzymatic conversion of a-D-gluco-hexodialdo-1,5-pyranose to a-D-glucopyranuronic acid (Step 10); the enzymatic conversion of a-D-glucopyranuronic acid to D-glucaric acid 1,5-lactone (Step 11); the enzymatic conversion of D-glucaric acid 1,5-lactone to D-glucarate (Step 1C); the enzymatic conversion of D-glucarate to DDG (Step 8). Route 4 also comprises sub-routes where glucose or any intermediate in the pathway as starting substrate is converted into any other downstream intermediate as final end product, and each sub-route is considered disclosed as if each is set forth herein in full.

[0056] Route 5 is illustrated in Figure 3c. The steps in Route 5 are the enzymatic conversion of D-galactose to D-galacto-hexodialdose (Step 9A); the enzymatic conversion of D-galacto-hexodialdose to galacturonate (Step 10A); the enzymatic conversion of galacturonate to galactarate (Step 11A); the enzymatic conversion of galactarate to DDG (Step 13). Route 5 also comprises sub-routes where galactose or any intermediate in the pathway as starting substrate is converted into any other downstream intermediate as final product, and each sub-route is considered disclosed as if each is set forth herein in full.

The Enzymatic Steps

[0057] There are disclosed a wide variety of enzymes (and nucleic acids that encode the enzymes) that can perform the steps of the methods outlined herein. In addition to the families and classes of enzymes disclosed herein for performing the steps of the invention, additional enzymes (or nucleic acids encoding the enzymes) having a sequence identity to any enzyme or member of a class of enzymes disclosed herein will also be useful in the invention that has a sequence identity of at least 40% or at least 50% or at least 60% or at least 70% or at least 80% or at least 90% or at least 95% or at least 97% or at least 98% or at least 99% to any enzyme or member of an enzyme class disclosed herein. Percent sequence identity or homology with respect to amino acid or nucleotide sequences is defined herein as the percentage of amino acid or nucleotide residues in the candidate sequence that are identical with the known polypeptides, after aligning the sequences for maximum percent identity and introducing gaps, if necessary, to achieve the maximum percent identity or homology. Homology or identity at the nucleotide or amino acid sequence level may be determined using methods known in the art, including but not limited to BLAST (Basic Local Alignment Search Tool) analysis using the

algorithms employed by the programs blastp, blastn, blastx, tblastn and tblastx (Altschul (1997), *Nucleic Acids Res*. 25, 3389-3402, and Karlin (1990), *Proc. Natl. Acad. Sci. USA* 87, 2264-2268), which are tailored for sequence similarity searching. Alternatively a functional fragment of any of the enzymes (or nucleic acids encoding such enzymes) disclosed herein may also be used. The term "functional fragment" refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, where the remaining amino acid sequence has at least about 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to the corresponding positions in the reference sequence, and that retains about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% of the activity of the full-length polypeptide. Functional fragments may comprise, *e.g.*, 90% or less, 80% or less, 70% or less, 60% or less, 50% or less, 40% or less, 30% or less, or 20% or less of the full-length polypeptide, and can include, for example, up to about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% of the full-length polypeptide. The EC numbers provided use the enzyme nomenclature of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology.

[0058] Step 1 – Conversion (oxidation or dehydrogenation) of glucose to 1,5-gluconolactone. This step can be performed with various enzymes, such as those of the family oxygen dependent glucose oxidases (EC 1.1.3.4) or NAD(P)-dependent glucose dehydrogenases (EC 1.1.1.118, EC 1.1.1.119). *Gluconobacter oxydans* has been shown to efficiently oxidize glucose to gluconic acid and 5-ketogluconate (5-KGA) when grown in a fermentor. Enzymes of the family of soluble and membrane-bound PQQ-dependent enzymes (EC 1.1.99.35 and EC 1.1.5.2) found in *Gluconobacter* and other oxidative bacteria can be used. Quinoprotein glucose is another enzyme that is useful in performing this step. The specific enzyme selected will be dependent on the desired reaction conditions and necessary co-factors that will be present in the reaction, which are illustrated in Table 1.

[0059] Step 1A – Conversion (e.g., hydrolysis) of 1,5-gluconolactone to gluconate. This step can be performed chemically in aqueous media and the rate of hydrolysis is dependent on pH (Shimahara, K, Takahashi, T., *Biochim. Biophys. Acta* (1970), 201, 410). Hydrolysis is faster in basic pH (e.g. pH 7.5) and slower in acid pH. Many microorgranisms also contain specific 1,5-glucono lactone hydrolases, and a few of them have been cloned and characterized (EC 3.1.1.17; Shinagawa, E *Biosci. Biotechnol. Biochem.* 2009, 73, 241-244).

[0060] Step 1B – Conversion of Guluronic acid lactone to guluronic acid. The chemical hydrolysis of guluronic acid lactone can be done by a spontaneous reaction in aqueous solutions. An enzyme capable of catalyzing this hydrolysis is identified amongst the large number of lactonases (EC 3.1.1. XX and more specifically 3.1.1.17, 3.1.1.25).

[0061] Step 2 – Conversion of gluconic acid to 3-dehydro gluconic acid (DHG): Several enzymes, such as gluconate dehydratases, can be used in the dehydration of gluconic acid to dehydro gluconic acid (DHG). Examples include those belonging to the gluconate dehydratase family (EC 4.2.1.39). A specific example of such a dehydratase has been shown to dehydrate gluconate (Kim, S. Lee, S.B. *Biotechnol. Bioprocess Eng.* (2008), 13, 436). Particular examples of enzymes from this family and their cloning are shown in Example 1.

[0062] Step 3: Conversion of 3-dehydro-gluconic acid (DHG) to 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH). Enzymes, 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase (or DHG dehydrogenases) (EC 1.1.1.127) for performing this conversion have been described.

[0063] Step 4: Conversion of 4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH) to 4-deoxy-L-threo-hexosulose uronate (DTHU). Enzymes of the family EC 5.3.1.12 can be used in this step, and Step 15 shows that five such enzymes were cloned and shown to have activity for the dehydration of 5-KGA. These enzyme will also show activity towards 2,5-DDH and DTHU.

[0064] Step 5: Conversion of DTHU to 5-dehydro-4-deoxy-glucarate (DDG). DDG can be produced from the chemical or enzymatic oxidation of DTHU, for example with a mild chemical catalyst capable of oxidizing aldehydes in the presence of alcohols. Aldehyde oxidases can be used to catalyze this oxidation. Oxidative bacteria such as *Acetobacter* and *Gluconobacter* (Hollmann *et al Green Chem.* 2011, 13, 226) will be useful in screening. Enzymes of the following families can perform this reaction: aldehyde oxidase EC1.2.3.1, aldehyde ferredoxin oxidoreductase (EC1.2.7.5), and in all the families of EC1.2.1.-XX. Enzymes of the family of uronate dehydrogenases (EC 1.1.1.203) (e.g. see Step 7) will also have this activity. Other enzymes with both alcohol and aldehyde oxidation activity can be used, including enzymes in the alditol oxidase family (see Steps 19 and 6). Other broad substrate oxidases include soluble and membrane bound PQQ-dependent alcohol/aldehyde oxidases. More specifically soluble periplasmic PQQ oxidases enzymes and their homologs belonging into Type I (EC 1.1.9.1) and II (EC 1.1.2.8) families as well as membrane bound PQQ oxidases belonging into EC 1.1.5.X families are useful. In other embodiments aldehyde dehydrogenases/oxidases that act on DTHU can be used.

[0065] Steps 6 and 6A: Conversion of gluconic acid to guluronic acid (6) and conversion of 3-dehydro-gluconic acid (DHG) to 4-deoxy-5-erythro-hexosulose uronate (DEHU)(6A). The enzymes described in Step 5 are useful for these conversions. Other useful enzymes include NAD(P)-dependent dehydrogenases in the EC 1.1.1.XX families and more specifically glucuronate dehydrogenase (EC 1.1.1.19), glucuronolactone reductase (EC 1.1.1.20). In addition,a large number O_2-dependent alcohol oxidases with broad substate range including sugars will be useful (EC 1.1.3.XX), including sorbitol/mannitol oxidases (EC 1.1.3.40), hexose oxidases (EC 1.1.3.5), alcohol oxidases (EC 1.1.3.13) and vanillin oxidase (EC 1.1.3.38). PQQ-dependent enzymes and enzymes present in oxidative bacteria can also be used for these conversions.

[0066] Steps 7 and 7B: Conversion of guluronic acid to D-glucaric acid (7) and conversion of L-Iduronic acid to Idaric acid (7B). These steps can be accomplished with enzymes of the family of uronate dehydrogenases (EC 1.1.1.203) or the oxidases, as described herein.

[0067] Step 7A: Conversion of 4-deoxy-5-erythro-hexosulose uronate (DEHU) to 3-deoxy-D-erythro-2-hexulosaric acid (DDH). The same enzymes described in Step 5 will be useful for performing this conversion.

[0068] Steps 8 and 8A: Conversion of D-glucaric acid to 5-dehydro-4-deoxy-glucarate (DDG) (Step 8) and conversion of Idaric acid to DDG (Step 8A). Enzymes of the family of glucarate dehydratases (EC 4.2.1.40) can be used to perform these steps. Enzymes of this family have been cloned and have been shown to efficiently convert glucarate to DDG. Two D-glucarate dehydratases (EC 4.2.1.40) were cloned as shown in the Table of cloned glucarate dehydratases below. Both enzymes showed very high activity for the dehydration of Glucarate to DDG using the semicarbazide assay, as described in Step 2.

Cloned glucarate dehydratases

Organism	**pSGI (Vector)**	**Gene ID**	**WT/SYN**
E. coli	353 (pET28)	P0AES2	WT
Pseudomonas (SGI)	244	#8114	WT

[0069] Step 9 and 9A: Conversion of D-glucose to α-D-gluco-hexodialdo-1,5-pyranose (9) and conversion of D-galactose to D-galacto-hexodialdose (9A). Oxidases such as those of the galactose oxidase family (EC 1.1.3.9) can be used in this step. Mutant galactose oxidases are also engineered to have activity on glucose and have been described (Arnold, F.H. et al *ChemBioChem*, 2002, 3(2), 781).

[0070] Step 10: Conversion of α-D-gluco-hexodialdo-1,5-pyranose to α-D-glucopyranuronic acid (step 10) and D-galacto-hexodialdose to galacturonate (10A). This step can be performed using an enzyme of the family of aldehyde dehydrogenases.

[0071] Step 11 and 11A: Conversion of α-D-glucopyranuronic acid to glucuronic acid 1,5-lactone. Aldehyde dehydrogenases and oxidases as described in Step 5 will be useful in performing this step. Uronate dehydrogenases described in Steps 7 and 7B can also be useful in performing this step. Step-11A is the conversion of galacturonate to galactarate. The uronate dehydrogenase (EC 1.1.1.203), for example those described in Steps 7 and 7B, will be useful in performing this step.

[0072] Step 12: Conversion of fructose to glucose. Glucose and fructose isomerases (EC 5.3.1.5) will be useful in performing this step.

[0073] Step 13: Conversion of galactarate to 5-dehydro-4-deoxy-D-glucarate (DDG). Enzymes of the family of galactarate dehydrogenases (EC 4.2.1.42) can be used to perform this step, and additional enzymes can be engineered for performing this step.

[0074] Step 14: Conversion of gluconate to 5-ketogluconate (5-KGA). A number of enzymes of the family of NAD(P)- dependent dehydrogenases (EC1.1.1.69) have been cloned and shown to have activity for the oxidation of gluconate or the reduction of 5KGA. For example, the NADPH-dependent gluconate 5-dehydrogenase from *Gluconobacter* (Expasy P50199) was synthesized for optimal expression in *E. coli* as shown herein and was cloned in pET24 (pSGI-383). The enzyme was expressed and shown to have the required activities. Additional enzymes useful for performing this step include those of the family of PQQ-dependent enzymes present in *Gluconobacter* (Peters, B. et al. *Appl. Microbiol Biotechnol.*, (2013), 97, 6397), as well as the enzymes described in Step 6. Enzymes from these families can also be used to synthesize 5KGA from gluconate.

[0075] Step 15: Conversion of 5-KGA to L-Iduronic acid. This step can be performed with various enzymes from different isomerase families, as further described in Example 4.

[0076] Step 16: Conversion of 5-KGA to (4S)-4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH). This dehydration can be performed with enzymes in the gluconate dehydratase family (EC 4.2.3.39), such as those described in Example 5 or Step 17.

[0077] Step 17 and 17A: L-Iduronate to 4-deoxy-5-threo-hexosulose uronate (DTHU) and Guluronate to 4-deoxy-5-hexoulose uronate (DHU).

[0078] Enzymes of the family of dehydratases are identified that can be used in the performance of this step. Enzymes from the families of gluconate or glucarate dehydratases will have the desired activity for performing these steps. Furthermore, many dehydratases of the family (EC 4.2.1.X) will be useful in the performance of these steps. In particular, enzymes that dehydrate 1,2-dyhydroxy acids to selectively produce 2-keto-acids will be useful, such as enzymes of the families: EC 4.2.1.6 (galactonate dehydratase) , EC 4.2.1.8 (mannonate dehydratase), EC 4.2.1.25 (arabonate dehydratase), EC 4.2.1.39 (gluconate dehydratase), EC 4.2.1.40 (glucarate dehydratase), EC 4.2.1.67 (fuconate dehydratase), EC 4.2.1.82 (xylonate dehydratase), EC 4.2.1.90 (rhamnonate dehydratase) and dihydroxy acid dehydratases (4.2.1.9). Since known enzyme selectivity is the production of an alpha-keto acid the identified enzymes will produce DEHU and DTHU, respectively, as the reaction products.

[0079] Step 19: Conversion of 1,5-gluconolactone to guluronic acid lactone. This step can be performed by enzymes of the family of alditol oxidases (EC 1.1.3.41) or the enzymes described in Step 6.

Methods of Converting DDG to FDCA and of making esterified DDG and FDCA

[0080] The present invention also provides novel methods of converting DDG to FDCA and FDCA esters. Esters of FDCA include diethyl esters, dibutyl esters, and other esters. The methods involve converting DDG into a DDG ester by contacting DDG with an alcohol, an inorganic acid, and optionally a co-solvent to produce a derivative of DDG. The alcohol can be methanol, ethanol, propanol, butanol, or any C1-C20 alcohol. The inorganic acid can be sulfuric acid. The co-solvent

can be any of or any mixture of THF, acetone, acetonitrile, an ether, butyl acetate, an dioxane, chloroform, methylene chloride, 1,2-dichloroethane, a hexane, toluene, and a xylene. The esterified DDG can then be converted into esterified FDCA. The DDG can be optionally purified as a step prior to performing the method. Purifying the DDG can comprise removing water from the solvent comprising the DDG, for example removing greater than 87% of the water or greater than 90% of the water or greater than 95% of the water or greater than 97% or greater than 98% or greater than 99% of the water from the solvent comprising the DDG. Yields of greater than 25% or 30% or 35% or 40% or 45% molar can be obtained.

DDG Purification

[0081] DDG purification for dehydration or esterification was performed by acidifying the DDG, e.g., by lowering the pH of the reaction with the addition of conc HCl to pH ~2.5. At this pH proteins and any residual glucarate precipitate are removed by filtration and the mixture is lyophilized to give a white powder consisting of DDG and the reaction salts. This DDG can be dehydrated to give 2,5-FDCA, or be esterified to dibutyl-DDG (or di-ethyl DDG) prior to dehydration. This method of purifying or esterifying DDG can be added as a step in any of the methods and pathways disclosed herein that produce DDG.

Methods for synthesizing FDCA and FDCA Derivatives

[0082] The invention also provides various methods of synthesizing FDCA. One method for synthesizing FDCA involves contacting DDG with an alcohol, an inorganic acid at a high temperature to form FDCA. The alcohol can be any alcohol, and examples include (but are not limited to) methanol, ethanol, propanol, and butanol. Diols can also be used. The high temperature can be a temperature greater than 70 °C or greater than 80 °C or greater than 90 °C or greater than 100 °C or greater than 110 °C or greater than 120 °C or greater than 130 °C or greater than 140 °C or greater than 150 °C to form FDCA. Reaction yields of greater than 20% or greater than 30% or greater than 35% or greater than 40% can be achieved.

[0083] The invention also provides methods for synthesizing derivatives of FDCA. The methods involve contacting a derivative of DDG with an inorganic acid to produce a derivative of FDCA. The inorganic acid can be, for example, sulfuric acid. Optionally, the derivative of DDG can be purified prior to contacting it with the second inorganic acid. Non-limiting examples of the derivative of DDG that can be used include methyl DDG, ethyl DDG, propyl DDG, butyl DDG, isobutyl DDG, di-methyl DDG, di-ethyl DDG, di-propyl DDG, di-butyl DDG. The derivative of FDCA produced can be methyl FDCA, ethyl FDCA, propyl FDCA, butyl FDCA, di-methyl FDCA, di-ethyl FDCA, di-propyl FDCA, di-butyl FDCA, and isobutyl FDCA. The derivate of FDCA produced corresponds to the derivative of DDG used in the method. The derivative of FDCA can then be de-esterified to produce FDCA. The method can also be conducted in the gas phase, e.g., using the parameters described below.

[0084] Another method for synthesizing FDCA or derivatives of FDCA involves contacting DDG or derivatives of DDG (any described herein) with an inorganic acid in a gas phase, which can be done with a short residence time, e.g., of less than 10 seconds or less than 8 seconds, or less than 6 seconds or less than 5 seconds or less than 4 seconds or less than 3 seconds or less than 2 seconds or less than 1 second. The residence time refers to the time that the sample is present in the reaction zone of the high temperature flow through reactor. The method can also be conducted at high temperatures, for example at temperatures greater than 150 °C, greater than 200 °C, greater than 250 °C, greater than 300 °C or greater than 350 °C. Yields of greater than 25% or greater than 30% or greater than 40% or greater than 45% or greater than 50% molar are obtainable. Another method for synthesizing FDCA involves contacting DDG with an inorganic acid at a temperature in excess of 80 °C or 90 °C or 100 °C or 110 °C or 120 °C. Another method for synthesizing FDCA involves contacting DDG with an inorganic acid under anhydrous reaction conditions. In various embodiments the anhydrous conditions can be established by lyophilizing the DDG in any method of synthesizing FDCA disclosed herein so that the DDG contains less than 10% or less than 9% or less than 8% or less than 7% or less than 6% or less than 5% or less than 4% or less than 3% water or less than 2% water, by weight.

[0085] The methods of the invention for synthesizing FDCA described herein provide a significantly higher yield than has been available. In different embodiments molar yields of FDCA (v. DDG) can be obtained of greater than 10% or greater than 15% or greater than 20% or greater than 25% or greater than 30% or greater than 35% or greater than 40% or greater than 45% or greater than 50%.

EXAMPLES

Example 1 – Step 2, Gluconic Acid to 3-dehydro-gluconic acid (DHG)

[0086] Enzymes with natural activity for the dehydration of gluconate have been discovered (EC 4.2.1.39). Three enzymes from this family were cloned as shown in Table 1. Enzyme pSGI-365 was cloned and shown to be a dehydratase with broad substrate range having strong activity for the dehydration of gluconate (Kim, S. Lee, S.B. *Biotechnol. Bioprocess Eng.* **2008**, 13, 436).

Table 1: Enzymes used in this experiment and identity homology. All expressed in *P. luorescens*

Organism	pSGI (Vector)	Gene ID	WT/SYN	Expression Host
Achromobacter	365 (pRANGER)	E3HJU7	Syn	*P.fluorescens*
Achromobacter	359 (pRANGER)	#0385	wt	*P.fluorescens*
Acinetobacter	360 (pRANGER)	#0336	wt	*P.fluorescens*

	359_Achromob	365_E3HJU7
pSGI-360_Acinetobacter (SGI)	78	79
pSGI-359_Achromobacter (SGI)		95
pSGI-365 Acromobacter		

[0087] Proteins 359, 360, and 365 showed 2-5 µmole/min per mg of crude enzyme lysate activity for the synthesis of dehydration of gluconate (gel not shown). pSGI-359 was isolated by precipitation with ammonium sulfate and re-dissolving in buffer and assayed by the semicarbazide assay. Activities of 46.2 U/mL or 5.3 U/mg (1 unit=µmole/min) for the dehydration of gluconate were calculated from semicarbazide assay plots. Reaction buffer (93 mL) containing Kpi 10 mM pH 8.0 with 2 mM MgCl2 and 3.5 gr (0.016 mole) of sodium gluconate was mixed with 7 mL of the previous gluconate dehydratase solution. The reaction was incubated at 45 °C for 16 h before one aliquot was analyzed by HPLC-MS (Figure 4). As shown in Figure 4 one new major product with the molecular weight of DHG was produced. The product was also shown to have activity with DHG dehydratases.

[0088] All proteins were cloned on the pRANGER™ (Lucigen, Middleton, WI) expression vector and were expressed in a *Pseudomonas fluorecens* strain. pRANGER™ is a broad host commercially available plasmid vector containing the pBBR1 replicon, Kanamycin resistance and an pBAD promoter for inducible expression of genes. For the enzyme assay a modification of the semicarbazide assay for the quantification of alpha keto acid was used to calculate the activity of each enzyme (Kim, S.; Lee, S.B. *Biochem J.* 2005, 387, 271). SEQ ID NOs: 30-32 and 33-35 show the

amino acid and nucleotide sequences, respectively, of the gluconate dehydratases #0385, #0336, and E3HJU7.

Example 2 – Step 3 - 3-dehydro-gluconic acid (DHG) to (4S)-4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH)

[0089] Enzymes of the family (EC 1.1.1.127) can be used to perform this step. Two examples are 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase and DHG dehydrogenases. Five enzymes from this family were cloned as shown in Table 2 below. pRANGER™ vector was used in every case.

Table 2: Cloned of DHG oxidoreductase (or 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase)

Organism	pSGI (Vector)	Gene ID	WT/SYN	Expression Host
Agrobacterium sp (SGI)	374	#9041	WT	*P. fluorescens*
Agrobacterium tumefaciens (SGI)	375	#8939	WT	*P. fluorescens*
E. coli	376	**P37769**	WT	*P. fluorescens*
Sphingomonas (SGI)	395	#5112	WT	*P. fluorescens*
Hoeflea phototrophica (SGI)	396	#7103	WT	*P. fluorescens*

[0090] The product prepared from the dehydration of gluconate in Step 2 was used as substrate for assaying the lysates of Table 2. As shown in the following Table 3, enzymes were identified showing activity for the oxidation of DHG in assays measuring NADH formation (absorbance increase at 340 nm).

Table 3: Activity calculations for oxidation of DHG to2,5-DDH using DHG oxidoreductase.

A unit = µmole/min of NADH

ENZ	U/mg (100 mM DHG)		
	pH=7.5	pH=8.5 (10 mM DHG)	pH=9.5
pSGI_395	0.012	0.070 (0.02)	0.120
pSGI_396	0.033	0.139 (0.018)	0.418
pSGI_374	0.007	0.043 (0.012)	0.091
pSGI_376	0.007	0.121 (0.01)	1.610

[0091] Further verification of the formation of 2,5-DDH by these enzymes was shown in Step 16 where the reduction of 2,5-DDH (made from the dehydration of 5KGA) with pSGI-395 at acidic pH was shown.

Example 3 – Steps 7 and 7B - Conversion of guluronic acid to D-glucaric acid (7) and conversion of L-Iduronic acid to Idaric acid (7B).

[0092] To demonstrate Steps 7 and 7B the following study was performed. Uronate dehydrogenases (EC 1.1.1.203) are enzymes that oxidize glucuronic and galacturonic acid. Three enzymes with sequence similarity to the known uronate dehydrogenase (Expasy: Q7CRQ0; Prather, K.J, et al., *J. Bacteriol*. 2009, 191, 1565) were cloned from bacterial strains as shown in Tables 4 & 5.

Table 4 – Cloned Uronate Dehydrogenases

Organism	pSGI (pET28)	Gene ID	Expression
Agrobacterium	474	#8807	BL21DE3
Rhizobium	475	#8958	BL21DE3
Pseudomonas	476	#1770	BL21DE3

Table 5 – Sequence Identity

	475	476	Q7CRQ0
474_Agrobacterium	73	49	90
475_Rhizobium		51	74
476_Pseudomonas			50

[0093] Each protein was expressed with a His tag from pET28 and was purified prior to their screening. Protein gels of the crude lysates and purified enzymes are shown in the gel of Fig. 1. After purification all enzymes were tested for activity against glucuronate, as well as against guluronate and iduronate. Kinetic measurements at different substrate concentrations were performed and the calculated activities and Km values for each enzyme are shown in Table 6. All enzymes showed good activity for glucuronate, and also for L-iduronate and guluronate.

Table 6: Activity and Km value for purified uronate dehydrogenases.

Enzyme	Vmax (μM/min/mg); and Km (mM)		
	Glucuronate	Iduronate	Guluronate (Vm only)
474	128.2 ; 0.37	0.96 ; 29.8	0.017
475	47.4 ; 0.22	0.59 ; 42.1	0.016
476	90.9 ; 0.34	1.36 ; 29.6	0.014

[0094] Each plasmid shown in Table 4 was transformed in BL21DE3 *E. coli* cells. Clarified lysates were mixed with equal volume of (25 mL) of equilibration buffer and purified on an Ni NTA column. Activity of each purified enzyme was measured in by mixing 0.050 mL of various dilutions of each purified enzyme with 0.95 mL of reaction buffer (100 mM TrisHCl, pH 8.0, 50 mM NaCl, 0.75 mM NAD+). The reaction progress was measured by monitoring of the formation of NADH at 340 nm. Figures 6a and 6b provide Lineweaver-Burk plots for the oxidation of glucuronate and iduronate, with all three enzymes shown in Figure 6. Clear positive slopes were obtained with all enzymes giving the activities shown in the table above. Protein sequences of the uronate dehydrogenases are shown as SEQ ID NOs: 1-3 and the genes as SEQ ID NO: 4-6.

Example 4 - Step-15: Conversion of 5-ketogluconate (5-KGA) to L-Iduronic acid (15) or guluronic acid (15A).

[0095] This example illustrates the identification of an enzyme capable of isomerizing 5-KGA to iduronic acid (Step 15) or guluronic acid (Step 15A). Thirteen enzymes from three different isomerase families were cloned as shown in Table 7, while their % sequence identity is shown in Table 8.

Table 7: Isomerases cloned

EC	Organism	pSGI (pET28)	Gene ID Archetype® or Expasy	WT/SYN
5.3.1.17	Rhizobium	433	#8938	WT
5.3.1.17	E. coli	434	Q46938 (Expasy)	WT
5.3.1.17	Rhizobium	435	#3891	WT
5.3.1.17	Pannonibacter	436	#7102	WT
5.3.1.n1	Lactobacillus	458	A5YBJ4 (Expasy)	**SYN**
5.3.1.n1	Acidophilum	440	F0J748 (Expasy)	**SYN**
5.3.1.n1	Bacillus	437	#9209	WT
5.3.1.n1	Ochrobactrum	438	#9732	WT
5.3.1.n1	Halomonas	439	#7403	WT
5.3.1.12	Sphingobacteria	478	#1874	WT
5.3.1.12	Thermotoga	479	Q9WXR9	**SYN**
5.3.1.12	Bacillus	480	Q9KFI6	**SYN**
5.3.1.12	Bacillus	481	O34808	**SYN**

Table 8: % Identities of isomerases

	EC	436	434	435	458	440	437	438	439	481	480	479	478
433	5.3.1.17	65	44	43	16	13	18	11	14	6	11	11	7
436	5.3.1.17		45	46	18	14	15	12	13	5	10	11	7
434	5.3.1.17			46	17	10	15	10	13	6	10	12	7
435	5.3.1.17				18	16	18	14	16	9	11	13	7
458	5.3.1.n1					37	57	41	44	6	7	9	5
440	5.3.1.n1						40	67	50	6	6	6	5
437	5.3.1.n1							46	51	8	7	10	6
438	5.3.1.n1								52	5	5	6	4

439	5.3.1.n1									6	7	8	5
481	5.3.1.12										7	36	54
480	5.3.1.12											7	7
479	5.3.1.12												37
478	5.3.1.12												

[0096] As shown in Table 8, enzymes with medium homology (underlined) within each family were selected for cloning. The data demonstrated that enzymes from all families showed activity for the isomerization of 5-KGA giving L-iduronate as the main product. Two enzymes from the 5.3.1.17 family (433 & 434) were also used in the example showing the formation of DDG from 5-ketogluconate (5KGA).

[0097] Activity for the isomerization of 5KGA and iduronate using enzymes from Table 7 was measured using an enzymatic method that detected the formation of products by their activity against two different enzymes. For example, isomerization of 5KGA was detected by measuring the activity of the product iduronate using uronate dehydrogenase (pSGI-476). Isomerization of iduronate was detected by measuring the activity 5KGA reductase (pSGI-383, EC 1.1.1.69) of the product 5KGA. Presence of the products was also detected by GC-MS.

[0098] Enzymes from all families showed varying activity for the isomerization of 5KGA and iduronate. Two enzymes from EC 5.3.1.12 were used in a cell free reaction to isomerize 5KGA and ultimately produce DDG as described in the example. The enzymes were also purified by gel electrophoresis and showed a single band. The purified isomerases were used in reactions using lysate and buffer containing 5KGA or Iduronate. Product formation was demonstrating using both HPLC and the previously described enzymatic methods. Results for 17h of incubation using both HPLC and enzyme assays are shown in Figure 7a. All enzymes showed good activity for the isomerization of both 5KGA and iduronate. Yields for iduronate isomerization by pSGI433, pSGI 434, pSGI 435, and p SGI 436 were 56%, 48% 42%, (436 not measured), respectively when measured enzymatically and 78.8%, 78.5%, 73.3% and 76.6%, respectively when measured by HPLC assay. Yields after 16h for 5KGA isomerization by the same enzymes were 18%, 17%, and 19% respectively (436 not measured) when measured by enzymatic assay, and 16.6%, 17.8%, 16.3%, and 16.9%, respectively, when measured by HPLC assay.

EC 5.3.1.12 enzymes

[0099] Enzymes from the EC 5.3.1.12 family (glucuronate isomerases) were also purified by gel electrophoresis, isolated, and used to prepare reactions by mixing with buffer (50 mM HEPES, 1 mM ZnCl2, pH 8.0) that contained 5 mM of 5KGA or Iduronate. The reactions were incubated at

30 °C and analyzed for product formation using both HPLC and enzymatic methods. Results are shown in Figure 7b.

5.3.1.17 Enzymes

[0100] Enzymes pSGI-478 and pSGI-479 (5-dehydro-4-deoxy-D-glucuronate isomerases) showed isomerization activity for both 5KGA and iduronate. This activity was also confirmed with the enzymatic assays as above. Yields for isomerization of iduronate by pSGI-478 and -479 were 50% and 37%, respectively, when measured enzymatically, and 20% and 18% when measured by HPLC. Yields for 5KGA isomerization were 23% and 26%, respectively, when measured enzymatically, and 24% and 16%, respectively when measured by HPLC. Results are shown in Figure 7a.

5.3.1.n1 Enzymes

[0101] Enzymes in this family were purified by gel electrophoresis. Product formation was measured using enzymatic assays as described above and the results are shown in Figure 8. All enzymes cloned in this family were shown to have activity for the isomerization of 5KGA and iduronate.

[0102] In each case plasmids were transformed in BL21DE3 and proteins purified on a Ni NTA column.

Example 5: Step 16 – 5-keto-gluconate (5KGA) to (4S)-4,6-dihydroxy 2,5-diketo hexanoate (2,5-DDH)

[0103] The three gluconate dehydratases described in Step 2 (Example 1) were expressed as described in Example 1, along with a purified glucarate dehydratase from Step 8. Enzymatic reactions for activity were performed and HPLC-MS analysis showed the formation of 2,5-DDH (Figure 9), which was also confirmed by the fact that formation of the new product was accompanied by the reduction of 5-KGA only in the samples containing gluconate dehydratases, as well as by enzymatic assays with DHG dehydratase (pSGI-395). Good slopes at 340 nm indicating large enzyme activity were obtained when NADH, pSGI-395 lysate and aliquots of the previous reactions were mixed (data not shown). This result in combination with the HPLC analysis prove that the gluconate dehydratases examined dehydrate 5KGA to 2,5-DDH.

Example 6: Step 19 – Conversion of 1,5-gluconolactone to guluronic acid δ-lactone.

[0104] 1,5-gluconolactone oxidation is a side activity of enzymes from the alditol oxidases (EC 1.1.3.41) family. These enzymes oxidize various alditols such as sorbitol, xylitol, glycerol and others. Enzymes were identified having activity for the oxidation of 1,5-gluconolacone, as shown in Table 6 below.

Table 6. Alditol oxidases with activity on 1,5-gluconolactone.

Enzyme	Enzyme Source	Sorbitol U/mg	1,5- Gluconolactone			
			U/mg	Reaction Setup		
				Enzyme mg	Substrate mg/mM	Yield
AO#13	*Terriglobuds roseus*	0.23	0.02	5.3	15 / 85	7%
AO#22	*Granulicella mallensis*	0.27	0.015	7.6	15 / 85	9%
AO#28	*Streptomyces acidiscabies*	1.30	0.010	15	15 / 85	8%
AO#36	*Actinomycetales (SGI)*	1.83	0.102	25	90 / 35	46%
AO#51	*Frankia sp*	0.59	0.019	NT	NT	NT
AO#57	*Propionibacteriacaeae (SGI)*	1.47	0.051	40	70 / 57	6%
AO#76	*Streptomyces sp.*	1.45	0.045	8.2	15 / 85	23%
AO#251*	*Paenibacillus sp.*	0.47	0.003	24	15 8.5	~2%

*crude lysate

[0105] Reactions were prepared using lysates of all the purified enzymes shown on Table 6. Reactions were prepared in 50 mM K-phosphate buffer, pH 7.0 with 0.5 mg/mL catalase and incubated at 30 °C. A new product was observed by HPLC-MS analysis showing the same retention time as guluronate after comparison with authentic standards (Figure 10). This was confirmed by GC-MS, where the product also had the same MS fingerprint as guluronate. It is therefore clear that all the alditol oxidases described in the Table oxidize the 6-OH of 1,5-gluconolactone to produce the guluronic acid lactone. All alditol oxidases were cloned in pET28a with a HisTag and were expressed in BL21DE3 and purified on a Ni NTA column.

Example 7 – Synthesis of FDCA and Other Intermediates

[0106] Purified DDG mono potassium salt was used for the dehydration to 2,5-FDCA. Sulfuric acid was added to DDG and the reaction stirred at 60°C. The in situ yield was calculated (by HPLC-MS) to be ~24% and ~27%.

[0107] The reaction solutions were combined and then diluted by pouring into ice (to neutralize the heat). Approximately equivalent volume of THF was added, and the solution transferred to a separation funnel. Sodium chloride salt was added until separation was achieved. The solution was agitated between additions for best possible dissolution. The aqueous layer was removed, and the THF layer washed 3x more with sat. NaCL solution. Sodium sulfate was added and the solution left sitting overnight. Two layers formed again overnight. The aqueous layer was

discarded and then silica gel was added to the solution. It was then concentrated down to solids via rotovap. The solids were loaded into a silica flash column and then separated via chromatographically. The fraction was concentrated and dried. The isolated yield was 173.9mg. Corrected yield: 24.9%. ^{1}H and ^{13}C NMR and HPLC-MS analysis confirmed the product

Dehydration of DDG Dibutyl-2,5-FDCA in BuOH/H_2SO_4

[0108] Dehydration of un-derivitized lyophilized DDG containing the dehydration salts in BuOH was done using a Dean-Stark apparatus. Under these conditions, DDG was added to BuOH, and then H2SO4 was added and the reaction heated at 140 °C. After stirring for 4 h HPLC-MS analysis shows the disappearance of DDG and the formation of dibutyl-2,5-FDCA. The in situ yield was calculated (by HPLC-MS) to be 36.5%.

[0109] The mixture was extracted with water, 1% NaOH, and again with water. Then the organic layer was concentrated to a final mass of 37.21g. A portion of this mass (3.4423g) was removed and 0.34 g of dibutyl-2,5-FDCA was purified using HPLC. Extrapolating the yield of the isolated product to the total amount of compound isolated from the reaction (37.21g) and taking into account the amount of salts present in the original DDG (~60% pure by weight) the reaction yield was calculated to be 42%. ^{1}H and ^{13}C NMR and HPLC-MS analysis confirmed the product

Synthesis of dibutyl DDG

[0110] In another aspect the invention provides a method for synthesizing a derivative of DDG. The method involves contacting DDG with an alcohol, an inorganic acid, and optionally a co-solvent to produce a derivative of DDG. Optionally the derivative of DDG can be purified. The reaction can have a yield of the derivative of DDG of at least 10% molar yield or at least 15% molar yield or at least 20% molar yield or at least 25% or at least 30% or at least 35% molar yield or at least 40% molar yield. The inorganic acid can be sulfuric acid and the alcohol can be methanol, ethanol, propanol, butanol, isobutanol, or any C1-C20 alcohol. In various embodiments the co-solvent can be any of THF, acetone, acetonitrile, an ether, butyl acetate, an dioxane, chloroform, methylene chloride, 1,2-dichloroethane, a hexane, toluene, and a xylene. When the alcohol is ethanol the DDG derivative will be DDG mono-ethyl ester and/or DDG diethyl ester. When the alcohol is butanol the DDG derivative will be DDG mono-butyl ester and/or DDG dibutyl ester.

[0111] DDG mono-potassium salt was used for derivatization according to the following protocol. In a 1L Morton type indented reaction vessel equipped with a mechanical stirrer and heating mantle was charged with 60:40 DDG:KCl (31.2 mmol), BuOH, and heptane. In a separate vial, sulfuric acid was added to water, and allowed to cool after dissolution. The solution was then added to the flask. The solution was kept at 30°C.

[0112] The precipitate was filtered off concentrated. The remaining gel was dissolved in EtOAc, and then TLC plates were spotted with the solutions and the plates were sprayed with a phosphomolybdic acid mixture, and then heated to at least 150°C on a hot plate to identify the DDG-

DBE fraction. Isolated yield: 4.62 g (15.2 mmol, 47% yield), > 98% purity. ^{1}H and ^{13}C NMR and HPLC-MS analysis confirmed the product.

[0113] Different solvents can be used in the synthesis of DDG esters, such as mixtures of BuOH (5%-95% v/v) with co-solvents such as THF, acetone, acetonitrile, ethers (dibutyl, ditheyl etc), esters such as Butyl-acetate, 1,6-dioxane, chloroform, methylene chloride, 1,2-dichloroethane, hexanes, toluene, and xylenes may be used as cosolvents. Reaction catalysts such as acids (sulfuric, hydrochloric, polyphosphoric or immobilized acids such as DOWEX) or bases (pyridine, ethyl-amine, diethyl-amine, boron trifluoride) or other catalysts commonly used for the esterification of carboxylic acids.

Dehydration of dibutyl-DDG to dibutyl-FDCA in n-BuOH/H_2SO_4

[0114] A stock solution of DDG-DBE (di-butyl ester) was made in butanol and transferred to a clean, dry 100mL round-bottomed flask equipped with a stir bar. To the flask, 25mL of conc. sulfuric acid was added. The flask was sealed and then stirred at 60°C for 2hrs. The in situ yield was calculated to be ~56%. The reaction solution was concentrated and the residue was dissolved in MTBE and transferred to a separation funnel, and then washed with water. The recovered organic layer was concentrated and then separated via HPLC for an isolated yield: 250.7 mg (~90% purity) and 35% isolated yield (corrected for purity). ^{1}C and ^{13}C NMR and HPLC-MS analysis confirmed the product.

Example 8 – Cell free synthesis of DDG and FDCA and derivatives from 5-KGA (Route 2A)

[0115] This example illustrates the enzymatic conversion of 5KGA to DDG using purified enzymes according to Scheme 6 (a sub-Scheme of 2B), and also illustrates the DDG produced being dehydrated to FDCA using chemical steps. The Scheme involves the steps of isomerization of 5KGA (Step 15) and the subsequent oxidation to idaric acid (Step 7B). DDG was also dehydrated under differing chemical conditions to FDCA. The last step (Step-8A) was performed using glucarate dehydratase from *E. coli*.

[0116] Scheme 6 is illustrated in Figure 11. The scheme was performed using a cell free enzymatic synthesis of DDG from 5-KGA. The Scheme involves the performance of steps 15, 7B and 8A. Two additional proteins were used to complete the reaction path, the first being NADH-oxidase (Step A) that is recycling the NAD+ cofactor in the presence of oxygen, and catalase (Step B) that decomposes the peroxide produced from the action of NADH oxidase. The enzymes are shown in the following Table 7. All enzymes contained a HisTag and were purified using an Ni-NTA column. Yields for this synthesis of DDG were calculated to be at least 88-97%.

[0117]

Table 7:

STEP	*Enzyme*	*EC*	*Organism*
15	pSGI-433 (DTHU_IS)	5.3.1.17	*Rhizobium* (SGI)
15	pSGI-434 (DTHU_IS)	5.3.1.17	*E. coli*
7B	pSGI-476 (UroDH)	1.1.1.203	*Pseudomonas* (SGI)
8A	pSGI-353 (GlucDH)	4.2.1.40	*E. coli*
A	pSGI-431 (NADH_OX)	1.6.3.1	*Thermus thermophiilus*
B	Catalase	1.11.1.6	*Corynbacterium*

[0118] 500 mL of liquid culture was purified for each isomerase for the reaction. Besides the enzymes shown on Table 7, each reaction contained 50 mM TrisHCl (pH 8.0), 50 mM NaCl, 1 mM $ZnCl_2$ and 2 mM $MgCl_2$, 1 mM $MnCl_2$ and 1 mM NAD^+. Reactions were analyzed by HPLC after 16 h of incubation and Figure 12 presents the chromatograms.

[0119] For dehydration to FDCA, the reaction mixtures of both samples were combined and lyophilized into a white powder, which was split into two samples and each dissolved in AcOH with 0.25M H_2SO_4 or in 4.5 mL BuOH with 0.25M H_2SO_4. Both reactions were heated in sealed vials for 2-4 h at 120 °C. Reaction products are shown in Figure 13.

[0120] Samples 1 and 2 represent authentic standard and the 3h time point from the reaction in AcOH/ H_2SO_4, respectively. Spiking of sample 2 with sample 1 gave a single peak further verifying the FDCA product. Samples 1 and 3 (Figure 13) represent authentic standard and the 4h time point from the reaction in BuOH/ H_2SO_4, respectively. The formation of FDCA from the enzymatic reactions further confirms the presence of DDG in these samples.

Example 9 – Synthesis of DDG from Glucose and Gluconate

[0121] This example shows the enzymatic conversion of glucose and gluconate to DDG. The reaction was conducted with purified enzymes, and crude lysates as a catalyst. Enzymes and substrates were combined in a bio-reactor as shown in the Table below:

	Substrate	ST-1	ST-14 pSGI-504	ST-15 pSGI-434	ST-7B pSGI-476	ST-8A pSGI-353	ST-A pSGI-431	ST-B
Rxn-1	Glucose 600 mg	2 mg	7 mL[1]	50 mL[2]	7.5 mL[1]	1 mL[3]	4 mL[4]	2 mg
Rxn-2	Gluconate 700 mg	-	7 mL	50 mL	7.5 mL	1 mL	4 mL	2 mg

1. Lysate from 500 mL liquid culture of recombinant E. coli with plasmid
2. Lysate from 2L liquid culture of BL21DE3/pSGI-434
3. Purified enzyme, ~30 Units of activity (or 3 mg of purified GlucD)
4. Lysate from 250 mL of culture

[0122] The reaction was incubated at 35 oC and dissolved oxygen and pH were kept at 20% and 8 respectively. Time points were analyzed by HPLC-MS and the results are shown in Figure 17b. Extracted chromatograms verified the DDG mass (not shown) and corresponding MS fragmentation. The results clearly showed production of DDG during incubation of the enzymes with either glucose or gluconate.

Example 10 - Construction of expression cassettes for recombinant glucarate dehydratases.

[0123] The following example describes the creation of recombinant nucleic acid constructs that contained coding sequence of a D-glucarate dehydratase activity (GDH, EC 4.2.1.40) for heterologous expression in *E. coli* cells.

[0124] Genes encoding D-Glucarate dehydratase from *E. coli* (Expasy: P0AES2;), *Acinetobacter ADP1* (Expasy: P0AES2), as well as a proprietary *Pseudomonas* bacterial strain (BP1MICT2128114) were PCR-amplified from genomic DNA.

[0125] Each of the PCR-amplified genes was subsequently cloned into the bacterial transformation vector pET24a(+), in which the expression of each of the GDH genes was placed under control of a T7 promoter. The nucleotide sequences of each of the PCR-amplified inserts were also verified by sequencing confirmation.

Example 11 - *E. coli* strains expressing recombinant glucarate dehydratases.

[0126] Each of the expression vectors constructed as described in Example 9 was introduced into NovaBlue(DE3) *E. coli* by heat shock-mediated transformation. Putative transformants were selected on LB agar supplemented with Kanamycin (50 μg/ml). Appropriate PCR primers were used in colony-PCR assays to confirm positive clones that contained each of the expression vectors.

[0127] For each expression vector, a bacterial colony was picked from transformation plates and allowed to grow at 30°C in liquid LB media supplemented with Kanamycin (50 μg/ml) for two days. The culture was then transferred into vials containing 15% glycerol and stored at -80°C as a frozen pure culture.

Example 11 - Demonstration of *in vitro* synthesis of DDG by using cell lysate of recombinant *E. coli* cells expressing a GDH enzyme

[0128] This Example describes how *in intro* synthesis of DDG intermediate was achieved using recombinant GDH enzymes produced in *E. coli* cells.

[0129] Preparation of cell lysates: Recombinant bacterial strains constructed as described previously in Example 2 were grown individually in 3 mL of liquid LB media supplemented with Kanamycin (50 μg/ml) at 30°C on a rotating shaker with rotation speed pre-set at 250 rpm for 1 day. This preculture was used to inoculate 100 mL of TB media containing Kanamycin (50ug/ml), followed by incubation at 30°C on a rotating shaker pre-set at 250 rpm for 2-3 hour until early log phase (OD_{600}~0.5-0.6) before isopropyl D-1 thiogalactopyranoside (IPTG; 0.25 mM final concentration) was added to induce protein expression. Cells were allowed to grow for another 18 hours at 30°C before they were harvested by centrifugation, resuspended in 15 mL of lysis buffer (10 mM phosphate buffer, pH 7.8, 2 mM $MgCl_2$) and were lysed by sonication. The production of recombinant enzymes in *E. coli* cells was quantified using standard pre-cast SDS-PAGE gels system (BioRad), and specific activity was measured according to a procedure described by Gulick *et al.* (*Biochemistry* 39, 4590-4602, 2000). Cell lysates were then tested for the ability to convert gram amounts of glucarate to DDG as described in greater details below.

[0130] Enzymatic dehydration of glucarate: Five grams of mono-potassium glucarate (~0.02 moles) were added to 85 mL of 5 mM potassium phosphate buffer containing 10 mM MgCl2. The substrate glucarate was found slowly dissolved following the addition of ~2 mL of 5M NaOH. The pH of the reaction was adjusted to about 7.8. Subsequently, 15 mL of a cell lysate containing each of the three recombinant dehydratases in 10 mM phosphate buffer, pH 7.8, as described in Example 3. After incubation with gentle stirring at 30°C for 1-2

hours, the reactions were analyzed using HPLC-MS techniques. HPLC-MS results indicated a new peak as the only major product with a molecular weight corresponding to predicted product DDG, and trace amounts of the mono-potassium glucarate substrate. No other byproducts were detected by HPLC-MS analysis, indicating that the conversion reaction catalyzed by each of the recombinant enzymes was very efficient and highly specific.

Purification of DDG product from enzymatic reactions:

[0131] DDG produced via enzymatic dehydration was purified by using either of the two following techniques.

[0132] The enzymatic dehydration reactions were acidified to pH~2.0 with 6M HCl, filtered to eliminate precipitated proteins, and subsequently lyophilized. Methanol (MeOH) was added to the lyophilized powders, followed by gentle stirring for 10-15 minutes to dissolve the DDG product but not the other salts in the dehydration reaction mixtures (such as KCl and phosphates). Substantially pure DDG acid was obtained following filtration of the suspensions and evaporation of MeOH.

[0133] In some instances, an alternative procedure was deployed for the purification of DDG salt, in which the first MeOH filtrate was condensed to a volume of ~15-25 mL, then mixed with an equal volume of MeOH containing 0.5M KOH. Potassium salt of DDG precipitated after addition of KOH was subsequently isolated by filtration.

[0134] Results of HPLC-MS analyses indicated that DDG product constituted at least 95% of the total products in the samples obtained from either of the two purification techniques.

Example 12 - Demonstration of *in vitro* synthesis of FDCA from DDG in one-step chemical reaction

[0135] Applicants have discovered that the synthesis of FDCA (*i.e.* the free acid form) could be achieved by a chemical conversion of DDG to FDCA in the presence of H_2SO_4. The reaction was performed as follows. Approximately 20 mg of DDG acid (crude lyophilized powder with salts previously purified as described in Example 3) and 0.25 M of H2SO4 were added into an air tight sealed tube containing 1 mL of water and 1 mL of DMSO. The DDG was found completely dissolved in this solution. The reaction was stirred at 105°C for 18 hours. Results of an HPLC-MS analysis performed on a crude reaction sample indicated the formation of FDCA free acid (FDCA: 2,5-furan dicarboxylic acid) as the major product, as well as insignificant amounts of some other unidentified byproducts. As a control in HPLC-MS analysis, a commercial FDCA was analyzed in the same conditions.

Example 13 - Demonstration of *in vitro* synthesis of FDCA-esters (dimethyl-, diethyl-, dibutyl-, and isopropyl- esters)

Synthesis of diethyl-2,5 FDCA from purified DDG:

[0136] In an air tight sealed tube, 18 mL of EtOH, 0.2 gram (1 mmole) of DDG acid, previously purified as described in Example 11, and 0.25 M of H_2SO_4 were added. The DDG acid was not completely dissolved in this solution. The reaction was gently stirred at 105°C for 18 hours. Results of a GC-MS analysis of a crude reaction sample indicated that the formation of diethyl-FDCA the major product. As a control, an authentic FDCA was chemically synthesized, esterified to diethyl-FDCA and analyzed in the same conditions.

Example 14 - Synthesis of dibutyl-2,5 FDCA from purified DDG

[0137] In an air tight sealed tube, 18 mL of n-BuOH, 0.2 gram (1 mmole) of DDG acid, previously purified as described in Example 11, and 0.25 M of H_2SO_4 were added. The DDG acid was not completely dissolved in this solution. The reaction was gently stirred at 105°C for 18 hours. As shown in FIGURE 15, results of the GC-MS analysis of a reaction sample indicated that diethyl-FDCA (FDCA: 2,5-furan dicarboxylic acid) was formed as the major product. As a control, an authentic FDCA was chemically synthesized, esterified to diethyl-FDCA, and analyzed in the same conditions.

Example 15 - Synthesis of dibutyl-2,5 FDCA from crude DDG (unpurified):

[0138] 0.2 gram (1 mmole) of crude DDG acid, which was an unpurified lyophilized powder obtained directly from the enzymatic dehydration of glucarate as described in Example 11, was added into an air tight sealed tube contanning 18 mL of n-BuOH, followed by addition of 0.25 M of H_2SO_4. The crude DDG acid was not completely dissolved in this solution. The reaction was gently stirred at 105°C for 18 hours. Results of a GC-MS analysis of a crude reaction sample indicated that diethyl-FDCA (FDCA: 2,5-furan dicarboxylic acid) was formed as the major product. The GC-MS result indicated that the present of contaminant salts in crude/unpurified lyophilized powder did not significantly affect the reaction outcome. As a control, an authentic FDCA was chemically synthesized, esterified to diethyl-FDCA, and analyzed in the same conditions.

Example 16 - *In vitro* production of FDCA and/or esters using immobilized acids

[0139] In industrial practices, immobilized acids offer many advantages for performing dehydrations since they can typically operate in several types of solvent (aqueous, organic or mixed, *etc.*). In addition, they can be easily recycled and be re-used. Following some examples of the synthesis of esters of FDCA using immobilized AMBERLYST®15 (Rohm and Haas, Philadelphia, PA) and DOWEX®50 WX8 (Dow Chemical Co, Midland, MI).

Synthesis of dibutyl-FDCA from crude DDG by using DOWEX®50 WX8

[0140] In an air tight sealed tube, 2 mL of n-Butanol, 20 mg of crude DDG acid (unpurified lyophilized powder containing salts) and 200 mg of DOWEX®50 WX8 were combined. The DDG

was not completely dissolved in this solution. The reaction was gently stirred at 105°C for 18 hours. Results of the GC-MS analysis of a crude reaction sample indicated that diethyl-FDCA (FDCA: 2,5-furan dicarboxylic acid) was formed as the major product. This GC-MS result indicated that the present of contaminant salts (phosphate and NaCl) in crude/unpurified lyophilized powder did not significantly affect the reaction outcome. As a control, an authentic FDCA was chemically synthesized esterified to diethyl-FDCA and analyzed in the same conditions.

Synthesis of dibutyl-FDCA from crude DDG by using AMBERLYST®15

[0141] In an air tight sealed tube, 2 mL of n-Butanol, 20 mg of crude DDG acid (crude lyophilized powder with salts) and 200 mg of AMBERLYST®15 (Rohm and Haas, Philadelphia, PA) were combined. The DDG was not completely dissolved in this solution. The reaction was gently stirred at 105°C for 18 hours. Results of the GC-MS analysis of a crude reaction sample indicated that diethyl-FDCA (FDCA: 2,5-furan dicarboxylic acid) was formed as the major product. This GC-MS result indicated that the present of contaminant salts (phosphate and NaCl) in crude/unpurified lyophilized powder did not significantly affect the reaction outcome. As a control, an authentic FDCA was chemically synthesized esterified to diethyl-FDCA and analyzed in the same conditions.

Synthesis of ethyl-FDCA from crude DDG by using AMBERLYST®15

[0142] In an air tight sealed tube, 2 mL of ethanol, 20 mg of crude DDG acid (unpurified lyophilized powder containing salts) and 200 mg of AMBERLYST®15 (Rohm and Haas, Philadelphia, PA) were combined. The DDG was not completely dissolved in this solution. The reaction was gently stirred at 105°C for 18 hours. Results of the GC-MS analysis of a crude reaction sample indicated that diethyl-FDCA (FDCA: 2,5-furan dicarboxylic acid) was formed as the major product. This GC-MS result indicated that the present of contaminant salts (phosphate and NaCl) in crude/unpurified lyophilized powder did not significantly affect the reaction outcome. As a control, a commercial FDCA was chemically esterified to diethyl-FDCA and analyzed in the same conditions.

Synthesis of diethyl-FDCA from crude DDG by using DOWEX®50 WX8

[0143] In an air tight sealed tube, 2 mL of ethanol, 20 mg of crude DDG acid (unpurified lyophilized powder containing salts) and 200 mg of DOWEX®50 WX8 were combined. The DDG was not completely dissolved in this solution. The reaction was gently stirred at 105°C for 18 hours. Results of the GC-MS analysis of a crude reaction sample indicated that diethyl-FDCA (FDCA: 2,5-furan dicarboxylic acid) was formed as the major product. This GC-MS result indicated that the present of contaminant salts (phosphate and NaCl) in crude/unpurified lyophilized powder did not significantly affect the reaction outcome. As a control, a commercial FDCA was chemically esterified to diethyl-FDCA and analyzed in the same conditions.

Example 17 - Production of FDCA derivatives

[0144] The synthesis of a number of high-value FDCA derivatives is described in Figure 16 in which dehydration of DTHU produces furfural-5-carboxylic acid, *i.e.* FCA, which is then

chemically or enzymatically oxidized to FDCA, be reduced to FCH, or be transaminated (using chemical reductive amination or transaminase) to amino acid-AFC.

Example 18 – Production of di-butyl FDCA in a gas phase reaction

[0145] In this example the inlet of the GC was used as a high temperature reactor to catalyze the dehydration of di-butyl DDG to di-butyl FDCA. The resulting products were chromatographically separated detected by mass spectrometry. A solution of di-butyl DDG (10 mM) and sulfuric acid (100 mM) in butanol was placed in a GC vial. The vial was injected into a GC and FDCA Dibutyl ester was observed. The reaction occurred in the 300 °C inlet (residence time = 4 seconds). The average yield of 6 injections was 54%.

GC Settings: Direct liquid inject / MS detector

Inlet: 300° C, total flow 29.51 ml/min, split ratio 10:1, split flow 24.1 ml/min, Septum Purge flow 3 mL/min.

GC liner: 4 mm, glass wool (P/N 5183-4647)

Column Flow: 2.41 ml/min He constant pressure control

Oven Program: At 40 °C hold for 2 min, then ramp 25 °C/min to 275 °C, then ramp 40 °C/min to 325 °C, hold for 2 min.

Column: HP-5MS, Agilent Technologies, 30m x 0.25mm x 0.25um.

Total Runtime: 14.65 minutes

MSD Transfer line: 290 °C

MS Source: 250 °C

MS Quad: 150 °C

Retention times:

2,3-FDCA Dibutyl ester: 9.3 min

2,5-FDCA Dibutyl ester: 9.7 min

[0146] All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

[0147] No admission is made that any reference constitutes prior art. The discussion of the references states what their authors assert, and the applicants reserve the right to challenge the accuracy and pertinence of the cited documents. It will be clearly understood that although a number of prior art publications are referred to herein, this reference does not constitute an admission that any of these documents forms part of the common general knowledge in the art.

[0148] It should also be understood that the foregoing examples are offered to illustrate, but not limit, the invention.

<SEQ ID NO: 1> protein #474

MAMKRLLVTGAAGQLGRVMRKRLASMAEIVRLADLAPLDPAGPNEECMQCDLAD
ADAVDAMVAGCDGIVHLGGISVEKPFEQILQGNIIGLYNLYEAARAHGQPRIIFASSN
HTIGYYPQTERLGPDVPFRPDGLYGVSKCFGESLARMYFEKFGQETALVRIGSCTPEP
LNYRMLSTWFSHDDFVSLIEAAFRAPVLGCPIVWGASANDASWWDNSHLGFIGWKP
KDNAEAFRRKIAETTPQPDARDPIVRFQGGVFVDNPIFKET*

<SEQ ID NO: 2> protein #475

MKRLLITGAAGALGRVMRERLAPMATILRLSDIAPIGAARQNEEIVQCDLADAKAVH
ALVEDCDGIVHLGGVSVERKFSQIVAGNIVGLYNLYEAARAHRMPRIVFASSNHTIGF
YPQTERLSVDHPYRPDGLYGVSKCFGESLAHMYHEKFGQETALVRIGSCVTEPVNH
RMLSTWLSYDDFVSLIEAVFRAPKLGCPVIWGASNNDAGWWDNSAAGFLGWKPKD
NAEIFRSKIEAACERPGSDDPAARWQGGLFTQDPIFPEDE*

<SEQ ID NO: 3> Protein #476

MTTAYTPFNRLLLTGAAGGLGKVLRESLRPYANVLRVSDIAAMSPATGAHEEVQVC
DLADKAAVHQLVEGVDAILHFGGVSVERPFEEILGANICGVFHIYEAARRHGVKRVI
FASSNHVIGFYKQDETIDANCPRRPDSYYGLSKSYGEDMASFYFDRYGIETVSIRIGSS
FPEPHNRRMMSTWLSFADLTQLLERALYTPNVGHTVVYGMSANKNVWWDNHLAA
HLGFQPKDSSEVFRAQIDAQPMPAADDPAMVFQGGAFVAAGPFGDD*

SEQ ID NO: 4 pSGI-474-#8807-DNA

ATGGCAATGAAACGGCTTCTTGTTACCGGTGCTGCGGGCCAGCTTGGCCGCGTTA
TGCGCAAACGCCTTGCATCGATGGCCGAGATCGTTCGCCTTGCCGATCTCGCCCC
GCTCGATCCGGCAGGCCCGAACGAGGAATGCATGCAATGCGACCTTGCGGATGC
AGACGCCGTTGACGCCATGGTTGCCGGTTGCGACGGCATCGTTCACCTCGGCGGC
ATATCGGTGGAGAAGCCTTTCGAACAAATCCTTCAGGGCAACATCATCGGGCTGT
ATAATCTCTATGAGGCCGCCCGCGCCCACGGCCAGCCGCGCATCATCTTCGCCAG
TTCGAACCATACGATCGGTTATTACCCGCAGACGGAGAGGCTTGGACCGGATGTT
CCCTTCCGCCCGGATGGGCTTTACGGCGTCTCCAAATGTTTCGGCGAGAGCCTTG
CCCGCATGTATTTCGAGAAATTCGGCCAGGAGACCGCACTTGTCCGCATCGGCTC
CTGCACGCCGGAACCCCTTAATTACCGCATGCTGTCCACCTGGTTTTCGCATGAC
GATTTCGTCTCGCTGATCGAGGCGGCGTTCCGCGCCCCCGTGCTCGGCTGCCCCA
TCGTCTGGGGGGCGTCGGCCAACGATGCGAGCTGGTGGGACAATTCGCATCTCG
GCTTTATTGGATGGAAACCGAAGGACAATGCCGAGGCCTTCCGCCGGAAGATTG
CCGAAACGACGCCGCAGCCGGACGCGCGCGACCCGATTGTCCGCTTTCAGGGTG
GCGTGTTTGTCGACAACCCGATCTTCAAGGAGACGTGA

SEQ ID NO: 5 pSGI-475-#7895- DNA

ATGAAGAGACTTCTGATTACCGGCGCAGCGGGTGCACTGGGCCGCGTGATGCGG
GAAAGGCTCGCACCCATGGCAACGATTCTGCGCCTTTCCGATATCGCCCCGATTG
GAGCGGCCCGCCAGAACGAGGAAATCGTCCAGTGCGATCTTGCCGATGCCAAAG
CAGTGCATGCTCTGGTCGAAGATTGCGACGGGATCGTCCATCTCGGTGGCGTCTC
AGTAGAGCGCAAGTTCTCGCAGATCGTCGCCGGCAACATCGTCGGCCTTTACAAT
CTCTACGAAGCCGCACGCGCGCATCGGATGCCGCGCATCGTCTTTGCAAGTTCCA
ATCACACCATCGGCTTTTATCCGCAAACCGAACGGTTGTCGGTGGACCATCCCTA
TCGTCCGGACGGGCTCTACGGCGTATCGAAATGTTTCGGCGAGTCTCTGGCGCAT
ATGTACCATGAGAAGTTCGGGCAGGAGACGGCACTCGTGCGCATCGGGTCCTGC
GTGACCGAACCGGTCAACCATCGCATGCTTTCCACCTGGCTTTCCTACGATGATT
TCGTCTCGCTTATCGAGGCCGTATTCCGTGCGCCGAAACTCGGCTGCCCCGTCAT
CTGGGGCGCGTCGAACAACGATGCAGGATGGTGGGACAATTCCGCCGCCGGCTT
TCTCGGCTGGAAGCCGAAAGACAATGCCGAAATCTTCCGTTCGAAGATCGAAGC
CGCTTGCGAACGCCCCGGTTCTGATGATCCGGCCGCCCGCTGGCAAGGCGGGCTC
TTCACGCAGGACCCGATCTTCCCAGAGGACGAGTAA

SEQ ID NO: 6 pSGI-476-#1770-DNA

ATGACCACAGCCTACACCCCCTTCAATCGCCTGCTACTCACCGGAGCGGCAGGCG
GCCTCGGCAAGGTCCTGCGCGAAAGCCTGCGACCTTATGCCAACGTCCTGCGCGT
CTCCGACATCGCGGCCATGAGCCCTGCCACAGGCGCCCATGAAGAAGTCCAGGT
CTGCGACCTCGCCGATAAAGCGGCGGTCCATCAACTGGTCGAAGGCGTCGACGC
AATCCTGCACTTCGGTGGCGTATCGGTGGAGCGGCCCTTCGAGGAAATCCTCGGG
GCCAATATCTGCGGCGTGTTTCATATCTATGAAGCGGCGCGCCGGCATGGCGTAA
AGCGGGTGATCTTCGCCAGCTCCAACCACGTCATCGGTTTTTATAAGCAGGACGA
AACCATCGACGCCAACTGCCCGCGCCGCCCCGACAGCTACTACGGTCTGTCCAA
GTCCTACGGCGAAGACATGGCCAGCTTCTACTTCGACCGCTACGGCATCGAGACC
GTGAGCATCCGCATCGGCTCCTCGTTCCCCGAGCCGCACAATCGCCGCATGATGA
GCACCTGGCTGAGCTTTGCCGACCTGACGCAGCTGCTCGAACGCGCGCTGTACAC
CCCCAACGTCGGCCACACCGTGGTCTACGGCATGTCCGCTAACAAGAACGTCTG
GTGGGACAACCACCTGGCCGCGCACCTGGGCTTCCAACCGAAGGACAGCTCCGA
GGTGTTCCGTGCGCAGATCGATGCCCAGCCGATGCCCGCCGCCGATGACCCGGC
GATGGTCTTTCAAGGCGGCGCCTTTGTCGCAGCCGGGCCGTTCGGCGACGACTGA

SEQ ID NO: 7 pSGI-433 #8938-Protein

MLNVETRHAVHADHARSLDTEGLRRHFLAQGLFAEGEIRLIYTHYDRFVMGGAVPD
GAPLVLDHVEETKTPGFLDRREMGIVNIGAEGSVHAGNESWSLNRGDVLYLGMGAG
PVTFEGAGRFYLVSAPAHRSLPNRLVTPADSKEVKLGALETSNKRTINQFIHPLVMES
CQLVLGYTTLEDGSVWNTMPAHVHDRRMEAYLYFGMDETSRVLHLMGEPQQTRH
LFVANEEGAISPPWSIHAGAGIGSYTFIWAMAGDNVDYTDMEFIQPGDLR*

SEQ ID NO: 8 pSGI-434_Q46938-Protein

MDVRQSIHSAHAKTLDTQGLRNEFLVEKVFVADEYTMVYSHIDRIIVGGIMPITKTVS
VGGEVGKQLGVSYFLERRELGVINIGGAGTITVDGQCYEIGHRDALYVGKGAKEVV
FASIDTGTPAKFYYNCAPAHTTYPTKKVTPDEVSPVTLGDNLTSNRRTINKYFVPDVL
ETCQLSMGLTELAPGNLWNTMPCHTHERRMEVYFYFNMDDDACVFHMMGQPQET
RHIVMHNEQAVISPSWSIHSGVGTKAYTFIWGMVGENQVFDDMDHVAVKDLR

SEQ ID NO: 9 pSGI-435; gene #3891-Protein

MTMKILYGAGPEDVKGYDTQRLRDAFLLDDLFADDRVSFTYTHVDRLILGGAVPVT
TSLTFGSGTEIGTPYLLSAREMGIANLGGTGTIEVDGQRFTLENRDVLYVGRGARQM
TASSLSAERPARFYMNSVPAGADFPHRLITRGEAKPLDLGDARRSNRRRLAMYIHPE
VSPSCLLLMGITDLAEGSAWNTMPPHLHERRMEAYCYFDLSPEDRVIHMMGRPDET
RHLVVADGEAVLSPAWSIHMGAGTGPYAFVWGMTGENQEYNDVAPVAVADLK*

SEQ ID NO: 10 pSGI-436; gene #7102-Protein

MLTVETRHAIDPQTAKRMDTEELRKHFHMGSLFAAGEIRLVYTHYDRMIVGAAVPS
GAPLVLDQVKECGTASILDRREMAVVNVGASGKVSAAGETYAMERGDVLYLPLGS
GKVTFEGEGRFYILSAPAHAAYPARLIRIGEAEKVKLGSAETSNDRTIYQFVHPAVMT
SCQLVVGYTQLHNGSVWNTMPAHVHDRRMEAYLYFDMKPEQRVFHFMGEPQETR
HLVMKNEDAVVSPPWSIHCGAGTGSYTFIWAMAGDNVDYKDVEMVAMEDLR*

SEQ ID NO: 11 pSGI-437; gene #9209 –Protein

MSYLLRKPQSNEVSNGVKLVHEVTKSNSDLTYVEFKVLDLASGSSYAEELKKQEICI
VAVTGNITVTDHESTFENIGTRESVFERKPTDSVYISNDRSFEITAVSDARVALCYSPS
EKQLPTKLIKAEDNGIEHRGKFSNKRTVHNILPDSDPSANSLLVVEVYTDSGNWSSYP
PHKHDQDNLPEESFLEETYYHELDPGQGFVFQRVYTDDRSIDETMTVENENVVIVPA
GYHPVGVPDGYTSYYLNVMAGPTRKWKFHNDPAHEWILER*

SEQ ID NO: 12 pSGI-438; gene #9732 -Protein

MANLLRKPNGTHGKVHDITPENAKWGYVGFGLFRLKSGESVSEKTGSTEVILVLVE
GKAKISASGEDFGEMGERLNVFEKLPPHCLYVPAESDWHATATTDCVLAVCTAPGK
PGRKAQKLGPESLTLEQRGKGANTRFIHNIAMESRDVADSLLVTEVFTPQGNWSSYP

PHRHDEDNFPDMTYLEETYYHRLNPAQGFGFQRVFTEDGSLDETMAVSDGDVVLVP
KGHHPCGAPYGYEMYYLNVMAGPLRKWRFKNHPDHDWIFKRDNP*

SEQ ID NO: 13 pSGI-439; gene #7403-Protein

MASLLVRPTAPDAQGTVIDVTPESAGWTHVGFRVHKLAKGQRLEASSDDQEVCLVL
LTGRATVTCGEHRFEDIGQRMDIFEQIPPYAVYLPDHVSYAVEATTDLELAVCTAPG
HGNHAPRLIAPDNIKQSTRGQGTNTRHVHDILPETEPADSLLVVEVFTPAGNWSSYPP
HKHDVDNLPHESHLEETYYHRINPEQGFAFQRVYTDDRSLDETMAVENGCCVLVPK
GYHPVGASHGYSLYYLNVMAGPKRAWKFHNDPDHEWLMNAG*

SEQ ID NO: 14 pSGI-440; gene F0J748-Protein

MPDLLRKPFGTHGKVHDITPAAAGWRHVGFGLYRLRAGEFAAEATGGNEVILVMV
EGKASIRAAGRDWGVLGERMSVFEKSPPHSLYVPNGAEWALVAETDCIVAVCSAPG
RGGHAARRIGPEGIVLTARGEGTNTRHINNIAMEAEDYCDALLVTEVFTPAGHWSSY
PSHRHDEDDDPRITYLEETYYHRLNPASGFGVQRVYTDDRALDQTMAVSDGDVVLV
PRGHHPCAAPYGIEMYYLNVMAGPLRKWRFLPDPELGIAK

SEQ ID NO: 15 pSGI-458; gene A5YBJ4-Protein

MSLLYHKQNQELSSGVRLIQDVNASNSPMKYTAVKVLEFSADSSYEETLEAFEAGIV
VLEGKVTITADDQTFEDVGQRTSIFDKIPTDSVYVSTGLAFGIRAKQAAKILIAYAPTN
QTFP

VRLIRGNIHQVEHRGKYNNKRLVQNILPDNLPFADKLLLVEVYTDSANWSSYPPHRH
DHDDLPAESLLEEIYYHEMRPKQGFVFQRVYTDDLSLDETMAVQNQDVVVVPKGY
HPVGVPDGYDSYYLNVMAGPTRVWHFHNAPEHAWIIDRQ

SEQ ID NO: 16 pSGI-478; gene #1874-Protein

MKKFMDENFLLQTETAQKLYHNHAANMPIFDYHCHINPKDIAEDRMFKTITEIWLY
GDHYKWRAMRTNGVDERFCTGDASDWEKFEKWAETVPHTLRNPLYHWTHLELKK
FFGINEILSPKNAREIYDACNEKLQTPAYSCRNIIRMANVHTICTTDDPVDTLEYHQQI
KEDGFEVAVLPAWRPDKAMMVEDPKFFNDYMDQLAEAAGIHIESFEDLMEALDTR
HQYFHDNGCRLSDHGLDTVFAEDYTEEEIKAIFKKIRGGSRLSETEILKFKSCMLYEY
GVMDHSRGWTQQLHIGAQRNNNTRLFKKLGPDTGFDSIGDKPIAEPLAKLLDRLDQ
ENKLCKTVLYNLNPRDNELYATMLGNFQDGSVPGKIQYGSGWWFLDQKDGMIKQ
MNALSNLGLLSRFVGMLTDSRSFLSYTRHEYFRRTLCNLLGNDVENGEIPADMELLG
SMVENICFNNAKNYFNF*

SEQ ID NO: 17 pSGI-479; gene Q9WXR9-Protein

MFLGEDYLLTNRAAVRLFNEVKDLPIVDPHNHLDAKDIVENKPWNDIWEVEGATDH
YVWELMRRCGVSEEYITGSRSNKEKWLALAKVFPRFVGNPTYEWIHLDLWRRFNIK

KVISEETAEEIWEETKKKLPEMTPQKLLRDMKVEILCTTDDPVSTLEHHRKAKEAVE
GVTILPTWRPDRAMNVDKEGWREYVEKMGERYGEDTSTLDGFLNALWKSHEHFKE
HGCVASDHALLEPSVYYVDENRARAVHEKAFSGEKLTQDEINDYKAFMMVQFGKM
NQETNWVTQLHIGALRDYRDSLFKTLGPDSGGDISTNFLRIAEGLRYFLNEFDGKLKI
VLYVLDPTHLPTISTIARAFPNVYVGAPWWFNDSPFGMEMHLKYLASVDLLYNLAG
MVTDSRKLLSFGSRTEMFRRVLSNVVGEMVEKGQIPIKEARELVKHVSYDGPKALFF
G

SEQ ID NO: 18 pSGI-480; gene Q9KFI6-Protein

MSINSREVLAEKVKNAVNNQPVTDMHTHLFSPNFGEILLWDIDELLTYHYLVAEVM
RWTDVSIEAFWAMSKREQADLIWEELFIKRSPVSEACRGVLTCLQGLGLDPATRDLQ
VYREYFAKKTSEEQVDTVLQLANVSDVVMTNDPFDDNERISWLEGKQPDSRFHAAL
RLDPLLNEYEQTKHRLRDWGYKVNDEWNEGSIQEVKRFLTDWIERMDPVYMAVSL
PPTFSFPEESNRGRIIRDCLLPVAEKHNIPFAMMIGVKKRVHPALGDAGDFVGKASM
DGVEHLLREYPNNKFLVTMLSRENQHELVVLARKFSNLMIFGCWWFMNNPEIINEM
TRMRMEMLGTSFIPQHSDARVLEQLIYKWHHSKSIIAEVLIDKYDDILQAGWEVTEE
EIKRDVADLFSRNFWRFVGRNDHVTSVKVEQQT

SEQ ID NO: 19 pSGI-481; gene O34808-Protein

MEPFMGKNFLLKNETAVSLYHNYAKDMPIIDYHCHLSPKEIYENKTFQNITEAWLYG
DHYKWRIMRANGIEETYITGDAPDEEKFMAWAKTVPMAIGNPLYNWTHLELQRFFG
IYEILNEKSGSAIWKQTNKLLKGEGFGARDLIVKSNVKVVCTTDDPVDSLEYHLLLK
EDKDFPVSVLPGFRPDKGLEINREGFPEWVQALEDAAAISITTYDEFLKALEKRVRFF
HSAGGRVSDHAIDTMVFAETTKEEAGRIFSDRLQGTEVSCEDEKKFKTYTLQFLCGL
YAELDWAMQFHINALRNTNTKMMKRLGPDTGYDSMNDEEIAKPLYKLLNSVEMKN
QLPKTILYSLNPNDNYVIASMINSFQDGITPGKIQFGTAWWFNDTKDGMLDQMKALS
NVGLFSRFIGMLTDSRSFLSYTRHEYFRRIVCNLIGEWVENGEVPRDMELLGSIVQGI
CYDNAKHYFQFQEEKANV

SEQ ID NO: 20 pSGI-433; gene #8938-DNA

ATGCTCAACGTGGAAACGAGGCACGCCGTTCACGCGGATCACGCGAGATCACTC
GACACAGAGGGCCTGCGCCGGCACTTCCTGGCCCAGGGCCTGTTTGCGGAGGGC
GAGATACGGCTGATCTATACGCATTATGATCGATTCGTCATGGGAGGCGCCGTGC
CGGACGGCGCGCCACTTGTGCTCGATCATGTCGAGGAGACGAAAACGCCGGGCT
TTCTCGACCGACGGGAGATGGGAATCGTCAATATCGGTGCTGAGGGCAGCGTGC
ATGCCGGCAACGAAAGCTGGTCGCTGAACCGTGGTGACGTACTTTATCTCGGCAT
GGGGGCGGGACCGGTCACCTTCGAAGGGGCTGGGCGCTTCTACCTCGTCTCGGC

ACCGGCGCATCGCAGCCTGCCCAACCGGCTCGTCACGCCGGCCGACAGCAAGGA
GGTCAAGCTTGGCGCTCTCGAGACTTCCAACAAACGCACCATCAATCAGTTCATT
CATCCCCTGGTCATGGAAAGCTGCCAGCTCGTGCTGGGATATACCACGCTGGAGG
ACGGCTCGGTCTGGAATACCATGCCCGCGCATGTGCACGACCGACGCATGGAGG
CCTATCTCTATTTCGGCATGGATGAGACATCGCGGGTTCTGCATCTGATGGGCGA
GCCGCAGCAAACGAGGCATCTCTTCGTCGCCAATGAGGAAGGGGCGATCTCTCC
GCCGTGGTCCATCCATGCGGGAGCAGGCATTGGCAGCTATACCTTCATCTGGGCC
ATGGCGGGCGACAATGTCGATTATACCGACATGGAGTTCATCCAGCCGGGAGAT
CTTCGATGA

SEQ ID NO: 21 pSGI-434; gene Q46938-Protein

ATGGACGTAAGACAGAGCATCCACAGTGCGCACGCAAAAACGCTGGATACCCAA
GGGCTGCGCAATGAATTTTTGGTTGAAAAGGTATTTGTCGCCGATGAGTACACCA
TGGTTTACAGCCACATTGACCGAATTATTGTTGGCGGCATTATGCCGATAACTAA
AACGGTTTCCGTTGGCGGGGAAGTTGGTAAACAACTCGGCGTAAGCTATTTCCTT
GAACGTCGCGAGTTAGGTGTTATCAATATTGGCGGTGCCGGTACGATTACTGTCG
ATGGCCAATGCTATGAAATCGGTCACCGCGACGCCCTGTATGTTGGTAAAGGTGC
AAAAGAAGTTGTCTTTGCCAGTATTGATACCGGCACTCCGGCGAAGTTTTATTAC
AATTGCGCACCCGCGCATACGACGTATCCCACCAAAAAAGTCACACCGGACGAA
GTATCTCCAGTCACGTTAGGCGATAACCTCACCAGTAACCGTCGCACGATTAACA
AATATTTTGTCCCGGATGTACTGGAAACCTGCCAATTGAGTATGGGGCTGACGGA
GCTGGCTCCGGGTAACTTGTGGAACACCATGCCGTGTCACACCCACGAGCGCCG
GATGGAAGTTTATTTCTATTTCAATATGGATGATGACGCCTGCGTTTTCCACATGA
TGGGGCAGCCGCAAGAAACGCGTCATATTGTGATGCATAACGAGCAGGCGGTGA
TCTCCCCGAGCTGGTCGATCCATTCCGGTGTCGGAACCAAAGCTTATACCTTTAT
CTGGGGCATGGTCGGTGAAAACCAGGTCTTTGATGATATGGACCATGTGGCCGTT
AAAGATTTGCGCTAG

SEQ ID NO: 22 pSGI-435; gene #3891-Protein

ATGACGATGAAGATACTCTACGGCGCCGGACCGGAGGATGTGAAAGGGTATGAC
ACGCAGCGCCTGCGCGACGCCTTCCTGCTGGACGACCTCTTCGCCGACGACCGGG
TCAGTTTCACATATACCCATGTCGATCGCCTCATCCTCGGCGGGGCCGTCCCGGT
GACGACGAGCCTCACCTTCGGCTCCGGCACGGAGATCGGAACGCCCTACCTGCTT
TCCGCCCGCGAGATGGGGATCGCCAATCTCGGCGGCACGGGCACGATCGAGGTG
GATGGCCAGCGCTTCACGCTCGAAAACCGCGACGTGCTCTATGTCGGTCGCGGC
GCCCGGCAGATGACCGCCTCCAGCCTGTCGGCGGAGAGGCCAGCCCGCTTCTAC

ATGAATTCCGTGCCCGCCGGCGCCGATTTCCCGCACCGTCTGATCACCCGCGGAG
AGGCCAAGCCCCTCGATCTCGGCGATGCGCGCCGCTCGAACAGGCGCCGGCTCG
CAATGTACATCCATCCGGAGGTCTCGCCGTCCTGCCTGCTGCTCATGGGCATCAC
CGATCTTGCCGAGGGCAGCGCCTGGAACACCATGCCGCCGCATCTGCACGAGCG
GCGGATGGAGGCCTATTGCTACTTCGATCTCTCGCCCGAGGACCGGGTCATCCAC
ATGATGGGTCGGCCGGACGAAACCCGCCACCTTGTCGTGGCCGACGGCGAGGCG
GTCCTCTCTCCCGCCTGGTCGATCCATATGGGTGCCGGGACGGGGCCCTACGCCT
TCGTCTGGGGCATGACCGGCGAAAACCAGGAATACAACGACGTCGCTCCCGTAG
CCGTGGCTGATCTCAAATGA

SEQ ID NO: 23 pSGI-436; gene #7102-Protein

ATGCTGACCGTCGAAACCCGCCACGCCATTGATCCGCAGACCGCAAAGCGGATG
GACACGGAAGAGCTGCGCAAGCATTTCCACATGGGCAGCCTGTTTGCTGCCGGT
GAAATCCGCCTCGTCTACACCCACTATGACCGCATGATCGTCGGCGCTGCCGTGC
CCTCGGGCGCGCCGCTGGTGCTGGATCAGGTCAAGGAATGCGGCACCGCCAGCA
TCCTCGACCGCCGCGAGATGGCTGTCGTCAACGTCGGCGCCAGCGGCAAGGTCT
CTGCAGCAGGCGAAACCTACGCCATGGAACGCGGCGACGTGCTCTATCTGCCGC
TGGGCTCCGGCAAGGTGACCTTCGAAGGCGAAGGCCGCTTCTACATTCTCTCCGC
TCCGGCCCACGCTGCTTACCCGGCCCGCCTGATCCGCATCGGCGAGGCCGAGAA
GGTCAAGCTCGGCTCGGCCGAGACCTCCAACGACCGCACCATCTACCAGTTCGTG
CATCCGGCGGTGATGACTTCCTGCCAACTCGTCGTCGGCTACACCCAGCTGCACA
ACGGCTCTGTCTGGAACACCATGCCCGCCCACGTGCATGACCGGCGCATGGAGG
CCTATCTCTATTTCGACATGAAGCCGGAGCAGCGCGTGTTCCACTTCATGGGCGA
GCCGCAGGAAACCCGCCATCTGGTCATGAAGAACGAGGATGCGGTGGTCTCCCC
GCCCTGGTCCATCCACTGCGGCGCAGGCACCGGCAGCTACACCTTCATCTGGGCC
ATGGCCGGCGACAACGTCGACTACAAGGACGTGGAAATGGTCGCCATGGAGGAT
CTGCGGTGA

SEQ ID NO: 24 pSGI-437; gene #9209 –DNA

ATGAGTTATTTGTTGCGTAAGCCGCAGTCGAATGAAGTGTCTAATGGGGTCAAAC
TGGTGCACGAAGTAACGAAATCCAACTCTGATCTCACCTATGTAGAGTTTAAAGT
GTTAGATCTCGCTTCCGGTTCCAGCTATGCAGAAGAATTGAAAAAACAGGAAAT
CTGTATTGTCGCGGTAACGGGAAACATTACAGTGACCGATCACGAGTCGACTTTT
GAGAATATCGGCACGCGTGAAAGCGTATTCGAACGAAAACCGACAGACAGCGTC
TATATTTCAAATGACCGTTCCTTTGAGATCACAGCGGTCAGCGACGCAAGAGTGG
CGCTTTGCTATTCTCCATCGGAAAAACAGCTTCCGACAAAGCTGATCAAAGCGGA

AGACAATGGCATTGAGCATCGCGGGAAGTTTTCAAACAAACGTACTGTTCACAA
CATTCTTCCGGATTCAGACCCTTCAGCTAACAGCCTATTAGTAGTTGAAGTCTAT
ACAGACAGCGGCAACTGGTCCAGCTATCCGCCTCATAAACATGATCAAGACAAT
TTGCCGGAGGAATCTTTTTTAGAAGAAACGTACTACCATGAGTTAGACCCGGGAC
AGGGCTTTGTGTTTCAGCGTGTATACACAGATGACCGCTCGATTGACGAGACAAT
GACTGTAGAAAATGAAAACGTTGTCATCGTTCCTGCAGGATACCACCCGGTAGG
CGTGCCGGACGGATACACATCCTACTATTTAAATGTCATGGCAGGGCCGACGCG
GAAATGGAAGTTTCATAATGACCCGGCGCATGAGTGGATTTTAGAACGTTAA

SEQ ID NO: 25 pSGI-438; gene #9732 -DNA

ATGGCCAATTTGTTGCGCAAGCCCAACGGCACGCATGGCAAGGTCCACGACATC
ACTCCGGAAAACGCCAAATGGGGTTATGTCGGGTTCGGGCTCTTTCGTCTCAAAT
CCGGCGAGAGTGTCTCCGAAAAGACCGGATCGACGGAGGTGATCCTTGTTCTTGT
GGAAGGCAAGGCAAAGATTTCCGCTTCTGGCGAGGATTTCGGCGAGATGGGTGA
ACGCTTAAACGTGTTCGAGAAACTGCCGCCACACTGCCTCTATGTGCCTGCTGAA
AGCGACTGGCATGCAACCGCCACGACAGATTGTGTTCTGGCTGTTTGCACCGCAC
CGGGCAAGCCAGGCCGCAAGGCACAGAAGCTTGGGCCGGAAAGCTTGACACTTG
AACAACGCGGAAAAGGTGCCAATACCCGCTTTATCCATAATATCGCAATGGAAA
GCCGCGATGTTGCCGATAGCCTTCTTGTTACCGAGGTATTCACACCGCAGGGAAA
CTGGTCGTCCTATCCACCCCACAGACACGACGAAGACAATTTTCCGGATATGACC
TATCTGGAAGAGACCTATTATCACCGTCTCAACCCGGCGCAGGGCTTCGGCTTCC
AGCGTGTTTTCACCGAAGACGGAAGCCTTGATGAAACCATGGCGGTCTCTGACG
GAGACGTCGTGCTTGTACCAAAAGGCCACCATCCATGTGGCGCGCCCTATGGCTA
CGAGATGTATTATCTCAATGTGATGGCCGGTCCCTTGCGCAAATGGCGCTTCAAG
AACCATCCCGACCATGACTGGATTTTCAAACGCGACAATCCGTAA

SEQ ID NO: 26 pSGI-439; gene #7403-DNA

ATGGCTTCCCTACTGGTACGCCCCACCGCCCCAGATGCCCAGGGCACCGTGATTG
ACGTTACCCCTGAATCTGCTGGCTGGACGCACGTTGGCTTTCGGGTGCATAAACT
CGCCAAGGGCCAGCGCCTGGAGGCCAGCAGCGATGATCAGGAAGTCTGCCTGGT
GCTGCTCACCGGTCGCGCCACGGTAACTTGCGGCGAGCACCGCTTTGAAGATATT
GGCCAGCGTATGGATATTTTTGAGCAGATCCCTCCCTATGCGGTTTACCTACCTG
ACCATGTTAGCTACGCGGTGGAAGCGACCACAGACTTAGAGCTAGCGGTGTGCA
CCGCCCCTGGGCATGGCAACCATGCCCCACGGCTCATCGCGCCTGACAACATCA
AGCAAAGCACCCGTGGCCAGGGCACCAACACCCGCCATGTTCACGATATTCTGC
CGGAAACCGAGCCCGCCGATAGCCTATTAGTAGTCGAAGTATTCACACCTGCGG

GTAACTGGTCGAGCTACCCGCCCCACAAACACGATGTGGATAACTTACCCCACG
AATCACATCTGGAAGAGACCTACTACCACCGCATTAACCCTGAACAAGGGTTCG
CCTTCCAGCGCGTTTACACCGATGACCGCAGCCTTGATGAAACCATGGCGGTGGA
AAACGGCTGCTGTGTGTTGGTTCCCAAGGGTTACCATCCGGTGGGCGCCTCCCAT
GGCTACTCGCTCTACTACTTAAATGTGATGGCGGGGCCCAAGCGGGCATGGAAA
TTTCACAACGACCCCGACCACGAATGGCTGATGAACGCTGGATAG

SEQ ID NO: 27 pSGI-440; gene F0J748-DNA

ATGCCGGACTTACTGAGAAAACCGTTTGGCACCCATGGCAAAGTGCACGATATT
ACCCCAGCAGCAGCAGGTTGGAGACATGTTGGTTTTGGCTTATATCGCTTAAGAG
CGGGCGAATTTGCAGCAGAAGCGACAGGCGGCAATGAAGTTATTCTGGTGATGG
TTGAGGGCAAAGCGTCTATTAGAGCAGCAGGCAGAGATTGGGGCGTTTTAGGCG
AACGTATGAGCGTCTTCGAAAAAAGTCCACCACATTCCCTGTATGTCCCGAATGG
TGCAGAATGGGCCTTAGTAGCCGAAACAGATTGCATTGTAGCAGTGTGTAGCGCT
CCGGGTAGAGGAGGTCATGCTGCAAGAAGAATTGGTCCTGAAGGTATTGTGTTA
ACCGCCAGAGGTGAAGGCACCAATACACGCCACATCAACAACATCGCCATGGAA
GCCGAAGATTATTGTGATGCCCTGTTAGTCACCGAAGTGTTCACCCCAGCCGGCC
ATTGGAGCTCTTATCCATCTCATCGTCATGATGAAGACGACGATCCGCGCATCAC
CTATTTAGAAGAGACCTACTATCATCGCTTAAATCCTGCCTCGGGCTTTGGCGTTC
AACGCGTCTATACCGATGATCGCGCCTTAGATCAAACCATGGCGGTTTCTGATGG
CGATGTTGTTTTAGTTCCTCGCGGCCATCATCCGTGTGCAGCCCCGTATGGTATTG
AAATGTATTACCTGAACGTCATGGCCGGCCCGTTACGTAAATGGCGCTTTTTACC
TGATCCTGAACTTGGCATTGCGAAATAA

SEQ ID NO: 28 pSGI-458; gene A5YBJ4-DNA

ATGTCTCTGCTGTACCACAAGCAGAACCAGGAACTGAGTAGTGGTGTGCGCCTG
ATCCAAGATGTTAATGCCAGCAATAGCCCGATGAAATATACCGCCGTGAAAGTG
CTGGAGTTTAGCGCCGATAGCAGCTATGAGGAAACCTTAGAGGCCTTTGAAGCC
GGCATTGTTGTGTTAGAGGGCAAAGTGACCATCACCGCCGACGATCAAACCTTCG
AAGATGTGGGTCAAAGAACCTCGATCTTCGACAAAATCCCGACCGATAGCGTTT
ATGTGTCTACCGGTTTAGCCTTCGGTATTCGCGCCAAACAAGCCGCCAAAATCTT
AATCGCGTATGCTCCGACCAATCAGACCTTCCCAGTTCGCTTAATTCGCGGCAAT
ATCCACCAGGTGGAACATCGCGGCAAGTACAACAACAAACGCTTAGTGCAGAAC
ATTCTCCCGGATAATCTCCCGTTCGCCGATAAATTACTGCTGGTTGAGGTGTACA
CCGATAGCGCCAATTGGAGCTCCTATCCGCCGCATAGACATGATCACGATGATTT
ACCGGCCGAAAGTCTGTTAGAGGAGATCTACTATCACGAAATGCGCCCGAAGCA

GGGCTTCGTCTTTCAACGCGTGTATACCGATGATCTGAGTCTGGATGAGACCATG
GCCGTTCAAAATCAAGATGTTGTCGTTGTCCCGAAAGGCTATCATCCGGTTGGTG
TCCCCGACGGCTATGATTCGTATTACCTGAACGTGATGGCCGGCCCGACAAGAGT
GTGGCATTTTCATAATGCTCCGGAACATGCCTGGATTATTGATCGCCAGTAA

SEQ ID NO: 29 pSGI-478; gene #1874-DNA

ATGAAAAAATTTATGGATGAAAATTTTCTGTTGCAAACCGAAACAGCGCAGAAA
TTGTATCATAATCACGCGGCAAACATGCCGATTTTCGATTACCACTGCCACATTA
ACCCCAAAGACATCGCGGAAGACCGGATGTTTAAAACCATCACCGAAATCTGGT
TGTACGGCGATCATTATAAATGGCGCGCCATGCGTACAAACGGCGTTGACGAGC
GCTTTTGCACCGGCGATGCAAGCGATTGGGAAAAGTTTGAAAAGTGGGCCGAAA
CGGTTCCTCATACCCTGCGTAATCCGCTTTATCACTGGACACACCTGGAGCTAAA
GAAATTTTTCGGGATTAACGAGATCCTGAGTCCGAAAAATGCCCGGGAAATTTAT
GATGCCTGTAACGAAAAACTGCAAACGCCCGCGTATAGTTGCCGCAACATCATC
CGGATGGCCAATGTGCATACAATCTGTACCACCGACGACCCGGTTGACACACTG
GAATATCATCAGCAAATTAAAGAAGACGGCTTTGAAGTGGCGGTTTTACCTGCCT
GGCGTCCGGATAAAGCGATGATGGTGGAAGACCCGAAGTTCTTTAACGACTATA
TGGACCAGTTGGCCGAAGCTGCCGGTATCCATATCGAATCGTTTGAGGATTTGAT
GGAAGCCTTGGATACGCGTCACCAGTATTTTCATGATAATGGTTGCCGTTTGTCC
GACCACGGGCTGGATACCGTTTTTGCTGAAGATTATACGGAGGAAGAAATTAAA
GCGATCTTCAAAAAAATCCGTGGCGGCAGCAGGCTTAGCGAAACGGAAATCCTG
AAATTCAAGTCCTGCATGTTGTACGAATATGGGGTGATGGACCATTCGCGCGGCT
GGACACAACAATTGCACATTGGCGCACAACGCAACAACAACACCCGTTTGTTCA
AAAAATTAGGTCCCGACACTGGTTTCGATTCGATTGGCGATAAGCCGATCGCTGA
ACCATTGGCCAAATTGCTCGACCGCCTGGATCAGGAAAACAAATTGTGCAAAAC
GGTTTTGTATAATCTGAATCCGCGTGATAACGAGTTGTACGCTACCATGTTGGGC
AACTTTCAGGACGGATCGGTTCCCGGGAAAATTCAATACGGCTCGGGTTGGTGGT
TTCTCGATCAGAAAGACGGCATGATTAAACAGATGAATGCCCTTTCCAATCTGGG
TTTGCTGAGCCGTTTCGTAGGCATGCTGACCGACTCAAGGAGCTTCCTTTCGTAC
ACCCGTCACGAATATTTCCGTCGTACCCTTTGCAACCTGCTTGGGAATGATGTTG
AAAACGGGGAGATTCCGGCAGATATGGAGCTTTTGGGCAGTATGGTTGAGAATA
TTTGTTTTAATAACGCGAAGAACTATTTTAATTTTTAG

SEQ ID NO: 30 pSGI-479; gene Q9WXR9-DNA

ATGTTTCTGGGCGAAGACTATCTGCTGACCAATCGTGCGGCAGTTCGTCTGTTCA
ACGAAGTGAAAGATCTGCCGATCGTTGATCCGCATAACCACCTGGATGCGAAAG

ATATCGTGGAAAACAAACCGTGGAACGACATCTGGGAAGTGGAAGGTGCGACCG
ATCACTATGTGTGGGAACTGATGCGTCGTTGTGGTGTTAGCGAAGAATATATTAC
CGGCTCTCGTAGCAACAAAGAAAAATGGCTGGCGCTGGCGAAAGTGTTTCCGCG
TTTTGTGGGTAATCCGACGTACGAATGGATCCACCTGGATCTGTGGCGTCGTTTC
AACATCAAAAAAGTCATCAGCGAAGAAACCGCGGAAGAAATCTGGGAAGAAAC
CAAAAAAAAACTGCCGGAGATGACCCCGCAGAAACTGCTGCGCGACATGAAAGT
GGAAATCCTGTGCACCACCGATGATCCGGTGTCTACCCTGGAACATCACCGTAAA
GCGAAAGAAGCCGTGGAAGGCGTGACCATTTTACCGACCTGGCGTCCGGATCGT
GCAATGAATGTTGATAAAGAAGGTTGGCGTGAATATGTTGAAAAAATGGGTGAA
CGCTATGGCGAAGATACCAGCACCCTGGATGGTTTTCTGAATGCCCTGTGGAAAA
GCCACGAACACTTCAAAGAACACGGCTGTGTGGCGAGCGATCATGCGCTGCTGG
AACCGAGCGTGTACTACGTGGATGAAAACCGCGCGCGTGCAGTTCATGAAAAAG
CATTTTCTGGTGAAAAACTGACTCAAGATGAAATCAACGACTATAAAGCGTTCAT
GATGGTGCAGTTCGGCAAAATGAACCAGGAAACCAACTGGGTGACCCAGCTGCA
CATTGGTGCCCTGCGCGATTACCGCGATAGCCTGTTCAAAACCCTGGGCCCGGAT
TCTGGTGGCGATATCAGCACCAACTTTCTGCGTATTGCTGAAGGTCTGCGTTATTT
TCTGAACGAATTTGATGGTAAACTGAAAATTGTGCTGTACGTGCTGGATCCGACC
CATTTACCGACCATTTCGACCATTGCACGTGCGTTCCCGAACGTGTATGTGGGTG
CACCGTGGTGGTTCAACGATAGCCCGTTCGGCATGGAAATGCACCTGAAATACCT
GGCGAGCGTTGATCTGCTGTACAATCTGGCTGGTATGGTTACCGATTCACGTAAA
TTACTGAGTTTTGGTTCTCGTACCGAAATGTTTCGTCGCGTTCTGTCTAATGTGGT
TGGCGAAATGGTGGAAAAAGGCCAGATCCCGATCAAAGAAGCGCGCGAACTGGT
GAAACACGTGAGCTACGACGGCCCGAAAGCCCTGTTCTTTGGCTGA

SEQ ID NO: 31 pSGI-480; gene Q9KFI6-DNA

ATGAGCATCAACAGCCGTGAAGTTCTGGCGGAAAAAGTGAAAAACGCGGTGAAC
AACCAGCCGGTTACCGATATGCATACCCACCTGTTTAGCCCGAACTTTGGCGAAA
TTCTGCTGTGGGACATCGATGAACTGCTGACCTATCACTACCTGGTTGCGGAAGT
TATGCGTTGGACCGATGTGAGCATTGAAGCGTTTTGGGCAATGAGCAAACGTGA
ACAGGCCGATCTGATTTGGGAAGAACTGTTCATCAAACGCAGCCCGGTGAGCGA
AGCATGTCGTGGCGTTCTGACCTGTTTACAAGGTTTAGGTCTGGATCCGGCAACT
CGTGATTTACAGGTGTATCGTGAATACTTCGCCAAAAAAACCAGCGAGGAACAG
GTGGATACCGTTCTGCAGCTGGCAAATGTGAGCGATGTGGTGATGACCAATGATC
CGTTCGATGATAATGAACGCATCAGCTGGCTGGAAGGCAAACAGCCGGATAGCC
GCTTTCATGCAGCGTTACGTCTGGATCCGCTGCTGAATGAATATGAACAGACCAA

ACATCGTCTGCGTGATTGGGGTTATAAAGTGAACGACGAATGGAACGAAGGCAG
CATCCAGGAAGTGAAACGCTTTCTGACCGACTGGATTGAACGTATGGATCCGGTG
TATATGGCGGTGAGCTTACCGCCGACCTTCAGCTTTCCGGAAGAATCGAACCGTG
GCCGCATTATCCGTGATTGTCTGTTACCGGTTGCAGAAAAACATAACATCCCGTT
TGCAATGATGATTGGCGTGAAAAAACGCGTGCATCCGGCGTTAGGTGATGCAGG
CGATTTTGTGGGTAAAGCAAGTATGGATGGCGTTGAACACCTGCTGCGCGAATAC
CCGAACAACAAATTCCTGGTGACCATGCTGAGCCGCGAAAACCAGCACGAACTG
GTGGTTCTGGCGCGTAAATTTAGTAACCTGATGATTTTTGGTTGTTGGTGGTTTAT
GAACAACCCGGAGATCATCAACGAAATGACCCGCATGCGCATGGAAATGCTGGG
TACCAGCTTTATCCCGCAGCACAGCGATGCCCGTGTTCTGGAACAGCTGATCTAT
AAATGGCACCACAGCAAAAGCATCATCGCGGAAGTCCTGATCGACAAATACGAC
GACATCCTGCAAGCAGGTTGGGAAGTTACCGAAGAAGAAATCAAACGTGATGTG
GCAGATCTGTTTAGCCGCAACTTTTGGCGCTTTGTGGGCCGTAACGATCACGTGA
CCAGCGTGAAAGTGGAACAGCAGACCTGA

SEQ ID NO: 32 pSGI-481; gene O34808-DNA

ATGGAACCGTTTATGGGCAAAAACTTCCTGCTGAAAAACGAGACCGCGGTGAGC
CTGTACCACAACTACGCGAAAGATATGCCGATCATCGACTACCATTGCCATCTGA
GCCCGAAAGAAATCTACGAGAACAAAACCTTCCAGAACATCACCGAAGCGTGGC
TGTACGGCGATCACTACAAATGGCGCATCATGCGTGCGAATGGCATCGAAGAAA
CCTATATTACCGGTGATGCACCGGACGAAGAAAAATTCATGGCGTGGGCGAAAA
CCGTGCCGATGGCCATTGGTAATCCGCTGTATAACTGGACCCATCTGGAACTGCA
ACGTTTTTTTGGCATCTACGAAATCCTGAACGAAAAAAGCGGCAGCGCGATCTGG
AAACAGACCAACAAACTGCTGAAAGGCGAAGGCTTTGGTGCGCGTGATCTGATC
GTGAAAAGCAACGTTAAAGTGGTGTGCACCACCGACGATCCGGTGGATTCTCTG
GAATACCATCTGCTGCTGAAAGAAGACAAAGACTTCCCGGTTAGCGTTTTACCGG
GTTTTCGTCCGGATAAAGGTCTGGAAATCAACCGTGAAGGCTTTCCGGAATGGGT
TCAAGCCCTGGAAGATGCGGCCGCAATTAGCATTACGACCTATGATGAATTTCTG
AAAGCGCTGGAAAAACGCGTGCGCTTCTTCCATAGTGCGGGTGGTCGTGTTAGCG
ATCATGCAATCGATACCATGGTTTTCGCCGAAACCACCAAAGAAGAAGCGGGTC
GCATTTTTAGTGATCGTCTGCAAGGCACCGAAGTTAGCTGCGAAGACGAGAAAA
AATTCAAAACCTACACCCTGCAGTTTCTGTGTGGCCTGTATGCCGAACTGGACTG
GGCAATGCAGTTTCACATCAACGCGCTGCGCAACACCAACACCAAAATGATGAA
ACGCCTGGGTCCGGATACCGGTTATGATAGCATGAACGATGAAGAAATCGCGAA
ACCGCTGTACAAACTGCTGAACAGCGTGGAAATGAAAAACCAACTGCCGAAAAC

CATCCTGTACAGCCTGAACCCGAACGACAACTACGTGATCGCGAGCATGATCAA
CAGCTTCCAGGATGGCATCACCCCGGGCAAAATTCAGTTTGGCACCGCATGGTGG
TTCAACGATACCAAAGATGGTATGCTGGATCAGATGAAAGCACTGAGCAATGTG
GGCCTGTTTAGCCGTTTTATTGGCATGCTGACCGATAGCCGTAGCTTTCTGAGCTA
TACCCGTCACGAATACTTTCGCCGCATTGTGTGTAACCTGATCGGCGAATGGGTG
GAAAACGGCGAAGTTCCGCGCGATATGGAACTGCTGGGTAGTATTGTGCAAGGT
ATTTGCTACGATAACGCGAAACATTACTTCCAGTTCCAGGAGGAAAAAGCGAAC
GTGTGA

SEQ ID NO: 33 pSGI-359-0385-Protein

MSQTPRKLRSQKWFDDPAHADMTAIYVERYLNYGLTRQELQSGRPIIGIAQTGSDLAPCNRH
HLALAERVKAGIRDAGGIPMEFPVHPLAEQGRRPTAALDRNLAYLGLVEILHGYPLDGVVLT
TGCDKTTPACLMAAATVDLPAIVLSGGPMLDGWHDGQRVGSGTVIWHARNLMAAGKLDY
EGFMTLATASSPSVGHCNTMGTALSMNSLAEALGMSLPTCASIPAPYRERAQMAYATGMRI
CDMVREDLRPSHILTRQAFENAIVVASALGASTNCPPHLIAMARHAGIDLSLDDWQRLGEDV
PLLVNCVPAGEHLGEGFHRAGGVPAVMHELFAAGRLHPDCPTVSGKTIGDIAAGAKTRDAD
VIRSCAAPLKHRAGFIVLSGNFFDSAIIKMSVVGEAFRRAYLSEPGSENAFEARAIVFEGPEDY
HARIEDPALNIDEHCILVIRGAGTVGYPGSAEVVNMAPPSHLIKRGVDSLPCLGDGRQSGTSG
SPSILNMSPEAAVGGGLALLRTGDKIRVDLNQRSVTALVDDAEMARRKQEPPYQAPASQTP
WQELYRQLVGQLSTGGCLEPATLYLKVIETRGDPRHSH

SEQ ID NO: 34 pSGI-360-0336-Protein

MSERIKKMNDQNKRIFLRSQEWFDDPEHADMTALYVERYMNYGLTRAELQSGRPIIGIAQTG
SDLTPCNRHHKELAERVKAGIRDAGGIPMEFPVHPIAEQTRRPTAALDRNLAYLGLVEILHGY
PLDGVVLTTGCDKTTPACLMAAATTDIPAIVLSGGPMLDGHFKGELIGSGTVLWHARNLLAT
GEIDYEGFMEMTTSASPSVGHCNTMGTALSMNALAEALGMSLPTCASIPAPYRERGQMAYM
TGKRICEMVLEDLRPSKIMNKQSFENAIAVASALGASSNCPPHLIAIARHMGIELSLEDWQRV
GENIPLIVNCMPAGKYLGEGFHRAGGVPAVLHELQKASVLHEGCASVSGKTMGEIAKNAKT
SNVDVIFPYEQPLKHGAGFIVLSGNFFDSAIMKMSVVGEAFKKTYLSDPNGENSFEARAIVFE
GPEDYHARINDPALDIDEHCILVIRGAGTVGYPGSAEVVNMAPPAELIKKGIDSLPCLGDGRQ
SGTSASPSILNMSPEAAVGGGIALLKTNDRLRIDLNKRSVNVLISDEELEQRRREWKPTVSSSQ
TPWQEMYRNMVGQLSTGGCLEPATLYMRVINQDNLPRHSH

SEQ ID NO: 35 pSGI-365 E3HJU7-Protein

MSQTPRKLRSQKWFDDPAHADMTAIYVERYLNYGLTRQELQSGRPIIGIAQTGSDLAPCNRH
HLALAERIKAGIRDAGGIPMEFPVHPLAEQGRRPTAALDRNLAYLGLVEILHGYPLDGVVLTT
GCDKTTPACLMAAATVDIPAIVLSGGPMLDGWHDGQRVGSGTVIWHARNLMAAGKLDYEG
FMTLATASSPSIGHCNTMGTALSMNSLAEALGMSLPTCASIPAPYRERGQMAYATGLRICDM
VREDLRPSHVLTRQAFENAIVVASALGASSNCPPHLIAMARHAGIDLSLDDWQRLGEDVPLL

VNCVPAGEHLGEGFHRAGGVPAVLHELAAAGRLHMDCATVSGKTIGEIAAAAKTNNADVIR
SCDAPLKHRAGFIVLSGNFFDSAIIKMSVVGEAFRRAYLSEPGSENAFEARAIVFEGPEDYHAR
IEDPTLNIDEHCILVIRGAGTVGYPGSAEVVNMAPPSHLLKRGIDSLPCLGDGRQSGTSASPSIL
NMSPEAAVGGGLALLRTGDRIRVDLNQRSVIALVDQTEMERRKLEPPYQAPESQTPWQELY
RQLVGQLSTGGCLEPATLYLKVVETRGDPRHSH

SEQ ID NO: 36 pSGI-359-0385-DNA

ATGTCTCAGACACCCCGCAAGTTGCGCAGCCAGAAATGGTTCGACGACCCTGCGCATGC
CGATATGACGGCGATTTACGTCGAGCGTTATCTGAATTACGGCCTGACGCGGCAAGAGTT
GCAGTCCGGGCGGCCGATCATCGGCATCGCCCAGACCGGCAGCGATCTGGCGCCCTGCA
ACCGCCATCACCTGGCGCTGGCCGAGCGCGTCAAAGCGGGCATCCGGGACGCGGGCGGC
ATCCCGATGGAGTTCCCCGTGCACCCGCTGGCCGAACAAGGCCGGCGGCCCACGGCCGC
GCTGGACCGCAACCTGGCCTATCTGGGCCTGGTCGAAATCCTGCACGGCTACCCCTTGGA
CGGGGTGGTGCTGACGACTGGCTGCGACAAGACCACGCCTGCCTGCCTGATGGCCGCCG
CCACGGTCGACCTGCCCGCCATCGTGCTGTCCGGCGGCCCCATGCTGGACGGCTGGCACG
ACGGCCAGCGCGTCGGTTCCGGCACCGTCATCTGGCACGCGCGCAACCTGATGGCGGCC
GGCAAGCTTGATTACGAAGGCTTCATGACGCTGGCCACCGCGTCTTCGCCGTCGGTCGGC
CACTGCAACACCATGGGCACGGCGTTGTCGATGAATTCGCTGGCCGAAGCGCTGGGCAT
GTCGCTGCCCACCTGCGCCAGCATTCCCGCCCCCTACCGCGAACGCGCCCAGATGGCCTA
CGCCACCGGCATGCGCATCTGCGACATGGTGCGCGAAGACCTGCGACCCTCCCACATCCT
GACACGGCAGGCATTCGAGAACGCCATCGTCGTGGCATCGGCGCTGGGCGCGTCCACCA
ATTGCCCGCCGCACCTGATCGCGATGGCCCGCCACGCCGGCATCGACCTTAGCCTGGACG
ACTGGCAGCGCCTGGGTGAAGACGTGCCGCTGCTGGTCAACTGCGTGCCGGCGGGCGAG
CATCTGGGCGAGGGCTTCCACCGCGCGGGCGGCGTCCCCGCGGTCATGCATGAACTGTTC
GCCGCCGGGCGCCTTCACCCCGACTGCCCCACCGTATCCGGCAAGACCATCGGGGACAT
CGCCGCGGGCGCCAAGACCCGCGACGCCGACGTCATCCGCAGCTGCGCCGCCCCGCTGA
AACACCGGGCAGGCTTCATCGTGCTGTCGGGCAATTTCTTCGACAGCGCCATCATCAAGA
TGTCGGTCGTAGGCGAAGCGTTCCGCCGCGCCTACCTGTCCGAACCCGGCTCAGAGAAC
GCCTTCGAGGCCCGCGCCATCGTGTTCGAAGGCCCCGAGGACTACCACGCGCGCATCGA
AGACCCGGCGCTGAACATCGACGAACACTGCATCCTTGTCATCCGCGGCGCCGGCACCG
TGGGCTACCCGGGCAGCGCCGAAGTGGTCAACATGGCGCCGCCGTCCCACCTGATCAAG
CGCGGCGTGGATTCCCTGCCGTGCCTGGGGGATGGCAGGCAAAGCGGCACTTCCGGCAG
CCCGTCCATTTTGAACATGTCCCCTGAAGCAGCAGTCGGGGGAGGATTGGCGCTGCTGCG
CACCGGCGACAAGATCCGTGTCGATCTGAACCAGCGCAGCGTCACCGCCTTGGTCGACG
ACGCGGAAATGGCAAGACGGAAGCAAGAACCGCCCTACCAGGCACCGGCCTCGCAAAC
GCCCTGGCAAGAGCTGTACCGGCAACTGGTCGGCCAGTTGTCGACGGGCGGCTGCCTGG
AGCCCGCGACGCTATATCTGAAAGTCATCGAAACGCGCGGCGATCCCCGGCACTCTCACT
GA

SEQ ID NO: 37 pSGI-360-0336-DNA

ATGAGTGAAAGGATCAAAAAAATGAATGATCAAAATAAACGGATTTTTTTACGTAGCCA
AGAATGGTTTGATGATCCTGAACATGCTGACATGACAGCACTCTATGTTGAGCGTTATAT
GAATTATGGCCTGACCCGTGCCGAGCTACAATCAGGCCGCCCGATTATTGGTATTGCACA
AACTGGCAGTGATTTAACTCCATGTAACCGTCACCACAAAGAACTTGCTGAACGGGTTAA
AGCAGGTATTCGAGATGCGGGAGGTATTCCCATGGAATTCCCCGTTCACCCGATTGCAGA
ACAAACCCGTCGCCCTACTGCTGCACTTGATAGAAATTTAGCTTACTTAGGCTTAGTTGA
AATATTGCATGGTTATCCGCTTGATGGTGTGGTGCTAACCACAGGTTGTGACAAAACTAC
ACCTGCTTGTTTAATGGCTGCCGCAACGACAGATATACCAGCCATTGTGTTGTCTGGTGG
ACCAATGCTAGATGGTCATTTTAAAGGTGAGTTAATTGGTTCTGGGACTGTGCTTTGGCA
TGCAAGAAATTTACTTGCCACGGGTGAAATTGATTATGAAGGGTTCATGGAAATGACCA
CTTCAGCATCGCCTTCGGTCGGACATTGCAACACCATGGGCACTGCACTTTCTATGAATG
CCTTGGCAGAAGCTTTGGGCATGTCTTTACCGACATGTGCAAGTATTCCAGCGCCGTATC
GCGAACGAGGGCAAATGGCCTATATGACAGGCAAAAGAATTTGTGAAATGGTTTTAGAA
GATTTACGCCCTTCTAAAATCATGAACAAACAATCATTTGAAAATGCCATCGCGGTAGCT
TCAGCATTAGGGGCATCAAGTAATTGCCCTCCTCACCTCATTGCAATTGCCCGTCATATG
GGCATTGAGCTCAGTTTAGAAGACTGGCAACGCGTTGGGGAGAACATTCCTCTCATTGTG
AACTGTATGCCTGCGGGTAAATATTTAGGTGAAGGTTTTCACCGTGCTGGCGGTGTTCCT
GCTGTTTTGCATGAATTACAAAAGGCCAGCGTTTTACATGAAGGCTGTGCATCAGTCAGC
GGTAAAACGATGGGAGAAATTGCTAAAAATGCTAAAACCTCCAATGTAGATGTTATTTTT
CCATATGAACAACCATTAAAACATGGTGCAGGTTTTATTGTGCTTAGTGGCAATTTCTTC
GACAGCGCCATTATGAAAATGTCTGTTGTGGGTGAAGCATTTAAGAAAACCTATTTATCT
GACCCAAATGGGGAAAATAGCTTTGAAGCACGGGCAATCGTTTTTGAAGGGCCAGAGGA
CTACCATGCACGAATTAATGATCCAGCCTTAGACATTGATGAACATTGTATTTTGGTCAT
TCGTGGCGCTGGAACAGTGGGCTATCCAGGTAGTGCAGAAGTTGTAAATATGGCTCCAC
CCGCAGAGTTAATTAAAAAAGGCATCGATTCACTGCCTTGCTTAGGAGATGGCCGCCAA
AGTGGTACGTCTGCCAGCCCTTCTATTTTAAATATGTCACCCGAAGCGGCGGTAGGCGGT
GGAATTGCATTATTAAAGACCAATGACCGTTTACGCATTGATCTCAATAAACGCTCCGTC
AACGTACTCATTTCTGACGAAGAGTTAGAACAACGCCGCCGTGAGTGGAAACCGACGGT
CTCTTCATCTCAAACACCTTGGCAAGAAATGTATCGCAACATGGTGGGTCAATTATCCAC
TGGCGGTTGTTTGGAACCTGCAACTTTATATATGCGAGTCATAAATCAAGACAACCTTCC
AAGACACTCTCATTAA

SEQ ID NO: 38 pSGI-365 E3HJU7-DNA

ATGAGCCAAACACCGCGTAAATTACGCAGCCAGAAGTGGTTTGACGATCCTGCACATGC
CGATATGACCGCCATCTATGTTGAACGCTACCTGAACTATGGCTTAACCCGCCAAGAACT
GCAAAGTGGTCGCCCGATTATTGGTATTGCCCAAACCGGCAGCGATTTAGCCCCGTGTAA
TCGCCATCATTTAGCCTTAGCCGAACGCATTAAAGCAGGCATTAGAGATGCAGGCGGCA

TTCCTATGGAATTTCCCGTTCATCCGCTGGCCGAACAAGGTAGACGTCCTACAGCAGCAT
TAGATCGCAATTTAGCCTATTTAGGCCTGGTGGAAATTTTACACGGCTATCCCCTGGACG
GTGTGGTGCTGACAACCGGTTGCGATAAAACAACACCGGCGTGTTTAATGGCAGCTGCA
ACAGTTGATATTCCGGCGATCGTGTTATCAGGTGGTCCGATGTTAGATGGCTGGCATGAT
GGCCAAAGAGTTGGCAGTGGTACCGTGATTTGGCATGCACGCAATTTAATGGCAGCAGG
CAAACTGGATTATGAAGGCTTCATGACCCTGGCGACAGCCTCTTCTCCGAGTATTGGACA
CTGTAATACCATGGGCACAGCCTTAAGCATGAATAGTCTGGCAGAAGCCCTGGGTATGTC
TTTACCGACCTGTGCGTCTATTCCAGCCCCGTATAGAGAACGCGGTCAAATGGCGTATGC
TACTGGTTTACGCATTTGCGATATGGTGCGCGAAGATTTACGCCCGTCACATGTTTTAAC
CCGCCAAGCCTTCGAAAATGCCATTGTTGTTGCCTCAGCCTTAGGTGCAAGCTCTAATTG
TCCCCCTCATTTAATTGCCATGGCCCGTCATGCCGGTATCGACTTAAGCCTGGATGACTG
GCAACGCTTAGGCGAAGATGTTCCGTTACTGGTCAATTGTGTGCCTGCCGGTGAACATTT
AGGTGAAGGATTTCATCGCGCGGGTGGTGTTCCTGCTGTTTTACATGAATTAGCTGCCGC
AGGTCGTTTACATATGGATTGTGCTACCGTTTCTGGCAAGACCATCGGCGAAATTGCAGC
TGCCGCAAAAACCAACAACGCAGACGTGATTCGCTCGTGTGATGCCCCGTTAAAACATA
GAGCCGGCTTTATTGTGTTAAGCGGCAATTTCTTCGACTCCGCCATCATCAAGATGTCCG
TTGTGGGTGAAGCCTTTCGCAGAGCCTATTTAAGTGAACCTGGCAGCGAAAATGCCTTTG
AAGCCCGTGCCATCGTGTTTGAAGGCCCGGAAGACTATCATGCCCGCATTGAAGATCCG
ACCCTGAATATTGATGAACACTGCATTCTGGTGATTCGCGGCGCAGGTACCGTTGGTTAT
CCTGGTAGTGCTGAAGTTGTGAATATGGCCCCGCCGAGCCATTTATTAAAACGCGGTATT
GATTCATTACCTTGCCTGGGAGATGGCCGCCAAAGTGGTACCTCAGCTAGTCCGTCTATC
CTGAATATGAGCCCTGAAGCCGCCGTTGGAGGAGGTTTAGCATTATTAAGAACCGGTGA
TCGCATTCGCGTCGATCTGAATCAACGCTCAGTCATTGCATTAGTCGACCAGACCGAAAT
GGAACGCCGCAAATTAGAACCACCGTATCAAGCACCTGAAAGCCAAACCCCGTGGCAAG
AACTGTATCGCCAATTAGTCGGTCAACTGTCAACAGGCGGCTGCCTGGAACCAGCCACCT
TATATTTAAAAGTCGTGGAAACCCGTGGAGATCCTCGTCATAGCCATTAA

SEQ ID NO: 39 - AO#13-0573

MDRRELLKTSALLMAAAPLARAANVPEDHANVPRTNWSKNFHYSTSRVYAPTTPEEVPAIV
LENGHLKGLGSRHCFNNIADSQYAQISMREVKGIQIDEAAQTVTVGAGIAYGELAPVLDKAG
FALANLASLPHISVGGTIATATHGSGVGNKNLSSATRAIEIVKADGSILRLSRDTDGERFRMA
VVHLGALGVLTKVTLDIVPRFDMSQVVYRNLSFDQLEHNLDTILSSGYSVSLFTDWQRNRVN
QVWIKDKATADAPQKPLPPMFYGATLQTAKLHPIDDHPADACTEQMGSVGPWYLRLPHFK
MEFTPSSGEELQTEYFVARKDGYRAIRAVEKLRDKITPHLFITEIRTIAADDLPMSMAYQRDS
MAIHFTWKPEEPTVRKLLPEIEAALAPFGVRPHWGKIFEIPPSYLHKQYPALPRFRAMAQALD
PGGKFRNAYLDRNIFGA

SEQ ID NO: 40 - AO#22-8001

MDKRDFLKGSATTAVALMMGLNESKAFADDSVPRTNWSGNYHYSTNKVLQPASVAETQD
AVRSVAGVRALGTRHSFNGIADSQIAQISTLKLKDVSLDAKSSTVTVGAGIRYGDLAVQLDA
KGFALHNLASLPHISVGGACATATHGSGMGNGNLATAVKAVEFVAADGSVHTLSRDRDGD
RFAGSVVGLGALGVVTHLTLQVQPRFEMTQVVYRDLPFSELEHHLPEIMGAGYSVSLFTDW
QNGRAGEVWIKRRVDQGGASAPPARFFNATLATTKLHPILDHPAEACTDQLNTVGPWYERL
PHFKLNFTPSSGQELQTEFFVPFDRGYDAIRAVETLRDVITPHLYITELRAVAADDLWMSMAY
QRPSLAIHFTWKPETDAVLKLLPQIEAKLAPFGARPHWAKVFTMKSSHVAPLYPRLKDFLVL
AKSFDPKGKFQNAFLQDHVDIA

SEQ ID NO: 41 - AO#28-9635.1

MTASVTNWAGNISFVAKDVVRPGGVEALRKVVAGNDRVRVLGSGHSFNRIAEPGADGVLV
SLDALPQVIDVDTERRTVRVGGGVKYAELARHVNESGLALPNMASLPHISVAGSVATGTHGS
GVNNGPLATPVREVELLTADGSLVTIGKDDARFPGAVTSLGALGVVVALTLDLEPAYGVEQ
YTFTELPLEGLDFEAVASAAYSVSLFTDWREAGFRQVWVKRRIDEPYAGFPWAAPATEKLHP
VPGMPAENCTDQFGAAGPWHERLPHFKAEFTPSSGDELQSEYLLPREHALAALDAVGNVRE
TVSTVLQICEVRTIAADTQWLSPAYGRDSVALHFTWTDDMDAVLPAVRAVESALDGFGARP
HWGKVFTTAPAALRERYPRLDDFRTLRDELDPAGKFTNAFVRDVLEG

SEQ ID NO: 42 - AO#36-7049

MTLERNWAGTHTFAAPRIVNATSIDEVRALVAEAARTGTRVRALGTRHSFTDLADSDGTLIT
VLDIPADPVFDEAAGSVTIGAGTRYGIAAAWLAEHGLAFHNMGSLPHISVGGAIATGTHGSG
NDNGILSSAVSGLEYVDATGELVHVRRGDPGFDGLVVGLGAYGIVVRVTVDVQPAYRVRQD
VYRDVPWDAVLADFEGVTGGAYSVSIFTNWLGDTVEQIWWKTRLVAGDDELPVVPESWLG
VQRDSLTAGNLVETDPDNLTLQGGVPGDWWERLPHFRLESTPSNGDEIQTEYFIDRADGPAA
ITALRALGDRIAPLLLVTELRTAAPDKLWLSGAYHREMLAVHFTWRNLPEEVRAVLPAIEEA
LAPFDARPHWGKLNLLTAERIAEVVPRLADARDLFEELDPAGTFSNAHLERIGVRLPR

SEQ ID NO: 43 - AO#51-9823

MRDAAAANWAGNVRFGAARVVAPESVGELQEIVAGSRKARALGTGHSFSRIADTDGTLIAT
ARLPRRIQIDDGSVTVSGGIRYGDLARELAPNGWALRNLGSLPHISVAGACATGTHGSGDRN
GSLATSVAALELVTASGELVSVRRGDEDFDGHVIALGALGVTVAVTLDLVPGFQVRQLVYE
GLTRDTLLESVQEIFAASYSVSVFTGWDPESSQLWLKQRVDGPGDDGEPPAERFGARLATRP
LHPVPGIDPTHTTQQLGVPGPWHERLPHFRLDFTPSAGDELQTEYFVAREHAAAAIEALFAIG
AVVRPALQISEIRTVAADALWLSPAYRRDVMALHFTWISAEGTVMPAVAAVERALAPFDPV
PHWGKVFALPPAAVRAGYPRAAEFLALAARRDPEAVFRNQYLDAYLPAA

SEQ ID NO: 44 - AO#57-0794

MTQRNWAGNVSYSSSRVAEPASVDDLTALVESEPRVRPLGSRHCFNDIADTPGVHVSLARLR
GEEPRLTAPGTLRTPAWLRYGDLVPVLREAGAALANLASLPHISVAGAVQTGTHGSGDRIGT
LATQVSALELVTGTGEVLRLERGEPDFDGAVVGLGALGVLTHVELDVSPARDVAQHVYEGV

RLDDVLADLGAVTGAGDSVSMFTHWQDPAVVSQVWVKSGGDVDDAAIRDAGGRPADGPR
HPIAGIDPTPCTPQLGEPGPWYDRLPHFRLEFTPSVGEELQSEYLVDRDDAVDAIRAVQDLAP
RIAPLLFVCEIRTMASDGLWLSPAQGRDTVGLHFTWRPDESAVRQLLPEIERALPASARPHW
GKVFTLPGHDVAARYPRWADFVALRRRLDPERRFANAYLERLGL

SEQ ID NO: 45 - AO#76-BAA19135

MTPAEKNWAGNITFGAKRLCVPRSVRELRETVAASGAVRPLGTRHSFNTVADTSGDHVSLA
GLPRVVDIDVPGRAVSLSAGLRFGEFAAELHARGLALANLGSLPHISVAGAVATGTHGSGVG
NRSLAGAVRALSLVTADGETRTLRRTDEDFAGAVVSLGALGVVTSLELDLVPAFEVRQWVY
EDLPEATLAARFDEVMSAAYSVSVFTDWRPGPVGQVWLKQRVGDEGARSVMPAEWLGAR
LADGPRHPVPGMPAGNCTAQQGVPGPWHERLPHFRMEFTPSNGDELQSEYFVARADAVAA
YEALARLRDRIAPVLQVSELRTVAADDLWLSPAHGRDSVAFHFTWVPDAAAVAPVAGAIEE
ALAPFGARPHWGKVFSTAPEVLRTLYPRYADFEELVGRHDPEGTFRNAFLDRYFRR

SEQ ID NO: 46 - AO#251-F3MC79

MGDKLNWAGNYRYRSMELLEPKSLEEVKDLVVSRTSIRVLGSCHSFNGIADTGGSHLSLRK
MNRVIDLDRVQRTVTVEGGIRYGDLCRYLNDHGYALHNLASLPHISVAGAVATATHGSGDL
NASLASSVRAIELMKSDGEVTVLTRGTDPEFDGAVVGLGGLGVVTKLKLDLVPSFQVSQTVY
DRLPFSALDHGIDEILSSAYSVSLFTDWAEPIFNQVWVKRKVGINGEDETSPDFFGALPAPEKR
HMVLGQSVVNCSEQMGDPGPWYERLPHFRMEFTPSAGNELQSEYFVPRRHAVEAMRALGK
LRDRIAPLLFISEIRTIASDTFWMSPCYRQDSVGLHFTWKPDWERVRQLLPLIERELEPFAARP
HWAKLFTMESEMIQARYERLADFRQLLLRYDPIGKFRNTFLDHYIMH

SEQ ID NO: 47 - AO#13-0573-DNA

ATGGATCGTCGTGAACTGCTGAAAACCTCTGCACTGCTGATGGCAGCAGCACCGTTAGCA
CGTGCAGCAAATGTTCCGGAAGATCATGCAAATGTTCCGCGTACCAATTGGAGCAAAAA
CTTCCACTATAGCACCAGCCGCGTTTATGCACCGACTACCCCGGAAGAAGTTCCGGCAAT
TGTTCTGGAAAATGGTCATCTGAAAGGTCTGGGTTCTCGTCACTGCTTCAACAACATCGC
CGATAGCCAGTATGCGCAGATCAGCATGCGCGAAGTTAAAGGCATTCAGATCGATGAAG
CCGCACAAACCGTTACCGTGGGTGCAGGTATTGCGTATGGTGAATTAGCACCGGTGCTGG
ATAAAGCGGGTTTTGCACTGGCAAATTTAGCAAGTTTACCGCATATCAGCGTGGGTGGCA
CCATTGCAACCGCAACACATGGCTCTGGCGTTGGTAACAAAAACCTGTCTTCTGCAACCC
GTGCAATTGAAATCGTGAAAGCGGATGGCAGCATTCTGCGTCTGTCGCGTGATACTGATG
GTGAACGTTTTCGTATGGCGGTGGTTCATCTGGGTGCATTAGGTGTTTTAACCAAAGTTA
CCCTGGATATCGTGCCGCGCTTCGATATGTCTCAGGTGGTGTATCGCAACCTGTCCTTTGA
TCAGCTGGAACACAACCTGGATACCATTCTGAGCTCTGGCTATAGCGTTAGCCTGTTCAC
CGACTGGCAGCGTAATCGTGTTAATCAGGTGTGGATCAAAGATAAAGCGACCGCGGATG
CACCGCAAAAACCGTTACCTCCGATGTTTTATGGTGCGACCCTGCAAACCGCAAAACTGC
ATCCGATCGATGATCATCCGGCAGATGCATGTACCGAACAAATGGGTAGTGTTGGTCCGT
GGTATTTACGTCTGCCGCATTTCAAAATGGAGTTTACCCCGAGCAGCGGTGAAGAATTAC

AGACCGAATACTTCGTGGCGCGCAAAGATGGCTATCGCGCAATTCGTGCCGTGGAAAAA
CTGCGCGATAAAATTACCCCGCACCTGTTTATCACCGAAATCCGCACCATTGCAGCAGAT
GATCTGCCGATGAGCATGGCATATCAACGTGACAGTATGGCGATTCATTTTACCTGGAAA
CCGGAAGAACCGACCGTGCGTAAATTACTGCCGGAAATCGAAGCAGCACTGGCGCCGTT
TGGTGTTCGTCCGCATTGGGGCAAAATTTTTGAAATTCCGCCGAGCTATCTGCATAAACA
GTATCCGGCACTGCCGCGTTTTCGCGCAATGGCACAGGCATTAGATCCTGGTGGCAAATT
TCGTAATGCATATCTGGATCGTAACATCTTTGGCGCGTAG

SEQ ID NO: 48 - AO#22-8001-DNA

ATGGACAAACGCGATTTCCTGAAAGGTAGCGCAACCACCGCAGTTGCACTGATGATGGG
TCTGAATGAAAGCAAAGCGTTTGCGGATGATAGCGTTCCGCGTACCAATTGGAGCGGCA
ACTACCATTATAGCACCAACAAAGTGCTGCAGCCGGCAAGTGTTGCAGAAACCCAAGAT
GCAGTTCGTAGTGTTGCAGGTGTTCGTGCATTAGGTACTCGTCATAGCTTTAACGGCATC
GCGGATAGCCAGATTGCCCAGATTAGTACCCTGAAACTGAAAGATGTGAGCCTGGATGC
GAAAAGCTCGACCGTGACCGTTGGTGCAGGTATTCGTTATGGTGATCTGGCGGTTCAGCT
GGATGCGAAAGGTTTTGCTCTGCATAATCTGGCAAGTCTGCCGCATATTTCTGTTGGTGG
TGCATGTGCAACTGCGACCCATGGTTCAGGTATGGGTAATGGTAATTTAGCAACCGCAGT
TAAAGCGGTGGAATTTGTTGCGGCGGATGGTAGCGTGCATACCCTGTCTCGTGATCGTGA
TGGTGATCGTTTTGCGGGCTCTGTTGTTGGTCTGGGTGCATTAGGTGTTGTTACCCATTTA
ACCCTGCAAGTTCAGCCACGTTTCGAAATGACCCAGGTGGTGTACCGTGATCTGCCATTT
AGTGAACTGGAACATCATCTGCCGGAAATTATGGGTGCCGGTTATAGCGTGTCCCTGTTT
ACCGATTGGCAGAATGGTCGTGCAGGTGAAGTGTGGATCAAACGTCGCGTGGATCAAGG
TGGTGCAAGTGCTCCTCCAGCTCGTTTTTTTAATGCAACCTTAGCAACCACCAAACTGCA
CCCGATCCTGGATCATCCTGCTGAAGCATGTACCGATCAGTTAAATACCGTAGGTCCGTG
GTATGAACGTTTACCGCACTTCAAACTGAACTTCACCCCGAGCAGTGGCCAAGAATTACA
GACCGAGTTTTTCGTGCCGTTCGATCGCGGCTATGACGCCATTCGTGCCGTTGAAACTTT
ACGTGATGTGATTACCCCGCACCTGTATATCACCGAACTGCGTGCAGTTGCAGCTGATGA
TTTATGGATGAGCATGGCATATCAACGTCCGAGTCTGGCAATCCATTTTACCTGGAAACC
GGAAACCGATGCAGTGCTGAAATTACTGCCGCAGATTGAAGCGAAACTGGCCCCGTTTG
GTGCTCGTCCGCATTGGGCAAAAGTTTTTACCATGAAAAGCAGCCATGTGGCACCGCTGT
ATCCGCGCCTGAAAGATTTTCTGGTTCTGGCAAAATCCTTTGATCCGAAAGGCAAATTCC
AAAACGCGTTTCTGCAGGACCATGTGGACATCGCATAG

SEQ ID NO: 49 - AO#28-9635-DNA

ATGACCGCATCTGTGACCAATTGGGCGGGTAACATCAGCTTTGTGGCGAAAGATGTTGTT
CGTCCGGGTGGTGTTGAAGCACTGCGTAAAGTTGTTGCGGGTAATGATCGTGTTCGTGTT
CTGGGTTCTGGTCATAGCTTTAACCGTATCGCTGAACCGGGTGCTGATGGTGTTCTGGTT
AGCCTGGATGCATTACCGCAAGTGATTGATGTTGATACCGAACGTCGTACCGTGCGTGTT
GGTGGTGGTGTTAAATACGCGGAACTGGCTCGTCATGTGAATGAATCTGGTCTGGCACTG

CCGAATATGGCATCTCTGCCGCATATTTCTGTTGCAGGTTCTGTTGCAACTGGTACCCATG
GTTCTGGTGTGAATAATGGCCCGTTAGCAACCCCGGTTCGTGAAGTTGAATTATTAACCG
CGGATGGCTCTCTGGTGACCATCGGTAAAGATGATGCGCGTTTTCCGGGTGCAGTTACTT
CTCTGGGTGCGCTGGGTGTTGTTGTTGCACTGACCTTAGATTTAGAACCGGCGTATGGTG
TTGAACAGTATACCTTTACCGAATTACCGCTGGAAGGTCTGGACTTCGAAGCAGTTGCGA
GTGCAGCATATTCTGTTAGCCTGTTCACCGATTGGCGTGAAGCTGGTTTTCGCCAAGTTTG
GGTGAAACGCCGCATTGATGAACCGTACGCGGGCTTTCCGTGGGCAGCACCGGCAACTG
AAAAATTACATCCGGTTCCGGGTATGCCAGCAGAAAATTGTACTGATCAATTTGGTGCAG
CAGGTCCATGGCATGAACGTTTACCGCATTTTAAAGCGGAATTTACCCCGTCTAGCGGTG
ATGAATTACAGAGCGAATATCTGCTGCCGCGTGAACATGCACTGGCGGCACTGGATGCA
GTGGGCAACGTGCGTGAAACCGTTTCTACCGTGCTGCAGATTTGCGAAGTTCGTACCATT
GCAGCAGATACCCAGTGGTTAAGTCCGGCTTATGGTCGTGATAGTGTTGCATTACATTTT
ACTTGGACCGATGATATGGATGCAGTTTTACCTGCAGTTCGTGCCGTTGAAAGCGCGCTG
GATGGCTTTGGTGCTCGCCCGCATTGGGGTAAAGTGTTTACCACCGCACCGGCAGCATTA
CGTGAACGTTATCCGCGTCTGGATGATTTTCGTACCCTGCGTGATGAATTAGATCCGGCA
GGCAAATTTACTAATGCATTTGTTCGTGATGTTCTGGAAGGTTAG

SEQ ID NO: 50 - AO#36-7049-DNA

ATGACCCTGGAACGTAATTGGGCAGGTACCCATACCTTTGCAGCACCGCGTATTGTTAAT
GCAACCAGCATCGATGAAGTTCGTGCGTTAGTGGCAGAAGCAGCACGTACCGGTACCCG
TGTTCGTGCATTAGGTACTCGTCATTCTTTTACCGATCTGGCAGATAGCGATGGTACCCTG
ATTACCGTGCTGGATATTCCGGCAGATCCAGTTTTCGATGAAGCAGCAGGTAGCGTTACC
ATTGGTGCAGGTACCCGTTATGGTATTGCAGCAGCATGGTTAGCAGAACATGGTCTGGCG
TTTCACAACATGGGTAGCCTGCCGCATATTAGCGTTGGTGGTGCAATTGCAACCGGTACC
CATGGTAGTGGTAATGATAACGGCATTCTGAGTAGCGCAGTTAGTGGTCTGGAATATGTT
GATGCGACCGGTGAACTGGTTCATGTGCGTCGTGGTGATCCTGGTTTTGATGGTCTGGTT
GTTGGTTTAGGCGCGTATGGTATTGTGGTTCGTGTGACGGTGGATGTTCAACCGGCATAT
CGTGTTCGCCAGGATGTGTATCGTGATGTTCCGTGGGATGCAGTTCTGGCAGATTTTGAA
GGTGTTACAGGTGGTGCGTATAGCGTTAGCATCTTTACCAACTGGCTGGGTGATACGGTG
GAACAGATTTGGTGGAAAACCCGTCTGGTTGCAGGTGATGATGAACTGCCGGTGGTTCC
GGAAAGCTGGCTGGGTGTTCAACGTGATTCTTTAACCGCAGGTAATCTGGTTGAAACCGA
TCCGGATAATTTAACCCTGCAAGGTGGTGTTCCGGGTGATTGGTGGGAACGTTTACCGCA
TTTTCGTCTGGAAAGTACCCCGTCTAATGGTGATGAAATCCAGACCGAATACTTCATCGA
TCGCGCGGATGGTCCGGCGGCAATTACCGCACTGCGTGCATTAGGTGATCGTATTGCTCC
GTTACTGTTAGTTACCGAATTACGTACCGCAGCTCCAGATAAACTGTGGCTGAGTGGCGC
ATATCATCGCGAAATGTTAGCGGTCCATTTTACCTGGCGTAATTTACCGGAAGAAGTGCG
TGCAGTTTTACCAGCGATCGAAGAAGCCCTGGCGCCGTTTGATGCTCGTCCGCATTGGGG
TAAACTGAATCTGTTAACCGCAGAACGTATTGCAGAAGTTGTTCCGCGTCTGGCTGATGC

ACGTGATCTGTTTGAAGAACTGGACCCGGCTGGTACCTTTTCTAATGCTCATCTGGAACG
TATTGGTGTTCGTTTACCGCGTTAG

SEQ ID NO: 51 - AO#51-9823-DNA

ATGCGTGATGCAGCAGCAGCAAATTGGGCAGGTAATGTGCGTTTTGGTGCAGCACGTGTT
GTTGCACCGGAAAGTGTTGGTGAACTGCAGGAAATTGTTGCAGGTAGCCGTAAAGCACG
TGCATTAGGTACCGGTCATAGCTTTAGCCGTATTGCAGATACCGATGGTACCCTGATTGC
TACCGCACGTTTACCACGTCGTATTCAGATCGATGATGGCAGCGTTACCGTTTCTGGTGG
TATCCGTTATGGCGATCTGGCCCGTGAATTAGCACCGAATGGTTGGGCATTACGTAATCT
GGGTTCTTTACCGCACATTTCAGTTGCAGGTGCATGTGCAACCGGTACCCATGGTTCAGG
TGATCGTAATGGTAGTCTGGCAACCTCTGTTGCAGCGTTAGAATTAGTTACCGCGTCTGG
TGAATTAGTGAGCGTTCGTCGTGGCGATGAAGATTTCGATGGCCATGTGATTGCGCTGGG
TGCACTGGGTGTTACTGTTGCAGTTACCCTGGATTTAGTTCCGGGTTTTCAGGTTCGTCAG
CTGGTGTATGAAGGTCTGACCCGTGATACCTTACTGGAAAGTGTGCAGGAAATCTTTGCT
GCGAGCTATAGTGTTAGCGTGTTTACCGGTTGGGACCCGGAAAGTTCTCAACTGTGGCTG
AAACAGCGCGTTGATGGTCCGGGCGATGATGGTGAACCACCGGCAGAACGTTTTGGTGC
ACGTTTAGCAACTCGTCCGTTACATCCAGTTCCGGGTATTGATCCGACTCATACTACTCA
ACAATTAGGTGTTCCAGGTCCGTGGCATGAACGTTTACCGCATTTTCGTCTGGATTTTACC
CCTTCTGCAGGTGATGAACTGCAAACCGAATACTTCGTGGCCCGCGAACATGCAGCGGC
GGCGATTGAAGCACTGTTTGCGATTGGTGCGGTTGTTCGTCCGGCATTACAAATTAGCGA
AATTCGTACCGTTGCAGCTGATGCATTATGGCTGTCTCCGGCATATCGTCGTGATGTTATG
GCGTTACATTTTACCTGGATTAGCGCAGAAGGTACCGTTATGCCAGCAGTTGCAGCAGTG
GAACGTGCACTGGCGCCGTTTGATCCGGTTCCTCATTGGGGTAAAGTTTTTGCGCTGCCG
CCAGCAGCAGTTCGTGCTGGTTATCCTCGTGCAGCAGAATTTTTAGCATTAGCAGCTCGT
CGTGATCCGGAAGCAGTTTTTCGTAATCAGTATTTAGATGCATATTTACCGGCAGCATAG

SEQ ID NO: 52 - AO#57-0794-DNA

ATGACCCAGCGTAATTGGGCGGGTAATGTGAGCTATAGTAGCAGCCGTGTTGCAGAACC
AGCAAGTGTGGATGATTTAACCGCACTGGTTGAAAGTGAACCGCGTGTTCGTCCGTTAGG
TAGTCGTCATTGCTTCAACGATATCGCCGATACCCCAGGTGTTCATGTTTCTCTGGCACGT
CTGCGTGGTGAAGAACCGCGTTTAACAGCACCGGGTACCTTACGTACTCCAGCTTGGTTA
CGTTATGGTGATTTAGTTCCGGTTCTGCGTGAAGCAGGTGCAGCATTAGCAAATTTAGCA
TCTCTGCCGCATATTAGCGTTGCAGGTGCAGTTCAAACCGGTACCCATGGTTCAGGTGAT
CGTATTGGCACTCTGGCAACCCAAGTTAGCGCCCTGGAATTAGTGACCGGCACCGGTGA
AGTTTTACGCTTAGAACGTGGTGAACCTGATTTTGATGGTGCGGTTGTTGGTTTAGGTGC
GTTAGGTGTTCTGACTCATGTGGAATTAGATGTTAGTCCGGCGCGTGATGTTGCACAGCA
CGTGTATGAAGGTGTTCGTCTGGATGATGTTCTGGCGGATTTAGGCGCGGTTACTGGCGC
AGGTGATTCGGTGAGCATGTTTACCCATTGGCAAGATCCGGCAGTTGTTAGTCAGGTTTG
GGTTAAAAGTGGCGGTGATGTGGATGATGCAGCAATTCGTGATGCAGGTGGTCGTCCGG

CAGATGGTCCGCGTCATCCAATTGCAGGTATTGATCCGACTCCATGTACTCCACAATTAG
GTGAACCAGGTCCGTGGTATGATCGTCTGCCGCATTTTCGTCTGGAATTTACCCCGAGTG
TTGGTGAAGAACTGCAAAGTGAATATCTGGTTGATCGCGATGATGCCGTTGATGCAATTC
GTGCGGTGCAGGATTTAGCCCCGCGTATTGCGCCGCTGCTGTTTGTTTGCGAAATTCGTA
CCATGGCAAGTGATGGTTTATGGCTGAGCCCGGCACAAGGTCGTGATACCGTTGGTCTGC
ATTTTACCTGGCGTCCTGATGAATCTGCAGTTCGTCAATTATTACCGGAAATTGAACGTG
CTTTACCGGCAAGTGCTCGTCCGCATTGGGGTAAAGTGTTTACCCTGCCGGGCCATGATG
TTGCAGCACGTTATCCGCGTTGGGCAGATTTTGTTGCATTACGTCGTCGTTTAGATCCGGA
ACGTCGTTTCGCGAATGCATACCTGGAACGTTTAGGTCTGTAG

SEQ ID NO: 53 - AO#76-BAA19135-DNA

ATGACTCCGGCGGAAAAAAATTGGGCGGGCAACATCACCTTTGGTGCAAAACGTCTGTG
TGTTCCGCGTTCTGTTCGTGAACTGCGTGAAACCGTTGCAGCATCTGGTGCAGTTCGTCC
GTTAGGTACTCGTCATAGCTTTAATACCGTTGCAGATACCAGTGGTGATCATGTTAGTCT
GGCAGGTTTACCGCGTGTTGTGGACATCGATGTTCCGGGTCGTGCAGTTTCTCTGTCTGCT
GGTCTGCGTTTTGGTGAATTTGCGGCTGAATTACATGCACGTGGTCTGGCGCTGGCAAAT
TTAGGTTCTCTGCCGCATATTAGCGTTGCAGGTGCAGTTGCAACCGGTACTCATGGTTCT
GGTGTTGGTAATCGTTCTTTAGCAGGTGCAGTTCGTGCTTTATCTCTGGTAACCGCCGATG
GTGAAACCCGTACCTTACGTCGTACCGATGAAGATTTTGCAGGTGCAGTGGTTTCTCTGG
GTGCACTGGGTGTTGTTACTTCTCTGGAACTGGATTTAGTTCCGGCGTTCGAAGTGCGTC
AGTGGGTGTACGAAGATCTGCCGGAAGCAACTTTAGCAGCTCGTTTTGATGAAGTTATGT
CAGCAGCGTATAGCGTGTCCGTGTTCACCGATTGGCGTCCGGGTCCTGTTGGTCAAGTTT
GGCTGAAACAACGTGTTGGTGATGAAGGTGCTCGTAGTGTTATGCCAGCAGAATGGTTA
GGTGCACGTTTAGCAGATGGTCCGCGTCATCCAGTTCCAGGTATGCCTGCAGGTAATTGT
ACAGCACAACAAGGTGTTCCAGGTCCGTGGCATGAACGTTTACCGCATTTTCGCATGGAA
TTTACCCCGTCTAACGGCGATGAACTGCAAAGCGAATATTTTGTGGCGCGTGCAGATGCA
GTTGCAGCGTATGAAGCATTAGCACGTCTGCGTGATCGTATTGCGCCGGTTCTGCAAGTT
AGCGAATTACGTACCGTTGCAGCAGATGATCTGTGGCTGAGTCCGGCACATGGTCGTGAT
AGTGTTGCGTTTCATTTTACCTGGGTTCCGGATGCAGCAGCAGTTGCACCGGTTGCAGGT
GCTATTGAAGAAGCATTAGCACCGTTTGGTGCACGTCCACATTGGGGTAAAGTTTTTAGC
ACCGCACCGGAAGTTTTACGTACCTTATATCCGCGTTATGCCGATTTCGAAGAACTGGTG
GGCCGCCATGATCCGGAAGGCACCTTTCGTAATGCATTTTTAGATCGCTACTTTCGTCGCT
AG

SEQ ID NO: 54 - AO#251-F3MC79-DNA

ATGGGCGATAAACTGAATTGGGCGGGCAACTATCGTTATCGCAGCATGGAACTGCTGGA
ACCGAAAAGCCTGGAAGAAGTGAAAGATCTGGTGGTTAGCCGTACCAGCATTCGTGTTC
TGGGTAGCTGTCATAGCTTTAACGGCATTGCGGATACCGGTGGTAGTCATCTGAGTCTGC
GCAAAATGAACCGCGTGATTGATCTGGATCGTGTTCAGCGTACCGTTACCGTTGAAGGTG

GTATTCGTTACGGTGATCTGTGCCGCTATCTGAACGATCATGGTTATGCCCTGCATAATCT
GGCAAGCTTACCGCACATCAGCGTTGCAGGTGCAGTTGCAACCGCAACCCATGGTTCTGG
TGATCTGAATGCAAGTCTGGCAAGCTCTGTTCGTGCAATTGAACTGATGAAAAGCGATGG
CGAAGTTACGGTTCTGACCCGTGGTACCGATCCGGAATTTGATGGTGCAGTTGTTGGTCT
GGGTGGTTTAGGTGTTGTGACCAAACTGAAACTGGATCTGGTTCCGAGCTTTCAGGTGTC
GCAGACCGTGTATGATCGTCTGCCGTTTAGCGCACTGGATCATGGCATCGATGAAATTCT
GAGTAGTGCATATAGCGTTAGCCTGTTCACCGATTGGGCGGAACCGATCTTTAATCAGGT
GTGGGTGAAACGCAAAGTGGGCATTAACGGCGAAGATGAAACCAGTCCGGATTTTTTTG
GCGCATTACCGGCACCGGAAAAACGCCACATGGTTCTGGGTCAGAGCGTGGTGAATTGC
AGCGAACAAATGGGTGATCCTGGTCCGTGGTATGAACGTTTACCGCATTTTCGCATGGAA
TTTACCCCGAGTGCAGGCAATGAATTACAGAGCGAATATTTTGTGCCGCGTCGTCATGCG
GTTGAAGCAATGCGTGCGTTAGGTAAACTGCGTGATCGTATTGCACCACTGCTGTTCATC
AGCGAAATCCGCACCATTGCGAGCGATACCTTCTGGATGAGCCCGTGTTATCGTCAGGAT
TCTGTTGGTCTGCATTTTACCTGGAAACCGGATTGGGAACGTGTTCGTCAGTTATTACCGC
TGATTGAACGTGAACTGGAACCGTTTGCGGCACGTCCGCATTGGGCGAAACTGTTTACCA
TGGAAAGCGAAATGATTCAGGCGCGCTATGAACGTCTGGCGGATTTTCGTCAGCTGCTGC
TGCGTTATGATCCGATTGGCAAATTCCGTAACACCTTTCTGGATCACTACATCATGCACT
AA

SEQ ID NO: 55 pSGI-431 Q72LK2-Protein

MEATLPVLDAKTAALKRRSIRRYRKDPVPEGLLREILEAALRAPSAWNLQPWRIVVVRDPAT
KRALREAAFGQAHVEEAPVVLVLYADLEDALAHLDEVIHPGVQGERREAQKQAIQRAFAA
MGQEARKAWASGQSYILLGYLLLLLEAYGLGSVPMLGFDPERVKAILGLPSHAAIPALVALG
YPAEEGYPSHRLPLERVVLWR

SEQ ID NO: 56 pSGI-431 Q72LK2-DNA

ATGGAAGCAACCTTACCGGTGTTAGACGCGAAAACCGCAGCACTGAAACGTCGTAGCAT
TCGCCGTTATCGCAAAGATCCAGTTCCGGAAGGTTTACTGCGCGAAATTCTGGAAGCAGC
ATTACGTGCACCGTCTGCATGGAATTTACAACCGTGGCGTATTGTGGTGGTTCGTGATCC
GGCAACTAAACGTGCATTACGTGAAGCAGCATTTGGTCAAGCCCATGTGGAAGAAGCAC
CGGTTGTTCTGGTTCTGTACGCAGATCTGGAAGATGCACTGGCACATCTGGATGAAGTGA
TTCATCCGGGCGTTCAAGGTGAACGTCGTGAAGCGCAGAAACAAGCAATTCAGCGTGCA
TTTGCAGCAATGGGTCAGGAAGCTCGTAAAGCTTGGGCAAGCGGTCAAAGTTATATTCTG
CTGGGTTATCTGCTGCTGCTGCTGGAAGCATATGGTCTGGGTTCTGTTCCGATGCTGGGTT
TTGATCCTGAACGTGTTAAAGCGATTCTGGGCCTGCCGTCACATGCAGCGATTCCGGCAT
TAGTTGCACTGGGTTATCCGGCTGAAGAAGGTTATCCGAGTCATCGTTTACCGCTGGAAC
GTGTTGTTTTATGGCGTTGA

SEQ ID NO: 57: pSGI-374 #9041 Protein

MKNPFSLQGRKALVTGANTGLGQAIAVGLAAAGAEVVCAARRAPDETLEMIASDGGKASA
LSIDFADPLAAKDSFAGAGFDILVNNAGIIRRADSVEFSELDWDEVMDVNLKALFFTTQAFAK
ELLAKGRSGKVVNIASLLSFQGGIRVPSYTAAKHGVAGLTKLLANEWAAKGINVNAIAPGYI
ETNNTEALRADAARNKAILERIPAGRWGRSEDIAGAAVFLSSAAADYVHGAILNVDGGWLA
R

SEQ ID NO: 58 pSGI-375 #8939 Protein

MIAGVGGEARELALDLSDPMAAKDVFAEGAYDLLINNAGIIRRADAVDFSEDDWDAVMDV
NLKAVFFTSQAFARALMSRNASGKIVNIASLLSFQGGIRVASYTAAKHGVAGITRLLANEWA
SRGINVNAIAPGYIATNNTEALRADEERNAAILARIPAGRWGRAEDIAGTAVYLCSPAADYV
HGAILNVDGGWLAR

SEQ ID NO: 59 pSGI-376 P37769-Protein

MILSAFSLEGKVAVVTGCDTGLGQGMALGLAQAGCDIVGINIVEPTETIEQVTALGRRFLSLT
ADLRKIDGIPALLDRAVAEFGHIDILVNNAGLIRREDALEFSEKDWDDVMNLNIKSVFFMSQA
AAKHFIAQGNGGKIINIASMLSFQGGIRVPSYTASKSGVMGVTRLMANEWAKHNINVNAIAP
GYMATNNTQQLRADEQRSAEILDRIPAGRWGLPSDLMGPIVFLASSASDYVNGYTIAVDGG
WLAR

SEQ ID NO: 60 pSGI-395 #5112 Protein

MPGMTTPFDLHGKTAIVTGANTGIGQAIALSLAQAGADIAAVGRTPAQDTVDQVRALGRRA
DIISADLSTIEPVQRVLDETLEKLGALDILVNNAGIIRRADSVDFTEEDWDAVIDTNLKTTFFLC
QAAGRHMLAQGAGKIINIASLLSFQGGIRVPSYTASKSGVAGLTKLLANEWAAKGVNVNAIA
PGYIATNNTAALQADETRNRQIQERIPAGRWGDPADIGGAAVFLASSAADYIHGHTLAVDGG
WLAR

SEQ ID NO: 61 pSGI-396 #7103-Protein

MNPFSLEGKTALVTGANTGIGQAIAMALGRAGADVICAGRSSCAETVALIAGSKGKARELVL
DFADPMAARDVFAAEPVDILVNNAGIIRRADAVDFTEADWDEVMDVNLKAVFFTCQAFGK
AVLGRGGNGKIVNIASLLSFQGGIRVPSYTASKHGVAGITKLLANEWAAKGINVNAIAPGYIE
TNNTEALRADPVRNKAILERIPAGRWGQASDIGEAAVFLASPAANYIHGAVLNVDGGWLAR

SEQ ID NO: 62 pSGI-374 #9041 DNA

ATGAAGAATCCCTTTTCGCTTCAGGGGCGTAAGGCGCTCGTCACCGGCGCGAATACGGGGCTTGGC
CAGGCGATTGCGGTTGGGCTCGCCGCGGCCGGTGCGGAGGTGGTCTGCGCCGCCCGCCGCGCGCC
GGATGAAACGCTGGAGATGATCGCCAGCGACGGCGGCAAGGCCAGCGCATTGTCCATCGATTTTG
CCGATCCGCTGGCGGCGAAGGACAGTTTTGCCGGCGCCGGTTTCGATATTCTCGTCAACAATGCCG
GTATCATCCGCCGTGCCGATTCCGTCGAGTTCTCCGAACTCGACTGGGACGAGGTGATGGACGTCA
ATCTCAAGGCGCTGTTTTTCACCACCCAGGCTTTTGCGAAAGAGCTGCTGGCGAAAGGCCGGTCCG
GCAAGGTGGTCAATATCGCTTCGCTCCTTTCCTTTCAGGGCGGTATTCGCGTGCCGTCCTATACGGC
GGCGAAACATGGTGTCGCCGGCCTAACCAAACTCCTGGCGAATGAATGGGCCGCCAAGGGCATCA

ATGTGAATGCCATTGCGCCCGGTTATATCGAAACCAACAATACCGAGGCGCTACGCGCCGATGCG
GCTCGTAACAAGGCCATTCTCGAGCGCATCCCGGCCGGCCGCTGGGGGCGCTCGGAAGACATCGC
CGGGGCGGCGGTTTTCCTGTCATCTGCGGCGGCGGACTATGTGCATGGCGCCATTCTCAACGTCGA
TGGCGGCTGGCTGGCGCGCTGA

SEQ ID NO: 63 pSGI-375 #8939 DNA

ATGATCGCCGGCGTGGGGGGGAGAAGCAAGGGAGCTGGCGCTCGATCTGTCCGATCCCATGGCGGC
AAAAGATGTTTTTGCTGAAGGCGCTTACGACCTCCTCATCAACAATGCCGGCATCATCCGCCGTGC
CGATGCAGTCGATTTCTCCGAGGATGACTGGGACGCGGTGATGGACGTGAACCTGAAAGCCGTCT
TCTTCACCTCGCAAGCCTTTGCGCGGGCTCTCATGTCCAGAAACGCAAGCGGAAAGATCGTTAACA
TTGCATCCCTTCTGTCGTTTCAAGGCGGCATTCGCGTTGCCTCCTACACGGCCGCCAAGCACGGTGT
GGCAGGCATCACCAGACTGTTGGCAAACGAATGGGCGTCCCGCGGCATCAACGTCAATGCGATAG
CGCCCGGTTACATTGCCACGAACAACACGGAAGCGCTTCGAGCCGACGAGGAGCGCAACGCGGCG
ATCCTCGCACGCATTCCGGCTGGCCGCTGGGGGCGGGCGGAGGATATTGCGGGTACTGCTGTCTAT
CTTTGTTCGCCGGCAGCCGATTATGTTCATGGCGCCATTCTAAACGTCGATGGCGGTTGGCTCGCG
CGCTGA

SEQ ID NO: 64 pSSI-376 P37769-DNA

ATGATTTTAAGTGCATTTTCTCTCGAAGGTAAAGTTGCGGTCGTCACTGGTTGTGATACTG
GACTGGGTCAGGGGATGGCGTTGGGGCTGGCGCAAGCGGGCTGTGACATTGTTGGCATT
AACATCGTTGAACCGACTGAAACCATCGAGCAGGTCACAGCGCTGGGGCGTCGTTTTTTA
AGCCTGACCGCCGATCTGCGAAAGATTGATGGTATTCCAGCACTGCTGGATCGCGCGGTA
GCGGAGTTTGGTCATATTGATATCCTGGTGAATAACGCCGGATTGATTCGCCGCGAAGAT
GCTCTCGAGTTCAGCGAAAAGGACTGGGACGATGTCATGAACCTGAATATCAAGAGCGT
ATTCTTCATGTCTCAGGCAGCGGCGAAACACTTTATCGCGCAAGGCAATGGCGGCAAGA
TTATCAATATCGCGTCAATGCTCTCCTTCCAGGGCGGGATCCGTGTGCCTTCTTATACCGC
ATCAAAAAGCGGCGTGATGGGTGTGACGCGATTGATGGCGAACGAATGGGCTAAACACA
ACATTAATGTTAATGCGATAGCCCCGGGTTACATGGCGACCAACAATACTCAACAACTAC
GGGCAGATGAACAACGTAGCGCGGAAATTCTCGACCGCATTCCAGCTGGTCGTTGGGGA
CTGCCGAGTGACCTGATGGGGCCGATAGTGTTCCTTGCCTCCAGCGCTTCAGATTATGTG
AATGGTTATACCATTGCCGTGGATGGCGGTTGGCTGGCGCGTTAA

SEQ ID NO: 65 pSGI-395 #5112 DNA

ATGCCCGGCATGACCACTCCTTTCGATCTTCATGGCAAGACCGCGATCGTCACCGGCGCCAATACC
GGCATCGGCCAGGCCATTGCCCTGTCGCTCGCGCAGGCCGGCGCGGATATCGCCGCCGTCGGCCG
CACGCCCGCACAGGACACGGTCGATCAGGTCCGCGCGCTCGGCCGCCGGGCGGACATTATCTCGG
CCGACCTTTCGACCATCGAACCGGTCCAGCGCGTCCTCGACGAAACGCTGGAAAAGCTTGGTGCCT
TGGACATACTGGTCAACAATGCCGGCATCATCCGCCGCGCCGACAGCGTCGATTTCACCGAGGAG
GATTGGGACGCGGTGATCGACACCAATCTCAAGACCACCTTCTTCCTCTGTCAGGCCGCCGGTCGC
CACATGCTTGCCCAAGGCGCTGGCAAGATCATCAACATCGCCTCGCTTCTTTCCTTCCAGGGCGGC
ATTCGCGTGCCGAGCTACACCGCGTCCAAAAGCGGCGTCGCGGGCCTGACCAAGCTGCTCGCCAA
CGAATGGGCGGCCAAGGGCGTCAATGTGAACGCCATCGCGCCGGGCTATATCGCCACCAACAACA

CCGCCGCGCTCCAGGCCGACGAAACCCGCAACCGCCAGATCCAGGAGCGCATCCCGGCTGGCCGC
TGGGGCGACCCCGCCGACATTGGCGGCGCGGCCGTGTTCCTGGCGTCCAGCGCCGCCGATTATATC
CATGGCCACACGCTCGCCGTCGACGGCGGCTGGCTCGCGCGCTGA

SEQ ID NO: 66 pSGI-396 #7103-DNA

ATGAACCCCTTCTCGCTTGAGGGCAAGACCGCCCTTGTGACCGGTGCCAATACGGGCATCGGTCAG
GCCATCGCCATGGCGCTTGGCCGCGCCGGGGCGGACGTCATCTGCGCGGGACGCTCGTCCTGTGCG
GAGACCGTTGCCCTCATCGCTGGCAGCAAGGGCAAGGCGCGCGAACTGGTGCTCGACTTCGCCGA
CCCGATGGCCGCCCGTGACGTGTTCGCCGCCGAACCGGTGGACATCCTCGTCAACAACGCGGGCA
TCATCCGGCGCGCCGATGCAGTGGATTTCACCGAGGCCGACTGGGATGAGGTGATGGACGTGAAC
CTGAAGGCCGTGTTCTTCACCTGCCAGGCCTTCGGCAAGGCCGTTCTTGGCCGTGGAGGAAACGGC
AAGATCGTCAACATTGCCTCGCTCCTGTCATTCCAGGGTGGTATCCGGGTGCCGTCCTACACGGCC
TCGAAGCATGGTGTTGCAGGCATCACCAAGCTTCTGGCCAACGAATGGGCGGCGAAGGGCATCAA
TGTGAATGCCATCGCCCCCGGTTACATCGAAACGAACAATACCGAAGCACTGCGGGCGGACCCGG
TGCGCAACAAGGCCATCCTTGAGCGTATCCCTGCCGGCCGCTGGGGCCAGGCCTCGGACATCGGC
GAAGCCGCCGTGTTCCTTGCCTCTCCGGCTGCCAATTACATCCATGGTGCAGTGCTGAATGTTGAC
GGAGGCTGGCTTGCCCGCTGA

SEQ ID NO: 67 pSGI-353 P0AES2

MSSQFTTPVVTEMQVIPVAGHDSMLMNLSGAHAPFFTRNIVIIKDNSGHTGVGEIPGGEKIRK
TLEDAIPLVVGKTLGEYKNVLTLVRNTFADRDAGGRGLQTFDLRTTIHVVTGIEAAMLDLLG
QHLGVNVASLLGDGQQRSEVEMLGYLFFVGNRKATPLPYQSQPDDSCDWYRLRHEEAMTP
DAVVRLAEAAYEKYGFNDFKLKGGVLAGEEEAESIVALAQRFPQARITLDPNGAWSLNEAIK
IGKYLKGSLAYAEDPCGAEQGFSGREVMAEFRRATGLPTATNMIATDWRQMGHTLSLQSVD
IPLADPHFWTMQGSVRVAQMCHEFGLTWGSHSNNHFDISLAMFTHVAAAAPGKITAIDTHW
IWQEGNQRLTKEPFEIKGGLVQVPEKPGLGVEIDMDQVMKAHELYQKHGLGARDDAMGMQ
YLIPGWTFDNKRPCMVR

SEQ ID NO: 68 pSGI-244 #8114

MTTAMSGTPRITELTVVPVAGQDSMLMNLSGAHGPWFTRNILILKDSAGHVGVGEVPGGEAI
RQTLDDARALLVGEPIGQYNALLGKVRRAFADRDAGGRGLQTFDLRIAIHAVTALESALLDL
LGQHLEVPVAALLGEGQQRDEVEMLGYLFFIGDRNRTDLGYRDESNSDDAWFRVRNEEAM
TPERIVRQAEAAYERYGFKDFKLKGGVLRGEEEVEAIRALAQRFPDARVTLDPNGAWSLDEA
SGLCRDLHGVLAYAEDPCGAENGYSGREVMAEFRRATGLPTATNMIATDWRQMSHAVCLH
SVDIPLADPHFWTMAGSVRVAQMCADFGLTWGSHSNNHFDISLAMFTHVAAAAPGRVTAID
THWIWQDGQHLTREPLKIVSGKVAVPQKPGLGVELDWDALEQAHAHYQEKGLGARDDAIA
MQYLIPNWTFNNKKPCMVR

SEQ ID NO: 69 pSGI-353 P0AES2-DNA

ATGAGTTCTCAATTTACGACGCCTGTTGTTACTGAAATGCAGGTTATCCCGGTGGCGGGTCATGAC
AGTATGCTGATGAATCTGAGTGGTGCACACGCACCGTTCTTTACGCGTAATATTGTGATTATCAAA
GATAATTCTGGTCACACTGGCGTAGGGGAAATTCCCGGCGGCGAGAAAATCCGTAAAACGCTGGA

AGATGCGATTCCGCTGGTGGTAGGTAAAACGCTGGGTGAATACAAAAACGTTCTGACGCTGGTGC
GTAATACTTTTGCCGATCGTGATGCTGGTGGGCGCGGTTTGCAGACATTTGACCTACGTACCACTA
TTCATGTAGTTACCGGGATAGAAGCGGCAATGCTGGATCTGCTGGGGCAGCATCTGGGGGTAAAC
GTGGCATCGCTGCTGGGCGATGGTCAACAGCGTAGCGAAGTCGAAATGCTCGGTTATCTGTTCTTC
GTCGGTAATCGCAAAGCCACGCCGCTGCCGTATCAAAGCCAGCCGGATGACTCATGCGACTGGTA
TCGCCTGCGTCATGAAGAAGCGATGACGCCGGATGCGGTGGTGCGCCTGGCGGAAGCGGCATATG
AAAAATATGGCTTCAACGATTTCAAACTGAAGGGCGGTGTACTGGCCGGGGAAGAAGAGGCCGAG
TCTATTGTGGCACTGGCGCAACGCTTCCCGCAGGCGCGTATTACGCTCGATCCTAACGGTGCCTGG
TCGCTGAACGAAGCGATTAAAATCGGTAAATACCTGAAAGGTTCGCTGGCTTATGCAGAAGATCC
GTGTGGTGCGGAGCAAGGTTTCTCCGGGCGTGAAGTGATGGCAGAGTTCCGTCGCGCGACAGGTC
TACCGACTGCAACCAATATGATCGCCACCGACTGGCGGCAAATGGGCCATACGCTCTCCCTGCAAT
CCGTTGATATCCCGCTGGCGGATCCGCATTTCTGGACAATGCAAGGTTCGGTACGTGTGGCGCAAA
TGTGCCATGAATTTGGCCTGACCTGGGGTTCACACTCTAACAACCACTTCGATATTTCCCTGGCGAT
GTTTACCCATGTTGCCGCCGCTGCACCGGGTAAAATTACTGCTATTGATACGCACTGGATTTGGCA
GGAAGGCAATCAGCGCCTGACCAAAGAACCGTTTGAGATCAAAGGCGGGCTGGTACAGGTGCCAG
AAAAACCGGGGCTGGGTGTAGAAATCGATATGGATCAAGTGATGAAAGCCCATGAGCTGTATCAG
AAACACGGGCTTGGCGCGCGTGACGATGCGATGGGAATGCAGTATCTGATTCCTGGCTGGACGTT
CGATAACAAGCGCCCGTGCATGGTGCGTTAA

SEQ ID NO: 70 pSGI-244 #8114

ATGACCACCGCCATGTCGGGCACGCCCCGCATCACCGAACTCACCGTCGTGCCCGTCGCCGGGCA
GGACAGCATGCTGATGAACCTCAGCGGCGCCCATGGGCCCTGGTTCACCCGCAACATCCTCATCCT
CAAGGACAGCGCCGGCCACGTCGGCGTCGGCGAAGTGCCGGGCGGCGAAGCCATCCGCCAGACCC
TCGACGATGCCCGTGCCCTGCTGGTCGGCGAACCGATCGGCCAGTACAACGCGCTGCTCGGCAAG
GTGCGCCGCGCCTTCGCCGACCGTGACGCCGGCGGCCGCGGCCTGCAGACCTTCGACCTGCGCATC
GCCATTCACGCCGTCACCGCGCTGGAGTCGGCGCTGCTCGACCTGCTCGGCCAGCACCTCGAGGTG
CCGGTCGCCGCCTTGCTCGGCGAAGGCCAGCAGCGTGACGAAGTGGAAATGCTCGGCTACCTGTT
CTTCATCGGCGATCGCAACAGGACCGACCTCGGCTACCGCGACGAATCCAACTCCGACGACGCCT
GGTTTCGCGTGCGCAACGAGGAGGCCATGACGCCGGAGCGCATCGTCCGCCAGGCCGAGGCGGCC
TACGAGCGCTACGGCTTCAAGGACTTCAAGCTCAAGGGCGGCGTACTGCGCGGCGAAGAGGAAGT
CGAGGCGATCCGCGCCCTGGCCCAGCGCTTCCCCGACGCCCGCGTGACTCTGGACCCCAACGGCG
CCTGGTCGCTGGACGAAGCCAGCGGCCTGTGTCGCGACCTGCACGGCGTGCTGGCCTATGCCGAA
GACCCCTGCGGTGCCGAGAACGGCTATTCCGGCCGCGAGGTGATGGCCGAGTTCCGCCGCGCCAC
CGGTCTGCCCACCGCGACCAACATGATCGCCACCGACTGGCGACAGATGAGTCACGCGGTGTGCC
TGCACTCGGTGGACATCCCGCTGGCCGACCCGCACTTCTGGACCATGGCCGGCTCTGTGCGCGTGG
CGCAGATGTGCGCCGACTTCGGCCTGACCTGGGGTTCGCACTCGAACAACCACTTCGACATCTCCC
TGGCGATGTTCACCCACGTGGCGGCCGCCGCGCCGGGTCGCGTCACCGCCATCGACACCCACTGG
ATCTGGCAGGACGGCCAGCACCTGACCCGCGAGCCGCTGAAGATCGTCAGCGGCAAGGTTGCGGT
GCCGCAGAAGCCGGGGCTGGGCGTCGAGCTGGACTGGGATGCCCTGGAGCAGGCGCATGCCCACT
ACCAAGAGAAAGGCCTGGGTGCCCGCGATGACGCCATCGCCATGCAGTACCTGATCCCCAACTGG
ACCTTCAACAACAAGAAGCCGTGCATGGTGCGCTGA

SEQ ID NO: 71 pSGI-383 P50199

MSHPDLFSLSGARALVTGASRGIGLTLAKGLARYGAEVVLNGRNAESLDSAQSGFEAEGLKA
STAVFDVTDQDAVIDGVAAIERDMGPIDILINNAGIQRRAPLEEFSRKDWDDLMSTNVNAVFF
VGQAVARHMIPRGRGKIVNICSVQSELARPGIAPYTATKGAVKNLTKGMATDWGRHGLQIN
GLAPGYFATEMTERLVADEEFTDWLCKRTPAGRWGQVEELVGAAVFLSSRASSFVNGQVL
MVDGGITVSL

SEQ ID NO: 72 pSGI-383 P50199-DNA

ATGTCTCACCCGGATCTGTTTAGCTTAAGTGGCGCACGCGCATTAGTTACTGGTGCCTCTCGTGGTA
TTGGTTTAACCCTGGCCAAAGGTTTAGCCCGTTATGGTGCCGAAGTGGTTTTAAATGGCCGTAATG
CCGAAAGCCTGGATTCTGCCCAAAGTGGCTTTGAAGCCGAAGGCTTAAAAGCATCTACCGCTGTGT
TTGACGTGACCGATCAAGATGCAGTCATTGACGGCGTGGCAGCAATTGAACGCGATATGGGTCCG
ATTGATATCCTGATCAACAATGCGGGCATTCAACGCAGAGCCCCGTTAGAAGAATTTTCTCGCAAA
GACTGGGACGATCTGATGAGCACCAACGTTAACGCCGTGTTCTTTGTGGGACAAGCCGTTGCCAGA
CACATGATTCCTAGAGGTCGCGGTAAAATCGTCAACATCTGTTCAGTGCAGAGCGAACTGGCAAG
ACCGGGTATTGCACCTTATACCGCCACAAAAGGAGCCGTCAAAAATCTGACCAAAGGTATGGCCA
CCGATTGGGGTCGTCATGGTTTACAGATTAATGGCTTAGCACCGGGCTATTTTGCCACCGAGATGA
CCGAACGCTTAGTTGCCGACGAAGAATTTACCGACTGGTTATGCAAACGCACCCCTGCAGGCAGA
TGGGGCCAAGTTGAAGAATTAGTAGGCGCAGCCGTGTTTTTAAGTAGTAGAGCCTCAAGCTTCGTG
AATGGCCAAGTCCTGATGGTTGATGGTGGAATTACTGTGAGCCTGTAA

2013317782 24 Aug 2016

CLAIMS

What is claimed is:

1. A method for synthesizing a derivative of FDCA comprising:
contacting DDG with an alcohol and an inorganic acid at a temperature in excess of 60°C to form a derivative of FDCA.

2. The method of claim 1 wherein the alcohol is butanol or ethanol and the derivative of FDCA is a butyl or ethyl derivative of FDCA, respectively.

3. The method of claim 1 having a yield of at least 25% molar.

4. A method of synthesizing a derivative of DDG comprising:
contacting DDG with an alcohol, an inorganic acid, and optionally a co-solvent to produce a derivative of DDG.

5. The method of claim 4 wherein:
a) the alcohol is ethanol or butanol;
b) the inorganic acid is sulfuric acid; and
c) the co-solvent is selected from the group consisting of: THF, acetone, acetonitrile, an ether, ethyl acetate, butyl acetate, an dioxane, chloroform, methylene chloride, 1,2-dichloroethane, a hexane, a heptane, toluene, carbon tetrachloride, petroleum ether, and a xylene.

6. A method for synthesizing a derivative of FDCA comprising:
contacting a derivative of DDG with an inorganic acid to produce a derivative of FDCA.

7. The method of claim 6 having a yield of greater than 25% molar.

8. The method of claim 6 wherein the derivative of DDG is selected from the group consisting: methyl-DDG, ethyl-DDG, butyl-DDG, di-methyl DDG, diethyl-DDG, and di-butyl DDG; and
the derivative of FDCA is a methyl, ethyl, butyl, dimethyl, diethyl, or dibutyl derivative of FDCA, respectively.

9. The method of claim 8 further comprising that the derivative of FDCA is de-esterified to yield FDCA.

2013317782 24 Aug 2016

10. The method of claim 6 further comprising a step of polymerizing the derivative of FDCA.

11. A method for synthesizing FDCA comprising:
contacting DDG with an inorganic acid at a temperature greater than 70 °C to synthesize FDCA.

12. A method for synthesizing FDCA comprising:
contacting DDG with an inorganic acid in a gas phase at a temperature in excess of 120 °C to synthesize FDCA.

13. A method for synthesizing FDCA comprising:
contacting DDG with an inorganic acid under anhydrous reaction conditions to synthesize FDCA.

14. The method of claim 1 wherein:
the alcohol is selected from: butanol, ethanol, methanol, and propanol;
the acid is sulfuric acid;
the contacting occurs at a temperature of greater than 70 °C; and
thereby synthesizing a butyl, ethyl, methyl, or propyl derivative of FDCA, respectively.

15. The method of claim 14 wherein the contacting occurs in a gas phase at a temperature of greater than 150 °C.

16. The method of claim 5 further comprising a step of removing water from a solvent comprising the DDG prior to performing the method.

17. The method of claim 16 wherein greater than 90% of the water is removed from the solvent comprising the DDG prior to performing the method.

18. The method of claim 8 wherein the contacting occurs in the gas phase at a temperature of at least 90°C.

19. The method of claim 11 wherein the inorganic acid is sulfuric acid.

20. The method of claim 12 wherein the inorganic acid is sulfuric acid.

21. The method of claim 13 wherein the contacting occurs at a temperature of greater than 80°C.

22. The method of claim 13 wherein the DDG is comprised in a solvent that contains less than 10% water (w/w).

23. The method of claim 22 wherein the DDG is comprised in a solvent that contains less than 5% water (w/w).

24. The method of claim 8 further comprising a step of polymerizing the derivative of FDCA.

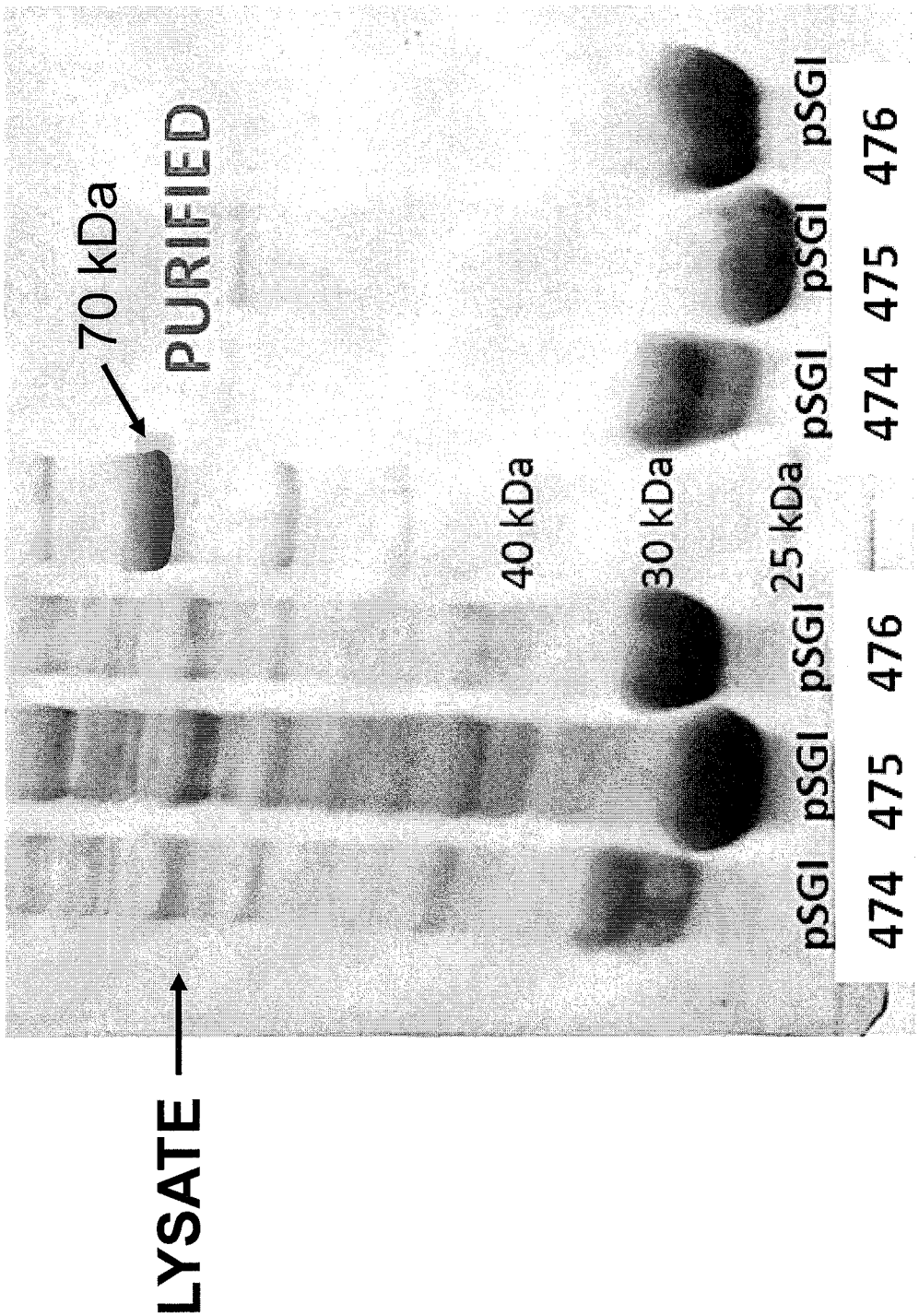

LYSATE
70 kDa
PURIFIED
40 kDa
30 kDa
25 kDa
pSGI 474
pSGI 475
pSGI 476
pSGI 474
pSGI 475
pSGI 476

FIG. 1

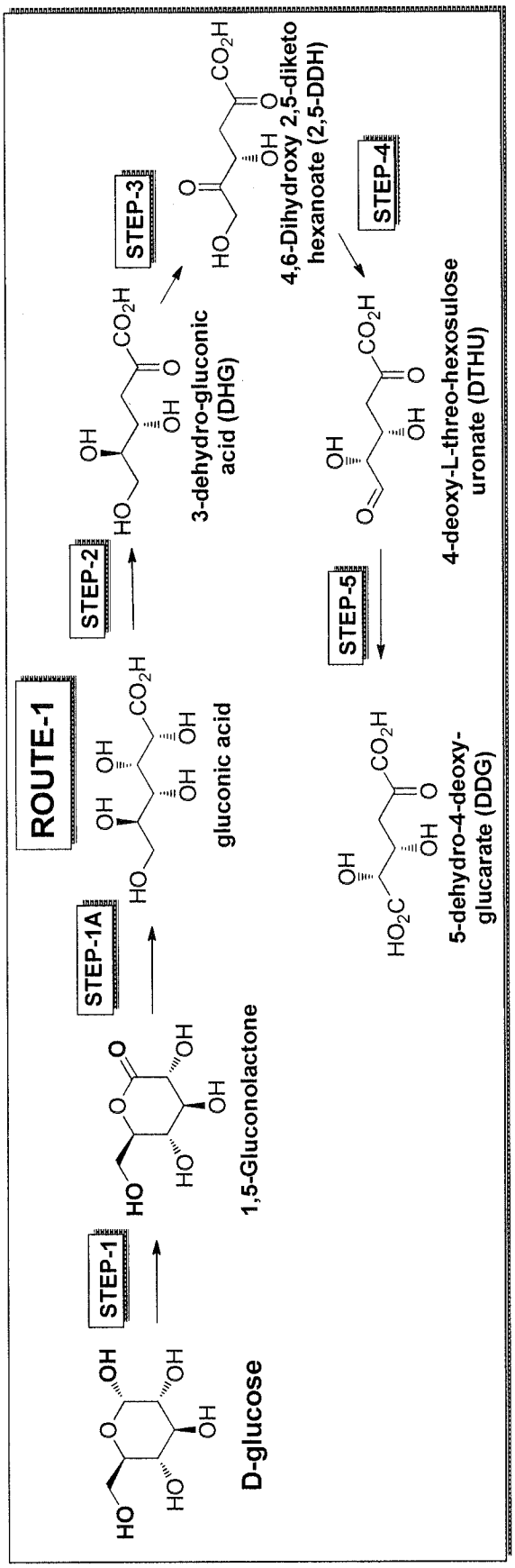

ROUTE-1
D-glucose
STEP-1
1,5-Gluconolactone
STEP-1A
gluconic acid
STEP-2
3-dehydro-gluconic acid (DHG)
STEP-3
4,6-Dihydroxy 2,5-diketo hexanoate (2,5-DDH)
STEP-4
4-deoxy-L-threo-hexosulose uronate (DTHU)
STEP-5
5-dehydro-4-deoxy-glucarate (DDG)

Fig. 2A

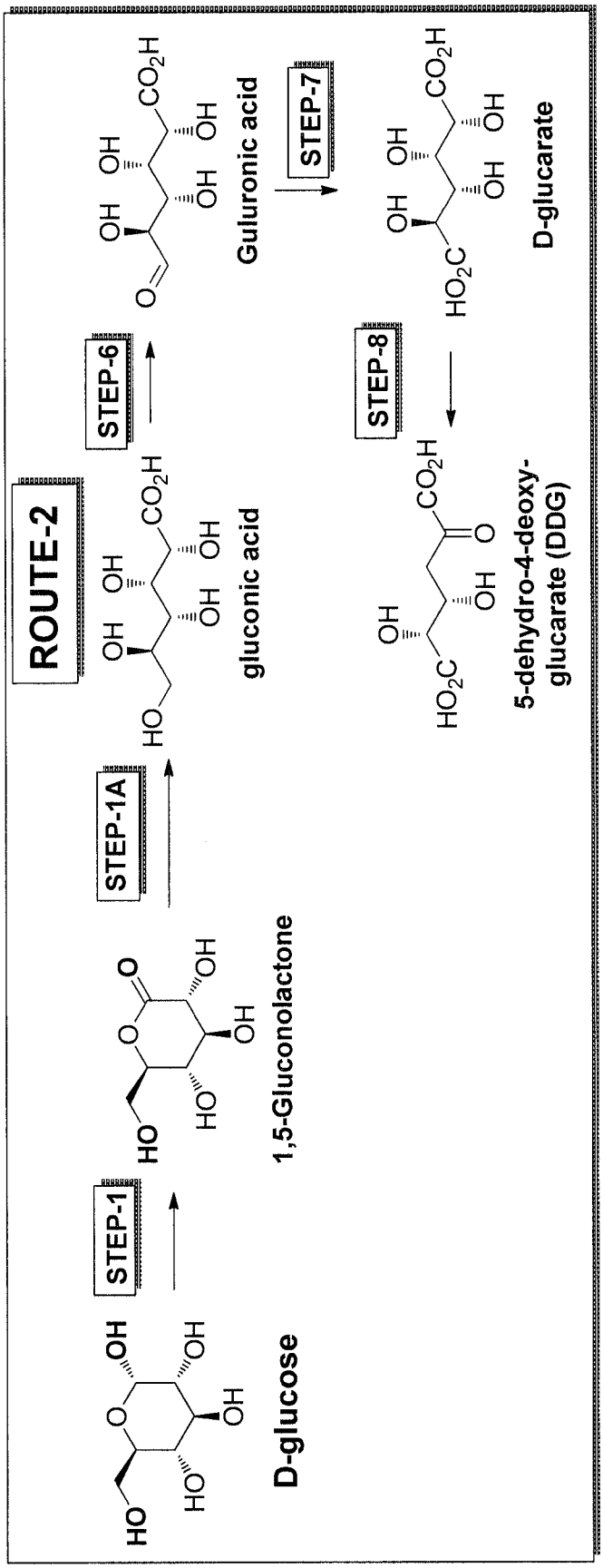

ROUTE-2
STEP-1
STEP-1A
STEP-6
STEP-7
STEP-8
D-glucose
1,5-Gluconolactone
gluconic acid
Guluronic acid
5-dehydro-4-deoxy-glucarate (DDG)
D-glucarate

Fig. 2B

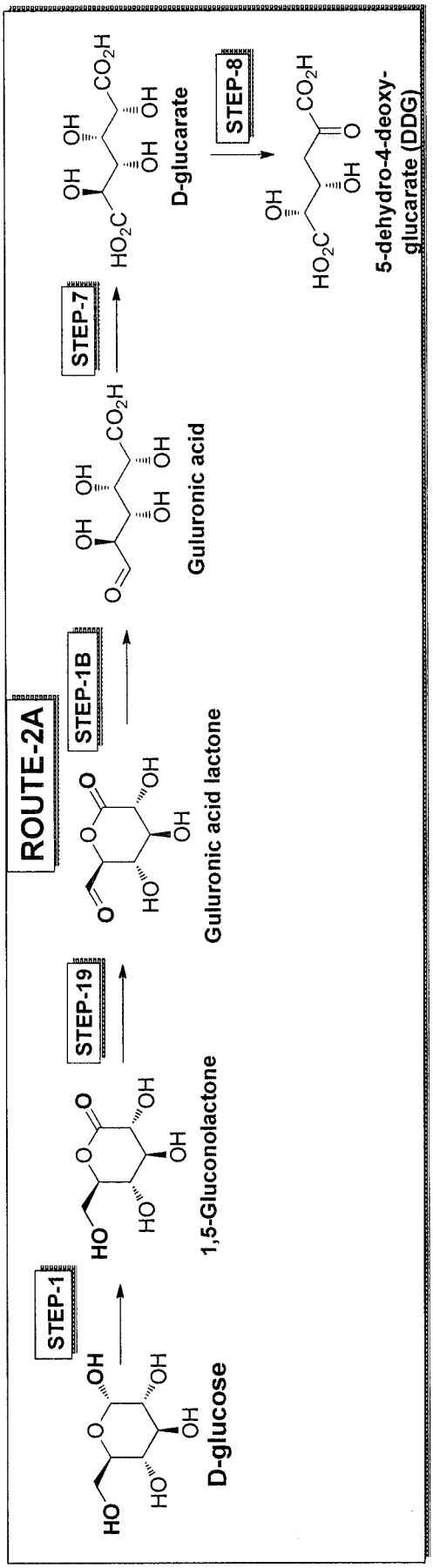

ROUTE-2A
STEP-1
STEP-19
STEP-1B
STEP-7
STEP-8
D-glucose
1,5-Gluconolactone
Guluronic acid lactone
Guluronic acid
D-glucarate
5-dehydro-4-deoxy-glucarate (DDG)

Fig. 2C

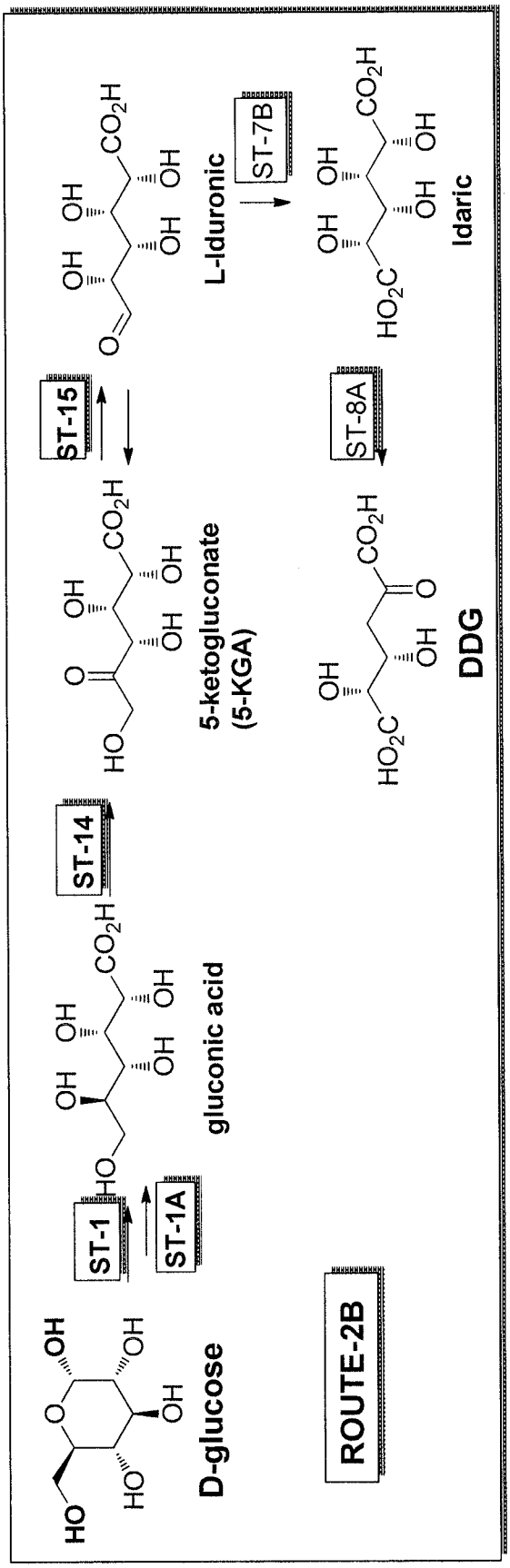

D-glucose
ST-1
ST-1A
gluconic acid
ST-14
5-ketogluconate
(5-KGA)
ST-15
L-Iduronic
ST-7B
Idaric
ST-8A
DDG
ROUTE-2B

Fig. 2D

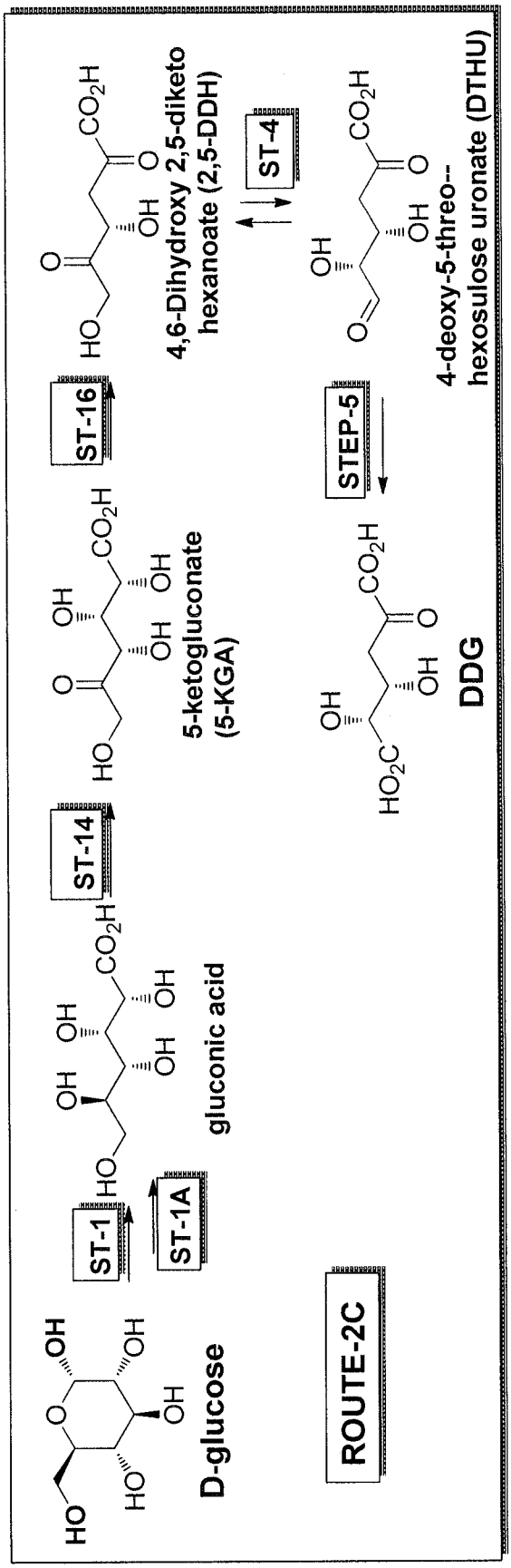

D-glucose
ST-1
ST-1A
gluconic acid
ST-14
5-ketogluconate
(5-KGA)
ST-16
4,6-Dihydroxy 2,5-diketo
hexanoate (2,5-DDH)
ST-4
4-deoxy-5-threo--
hexosulose uronate (DTHU)
STEP-5
DDG
ROUTE-2C

Fig. 2E

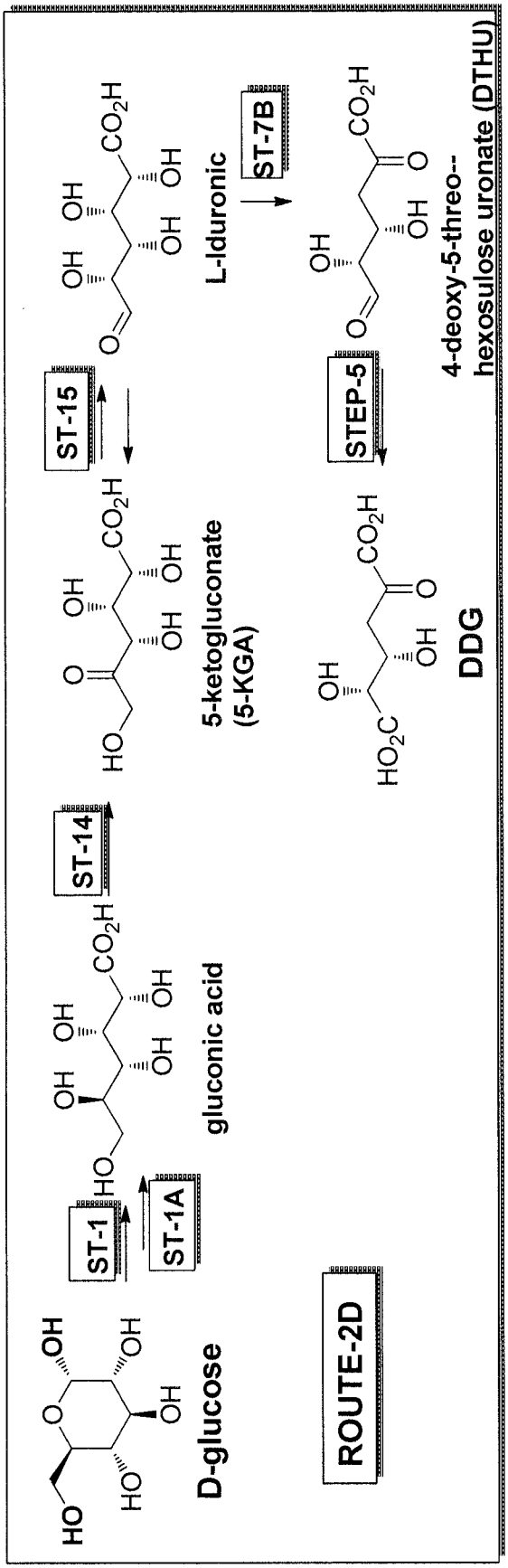

D-glucose
ST-1
ST-1A
gluconic acid
ST-14
5-ketogluconate
(5-KGA)
ST-15
L-Iduronic
ST-7B
4-deoxy-5-threo--
hexosulose uronate (DTHU)
STEP-5
DDG
ROUTE-2D

Fig. 2F

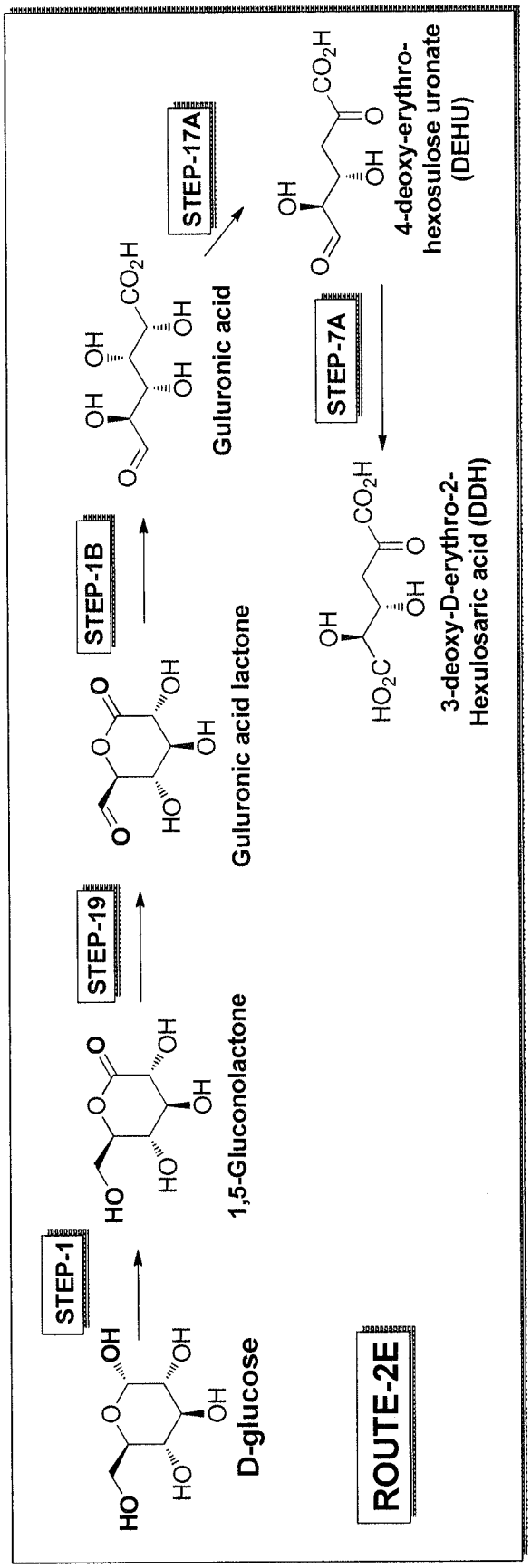

ROUTE-2E
D-glucose
STEP-1
1,5-Gluconolactone
STEP-19
Guluronic acid lactone
STEP-1B
Guluronic acid
STEP-17A
3-deoxy-D-erythro-2-Hexulosaric acid (DDH)
STEP-7A
4-deoxy-erythro-hexosulose uronate (DEHU)

Fig. 2G

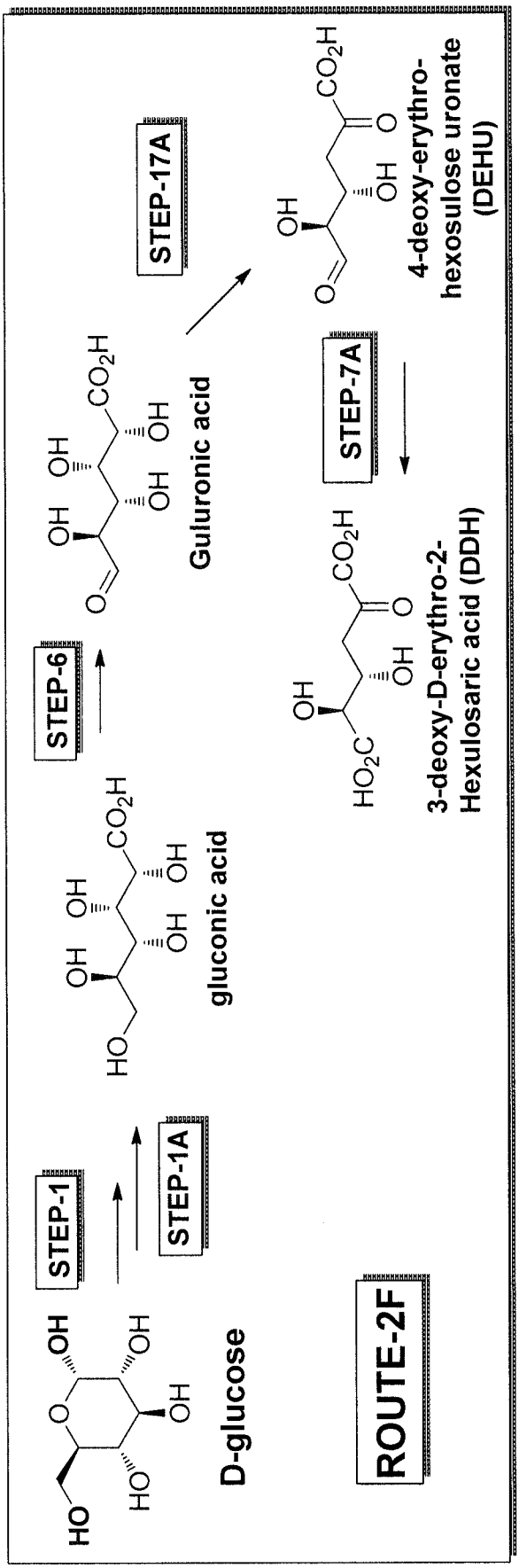

ROUTE-2F
D-glucose
STEP-1
STEP-1A
gluconic acid
STEP-6
Guluronic acid
STEP-17A
3-deoxy-D-erythro-2-
Hexulosaric acid (DDH)
STEP-7A
4-deoxy-erythro-
hexosulose uronate
(DEHU)

Fig. 2H

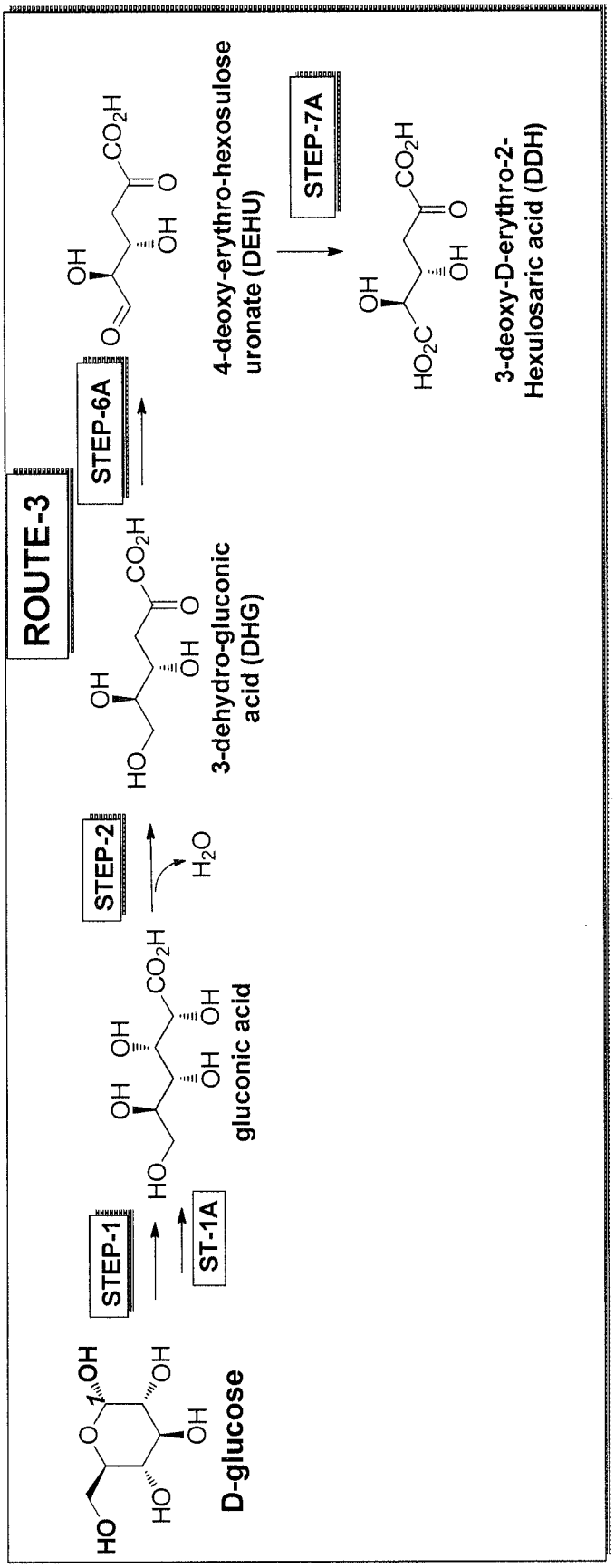

ROUTE-3
D-glucose
STEP-1
ST-1A
gluconic acid
STEP-2
H_2O
3-dehydro-gluconic acid (DHG)
STEP-6A
4-deoxy-erythro-hexosulose uronate (DEHU)
STEP-7A
3-deoxy-D-erythro-2-Hexulosaric acid (DDH)

Fig. 3A

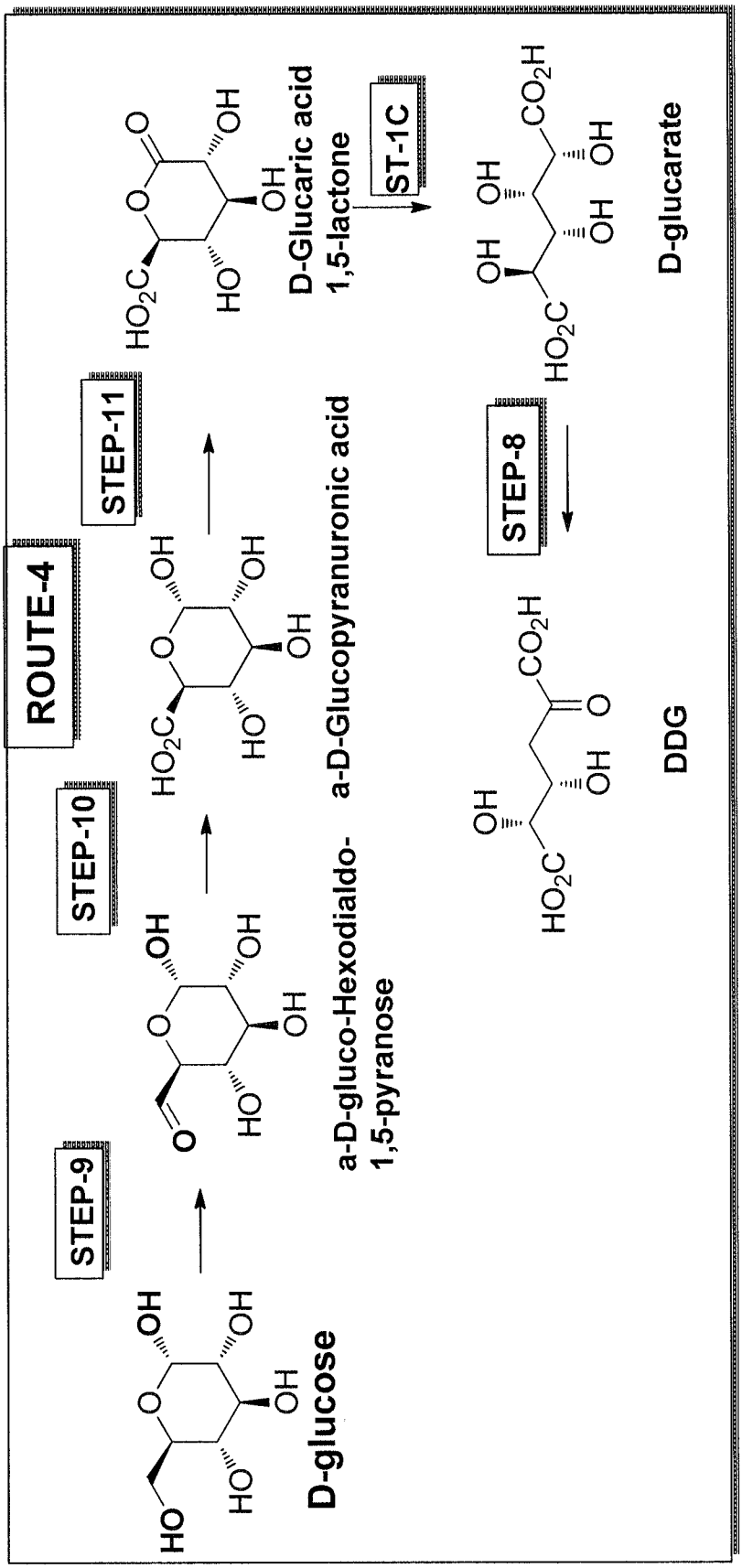

ROUTE-4
STEP-9
STEP-10
STEP-11
ST-1C
STEP-8
D-glucose
a-D-gluco-Hexodialdo-1,5-pyranose
a-D-Glucopyranuronic acid
D-Glucaric acid 1,5-lactone
DDG
D-glucarate

Fig. 3B

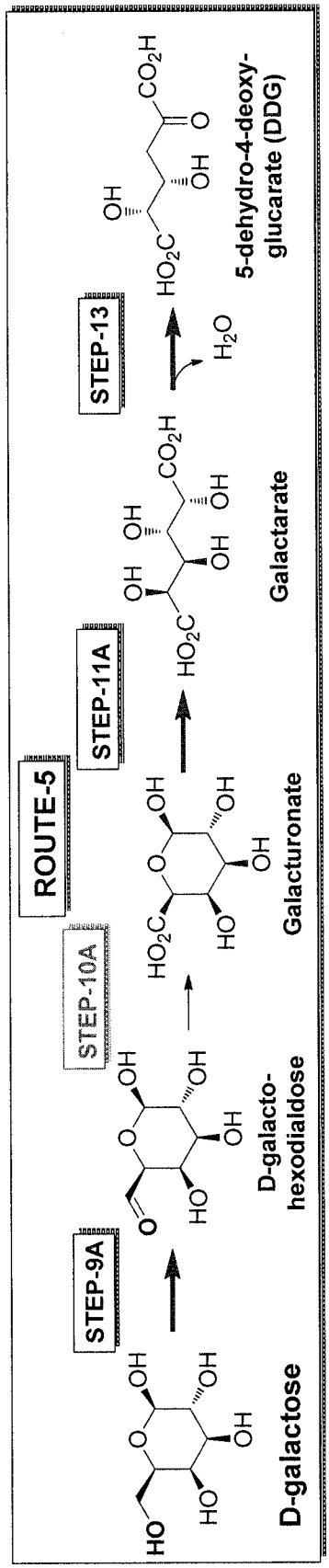

ROUTE-5
STEP-9A
STEP-10A
STEP-11A
STEP-13
D-galactose
D-galacto-hexodialdose
Galacturonate
Galactarate
5-dehydro-4-deoxy-glucarate (DDG)
H_2O

Fig. 3C

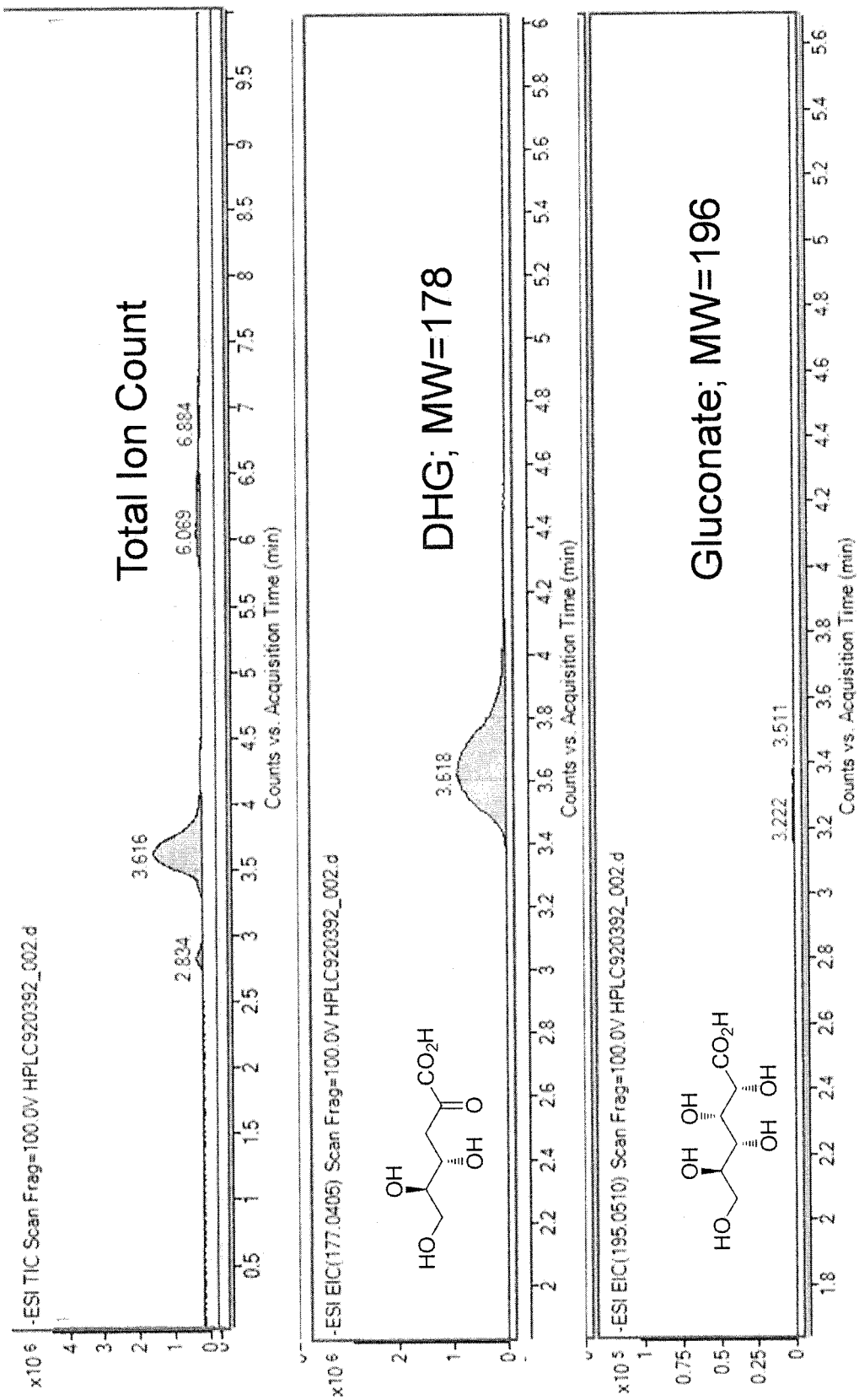

Total Ion Count
-ESI TIC Scan Frag=100.0V HPLC920392_002.d
2.834
3.616
6.069
6.884
Counts vs. Acquisition Time (min)
DHG; MW=178
-ESI EIC(177.0405) Scan Frag=100.0V HPLC920392_002.d
3.618
Counts vs. Acquisition Time (min)
Gluconate; MW=196
-ESI EIC(195.0510) Scan Frag=100.0V HPLC920392_002.d
3.222
3.511
Counts vs. Acquisition Time (min)

Fig. 4

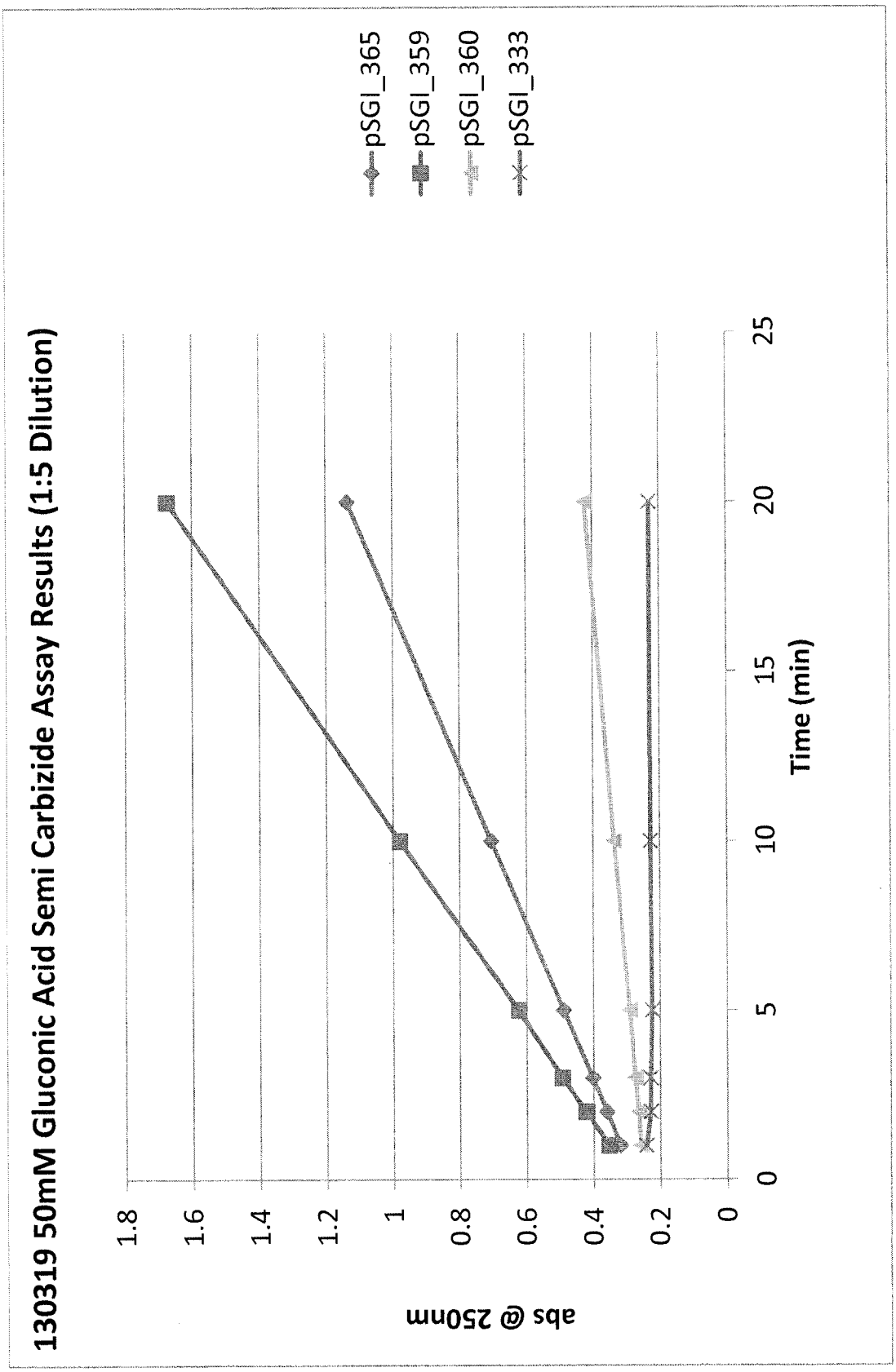

130319 50mM Gluconic Acid Semi Carbizide Assay Results (1:5 Dilution)
pSGI_365
pSGI_359
pSGI_360
pSGI_333
abs @ 250nm
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
5
10
15
20
25
Time (min)

Fig. 5

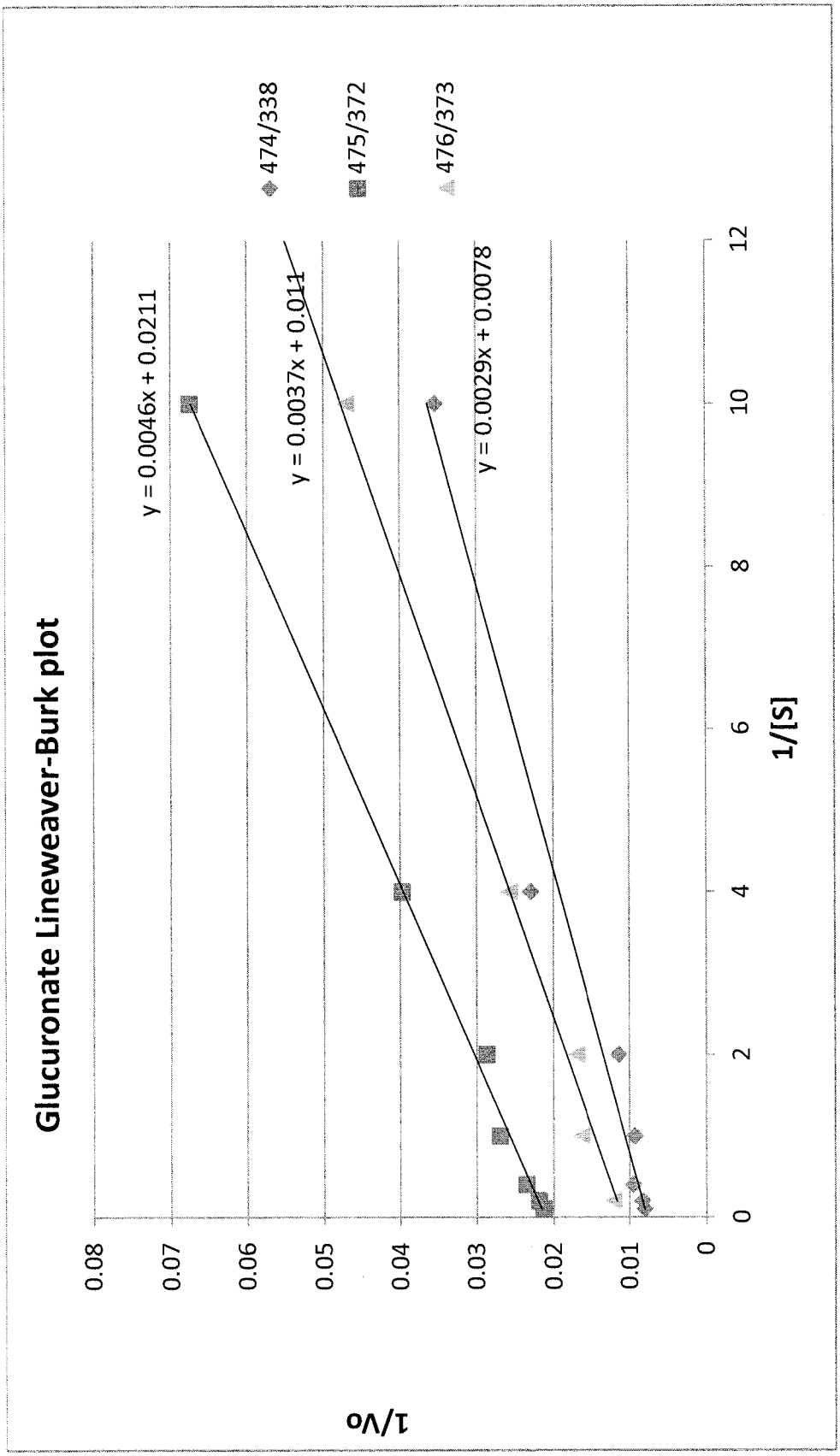

Glucuronate Lineweaver-Burk plot
y = 0.0046x + 0.0211
y = 0.0037x + 0.011
y = 0.0029x + 0.0078
474/338
475/372
476/373
1/Vo
1/[S]
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
2
4
6
8
10
12

Fig. 6A

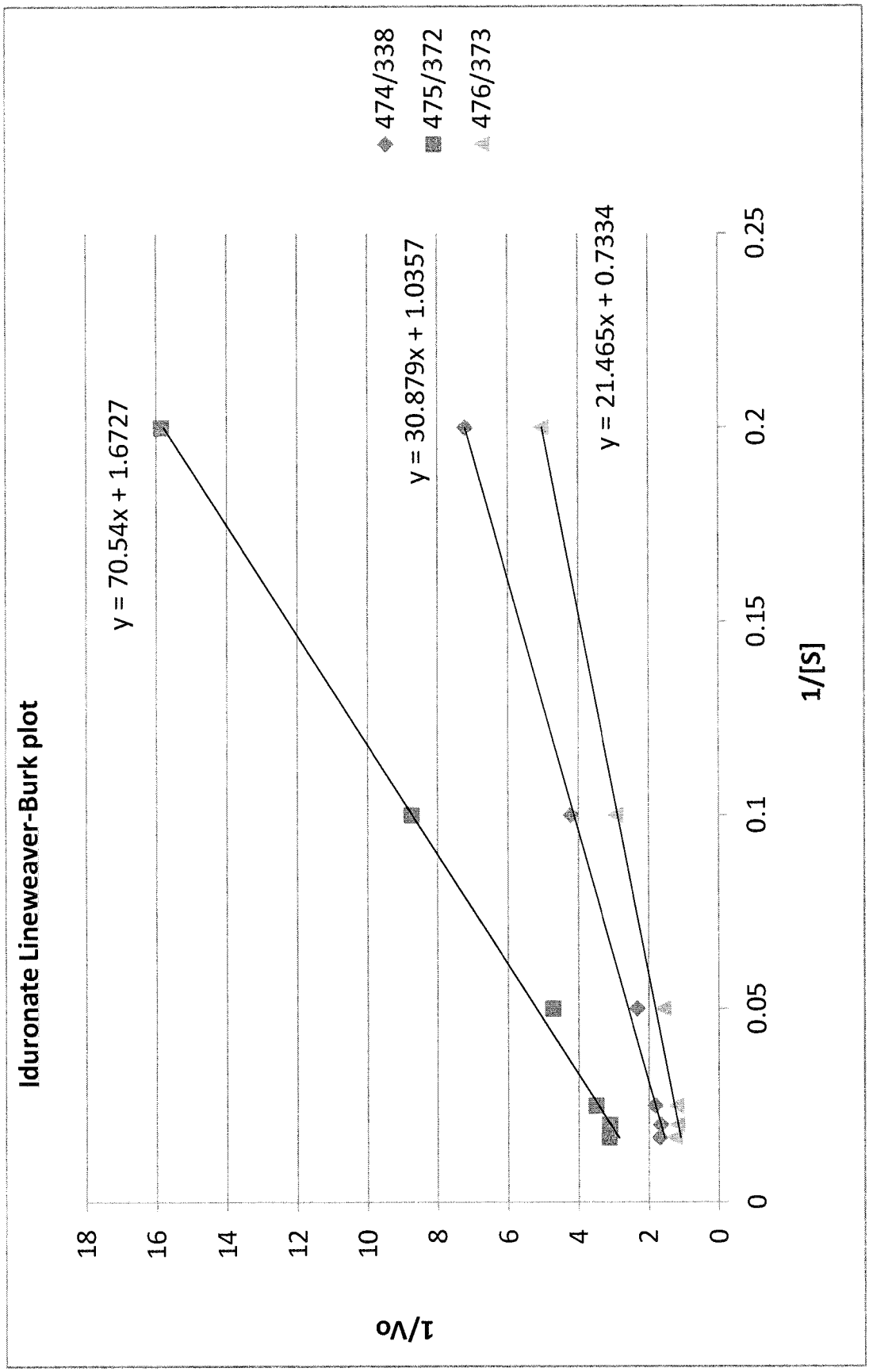

Iduronate Lineweaver-Burk plot
y = 70.54x + 1.6727
y = 30.879x + 1.0357
y = 21.465x + 0.7334
474/338
475/372
476/373
1/Vo
1/[S]
0
2
4
6
8
10
12
14
16
18
0
0.05
0.1
0.15
0.2
0.25

Fig. 6B

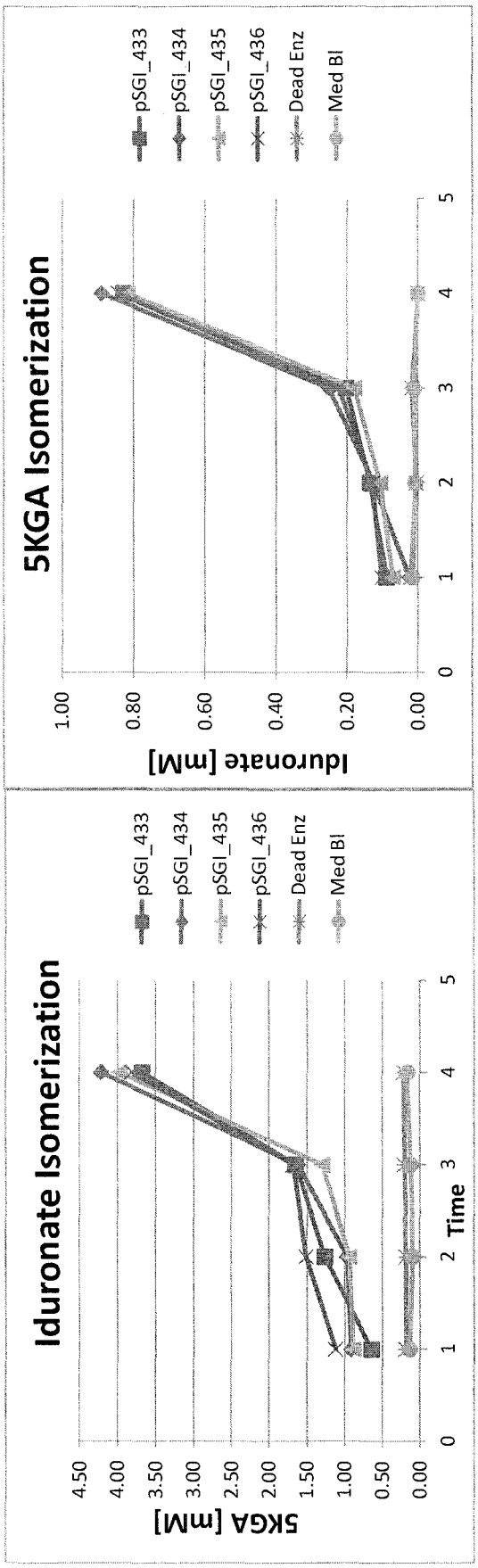

Iduronate Isomerization
5KGA [mM]
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
0
1
2
3
4
5
Time
pSGI_433
pSGI_434
pSGI_435
pSGI_436
Dead Enz
Med Bl
5KGA Isomerization
Iduronate [mM]
1.00
0.80
0.60
0.40
0.20
0.00
0
1
2
3
4
5
pSGI_433
pSGI_434
pSGI_435
pSGI_436
Dead Enz
Med Bl

Fig. 7A

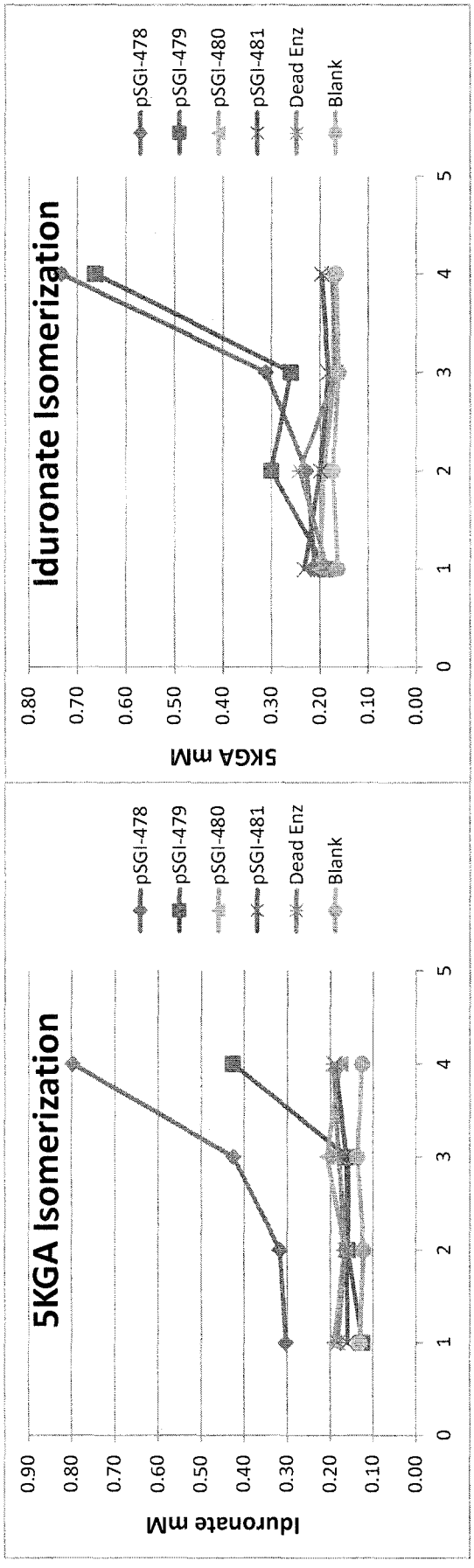

5KGA Isomerization
Iduronate mM
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
1
2
3
4
5
pSGI-478
pSGI-479
pSGI-480
pSGI-481
Dead Enz
Blank
Iduronate Isomerization
5KGA mM
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0
1
2
3
4
5
pSGI-478
pSGI-479
pSGI-480
pSGI-481
Dead Enz
Blank

Fig. 7B

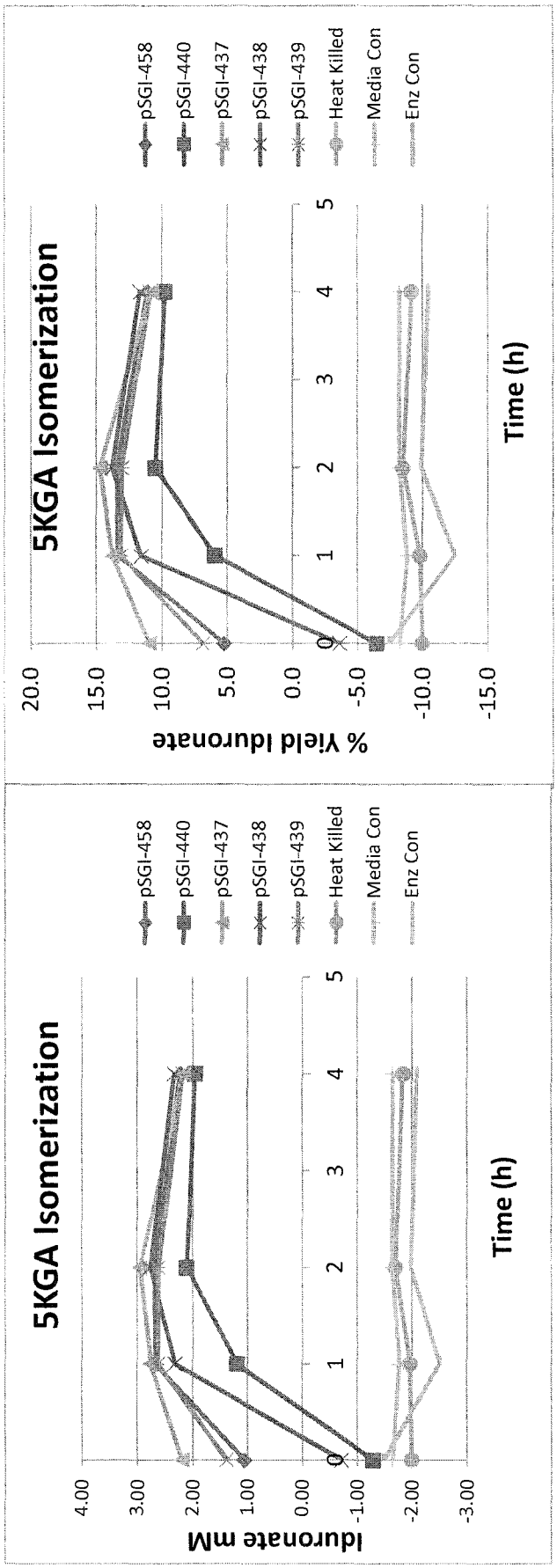
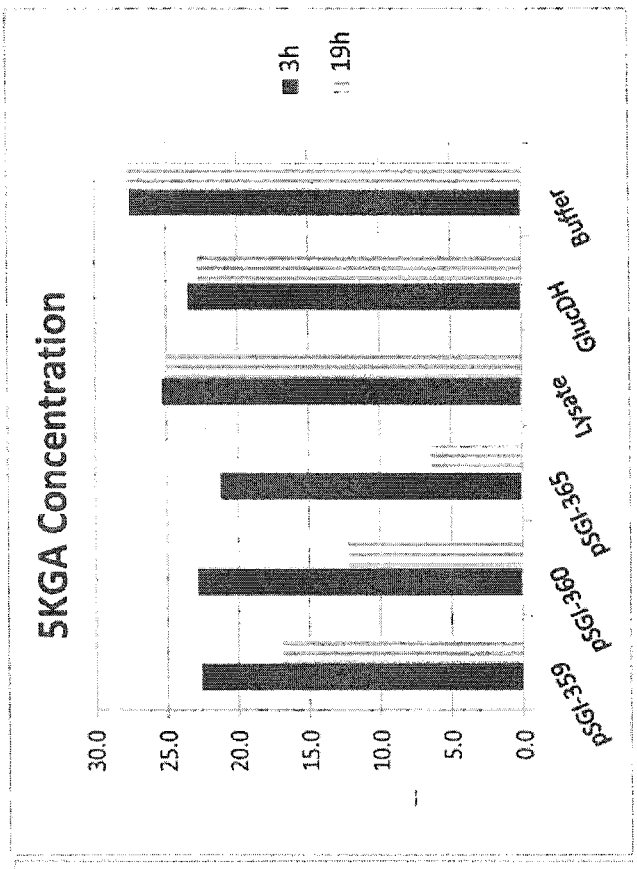

5KGA Isomerization
Iduronate mM
4.00
3.00
2.00
1.00
0.00
-1.00
-2.00
-3.00
0
1
2
3
4
5
Time (h)
pSGI-458
pSGI-440
pSGI-437
pSGI-438
pSGI-439
Heat Killed
Media Con
Enz Con
5KGA Isomerization
% Yield Iduronate
20.0
15.0
10.0
5.0
0.0
-5.0
-10.0
-15.0
0
1
2
3
4
5
Time (h)
pSGI-458
pSGI-440
pSGI-437
pSGI-438
pSGI-439
Heat Killed
Media Con
Enz Con

Fig. 8


5KGA Concentration
3h
19h
30.0
25.0
20.0
15.0
10.0
5.0
0.0
pSGI-359
pSGI-360
pSGI-365
Lysate
GlucDH
Buffer

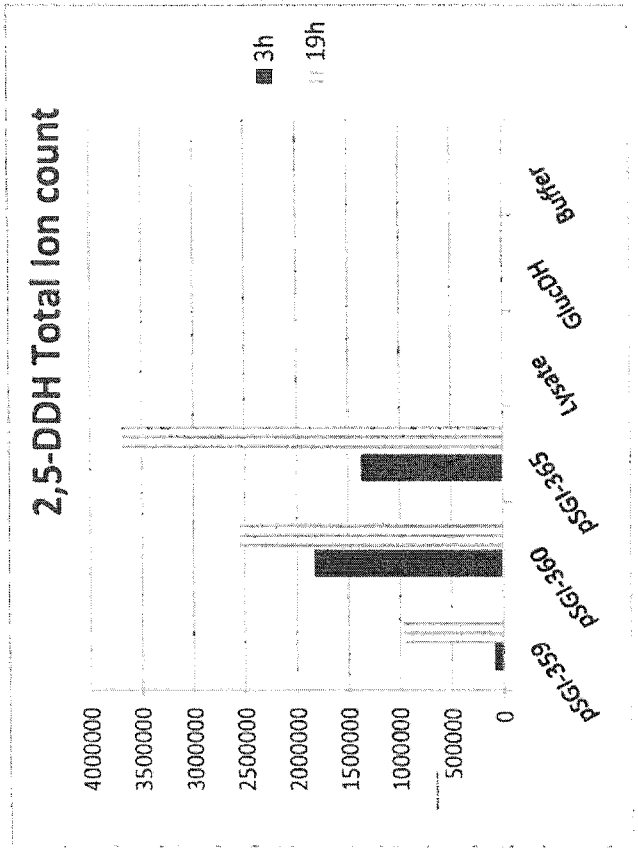

2,5-DDH Total Ion count
3h
19h
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
pSGI-359
pSGI-360
pSGI-365
Lysate
GlucDH
Buffer

Fig. 9

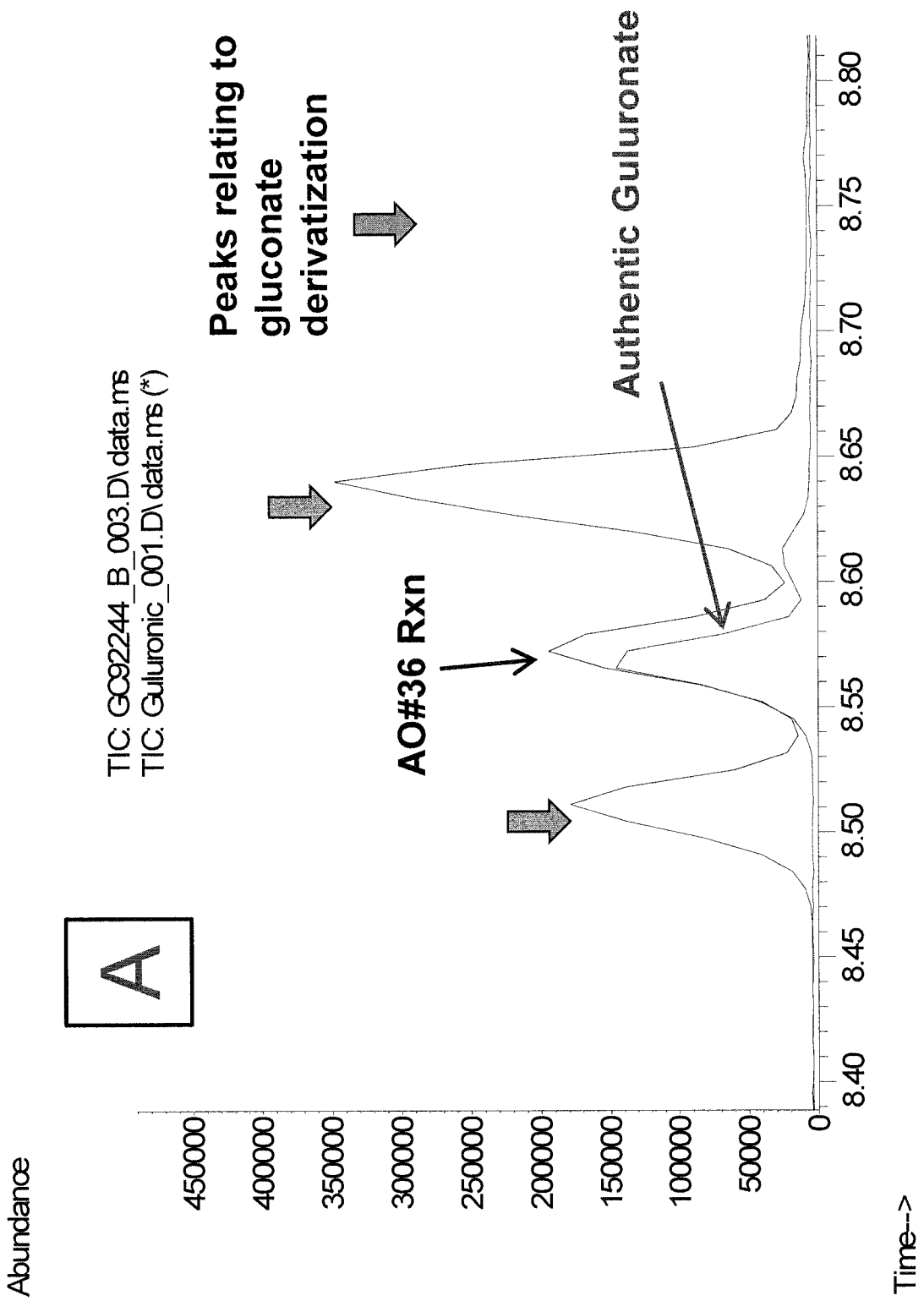
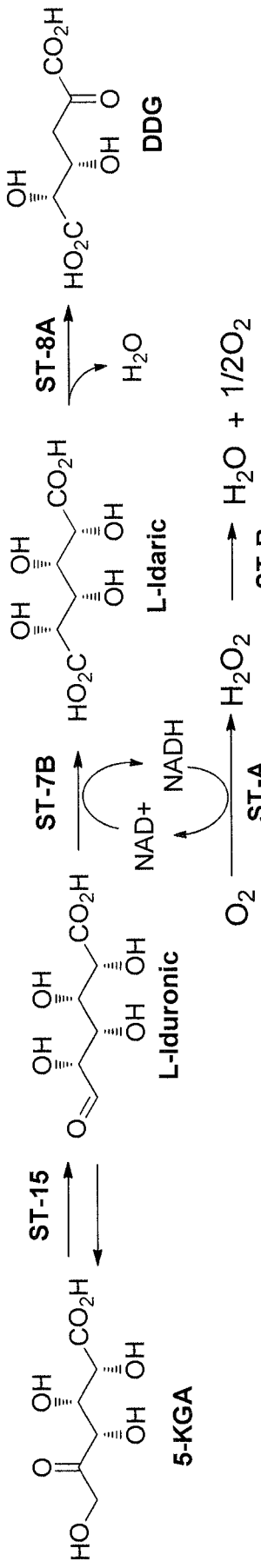

A
TIC: GC92244_B_003.D\data.ms
TIC: Guluronic_001.D\data.ms (*)
Peaks relating to gluconate derivatization
AO#36 Rxn
Authentic Guluronate
Abundance
Time-->
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
8.40
8.45
8.50
8.55
8.60
8.65
8.70
8.75
8.80

Fig. 10



Scheme 6

Fig. 11

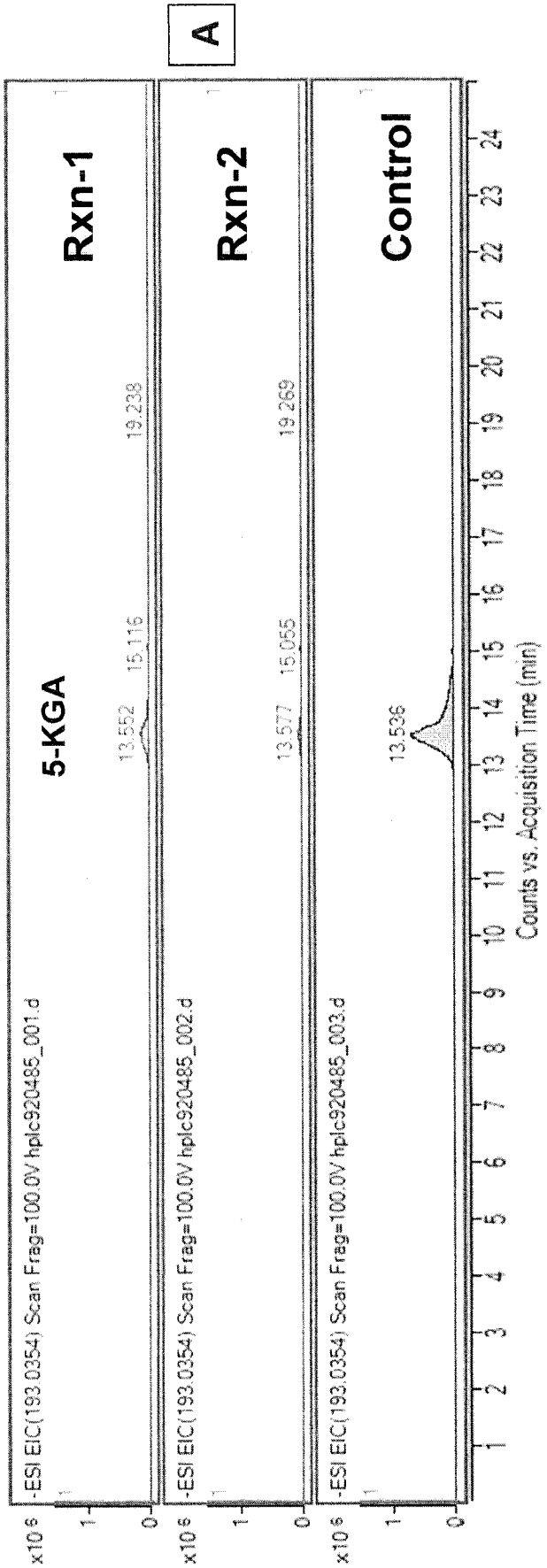

A
5-KGA
Rxn-1
Rxn-2
Control
-ESI EIC(193.0354) Scan Frag=100.0V hplc920485_001.d
-ESI EIC(193.0354) Scan Frag=100.0V hplc920485_002.d
-ESI EIC(193.0354) Scan Frag=100.0V hplc920485_003.d
x10 6
13.552
15.116
19.238
13.577
15.055
19.269
13.536
Counts vs. Acquisition Time (min)

Fig. 12A

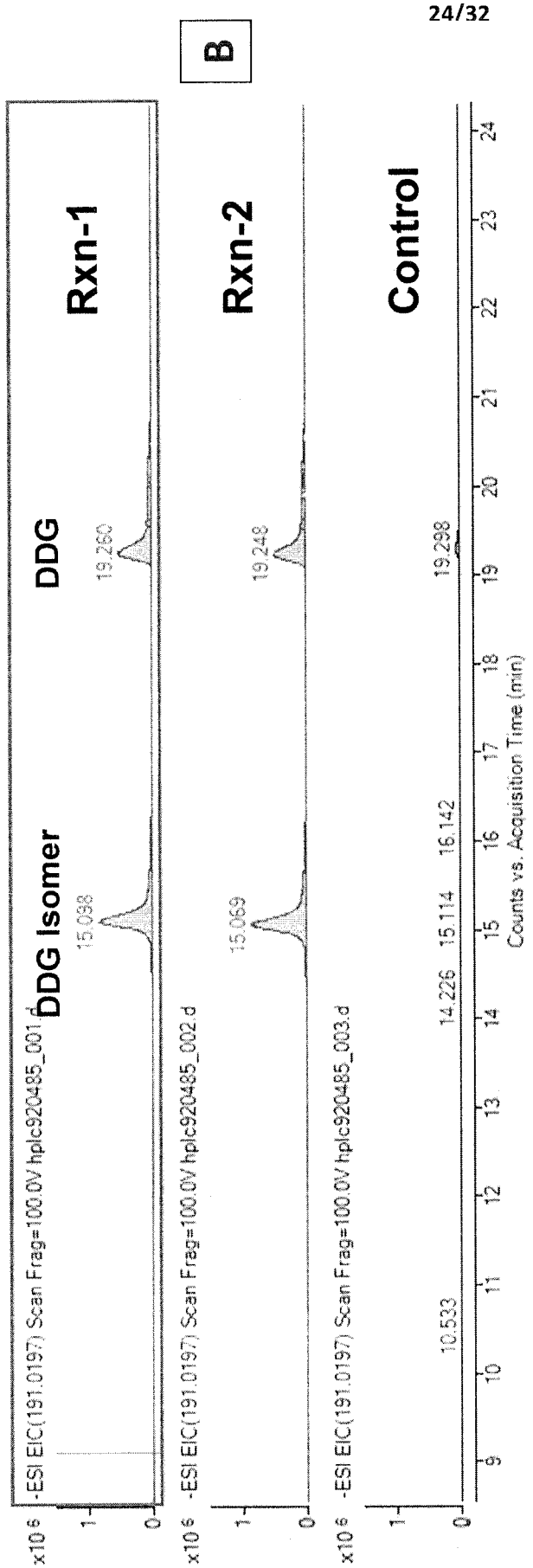

B
DDG
DDG Isomer
Rxn-1
Rxn-2
Control
-ESI EIC(191.0197) Scan Frag=100.0V hplc920485_001.d
-ESI EIC(191.0197) Scan Frag=100.0V hplc920485_002.d
-ESI EIC(191.0197) Scan Frag=100.0V hplc920485_003.d
x10 6
15.088
19.260
15.069
19.248
10.533
14.226
15.114
16.142
19.298
Counts vs. Acquisition Time (min)

Fig. 12B

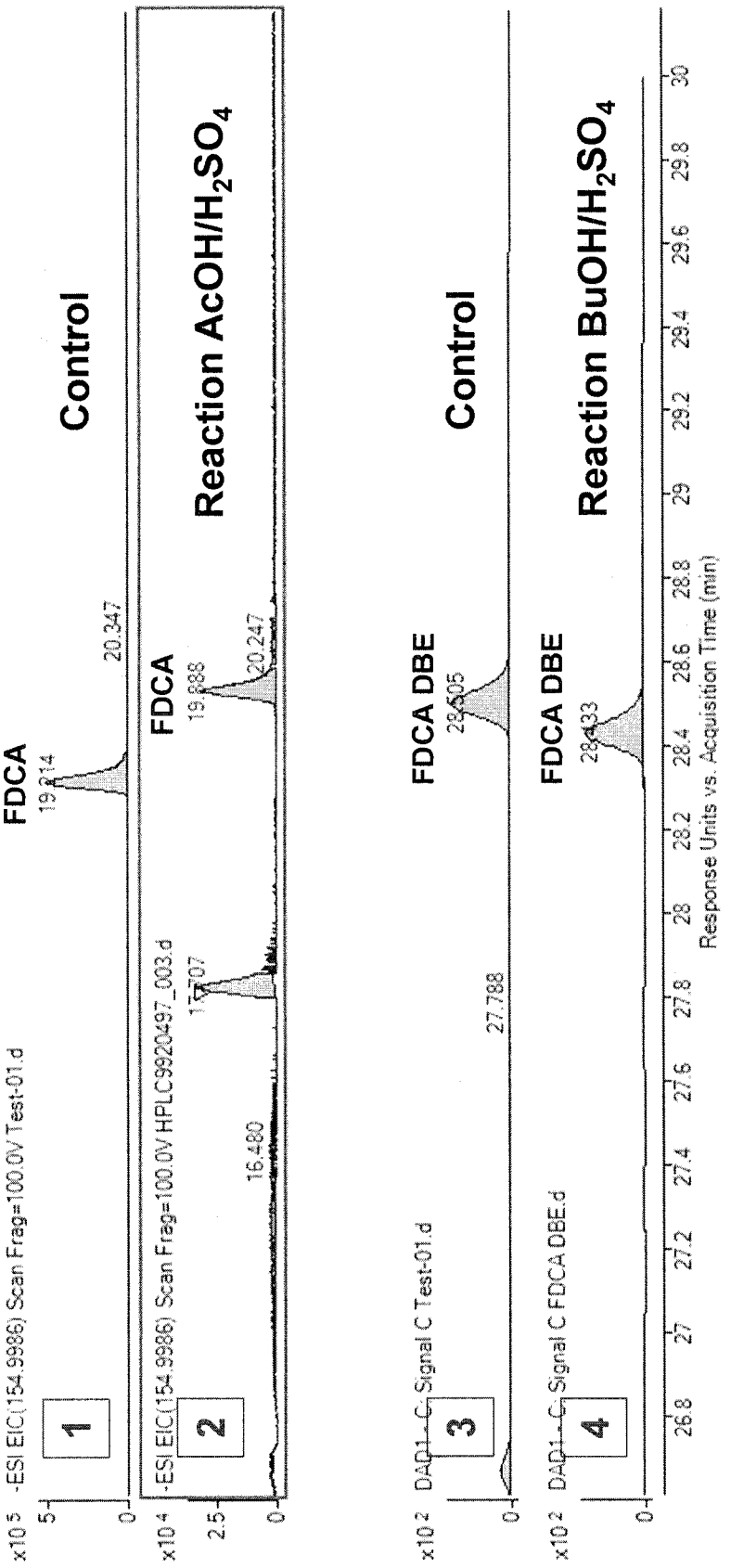

Control
FDCA
19.814
20.347
Reaction AcOH/H2SO4
FDCA
19.888
20.247
16.480
15.707
-ESI EIC(154.9986) Scan Frag=100.0V Test-01.d
-ESI EIC(154.9986) Scan Frag=100.0V HPLC9920497_003.d
Control
FDCA DBE
28.505
27.788
DAD1 - C, Signal C Test-01.d
Reaction BuOH/H2SO4
FDCA DBE
28.433
DAD1 - C, Signal C FDCA DBE.d
Response Units vs. Acquisition Time (min)

Fig. 13

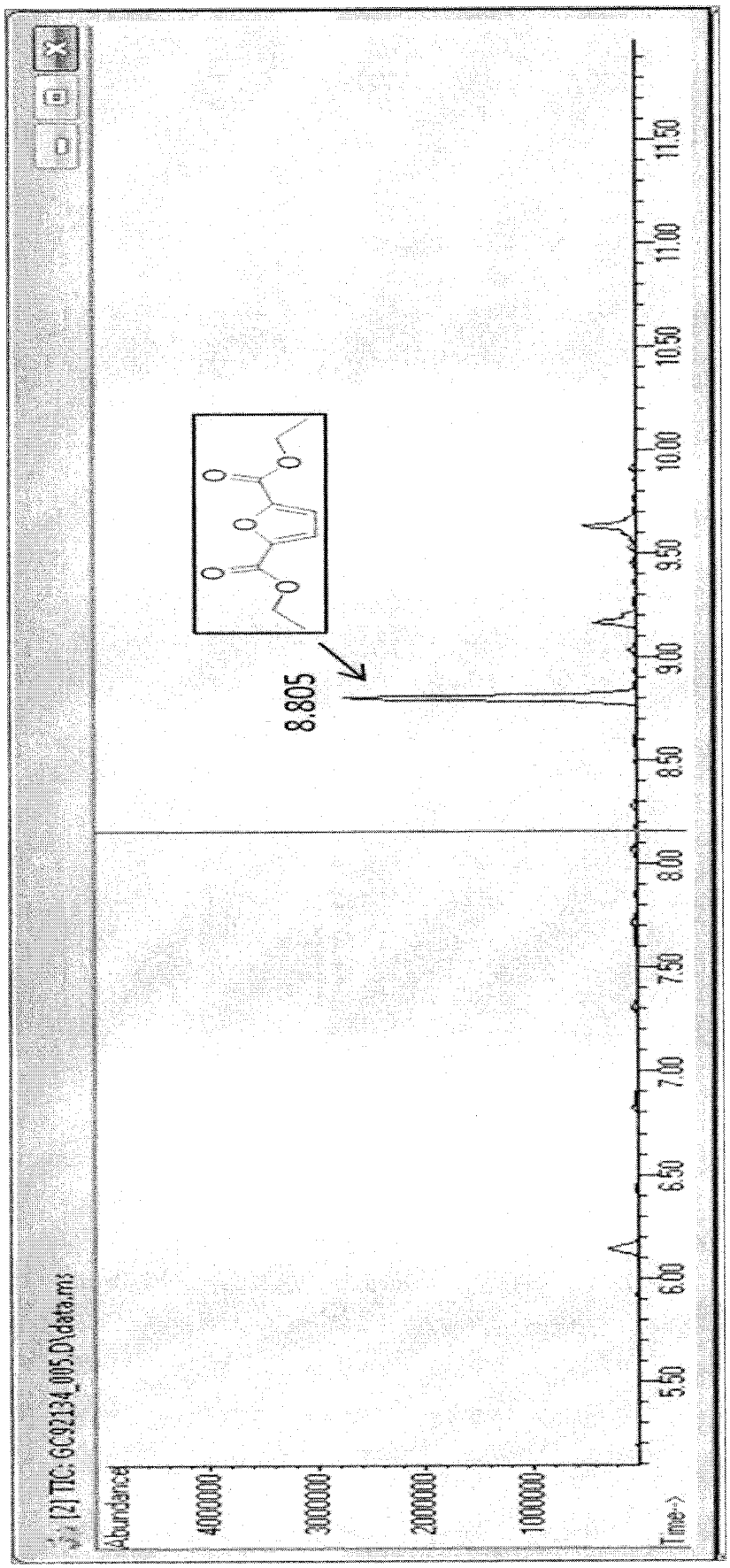

[2] TIC: GC92134_005.D\data.ms
Abundance
4000000
3000000
2000000
1000000
Time-->
5.50
6.00
6.50
7.00
7.50
8.00
8.50
9.00
9.50
10.00
10.50
11.00
11.50
8.805

Fig. 14A

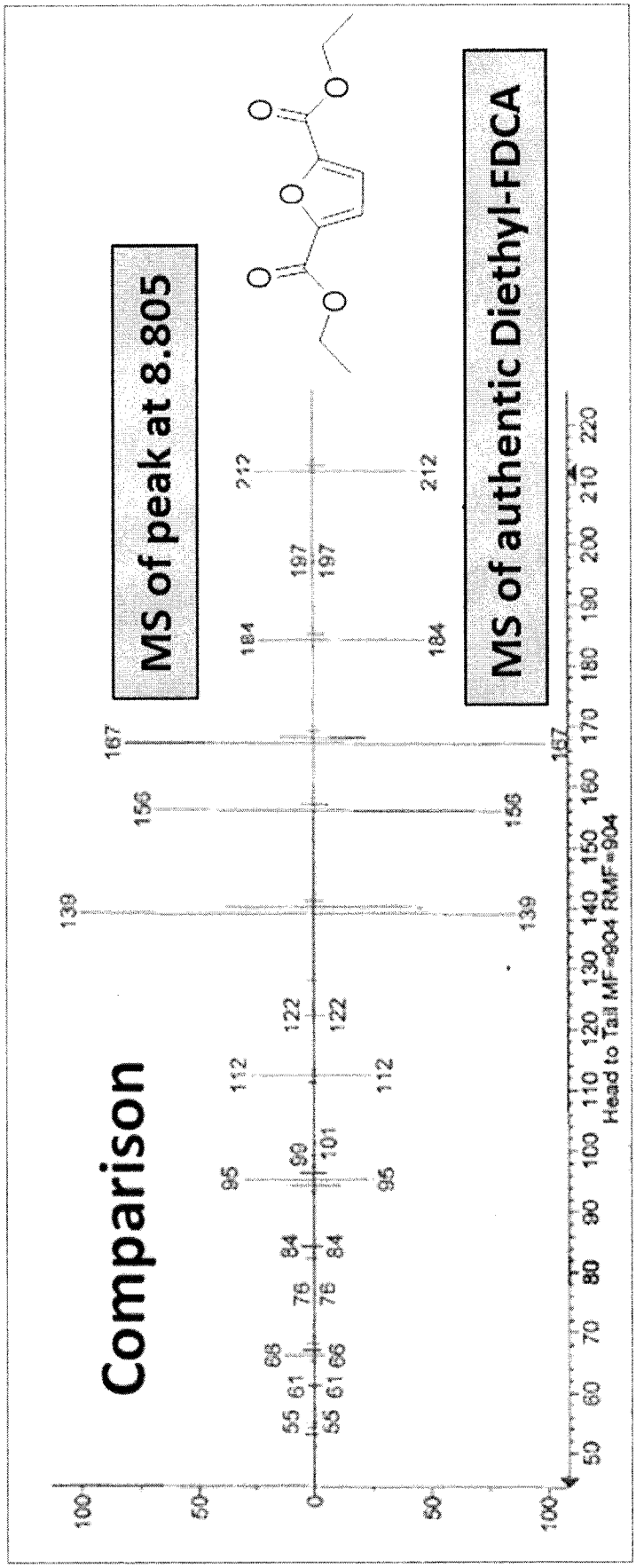

MS of peak at 8.805
MS of authentic Diethyl-FDCA
Comparison
Head to Tail MF=904 RMF=904

Fig. 14B

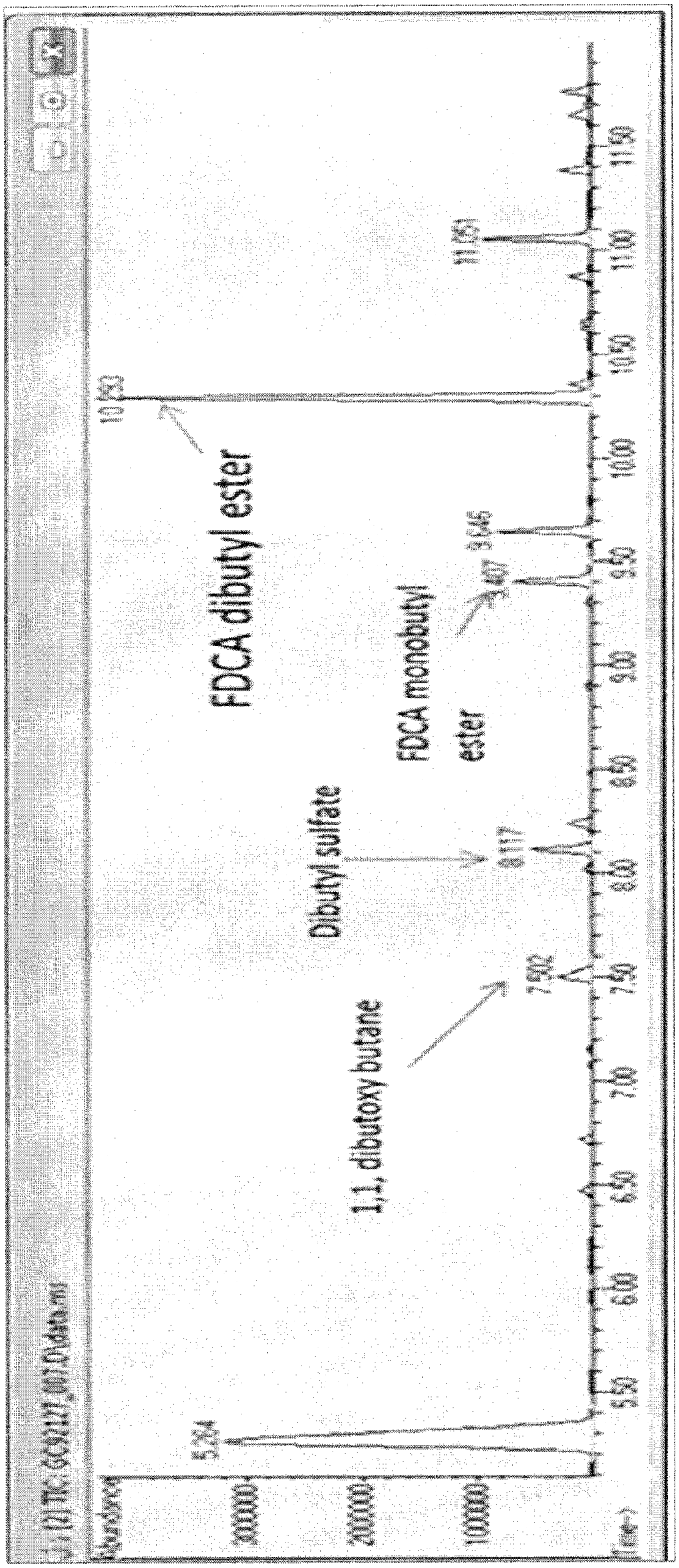

FDCA dibutyl ester
Dibutyl sulfate
FDCA monobutyl ester
1,1, dibutoxy butane
10.453
11.951
9.046
8.117
7.502
5.264
Abundance
Time

Fig. 15A

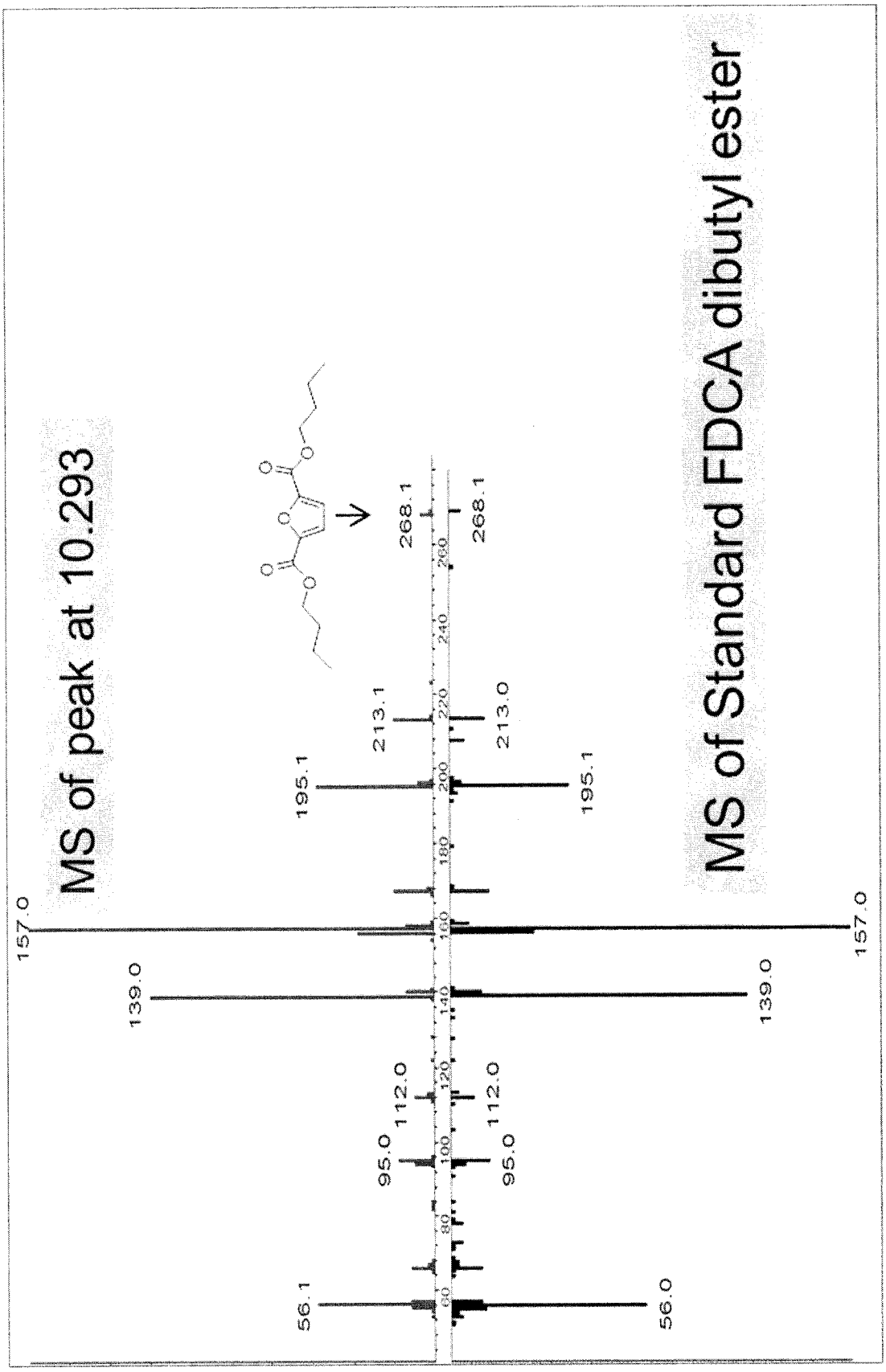

MS of peak at 10.293
MS of Standard FDCA dibutyl ester
157.0
139.0
112.0
95.0
56.1
195.1
213.1
268.1
213.0
268.1
195.1
157.0
139.0
112.0
95.0
56.0

Fig. 15B

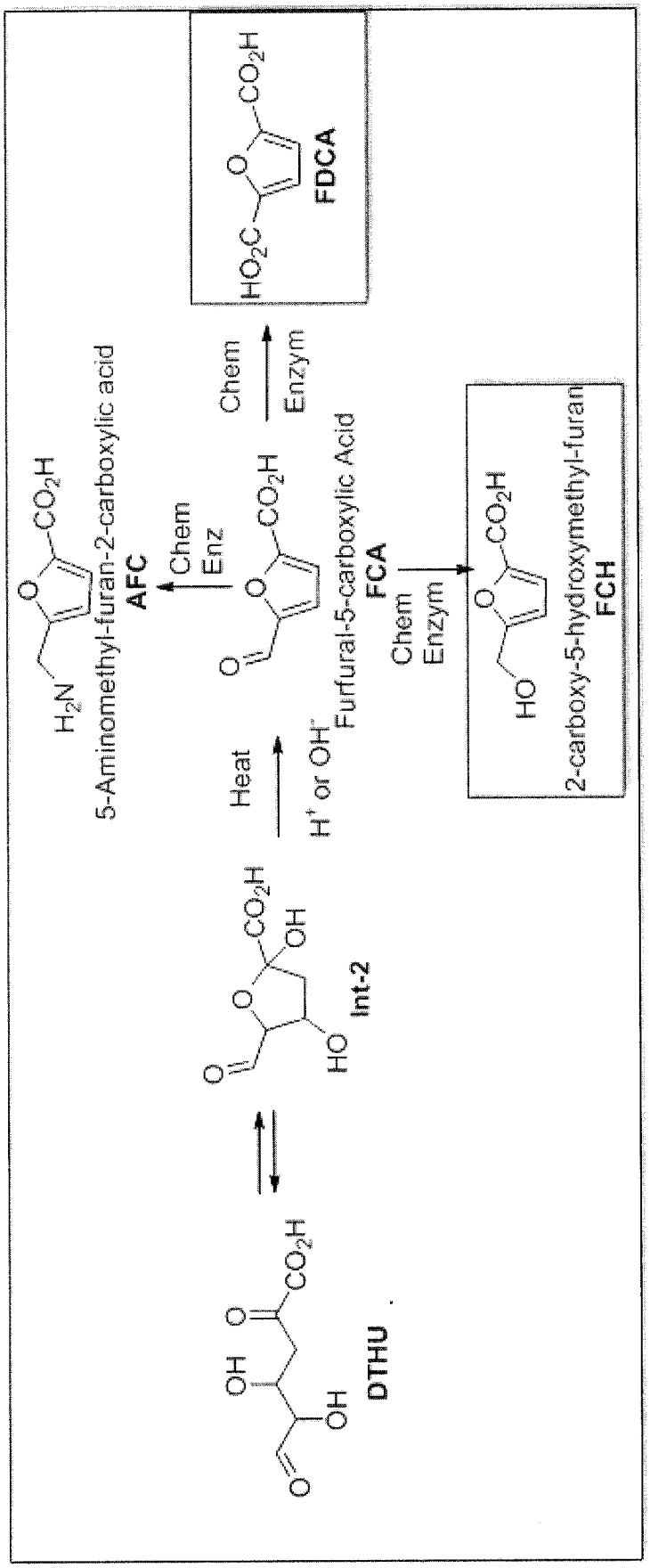

DTHU
Int-2
Heat
H+ or OH-
Furfural-5-carboxylic Acid
FCA
Chem
Enzym
2-carboxy-5-hydroxymethyl-furan
FCH
Chem
Enz
5-Aminomethyl-furan-2-carboxylic acid
AFC
Chem
Enzym
FDCA

Fig. 16

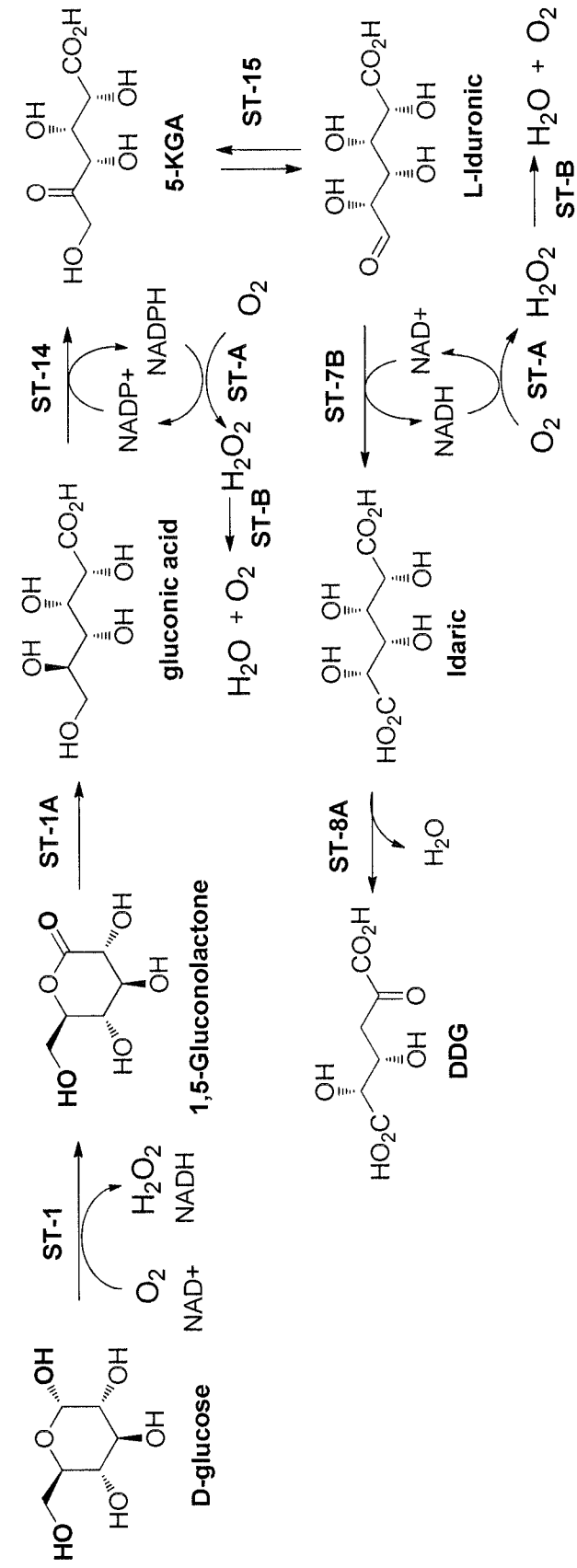

D-glucose
ST-1
O2
H2O2
NAD+
NADH
1,5-Gluconolactone
ST-1A
HO
gluconic acid
ST-14
NADP+
NADPH
5-KGA
ST-15
L-Iduronic
ST-7B
NAD+
NADH
Idaric
ST-8A
H2O
DDG
ST-A
O2
H2O2
ST-B
H2O + O2

Fig. 17A

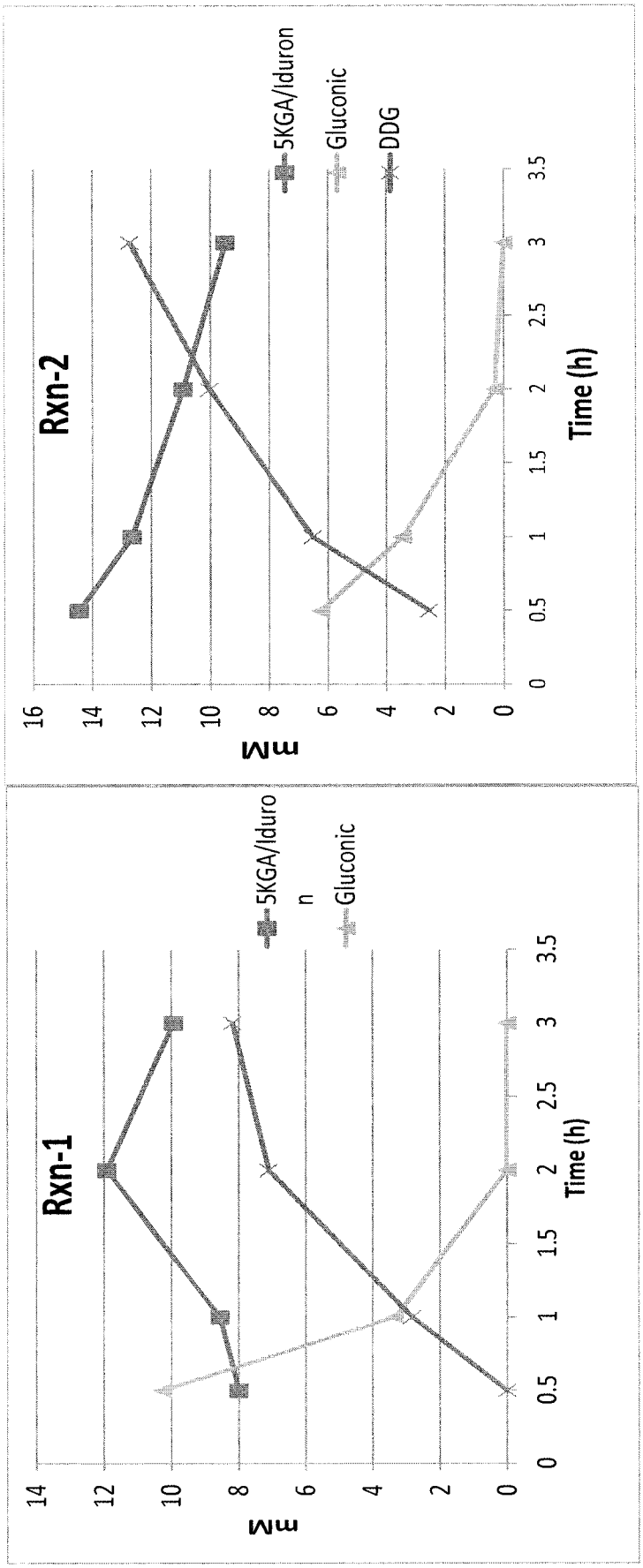

Rxn-2
mM
Time (h)
5KGA/Iduron
Gluconic
DDG
Rxn-1
mM
Time (h)
5KGA/Iduro
n
Gluconic

Fig. 17B