

 Open access • Journal Article • DOI:10.1042/BJ20120980

Inferring the metabolism of human orphan metabolites from their metabolic network context affirms human gluconokinase activity. — [Source link](#)

Ottar Rolfsson, Giuseppe Paglia, Manuela Magnusdottir, Bernhard O. Palsson ...+1 more authors

Institutions: University of Iceland

Published on: 15 Jan 2013 - Biochemical Journal (Portland Press)

Topics: Metabolic network, Gluconokinase activity and Gluconokinase

Related papers:

- [The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions](#)
- [A community-driven global reconstruction of human metabolism](#)
- [Global reconstruction of the human metabolic network based on genomic and bibliomic data](#)
- [What is flux balance analysis](#)
- [A protocol for generating a high-quality genome-scale metabolic reconstruction.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/inferring-the-metabolism-of-human-orphan-metabolites-from-1cbcijb7pn>



Inferring the metabolism of human orphan metabolites from their metabolic network context affirms human gluconokinase activity

Óttar ROLFSSON*†, Giuseppe PAGLIA*, Manuela MAGNUSDÓTTIR*, Bernhard Ø. PALSSON* and Ines THIELE*‡¹

*Center for Systems Biology, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland, †University of Iceland Biomedical Center, Laeknagardur, 101 Reykjavik, Iceland, and ‡Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, 101 Reykjavik, Iceland

Metabolic network reconstructions define metabolic information within a target organism and can therefore be used to address incomplete metabolic information. In the present study we used a computational approach to identify human metabolites whose metabolism is incomplete on the basis of their detection in humans but exclusion from the human metabolic network reconstruction RECON 1. Candidate solutions, composed of metabolic reactions capable of explaining the metabolism of these compounds, were then identified computationally from a global biochemical reaction database. Solutions were characterized with respect to how metabolites were incorporated into RECON 1 and their biological relevance. Through detailed case studies we show that biologically plausible non-intuitive hypotheses regarding the metabolism of these compounds can be proposed

in a semi-automated manner, in an approach that is similar to *de novo* network reconstruction. We subsequently experimentally validated one of the proposed hypotheses and report that C9orf103, previously identified as a candidate tumour suppressor gene, encodes a functional human gluconokinase. The results of the present study demonstrate how semi-automatic gap filling can be used to refine and extend metabolic reconstructions, thereby increasing their biological scope. Furthermore, we illustrate how incomplete human metabolic knowledge can be coupled with gene annotation in order to prioritize and confirm gene functions.

Key words: gap filling, human metabolism, metabolic network, RECON 1.

INTRODUCTION

The wealth of genomic and metabolic information that has amounted in the past decade has highlighted a need for methods that facilitate the functional annotation of biochemicals that take part in human metabolism [1]. Components of the human metabolome can be categorized on the basis of their degree of functional annotation into well-annotated, incompletely annotated and unknown components. The first category provides the basis for BiGG (biochemically, genetically and genomically) structured metabolic networks that have found widespread use in high-throughput data analysis, biochemical engineering and pathway discovery [2,3]. The latter two categories reflect that our knowledge of human metabolism is incomplete. Underlying these are (i) genes and proteins of unknown function, (ii) proteins that are not associated with a sequence (orphan enzymes) and (iii) metabolites, for which metabolic knowledge is incomplete. Estimates report that approximately 30–50% of all enzymes associated with EC (Enzyme Commission) numbers are not associated with genes [4–6] and that the function of over 50% of genes in higher organisms have not been experimentally determined [7]. Dead-end metabolites, i.e. compounds whose anabolic or catabolic route within humans are not completely defined, are also numerous [8].

BiGG metabolic network reconstructions contain components of an organism's metabolome, from information on genes and metabolic enzymes to metabolic reactants and products, and on how these components interact [9]. Incomplete genetic, genomic and biochemical information can be identified as network gaps in BiGG reconstructions. The two-step procedure of identifying

knowledge gaps and, subsequently, suggesting candidate gap-filling reactions is referred to as network gap filling [10–12]. Because BiGG reconstructions can be converted into a mathematical format, gap filling is carried out computationally. The SMILEY gap-filling algorithm, for example, has been successfully used to unravel novel metabolic functions in *Escherichia coli* [13] and to propose novel metabolic functions in humans [8]. Similar metabolic pathway discovery algorithms have found widespread use in biotechnology for pathway discovery [14].

The global human genome-scale metabolic reconstruction RECON 1 represents one of the most comprehensive BiGG metabolic reconstructions generated to date [15]. RECON 1 has found diverse applications in human metabolic research (reviewed in [16]) and has the potential to facilitate novel metabolic reaction discovery through a gap-filling approach. Although successful examples exist for bacterial metabolism [9,17], the use of eukaryotic metabolic models to this end has received limited attention [8].

Although comprehensive, RECON 1, and indeed all genome-scale reconstructions, is incomplete as knowledge of metabolism accumulates over time. Incorporation of acquired biological data can be performed using a gap-filling approach. If the incorporated biological components are incompletely annotated, then hypotheses of their biological functions can be generated from the biological network context. In the present study, we set out to identify and characterize human orphan metabolites that were defined here as metabolites that have been detected in human biofluids but were not accounted for in RECON 1. The present study aimed to elucidate the metabolism of these orphan metabolites in humans and simultaneously expand the biological

Abbreviations used: BiGG, biochemically, genetically and genomically; EC, Enzyme Commission; GFP, green fluorescent protein; HMDB, Human Metabolome Database; IS, internal standard; LC-MS, liquid chromatography-MS; 6PGDH, 6-phosphogluconate dehydrogenase; PLP, pyridoxal 5-phosphate; PPP, pentose phosphate pathway.

¹ To whom correspondence should be addressed (email ines.thiele@gmail.com).

accuracy of RECON 1. To our knowledge the systematic identification and characterization of human orphan metabolites has not been carried out previously. We used a gap-filling algorithm to compute reaction hypotheses that could account for these metabolites in humans and show that the approach not only helps to reconstruct known biochemical pathways, as expected, but is also capable of generating biologically plausible non-intuitive reaction hypotheses. We exemplified this by validating one of our predictions experimentally, which resulted in the characterization of a human gluconokinase.

EXPERIMENTAL

Pre-processing to metabolite integration

RECON 1 was downloaded from the BiGG database [18]. The reconstruction was imported into Matlab (MathWorks) using the COBRA toolbox [19]. Subsequently, we decompartmentalized all reactions in RECON 1 by placing all intracellular compartment reactions into the cytosol and removing duplicates. Extra-organism-located reactions were kept. We downloaded the KEGG Ligand database (version as of 10/2009) [20] to obtain a universal reference reaction list. For each metabolite present in RECON 1 and in the universal database, we created a transport reaction from cytosol to the extra-organism space. Furthermore, we downloaded the HMDB (Human Metabolome Database, release 2.5) [21]. We mapped the entries from the HMDB on to RECON 1 using the BiGG identifiers provided by the HMDB.

Computation of candidate solutions using gap filling

The gap-filling algorithm has been described previously [9]. We used the implementation provided by the COBRA toolbox. Briefly, RECON 1 served as the model to which reactions from either the universal reaction list and/or the transport reaction list should be added. In order to identify candidate anabolic and catabolic solutions for each metabolite, we added a sink (sink_A: $\emptyset \rightarrow 1$ A) or demand reaction (DM_A: 1 A $\rightarrow \emptyset$) to the model and performed the gap-filling analysis by requiring the flux through either DM_A or sink_A to be greater than zero. Up to ten solutions for each metabolite were computed. The gap-filling algorithm requires metabolites in the reconstruction to have KEGG identifiers in order to calculate metabolic reaction solutions. This restrained the metabolites identified to metabolites from the HMDB with KEGG identifiers. Note that not all metabolites in RECON 1 have a KEGG identifier, which may also limit the identification of candidate solutions from the KEGG database.

Gap analysis and validation of the gap-filling solution output

The gap-filling algorithm generated a total of 1131 solutions for the 343 orphan metabolites identified. In the majority of cases metabolites either had five solutions or only one (Supplementary Table S1 at <http://www.biochemj.org/bj/449/bj4490427add.htm>). Solutions were filtered with the following criteria in mind: (i) only solutions containing gene-associated resolving reactions were picked; (ii) if no solution met this criteria, solutions with the least number of resolving reactions and leading to integration of the metabolite within the network as opposed to transport were picked; and (iii) if no such solution was found, then the solution involving the least number of resolving reactions was picked, irrespective of the resolving reaction type. In cases where two solutions met these criteria an informed decision on

the basis of a literature review was made to select the most biologically relevant solution. These criteria generated a filtered candidate solution output (S1 and S2), which is reported in the Results section (Supplementary Table S2 at <http://www.biochemj.org/bj/449/bj4490427add.htm>).

Candidate reactions comprising the S1 and S2 solutions were investigated individually to generate plausible hypotheses for the gap filling of RECON 1. The orphan metabolites' biological origin was determined and the reactions composing the gap-filling solutions were assessed on the basis of the underlying literature and their annotations within multiple reaction and metabolite databases. Biochemical and metabolic information was obtained from the HMDB [21], the HSDB (Hazardous Substances Database; <http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB>) and the KEGG database [20]. For resolving reactions, not annotated in humans, BLAST homology searches of genes encoding the resolving reaction activity were performed against the human sequence (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and using STRING [22]. On the basis of literature and electronic data, solutions were classified as: (i) accepted, solutions were composed of reactions for which genetic or biochemical data had been reported in humans or closely related species; (ii) hypothesis, solutions were biochemically plausible but genetic or biochemical information was incomplete and requires experimental investigation; (iii) discarded, solutions were composed of reactions for which no evidence could be found in humans or closely related species; and (iv) no solution, the gap-filling algorithm failed to generate a solution. All metabolic reaction and subcellular compartment abbreviations are as described in the BiGG database [18].

Overexpression of C9orf103 and enzyme assays

C9orf103 isoform I (BI826543) and isoform II (BC142991) were directionally cloned into a pT7CFE1-CHis expression vector (Thermo Scientific) at the BamHI and XhoI restriction sites following amplification using the forward primers 5'-GG-AATTCATATGGGATCCGCGGCGCCGGGCGCGCTGCTG-3' and 5'-GGAATTCCATATGGGATCCAAGAAACCGAGGC-CCGCGGCTGC-3' for isoforms I and II respectively, and the reverse primer 5'-TATCTCGCTCGAGTTTCATTTTAGGGTT-TCCATAATTG-3'. This generated two vector constructs, pT7CFE1-GlnKIsoI and pT7CFE1-GlnKIsoII, which, following linearization with SpeI, were used in 25 μ l coupled *in vitro* transcription/translation reactions using the Pierce[®] Human *In Vitro* Protein Expression kit for 6 h at 30°C. Confirmation of overexpressed products was performed by Western blotting using the HisProbe-HRP reagents and kit (Thermo Scientific). GFP (green fluorescent protein) expressed from pCFE-GFP (Thermo Scientific) was used as a positive control in *in vitro* translation reactions and the gluconokinase assay. Following translation, lysates were diluted in 1 \times kinase reaction buffer [100 mM phosphate, 40 mM NaCl and 2.5 mM MgCl₂ (pH 7.2)] to a volume of 50 μ l and desalted using spin columns with a 7.5 kDa molecular mass cut-off. A 12 μ l volume of this diluted desalted lysate was then used in gluconokinase assays. The gluconokinase assay was carried out in 33 μ l reaction volumes in kinase reaction buffer with 0.1 mM or 1 mM gluconate and 1 mM ATP. Purified recombinant *E. coli* gluconokinase (50 ng; Megazyme, specific activity 31.8 units/mg) was added and used as a positive control. Following mixing by pipetting, reaction mixtures were immediately split into 20 μ l and 13 μ l aliquots. The 20 μ l aliquot was incubated for 2 h at room temperature (20°C), and then frozen at -80°C for MS analysis. To the 13 μ l aliquot, 2 μ l

of 6PGDH (6-phosphogluconate dehydrogenase) master mix was added and incubated at room temperature for 2 h. NADPH production was measured by monitoring fluorescence at 445 nm following excitation at 340 nm and interpolating values to that of an NADPH standard curve. The 6PGDH master mix resulted in the addition of 0.015 units of 6PGDH (Megazyme) and a 625 μM final concentration of NADP^+ . Following fluorescence measurements, a 3 μl aliquot was removed from the kinase reaction, diluted to 30 μl with kinase reaction buffer and the ATP content of the reaction was determined with a CellTiter-Glo assay (Promega) according to the manufacturer's instructions. All data are results of averaging over three replicate samples and were similar in $n = 3$ repeats.

LC-MS (liquid chromatography-MS)

A UPLC system (UPLC Acquity, Waters) coupled in-line with a quadrupole-time-of-flight hybrid mass spectrometer (Synapt G2, Waters) was used. Separation was achieved by HILIC (hydrophilic interaction chromatography) analysis on an 1.7 μm Acquity amide column (2.1 mm \times 150 mm) (Waters). An electrospray ionization interface was used in negative mode to direct column eluent to the mass spectrometer. The flow rate was 0.4 ml/min. Mobile phase A was 10 mM acetonitrile/sodium bicarbonate (95:5) and mobile phase B contained 10 mM acetonitrile/sodium bicarbonate (5:95). The following elution gradient was used: 0 min at 99% A, 3 min at 20% A, 4 min at 99% A and 5 min at 99% A. A capillary voltage of 1.8 kV and a cone voltage of 35 V was used. MS spectra were acquired in centroid mode from m/z 50 to 1000 using a scan time of 0.35 s. Leucine enkephalin (2 ng/ μl) was used as the lock mass (m/z 554.2615). [$^{13}\text{C}_6$]Gluconate was used as the IS (internal standard) for enzyme assays when gluconate was used as the substrate, whereas gluconate was used as the IS for enzyme assays when [$^{13}\text{C}_6$]gluconate was used as the substrate. A 5 μl volume of IS (0.6 mM) was added to 20 μl of sample. Successively 300 μl of methanol was added and samples were vortex-mixed for 5 min and centrifuged for 10 min at 10000 g. Supernatants were recovered, dried under nitrogen and reconstituted in 80 μl of acetonitrile/water (50:50, v:v). Samples (5 μl) were injected into the UPLC-MS system.

RESULTS

Identification and characterization of human orphan metabolites

We aimed to identify human orphan metabolites. Therefore we compared metabolites in RECON 1 with those that have been detected in human biofluids as reported within the HMDB database [21] (Figure 1A). A total of 343 metabolites were identified whose metabolic fate was not captured in RECON 1 but can be assumed to have a metabolic degradation or production route within humans on account of their detection in biofluids.

The metabolites were characterized on the basis of their chemical taxonomy super-classes (Figure 2A). Organic acids, both aromatic and linear, followed by various alcohols and amino acid derivatives were found to be the most numerous. These compound classes complement the gaps caused by blocked reactions identified in RECON 1 reported previously, where the largest proportion of blocked reactions have been observed in amino acid metabolism, carbohydrate metabolism and lipid metabolism [8]. The majority of the metabolites were detected in either blood or urine and reflects that these biofluids are most commonly investigated due to ease of sample access (Figure 2B). Numerous metabolites were identified that would

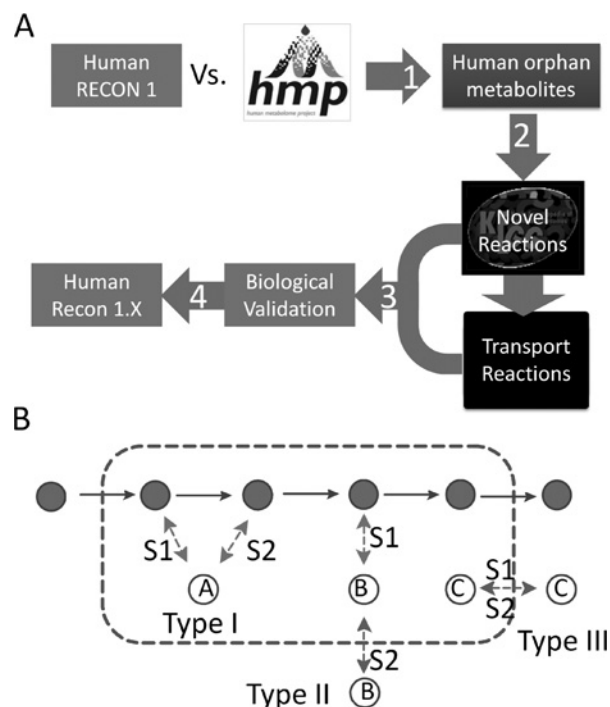


Figure 1 Identification of human orphan metabolites and of reactions capable of explaining their metabolic role

(A) Orphan metabolites were defined and identified as metabolites that have been detected in human biofluids but not accounted for in RECON 1 (1). A modified network gap-filling algorithm was used to propose the manner in which metabolites could be incorporated into RECON 1 using non-organism-specific metabolic reactions from the KEGG database or transport reactions (2–3). Following validation of the biological relevance of proposed solutions (3), metabolites could be incorporated into an updated metabolic reconstruction (4). (B) A toy network showing the computed candidate solution types. Metabolite A is incorporated into the toy network with two solution routes (S1 and S2), which couple the production and consumption of A to metabolites already accounted for in the network (dark grey). A type II solution involves transport into the cell and incorporation into the network as shown for metabolite B. A type III solution involves transport into and out of the cell as shown for metabolite C. Note that solution routes S1 and/or S2 can be composed of multiple reactions.

not normally be associated with basal metabolism, for example, drug derivatives and plant-derived antioxidants. Some of these metabolites, therefore, represent metabolic scope gaps [23], as they lie outside the biological scope of RECON 1 and/or are involved in non-metabolic pathways.

Incorporation of orphan metabolites into RECON 1

In order to incorporate the identified metabolites into RECON 1, we computed candidate solutions, composed of non-organism-specific metabolic reactions from the KEGG database that could account for the metabolic production and consumption of each metabolite within RECON 1 (see the Experimental section). Metabolites that could not be incorporated using reactions from KEGG were connected to RECON 1 using an artificial transport function (Figure 1A). Up to ten solutions were generated for each metabolite (Supplementary Table S1) of which we chose two solutions routes (S1 and S2) corresponding to its metabolic production and consumption route. These were categorized on how they incorporated the metabolites into RECON 1 (Figure 1B) and whether they corresponded to biologically plausible gap-filling solutions (Figure 2C) (defined in the Experimental section). Owing to the reversibility of enzymatic reactions, we did not assign solution routes specifically as production or consumption

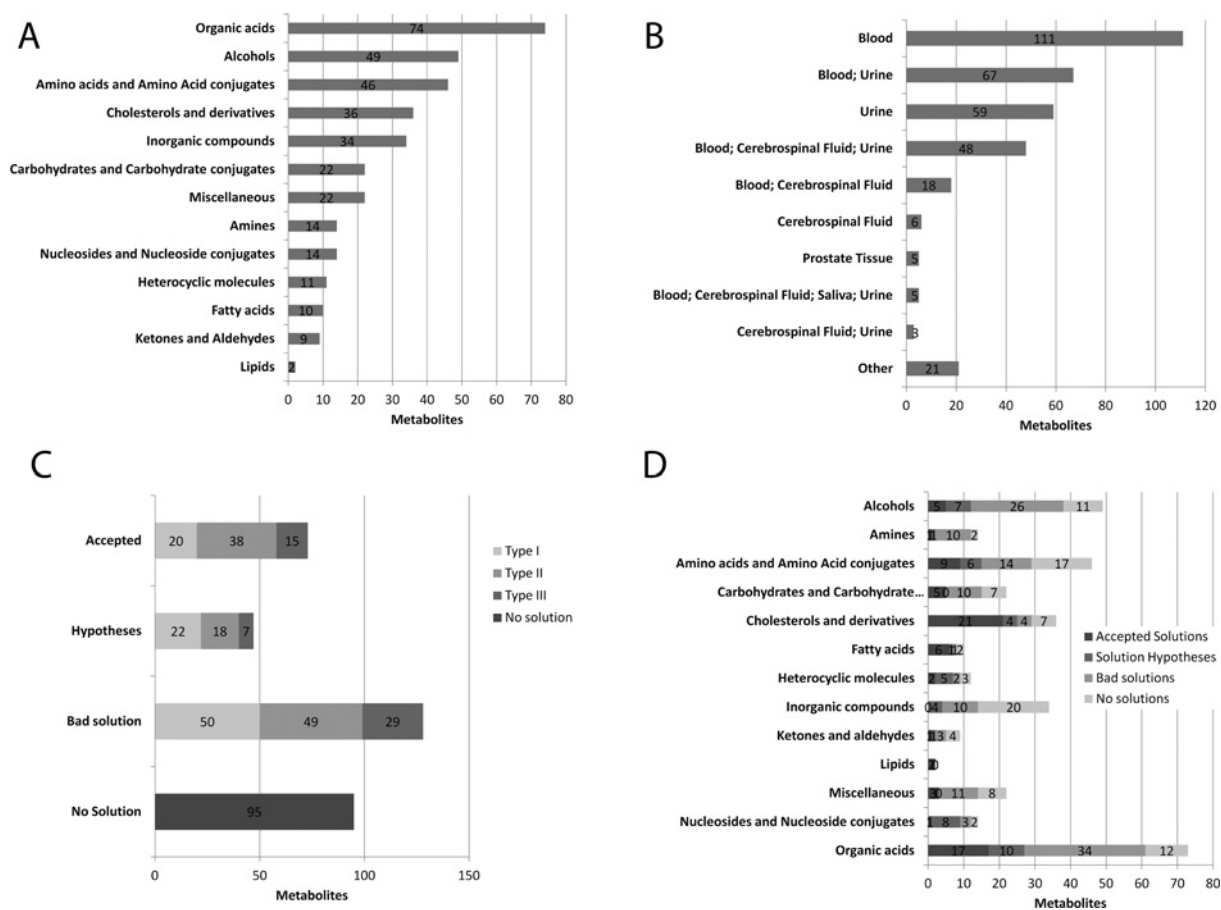


Figure 2 Characterization of identified orphan metabolites and the properties of their computed solutions

(A) The distribution of identified metabolites according to their chemical compound superclass. (B) The human biofluids in which the orphan metabolites were detected. (C) Computed candidate solutions for each metabolite were accepted, categorized as a valid hypotheses or rejected. Solutions were further characterized as type I, type II or type III solutions as outlined in Figure 1(B). (D) Computed candidate solutions with respect to their metabolite class.

routes. A total of 119 metabolites had candidate solutions that were either accepted or corresponded to hypotheses. The remaining metabolites had solutions that were discarded (128) or had no solution at all (95). This filtered gap-filling output was the main focus of the following report (Supplementary Table S2).

Accepted candidate solutions

Analysis of the candidate solutions revealed that not all of the metabolites identified corresponded to orphan metabolites. In fact, solutions for 73 metabolites could be added directly to RECON 1 (Figure 2C). In all cases, the reactions associated with these metabolites, with the exception of added transport reactions, were either associated with known human genes or biochemically identified orphan enzyme activities in humans or closely related mammals. With these metabolites, the gap-filling algorithm expanded the metabolic reconstruction through addition of characterized human reactions and metabolites. For example, cholesterols and derivatives represented the most populous group of compounds with accepted solutions (Figure 2D). Five oestrogens were identified that could be added directly to RECON 1 using type I or II solutions, thereby expanding the C_{18} steroid biosynthesis pathway of RECON 1 (Supplementary Figure S1 at <http://www.biochemj.org/bj/449/bj4490427add.htm>). A total of 58 out of the 73 metabolites with accepted solutions had type I or II solutions. Having been identified in biofluids,

metabolites with type I solutions suggest that at least a portion of these can be produced and consumed within the cell.

Metabolites with type III solutions do not add to a reconstruction knowledgebase in the same manner as type I and II solutions. Type III solutions introduce flux loops into the reconstruction. The plant metabolite phytosterol, for example, was connected to RECON 1 in three reactions: transport into the cell, conversion by Δ -sterol reductase (EC 1.3.1.72) activity into isofucosterol and subsequent transport out of the cell. Contextualization of this type of knowledge is, however, important in terms of modelling sterol metabolism, where biological effects of these compounds could be coupled to knowledge of their metabolic interconversions. Any abnormalities of Δ -sterol reductase activity could then be associated with biological effects of these compounds.

Computed candidate solutions direct hypothesis generation

As a next step we analysed how many biologically plausible hypotheses could be obtained using the gap-filling algorithm by a thorough literature review. We identified 47 orphan metabolites with plausible hypotheses. Most of these applied to organic acids, alcohols, nucleoside derivatives and amino acids (Figure 2D). A total of 22 of these metabolites could be connected to RECON 1 using type I solutions. One of these was cysteine *S*-sulfate, an NMDA (*N*-methyl-D-aspartate) receptor agonist excreted into

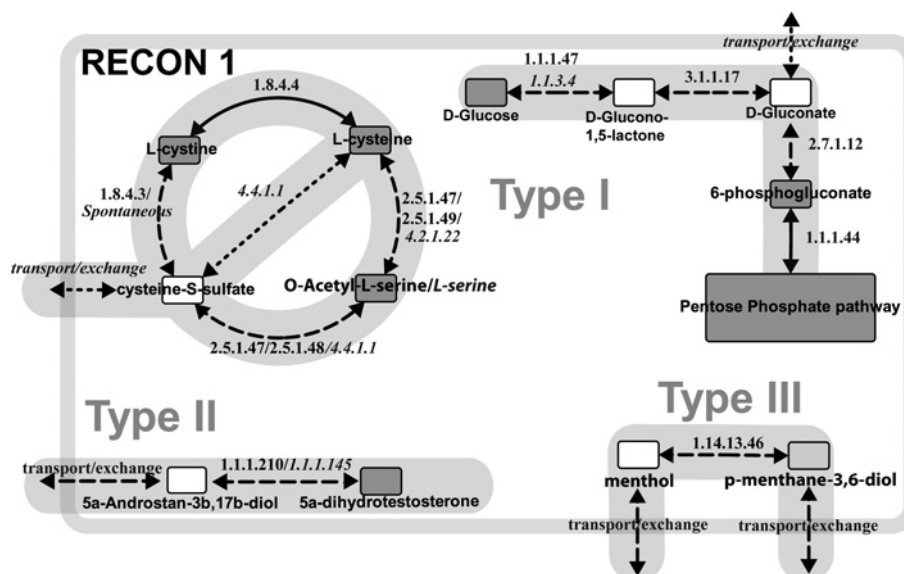


Figure 3 Orphan metabolites and examples of gap-filling candidate solutions

Cysteine *S*-sulfate and *D*-gluconate were incorporated via type I solutions, 5α -androstan- 3β , 17β -diol via a type II solution and menthol with a type III solution. Solid arrows and metabolites in dark grey represent metabolic reactions and metabolites within RECON 1 before the present study. Dashed arrows represent reactions added by the gap-filling algorithm to incorporate the human orphan metabolite (white). EC numbers are indicated in bold. Metabolites in light grey were not in RECON 1 and were not detected as orphan metabolites, but were added in order to incorporate the orphan metabolite. EC numbers in italic and dotted arrows represent functions identified following a review of the literature.

the urine of individuals with sulfite oxidase deficiency [24]. Cysteine *S*-sulfate was connected to RECON 1 through the *L*-cystine and *L*-cysteine by the addition of three enzymatic reactions and the metabolite *O*-acetyl-*L*-serine (Figure 3). The formation of cysteine *S*-sulfate from *L*-cystine was deemed plausible as glutathione-CoA-glutathione transhydrogenase (EC 1.8.4.3) is an orphan reaction in rat and bovine [25] and has also been shown to be spontaneous and reversible [26]. Conversion of cysteine *S*-sulfate into *L*-cysteine involved cysteine synthase (EC 2.5.1.47), cystathionine δ -synthase (EC 2.5.1.48) and *O*-acetylhomoserine aminocarboxypropyl-transferase (EC 2.5.1.49). These three enzymes have sequence similarity to two PLP (pyridoxal 5-phosphate)-dependent enzymes of the human *trans*-sulfuration pathway, cystathionine β -synthase (EC 4.2.1.22) and cystathionine δ -lyase (EC 4.4.1.1), which have recently been shown to have broad substrate specificity in the β -replacement reactions and in the elimination reactions that these enzymes catalyse respectively [27,28]. Cystathionine β -synthase is, for example, known to catalyse the α,β -elimination reaction of *L*-cysteine to produce *L*-serine [27,28]. Whether these two enzymes are capable of eliminating thiosulfide from cysteine *S*-sulfate, to generate serine and thiosulfate, has not been experimentally determined.

Type I solutions were also obtained for compounds originating from coffee or tea, such as theophyllin, theobromine and caffeine. The metabolism of these compounds is fairly well characterized, although the enzymes associated with their degradation have not been fully characterized and the resulting excreted metabolites are heterogeneous. Accordingly, the solution output consisted of multiple degradation routes, none of which could be discriminated against and highlights a shortcoming of automated hypothesis generation when multiple solutions are correct or can be obtained (Supplementary Figure S2 at <http://www.biochemj.org/bj/449/bj4490427add.htm>).

Type II solutions were obtained for 18 out of the 47 metabolites. Among these was 5α -androstan- 3β , 17β -diol, a steroid that has

been shown to exert an inhibitory effect on prostate cancer cell migration via activation of oestrogen receptor β [29]. The candidate solution involved production from dihydrotestosterone with 3β (or 20α)-hydroxysteroid dehydrogenase (EC 1.1.1.210). Its activity has been detected in ox and rat but has not been associated with a gene in mammals [30]. A recombinant 3β -hydroxy- Δ^5 -steroid dehydrogenase (EC 1.1.1.145) was, however, previously shown to carry out this reaction (Figure 3) [31].

A total of seven metabolites had type III solutions. For example, the metabolic route of menthol was predicted to involve intracellular transport, oxidation of menthol to generate *p*-menthane-3,8-diol followed by extracellular transport of *p*-menthane-3,8-diol (Figure 3). Menthol is known to enter either enterohepatic circulation as a biliary metabolite or be oxidized by cytochrome P450 generating menthol diols that are then excreted into the urine [32,33]. Type III solutions were also obtained for valproic acid, an epileptic drug. These two examples highlight the need for a detailed reconstruction of xenobiotics, which are currently not captured in RECON 1. Type III solutions were also obtained for methylated nucleosides, well-known cancer markers, which are also not represented in RECON 1 as they are derived from the degradation of RNA and therefore are beyond its scope.

Experimental validation of a computed hypothesis

Candidate reaction solutions remain hypotheses until experimentally proven. We performed experimental confirmations of the type II solution obtained for gluconate. Gluconate is the C-1-oxidized linear derivative of glucose and a common food and drug constituent [34]. The proposed metabolism involved the oxidation of glucose followed by hydrolysis of *D*-glucono-1,5-lactone to generate *D*-gluconate that is subsequently metabolized upon phosphorylation in the PPP (pentose phosphate pathway) (Figure 3). Hexose-1-dehydrogenase (EC 1.1.1.47) and gluconolactonase (EC 3.1.1.17) activity have previously been

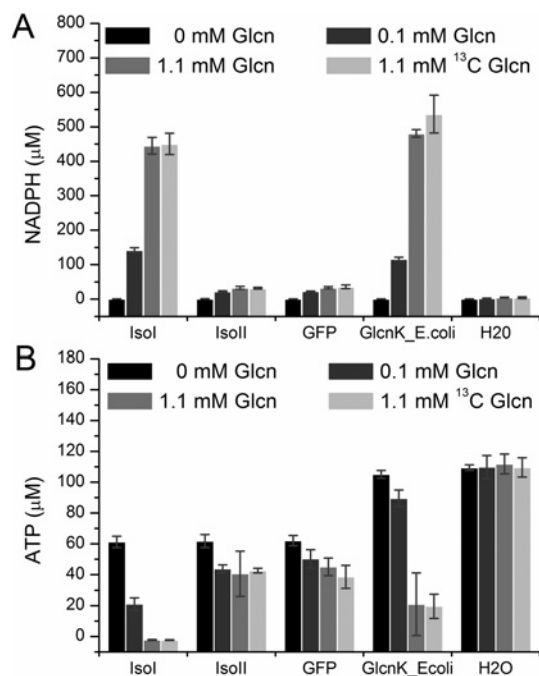


Figure 4 End point enzyme assays of overexpressed *C9orf103* in HeLa lysates confirms gluconokinase activity of isoform I of *C9orf103*

A 6PGDH-coupled enzyme assay of gluconokinase activity in HeLa lysates overexpressing isoform I or isoform II of *C9orf103* or GFP. More NADPH is generated in the presence of isoform I as compared with isoform II and the negative control GFP (A). ATP end point measurements in the gluconokinase reactions show a drop in ATP concentration consistent with ATP-dependent kinase activity for isoform I (B). A similar effect on NADPH and ATP concentrations was observed in the presence of the positive control *E. coli* gluconate kinase (*GlcK_E.coli*). All reaction mixtures contained ATP at a final concentration of 1 mM and were diluted 10× before performing the ATP assay.

reported in humans [35,36]. An enzyme with gluconokinase activity has also been previously purified and characterized from pig kidney extracts [37]. Performing a homology search of bacterial enzymes having gluconokinase activity (EC 2.7.1.12), we identified *C9orf103* as a plausible candidate gene, which may be capable of providing a catabolic route for gluconate in humans. A BLAST query of isoform I of *C9orf103* against the *E. coli* genome revealed a 40% sequence identity to gluconokinase (gene symbol *GNTK*) with an E value of 8^{-40} and over 90% sequence coverage. Furthermore, 12 out of the 14 amino acid residues that have been shown to be important for *E. coli* gluconokinase substrate binding are conserved in *C9orf103* (Supplementary Figure S3 at <http://www.biochemj.org/bj/449/bj4490427add.htm>).

In order to confirm gluconokinase activity of *C9orf103*, we overexpressed isoforms I and II of *C9orf103* in human HeLa cell lysates (Supplementary Figure S4 at <http://www.biochemj.org/bj/449/bj4490427add.htm>) and assayed for gluconokinase activity using a 6PGDH-coupled enzyme assay. Following incubation of cell lysates with added gluconate, we detected more NADPH in lysates containing overexpressed isoform I than with isoform II or the negative control containing GFP. The NADPH concentration was similar to that observed for the *E. coli* gluconokinase positive control. Little or no NADPH was detected in the water negative control (Figure 4A). These data suggest that *C9orf103* encodes an active gluconokinase. The difference in NADPH concentration can be largely attributed to the reduction of NADP^+ that accompanies the oxidation of 6-phosphogluconate produced by the *C9orf103* gene product. A

drop in ATP concentration upon gluconate addition was also observed in the lysate expressing isoform I, consistent with ATP-dependent kinase activity (Figure 4B). A change in both NADPH and ATP concentration was also observed upon gluconate addition in isoform II and GFP samples, perhaps hinting at background gluconokinase activity in these lysates.

To further confirm the gluconokinase activity of isoform I and completely rule out activity of isoform II or background activity, we quantified gluconate and 6-phosphogluconate in the cell lysates using LC-MS (Figure 5 and Supplementary Figure S5 at <http://www.biochemj.org/bj/449/bj4490427add.htm>). Component analysis showed little or no gluconate in isoform I and in the positive-control samples, whereas 6-phosphogluconate was readily detected. The opposite was true for isoform II and the negative-control samples. An experiment where uniformly labelled [^{13}C]gluconate was used as the substrate unequivocally demonstrates that the added gluconate is indeed converted into 6-phosphogluconate in the presence of overexpressed isoform I. No kinase activity was observed for isoform II above the negative control. This is most likely due to the absence of the phosphate-binding loop binding domain that is present in the N-terminal domain of isoform I but missing in isoform II (Supplementary Figure S3), suggesting a different function for this isoform. These data confirm gluconokinase activity in humans and support a gluconate degradation route through the PPP similar to the one observed in lower organisms [38].

Metabolites with discarded or no solutions

A large proportion of the metabolites identified had either candidate solutions that were discarded or had no solutions at all (Figure 2C). Approximately 60% of the metabolites having discarded solutions were organic acids, alcohols or amino acid derivatives, which also represented the most numerous metabolite classes. In the majority of the cases, no evidence was found to support the occurrence of the proposed solutions in humans. Some metabolites were also found to not correspond to orphan metabolites. Exogenous compounds, such as rosmarinic acid and curcumin, have been shown to be excreted following glucuronidation or glycosylation [39,40]. Similarly, compounds, such as secondary bile acids, tartarate, raffinose, melbiose and 6-hydroxynicotinic acid, which are known to be indirectly metabolized in humans by the gut microbiome, do not need to be accounted for in RECON 1. Compounds that had no solutions were mostly inorganic compounds, amino acid derivatives, organic acids (Figure 2D) and also compounds which are not associated with basal metabolism, such as arsenic, lead, ibuprofen and naproxen.

DISCUSSION

In the present study, RECON 1 was used as a discovery tool alongside a library of chemical reactions to computationally propose metabolic routes for human orphan metabolites found in biofluid samples. The key results include (i) the identification and characterization of 343 metabolites, (ii) metabolic hypotheses for multiple orphan metabolites and the biochemically detailed description of five of these, (iii) the experimental validation of one of the computationally proposed hypotheses leading to the characterization of *C9orf103* as a human gluconokinase, and (iv) a large proportion of the identified metabolites could not be accounted for using our approach and may focus future gap-filling efforts. The results of the present study demonstrate how gap filling can focus the expansion of metabolic networks and

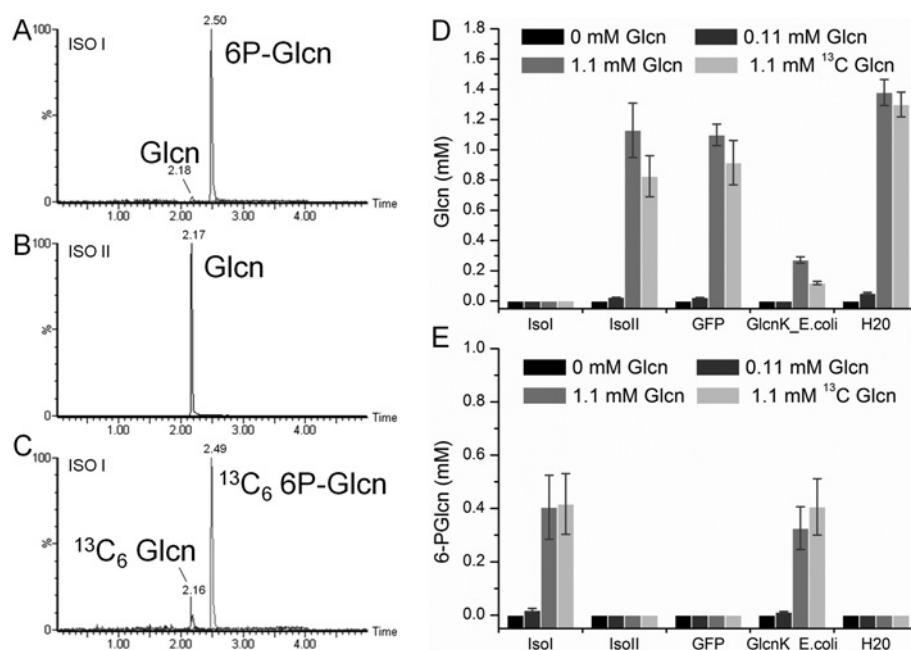


Figure 5 LC-MS analysis of gluconate and 6-phosphogluconate composition in HeLa lysates expressing C9orf103

Overlapped extracted ion chromatograms of gluconate (GlcN) and 6-phosphogluconate (6P-GlcN) from gluconokinase assays containing overexpressed C9orf103 isoform I (A and C) or isoform II (B) in the presence of 1 mM Glcn (A and B) or 1 mM [¹³C₆]gluconate (C). Quantification of the reaction components using ISs showed that gluconate was consumed and 6-phosphogluconate was produced only in the presence of overexpressed isoform I and of the positive control. Similarly, when [¹³C₆]gluconate was used as the substrate, [¹³C₆]6-phosphogluconate was only detected in lysates overexpressing isoform I and the positive control (D and E).

facilitate metabolite and gene annotation through a metabolite-centred approach.

Our approach assumed that RECON 1 correctly captured current knowledge and, therefore, that all metabolites identified were orphan metabolites. Clearly this was not the case as more than one-fifth of the metabolites could be added directly to RECON 1, thus not being true orphan metabolites. This result demonstrates that the metabolic network reconstruction process is iterative requiring periodical refinement so that the network model sustains its biological accuracy. With respect to the accepted solutions, the gap-filling algorithm therefore fulfils its role as a network gap-filling tool [9].

We generated biologically plausible hypotheses for almost 50 orphan metabolites. The notion that almost half of these have proposed anabolic and catabolic routes infers that their detection in biofluids may be associated with abnormal metabolic states, as is the case for cysteine S-sulfate (OMIM# 272300). The computed hypotheses (Figure 3) conforms with the previously proposed metabolism of cysteine S-sulfate from L-cysteine in the pathogenesis of sulfite oxidase deficiency [38] albeit in biochemical detail. The second computed solution, through L-serine, is also plausible on the account of the catalytic promiscuity of the PLP-dependent enzymes involved towards sulfur-containing amino acids [27,28]. PLP-dependent enzymes have broad catalytic activity on account of the diverse reactivity of the PLP Schiff base intermediate [41]. For modelling of disease-associated metabolic phenotypes, such as inborn errors of metabolism [42], these types of discoveries are valuable. The remainder of the hypotheses involved type II and III solutions. Interestingly, during the preparation of the present paper the computed candidate reaction for menthol was confirmed to be catalysed by CYP2A6 in human liver microsomes by Nakanishi and co-workers [43]. Multiple approaches have been devised to derive gene or protein function [44]. With the recent surge in non-

targeted metabolite profiling, the number of metabolites whose biological roles are not defined will surely increase [45]. The three solution types highlight how network gap filling directs metabolic hypotheses of orphan metabolites through contextualization of biological data. The linkage between metabolites and genes can therefore be successfully done in a semi-automated manner using a high-quality genome-scale metabolic reconstruction.

Our network gap-filling approach suffers a drawback in that it is not fully automated. Since the universal reaction database (KEGG) used captures biochemical reactions identified in all domains of life, manual ranking and curation of candidate solutions was required before addition of computed reaction solutions to the network, similar to the laborious bottom-up network reconstruction procedure [46]. The prioritization of predicted pathways between metabolites has been a focus point in synthetic pathway engineering, where candidate reactions between metabolites have been ranked according to thermodynamic feasibility, pathway length, chemical similarity and organism specificity [14]. Automated ranking and filtering of the gap-filling output would be valuable when faced with multiple candidate solutions.

We confirmed the computationally proposed hypotheses for gluconate consumption and showed that C9orf103 encodes a functional gluconokinase. The proposed metabolism of gluconate is identical with what is observed in lower organisms where it has been extensively characterized [38]. The molecular details of gluconate metabolism in mammals are less well understood, although early biochemical experiments suggest a similar degradation route. Using ¹⁴C-labelled gluconate, Stetten and Topper [47] demonstrated that the carboxyl carbon of gluconate yields CO₂ abundantly in intact rats. Because reduction of gluconate directly to glucose had been shown to not take place [47], the finding was consistent with omission of CO₂ from 6-phosphogluconate by 6PDGH (EC 1.1.1.44) [47,49].

C-6 of gluconate was, however, observed to contribute to fatty acid synthesis, presumably following integration into glycolysis through the non-oxidative branch of the PPP [50]. These studies of gluconate metabolism took part in pioneering investigations into oxidative glucose metabolism that aided in the discovery of the PPP [51]. An active gluconokinase was later characterized and purified from rat and pig kidney [37].

We have demonstrated for the first time that gluconokinase activity is encoded within the human genome. *C9orf103* was initially identified during cDNA library construction [52] and later shown to be located within the commonly deleted region of del(9q) acute myeloid leukaemia [53]. These studies did not report on the possible function of the gene. *C9orf103* was, however, electronically annotated prior to the present study as a putative gluconokinase. A similarity in protein sequence does not always transfer to enzyme function [54]. Isoform II was, for example, shown in the present study to be catalytically inactive. Also, an 18 amino acid insertion within both isoforms of *C9orf103* as compared with *E. coli* GlnK (Supplementary Figure S3) does not appear to abolish the gluconokinase activity of isoform I. The location of this peptide insertion within the amino acid sequence of isoform I hints at a structural difference between isoform I and *E. coli* GntK previously observed between *E. coli* GlnK and various NMP kinases that are known to catalyse the phosphorylation of a broad range of substrates [55]. Although this might suggest additional functions for the *C9orf103* protein product, the gluconokinase purified from pig kidney was found to exhibit strict specificity for D-gluconate [37]. Ultimately, the result of the present study provides a missing genetic link between mammalian gluconokinase activity and our computationally proposed gluconate catabolic route previously proposed through biochemical pulse–chase experiments in rats.

An anabolic route for gluconate through the oxidation of glucose was also computed. Although both activities for this conversion are found in humans, cell biology experiments will be required to confirm whether these act synergistically to generate gluconate *in vivo*. There is no apparent gain of oxidation of glucose through gluconate compared with glucose 6-phosphate. Both presumably yield the same pentose phosphates and two molecules of NADPH. This result does, however, not rule out a metabolic contribution of gluconate as a nutrient with a catabolic role via the PPP following, for example, cellular uptake or as a degradation product of intracellular carbohydrates. Similar experiments would also shed light on the effect of down-regulation of *C9orf103* associated with del(9q) acute myeloid leukaemia [53].

We identified more than 200 orphan metabolites for which no solution could be obtained as their metabolism was beyond the scope of RECON 1. These metabolites point to reaction pathways that should be incorporated into RECON 1 to account for metabolic phenotypes associated with, for example, phase I and II xenobiotic detoxification routes, the contribution of the human microbiome, signalling cascades, and transcription and translation. Integration of such non-metabolic reconstructions would help address orphan metabolites, whose metabolic fate lies beyond the scope of RECON 1 and subsequently better define a list of true human orphan metabolites.

The finding that RECON 1 can be used to generate computational hypotheses of human biological functions is intriguing as it implies that metabolic models can be used to prioritize the discovery of human metabolic functions albeit by focusing on metabolites rather than genes. This prioritization stems from filling gaps in knowledge of human metabolism, thereby systematically expanding metabolic knowledge through a bottom-up approach. This approach does, however, not unearth functions that are globally uncharacterized as the method relies on

annotated functions and genes. The manner in which incomplete biological data is put into context with current metabolic knowledge is therefore the driving force behind this approach of biological target identification and subsequent functional validation. The result is an expanded metabolic reconstruction, which better captures and defines the known and unknowns of human metabolism.

AUTHOR CONTRIBUTION

Óttar Rolfsson designed the study, performed the experiments, analysed the data and wrote the paper. Giuseppe Paglia and Manuela Magnúsdóttir performed MS analysis. Bernhard Palsson contributed to experimental design and writing of the paper. Ines Thiele designed the study, performed the experiments and contributed to the writing of the paper.

FUNDING

This work was supported by the European Research Council [grant number 232816].

REFERENCES

- Furnham, N., Garavelli, J. S., Apweiler, R. and Thornton, J. M. (2009) Missing in action: enzyme functional annotations in biological databases. *Nat. Chem. Biol.* **5**, 521–525
- Reed, J. L., Famili, I., Thiele, I. and Palsson, B. O. (2006) Towards multidimensional genome annotation. *Nat. Rev. Genet.* **7**, 130–141
- Oberhardt, M. A., Palsson, B. O. and Papin, J. A. (2009) Applications of genome-scale metabolic reconstructions. *Mol. Syst. Biol.* **5**, 320
- Karp, P. D. (2004) Call for an enzyme genomics initiative. *Genome Biol.* **5**, 401
- Lespinet, O. and Labeledan, B. (2005) Orphan enzymes? *Science* **307**, 42
- Chen, L. and Vitkup, D. (2007) Distribution of orphan metabolic activities. *Trends Biotechnol.* **25**, 343–348
- Hanson, A. D., Pribat, A., Waller, J. C. and de Crécy-Lagard, V. (2010) 'Unknown' proteins and 'orphan' enzymes: the missing half of the engineering parts list: and how to find it. *Biochem. J.* **425**, 1–11
- Rolfsson, O., Palsson, B. Ø. and Thiele, I. (2011) The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions. *BMC Syst. Biol.* **5**, 155
- Reed, J. L., Patel, T. R., Chen, K. H., Joyce, A. R., Applebee, M. K., Herring, C. D., Bui, O. T., Knight, E. M., Fong, S. S. and Palsson, B. O. (2006) Systems approach to refining genome annotation. *Proc. Natl. Acad. Sci., U.S.A.* **103**, 17480–17484
- Kharchenko, P., Vitkup, D. and Church, G. M. (2004) Filling gaps in a metabolic network using expression information. *Bioinformatics* **20** (Suppl. 1), 178–185
- Kumar, V. S. and Maranas, C. D. (2009) GrowMatch: an automated method for reconciling *in silico/in vivo* growth predictions. *PLoS Comput. Biol.* **5**, e1000308
- Yamada, T., Waller, A. S., Raes, J., Zelezniak, A., Perchat, N., Perret, A., Salanoubat, M., Patil, K. R., Weissenbach, J. and Bork, P. (2012) Prediction and identification of sequences coding for orphan enzymes using genomic and metagenomic neighbours. *Mol. Syst. Biol.* **8**, 851
- Orth, J. D. and Palsson, B. Ø. (2012) Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions. *BMC Syst. Biol.* **6**, 30
- Medema, M. H., Raaphorst, R. V., Takano, E. and Breitling, R. (2012) Computational tools for the synthetic design of biochemical pathways. *Nat. Rev. Microbiol.* **10**, 191–202
- Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., Srivas, R. and Palsson, B. Ø. (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1777–1782
- Bordbar, A. and Palsson, B. O. (2012) Using the reconstructed genome-scale human metabolic network to study physiology and pathology. *J. Intern. Med.* **271**, 131–141
- Nakahigashi, K., Toya, Y., Ishii, N., Soga, T., Hasegawa, M., Watanabe, H., Takai, Y., Honma, M., Mori, H. and Tomita, M. (2009) Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. *Mol. Syst. Biol.* **5**, 306
- Schellenberger, J., Park, J., Conrad, T. and Palsson, B. (2010) BIGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinf.* **11**, 213–213
- Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., Zielinski, D. C., Bordbar, A., Lewis, N. E., Rahmami, S. et al. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* **6**, 1290–1307

- 20 Goto, S., Okuno, Y., Hattori, M., Nishioka, T. and Kanehisa, M. (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* **30**, 402–404
- 21 Wishart, D. S., Tzur, D., Knox, C., Eisner, R., Guo, A. C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S. et al. (2007) HMDB: the Human Metabolome Database. *Nucleic Acids Res.* **35**, D521–D526
- 22 Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M. et al. (2009) STRING 8: a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* **37**, D412–D416
- 23 Orth, J. D. and Palsson, B. Ø. (2010) Systematizing the generation of missing metabolic knowledge. *Biotechnol. Bioeng.* **107**, 403–412
- 24 Tan, W.-H., Eichler, F. S., Hoda, S., Lee, M. S., Baris, H., Hanley, C. A., Grant, P. E., Krishnamoorthy, K. S. and Shih, V. E. (2005) Isolated sulfite oxidase deficiency: a case report with a novel mutation and review of the literature. *Pediatrics* **116**, 757–766
- 25 Chang, S. H. and Wilken, D. R. (1966) Participation of the unsymmetrical disulfide of coenzyme a and glutathione in an enzymatic sulfhydryl-disulfide interchange. *J. Biol. Chem.* **241**, 4251–4260
- 26 Stricks, W. and Kolthoff, I. M. (1951) Equilibrium constants of the reactions of sulfite with cystine and with dithiodiglycolic acid. *J. Am. Chem. Soc.* **73**, 4569–4574
- 27 Chiku, T., Padovani, D., Zhu, W., Singh, S., Vitvitsky, V. and Banerjee, R. (2009) H2S biogenesis by human cystathionine γ -lyase leads to the novel sulfur metabolites lantionine and homolantionine and is responsive to the grade of hyperhomocysteinemia. *J. Biol. Chem.* **284**, 11601–11612
- 28 Singh, S., Padovani, D., Leslie, R. A., Chiku, T. and Banerjee, R. (2009) Relative contributions of cystathionine β -synthase and γ -cystathionase to H2S biogenesis via alternative trans-sulfuration reactions. *J. Biol. Chem.* **284**, 22457–22466
- 29 Guerini, V., Sau, D., Scaccianoce, E., Rusmini, P., Ciana, P., Maggi, A., Martini, P. G. V., Katzenellenbogen, B. S., Martini, L., Motta, M. and Poletti, A. (2005) The androgen derivative 5 α -androstane-3 β ,17 β -diol inhibits prostate cancer cell migration through activation of the estrogen receptor β subtype. *Cancer Res.* **65**, 5445–5453
- 30 Sharaf, M. A. and Sweet, F. (1982) Dual activity at an enzyme active site: 3 β ,20 α -hydroxysteroid oxidoreductase from fetal blood. *Biochemistry* **21**, 4615–4620
- 31 Matsunaga, T., Endo, S., Maeda, S., Ishikura, S., Tajima, K., Tanaka, N., Nakamura, K. T., Imamura, Y. and Hara, A. (2008) Characterization of human DHRS4: an inducible short-chain dehydrogenase/reductase enzyme with 3 β -hydroxysteroid dehydrogenase activity. *Arch. Biochem. Biophys.* **477**, 339–347
- 32 Austin, C. A., Shephard, E. A., Pike, S. F., Rabin, B. R. and Phillips, I. R. (1988) The effect of terpenoid compounds on cytochrome P-450 levels in rat liver. *Biochem. Pharmacol.* **37**, 2223–2229
- 33 Gelal, A., Jacob, 3rd, P., Yu, L. and Benowitz, N. L. (1999) Disposition kinetics and effects of menthol. *Clin. Pharmacol. Ther.* **66**, 128–135
- 34 Singh, O. V. and Kumar, R. (2007) Biotechnological production of gluconic acid: future implications. *Appl. Microbiol. Biotechnol.* **75**, 713–722
- 35 Senesi, S., Csala, M., Marcolongo, P., Fulceri, R., Mandl, J., Banhegyi, G. and Benedetti, A. (2010) Hexose-6-phosphate dehydrogenase in the endoplasmic reticulum. *Biol. Chem.* **391**, 1–8
- 36 Kondo, Y., Inai, Y., Sato, Y., Handa, S., Kubo, S., Shimokado, K., Goto, S., Nishikimi, M., Maruyama, N. and Ishigami, A. (2006) Senescence marker protein 30 functions as gluconolactonase in L-ascorbic acid biosynthesis, and its knockout mice are prone to scurvy. *Proc. Natl. Acad. Sci., U.S.A.* **103**, 5723–5728
- 37 Leder, I. G. (1957) Hog kidney gluconokinase. *J. Biol. Chem.* **225**, 125–136
- 38 Peekhaus, N. and Conway, T. (1998) What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. *J. Bacteriol.* **180**, 3495–3502
- 39 Holder, G. M., Plummer, J. L. and Ryan, A. J. (1978) The metabolism and excretion of curcumin (1,7-Bis-(4-hydroxy-3-methoxyphenyl)-1,6-heptadiene-3,5-dione) in the rat. *Xenobiotica* **8**, 761–768
- 40 Baba, S., Osakabe, N., Natsume, M., Yasuda, A., Muto, Y., Hiyoshi, K., Takano, H., Yoshikawa, T. and Terao, J. (2005) Absorption, metabolism, degradation and urinary excretion of rosmarinic acid after intake of *Perilla frutescens* extract in humans. *Eur. J. Nutr.* **44**, 1–9
- 41 Toney, M. D. (2011) Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochem. Biophys. Acta* **1814**, 1407–1418
- 42 Sahoo, S., Franzson, L., Jonsson, J. J. and Thiele, I. (2012) A compendium of inborn errors of metabolism mapped onto the human metabolic network. *Mol. Biosyst.* **10**, 2545–2558
- 43 Miyazawa, M., Marumoto, S., Takahashi, T., Nakahashi, H., Haigou, R. and Nakanishi, K. (2011) Metabolism of (+)- and (–)-menthols by CYP2A6 in human liver microsomes. *J. Oleo Sci.* **60**, 127–132
- 44 Eisenthal, R. and Danson, M. J. (2002), *Enzyme Assays: A Practical Approach*. Oxford University Press
- 45 Patti, G. J., Yanes, O. and Siuzdak, G. (2012) Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269
- 46 Thiele, I. and Palsson, B. Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121
- 47 Stetten, M. R. and Topper, Y. J. (1953) Pathways from gluconic acid to glucose *in vivo*. *J. Biol. Chem.* **203**, 653–664
- 48 Reference deleted
- 49 Stetten, M. R. and Stetten, Jr, D. (1950) The metabolism of gluconic acid. *J. Biol. Chem.* **187**, 241–252
- 50 Bloom, B. and Stetten, Jr, D. (1955) The fraction of glucose catabolized via the glycolytic pathway. *J. Biol. Chem.* **212**, 555–563
- 51 Horecker, B. L. (2002) The pentose phosphate pathway. *J. Biol. Chem.* **277**, 47965–47971
- 52 Strausberg, R. L., Feingold, E. A., Grouse, L. H., Derge, J. G., Klausner, R. D., Collins, F. S., Wagner, L., Shenmen, C. M., Schuler, G. D., Altschul, S. F. et al. (2002) Generation and initial analysis of more than 15000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16899–16903
- 53 Sweetser, D. A., Peniket, A. J., Haaland, C., Blomberg, A. A., Zhang, Y., Zaidi, S. T., Dayyani, F., Zhao, Z., Heerema, N. A., Boulwood, J. et al. (2005) Delineation of the minimal commonly deleted segment and identification of candidate tumor-suppressor genes in del(9q) acute myeloid leukemia. *Genes Chromosomes Cancer* **44**, 279–291
- 54 Rost, B. (2002) Enzyme function less conserved than anticipated. *J. Mol. Biol.* **318**, 595–608
- 55 Kraft, L., Sprenger, G. A. and Lindqvist, Y. (2002) Conformational changes during the catalytic cycle of gluconate kinase as revealed by X-ray crystallography. *J. Mol. Biol.* **318**, 1057–1069

SUPPLEMENTARY ONLINE DATA

Inferring the metabolism of human orphan metabolites from their metabolic network context affirms human gluconokinase activity

Óttar ROLFSSON*[†], Giuseppe PAGLIA*, Manuela MAGNUSDÓTTIR*, Bernhard Ø. PALSSON* and Ines THIELE*[‡]¹

*Center for Systems Biology, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland, [†]University of Iceland Biomedical Center, Laeknagardur, 101 Reykjavik, Iceland, and [‡]Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, 101 Reykjavik, Iceland

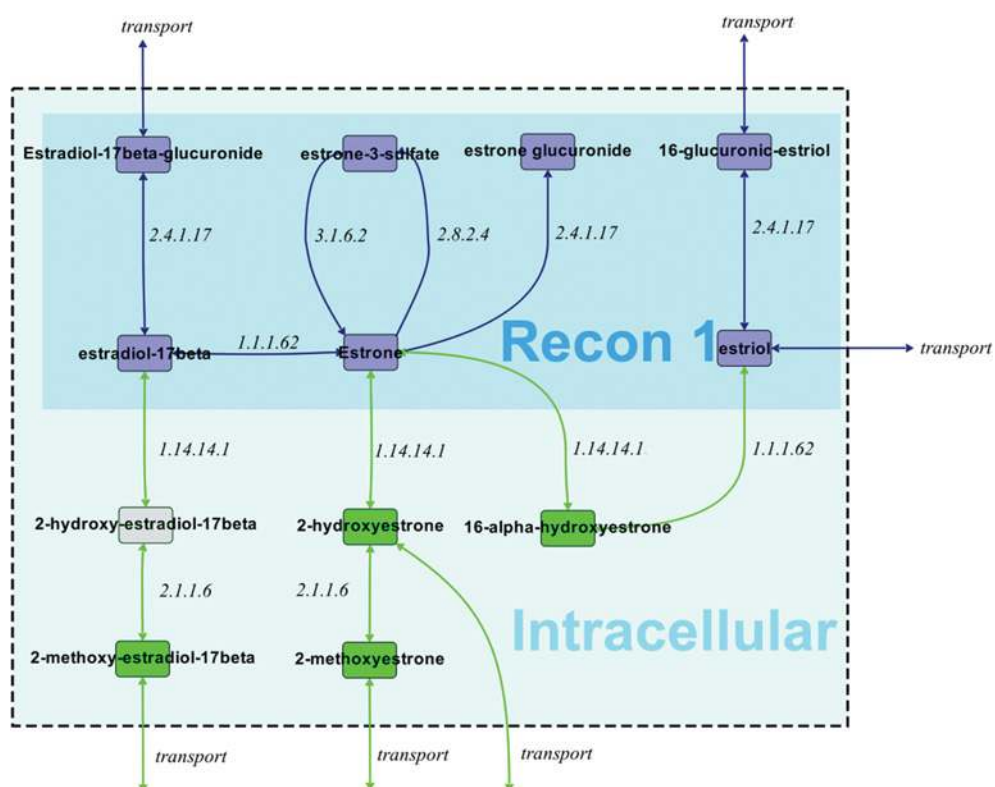


Figure S1 Four metabolites derived from oestrone that have been identified in human biofluids and that were incorporated into RECON 1 using the gap-filling algorithm

Selected reactions and metabolites in C_{18} steroid metabolism, as they appear in a decompartmentalized model of RECON 1, are shown in blue. The gap-filling solutions for 2-methoxy-estradiol-17 β , 2-methoxyestrone, 2-hydroxyestrone and 16- α -hydroxyestrone are shown with green arrows. 16- α -Hydroxyestrone has a type I solution, whereas the remainder have type II solutions. Note that the type II solution was also a gap-filling solution in RECON 1 in that it connected oestriol to the network, which would otherwise be caught in a loop and did not carry flux.

¹ To whom correspondence should be addressed (email ines.thiele@gmail.com).

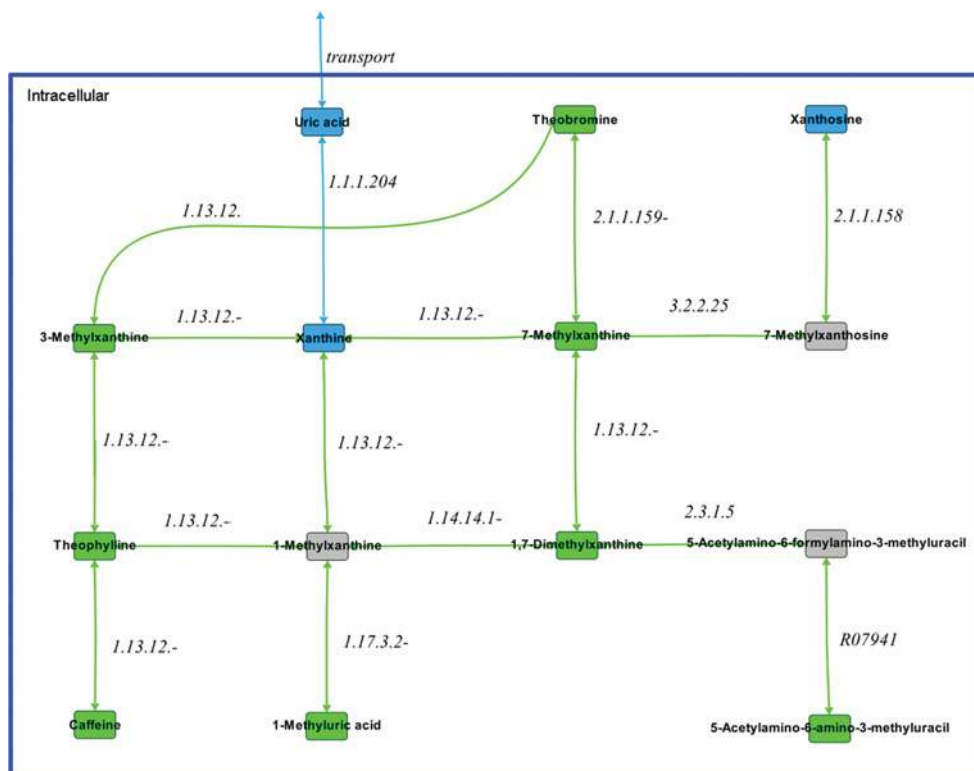


Figure S2 Reconstruction of caffeine and theobromine metabolism

Manual curation is required in order to generate biologically plausible metabolic pathways for highly interconnected metabolites. An example of where automated gap filling and hypotheses generation fell short. The primary metabolic degradation routes of caffeine, theobromine and its derivatives were through xanthine, which was then excreted as uric acid, and through the excretion of 5-acetylamino-6-amino-3-methyluracil. The gap-filling algorithm is capable of proposing multiple degradation routes, the manual curation of which are required in order to accurately reflect caffeine metabolism.

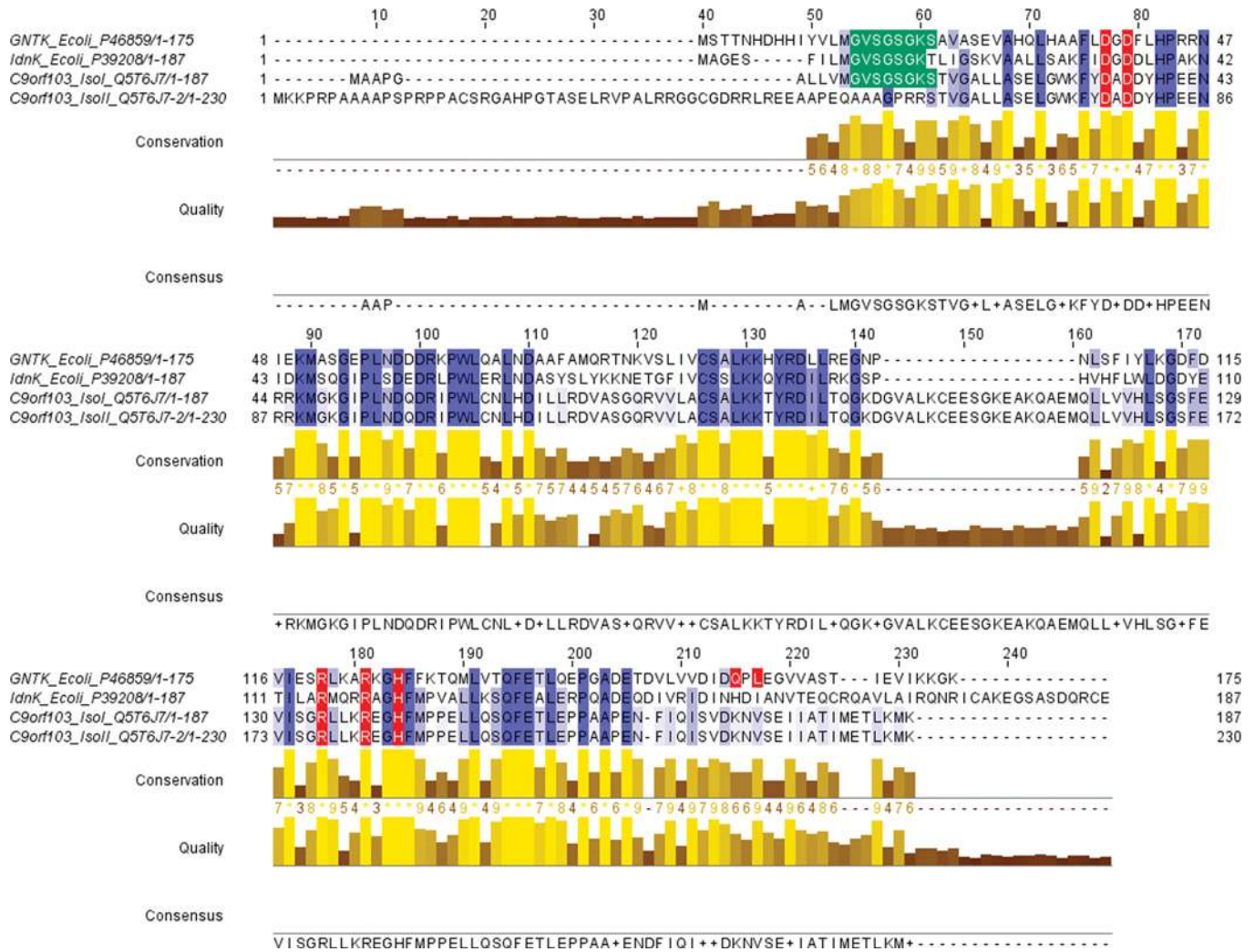


Figure S3 Sequence alignment of isoform I and II of *C9orf103* to the amino acid sequences of the gluconokinases of *E. coli*

Gluconokinase is encoded in *E. coli* by two genes, which encode a thermoresistant (*GNTK*) and thermosensitive (*IdnK*) gluconokinase. *C9orf103* has 40% and 39% sequence identity to these genes respectively (BLAST data not shown). Residues that have been shown to be important for gluconic acid binding in *E. coli* *GNTK* [1] are highlighted in red. The phosphate-binding domain P-loop is highlighted in green and is absent from *C9orf103* isoform II. Colouring is according to percentage identity. The alignment and Figure were generated with Jalview.

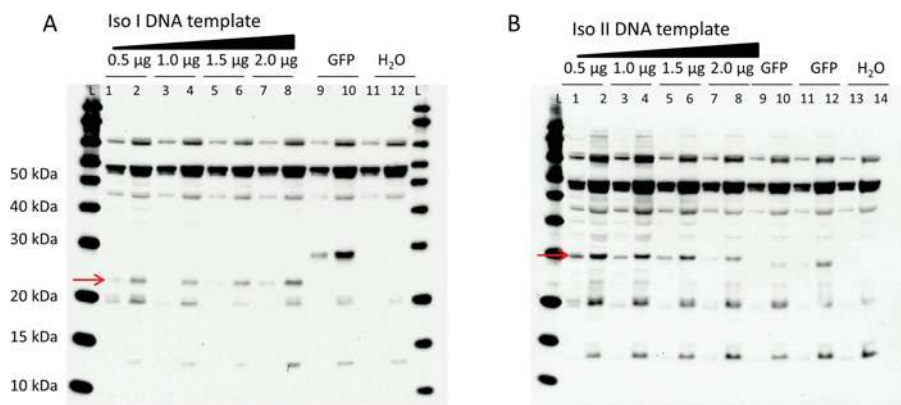


Figure S4 Western blots of coupled *in vitro* transcription/translation reactions of C9orf103 isoform I (pT7CFE1-GlncIIsol) (A) and C9orf103 isoform II (pT7CFE1-GlncIIsol) (B) using 0.5–2.0 μg of linearized plasmid template (lanes 1–8)

GFP template DNA (0.5 μg) was used in positive-control reactions. Water was used as a negative control. In total 1 μl and 3 μl from each 25 μl reaction mixture were loaded and run on 10 % Bis/Tris NuPAGE gels for 50 min. Using a supersignal West HisProbe Kit, both isoforms could be detected irrespective of DNA template amount used following transfer on to a nitrocellulose membrane. These are highlighted with red arrows. A BenchMark His-tagged Protein Standard (Invitrogen) was used to assess molecular mass using Image Lab software (Bio-Rad Laboratories). C9orf103 isoform I was observed to have a molecular mass of 22.4 ± 0.17 kDa (calculated molecular mass 22.328 kDa). C9orf103 isoform II was observed to have a molecular mass of 29.1 ± 0.22 kDa (calculated molecular mass 27.047 kDa).

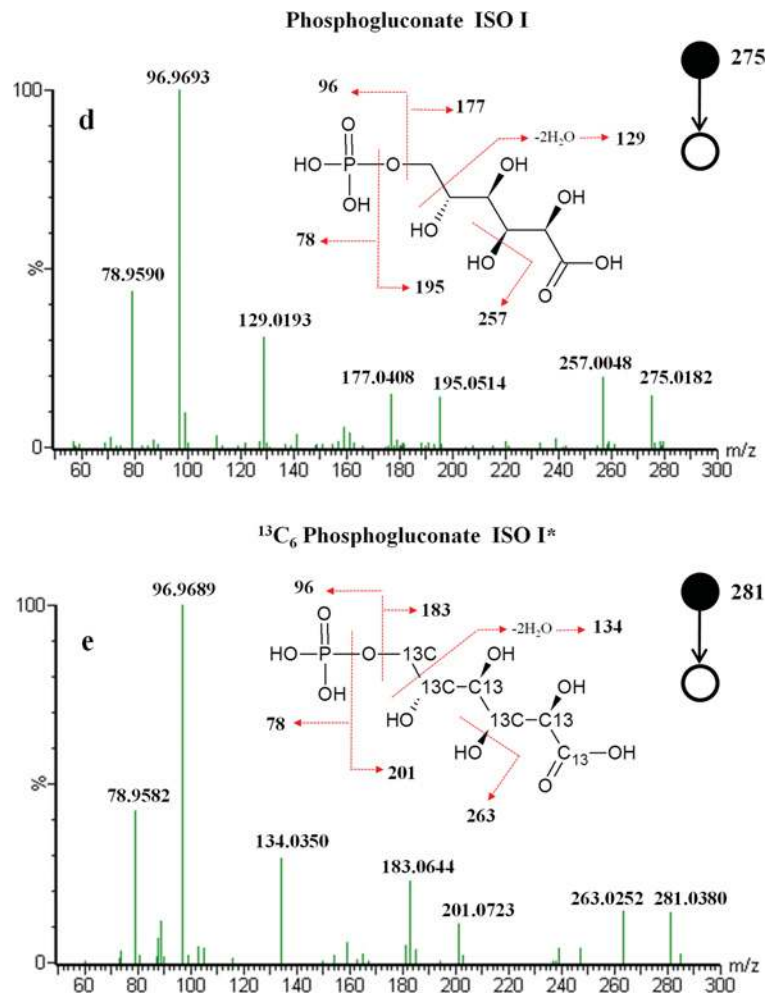


Figure S5 MS/MS (tandem MS) spectra of gluconate and 6-phosphogluconate

(A) MS/MS spectrum of 6-phosphogluconate during ISO I assay. (B) MS/MS spectrum of [¹³C₆]-6-phosphogluconate during ISO I* assay using [¹³C₆]-gluconate.

REFERENCE

- Kraft, L., Sprenger, G. A. and Lindqvist, Y. (2002) Conformational changes during the catalytic cycle of gluconate kinase as revealed by X-ray crystallography. *J. Mol. Biol.* **318**, 1057–1069

Received 18 June 2012/13 September 2012; accepted 15 October 2012
Published as BJ Immediate Publication 15 October 2012, doi:10.1042/BJ20120980