

RESEARCH ARTICLE

Open Access

The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions

Ottar Rolfsson¹, Bernhard Ø Palsson¹ and Ines Thiele^{1,2*}

Abstract

Background: Metabolic network reconstructions formalize our knowledge of metabolism. Gaps in these networks pinpoint regions of metabolism where biological components and functions are “missing.” At the same time, a major challenge in the post genomic era involves characterisation of missing biological components to complete genome annotation.

Results: We used the human metabolic network reconstruction RECON 1 and established constraint-based modelling tools to uncover novel functions associated with human metabolism. Flux variability analysis identified 175 gaps in RECON 1 in the form of blocked reactions. These gaps were unevenly distributed within metabolic pathways but primarily found in the cytosol and often caused by compounds whose metabolic fate, rather than production, is unknown. Using a published algorithm, we computed gap-filling solutions comprised of non-organism specific metabolic reactions capable of bridging the identified gaps. These candidate solutions were found to be dependent upon the reaction environment of the blocked reaction. Importantly, we showed that automatically generated solutions could produce biologically realistic hypotheses of novel human metabolic reactions such as of the fate of iduronic acid following glycan degradation and of N-acetylglutamate in amino acid metabolism.

Conclusions: The results demonstrate how metabolic models can be utilised to direct hypotheses of novel metabolic functions in human metabolism; a process that we find is heavily reliant upon manual curation and biochemical insight. The effectiveness of a systems approach for novel biochemical pathway discovery in mammals is demonstrated and steps required to tailor future gap filling algorithms to mammalian metabolic networks are proposed.

Background

An *in silico* model of a genome-scale metabolic network reconstruction is based upon a biochemically, genetically and genomically (BiGG) structured knowledge base [1,2]. It is subject to research that, in many cases, entails predicting an organism’s phenotypic response to gene deletions and/or environmental perturbations *in silico*. These properties have resulted in widespread applications of metabolic models in microbial bioengineering, contextualisation of high-throughput data, and biochemical pathway discovery [2,3]. While the number of

microbial genome-scale metabolic networks has increased exponentially over the past 10 years [4], fewer have been reconstructed for higher eukaryotes as their inherent complexity results in larger and more complex models which are harder to experimentally validate [5,6]. To date, only mouse [7-9], bovine [10], and human [11-13] genome-scale metabolic networks have been reconstructed. These latter ones have successfully been applied to systems driven eukaryotic metabolic research. For example, the human genome-scale metabolic network RECON 1 has been used to reveal transcriptional regulatory signatures of type 2 diabetes [14], to create tissue-specific models [15], to predict drug-off target effects [16], and to simulate cell specific metabolic changes upon pathogen infection [17]. Their potential to

* Correspondence: ithiele@hi.is

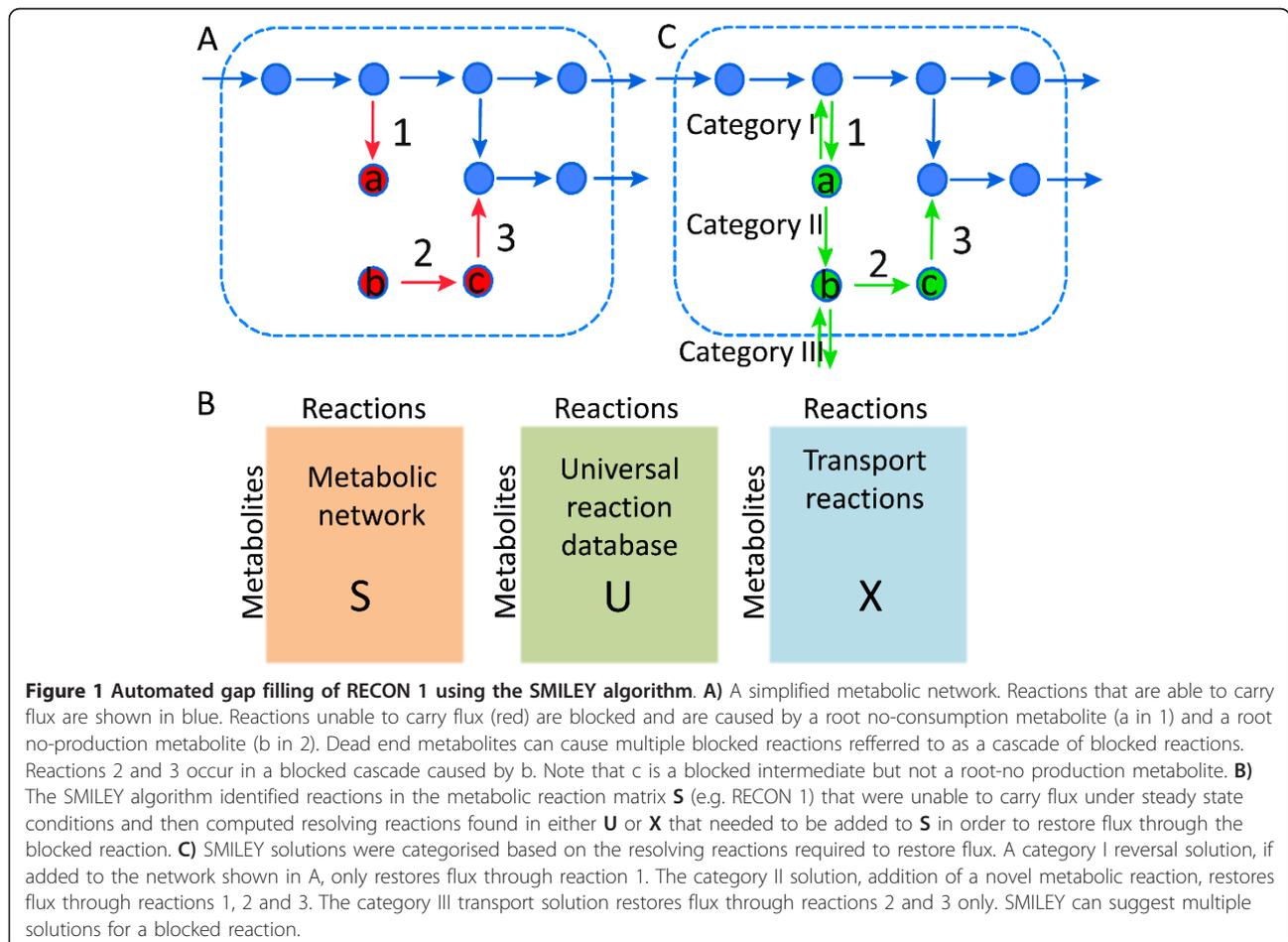
¹Center for Systems Biology, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland

Full list of author information is available at the end of the article

discover novel metabolic functions has however not been demonstrated.

RECON 1 accounts for 1496 ORFs, 2004 proteins, 2766 metabolites, and 3311 metabolic reactions. It represents one of the most comprehensive biological network reconstructions published to date [11]. RECON 1 is bipartite in nature. First, it is a BiGG knowledge base containing information on the components of human metabolism, mined by extensive literature review [4] and accessible online at <http://bigg.ucsd.edu>[18]. Second, it is a mathematical model in which reactions and their metabolites are presented in matrix format using metabolite stoichiometries. RECON 1 is therefore also a stoichiometric matrix amenable to constraint-based analysis using linear programming as the reaction stoichiometries impose constraints on the flow of metabolites through the network. As an additional constraint, a quasi steady state is assumed, meaning that the total amount of any metabolite being produced must equal the total amount of that metabolite being consumed. Using constraint-based analysis, the flow of metabolites through the network reactions can be calculated. In

addition, and perhaps most importantly, reactions can be identified which cannot carry flux due to metabolites being either only produced or consumed in the network [4,19]. These latter reactions and metabolites represent gaps in the metabolic network and are referred to as blocked reactions and dead-end metabolites, respectively. Dead-end metabolites can result in multiple blocked reactions, referred to as blocked reaction cascades. Consequently they have been further characterised as root-no production or root-no consumption metabolites to distinguish them from other blocked metabolites within the blocked cascade [20] (Figure 1a). In this manner RECON 1 can be utilised to pinpoint parts of metabolic pathways where knowledge is incomplete. A recent investigation highlighted that novel biochemical reactions can still be uncovered in the central metabolism of *E. coli* by employing a systems approach [21]. Considering that 30-50% of all known enzyme activities are not associated with genes [22-24] and over 50% of all genes in higher organisms are not associated with protein function [25], it appears that many human biological processes have yet to be discovered. Enzymes



with no known genes (e.g. orphan enzymes) and genes of unknown function underlie metabolic network gaps that manifest themselves as blocked reactions and dead-end metabolites in RECON 1. As the function and/or substrate specificity of many known human enzymes has not been fully elucidated, these similarly serve as candidates for filling network gaps.

Multiple computational algorithms have been devised to address gaps in metabolic network models in an automated manner [26]. These methods utilise metabolic network analysis, such as flux variability analysis (FVA) [27], in order to identify network gaps, alongside comparative genomics/metabolomics to suggest candidate metabolic reactions capable of restoring flux through the blocked reaction and/or dead-end metabolite. Experimental data, such as organism growth profiles [28,29] or metabolic flux data [30] can be integrated with metabolic models in order to highlight model gaps, although this is not always a requirement [26]. One such algorithm is the SMILEY algorithm, which has been successfully used to uncover novel metabolic functions required to explain discrepancies between the observed and model predicted growth phenotypes of *E. coli* on various substrates [28].

Because of the difference between humans and microorganisms in terms of the number of related organisms for which biological, biochemical, and genetic information exists, it was not known whether an automated approach to gap-filling of RECON 1 could yield biologically plausible hypotheses. Here we investigated the potential of RECON 1 for discovery of novel reactions involved in human metabolism. We used FVA to identify dead-end metabolites and blocked reactions in RECON 1 and SMILEY to propose reactions capable of restoring flux through the identified dead-end metabolites. We then characterised the metabolic pathway distribution of identified gaps and their solution types in order to get an idea of how gaps are distributed within RECON 1 and in what manner these gaps are resolved by SMILEY. Finally we validated the automatically generated metabolic reaction hypotheses by manual literature review in order to assess the biological relevance of the proposed solutions with the goal of identifying suitable experimental targets.

Results

In this study we first identified blocked reactions and dead-end metabolites occurring in RECON 1 using FVA [27]. We then employed SMILEY [28] to compute reactions that could be added to the RECON 1 (**S**) from universal reaction databases (**U**, **X**) to enable flux through a blocked reaction (Figure 1b). The matrix **U** was compiled from an extensive list of known metabolic reactions obtained from the KEGG database [31] while

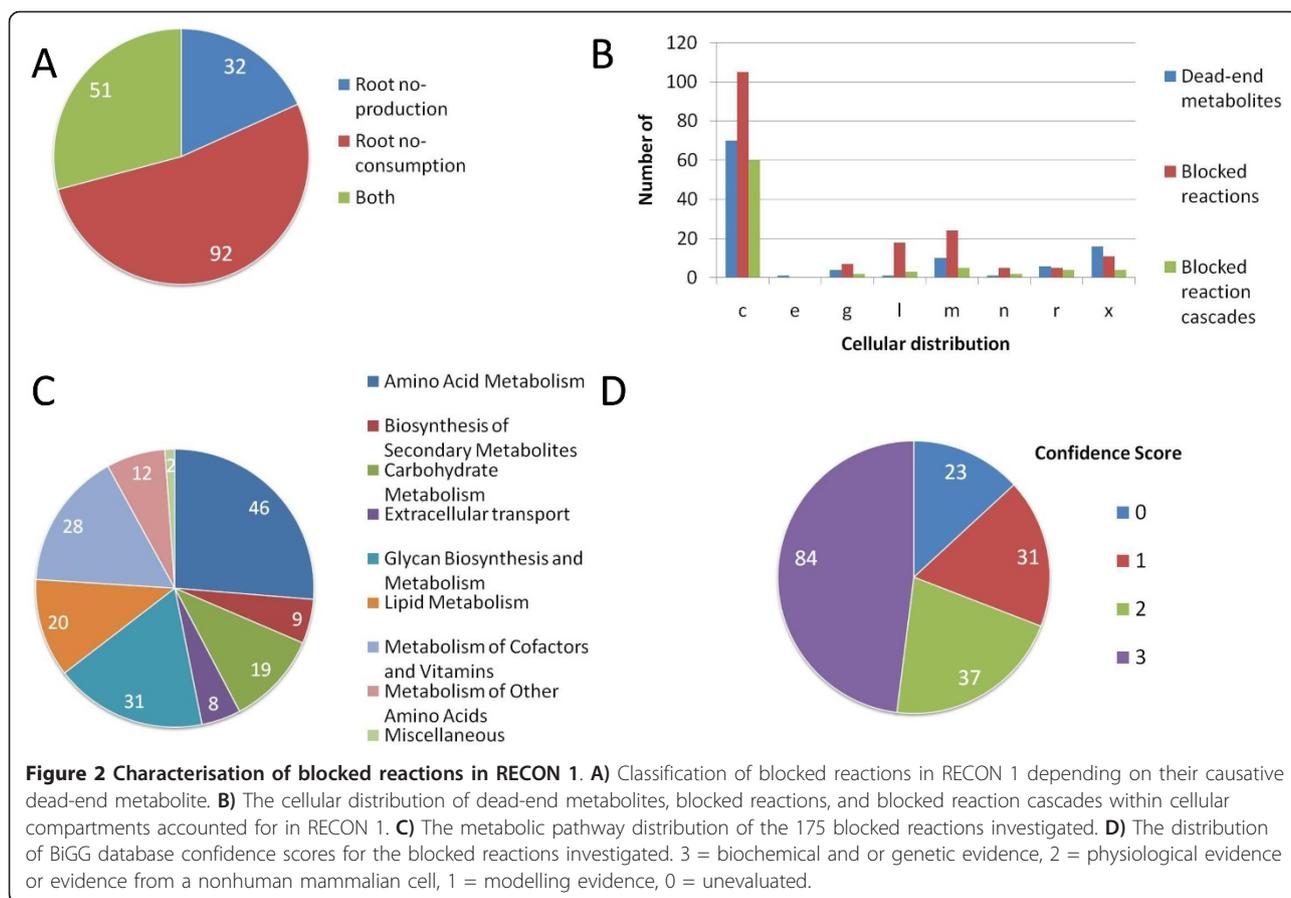
the matrix **X** contained transport reactions in and out of the system for every metabolite contained within **S** and **U**. SMILEY therefore proposed what reactions, irrespective of organism, needed to be added to RECON 1 in order to fill a network gap. If none were identified, SMILEY suggested transport of the dead-end metabolite into or out of the cell. Up to twenty solutions were suggested for each knowledge gap and each solution could be composed of multiple resolving reactions. Note that the number of solutions returned by SMILEY is user defined. Inspired by Satish Kumar et al. [20] we split the computed solutions into three categories based upon whether the complete solution involved a reversal of the directionality of the blocked reaction, addition of novel reaction(s), or addition of a transport reaction, defined as category I-III type solutions respectively (Figure 1c). The next four sections deal with the characterisation and metabolic pathway distribution of the gaps that we addressed in RECON 1 while the remainder of the results chapter reports analysis of SMILEY solutions and specific case studies.

Gap analysis basis

We identified 175 blocked reactions in 80 reaction cascades that were caused by 109 dead-end metabolites found in RECON 1. These numbers corresponds to 5% and 4% of the total number of reactions and metabolites accounted for in RECON 1, respectively. We observed that over half of the blocked reactions were caused by root no-consumption metabolites while roughly a quarter were due to root no-production metabolites with the remainder caused by both types of dead-end metabolites (Figure 2a). Reactions that have both types of dead-end metabolites represent metabolic reactions which are entirely uncoupled from the metabolic network, and whose effect on global metabolism can therefore not be described accurately by the metabolic model.

Sub-cellular distribution of gaps

The sub-cellular distribution of the dead-end metabolites and blocked reactions showed that the majority were found in the cytosol with the remainder distributed within the various cellular compartments, most notably in the lysosome, mitochondria, and peroxisome (Figure 2b). This observation agrees with the distribution of dead-end metabolites found in the eukaryotic metabolic reconstruction iND750 for *Saccharomyces cerevisiae* [20,32]. Because some of the dead-end metabolites were responsible for multiple blocked reactions, which themselves could be associated with both types of dead-end metabolite, there was not a direct relationship between the number of dead-end metabolites and the number of blocked reactions within cellular compartments. In addition, blocked reactions can take part in metabolic



subsystems/reaction cascades that span more than one cellular compartment. For example, in the mitochondria we found 24 blocked reactions participating in six distinct reaction cascades and ten dead-end metabolites. However, only seven of the dead-ends were responsible for the 24 blocked reactions while the remaining three metabolites caused blocked reactions in the cytosol. Gap distributions can be better understood by looking at the distribution of blocked reaction cascades caused by the dead-end metabolites. These data suggest that the number of knowledge gaps in cellular compartments other than the cytosol is around two to five knowledge gaps per compartment (Figure 2).

Metabolic pathway distribution of gaps

The metabolic pathway distribution of the blocked reactions is shown in Figure 2C and reflects their sub-cellular distribution. All of the metabolic pathways shown had blocked reactions in the cytosol. Most of the blocked reactions were observed in amino acid metabolism with the remainder distributed within the metabolic pathways shown. Of those involved in amino acid metabolism, nearly half were part of multiple reaction cascades in tryptophan metabolism (Additional file 1).

These reactions were blocked due to root no-consumption heterocyclic derivatives of tryptophan, such as anthranilate, kynurenic acid, 5-methoxy-indole acetate, and more, suggesting missing information on the metabolic fate of these compounds. Apart from knowledge gaps in tryptophan metabolism, blocked reactions were also identified in metabolic subsystems associated with eleven of the twenty most common amino acids. These gaps were due to single blocked reactions or reaction cascades implying that current information concerning the metabolism of these amino acids is fairly complete.

Multiple blocked reactions were also observed in glycan biosynthesis, the metabolism of cofactors/vitamins and lipid metabolism. As opposed to those observed in amino acid metabolism, these blocked reactions were involved in relatively few reaction cascades. Many of the blocked reactions in glycan metabolism, for example, were part of two reaction cascades involved in the degradation of heparan sulfate and dermatan sulfate in the lysosome, respectively. The root no-consumption metabolite glucose-1, 3-mannose and derivatives thereof were also the cause of several blocked reactions in glycan biosynthesis. Many blocked reactions involved in the metabolism of cofactors/vitamins took place in the

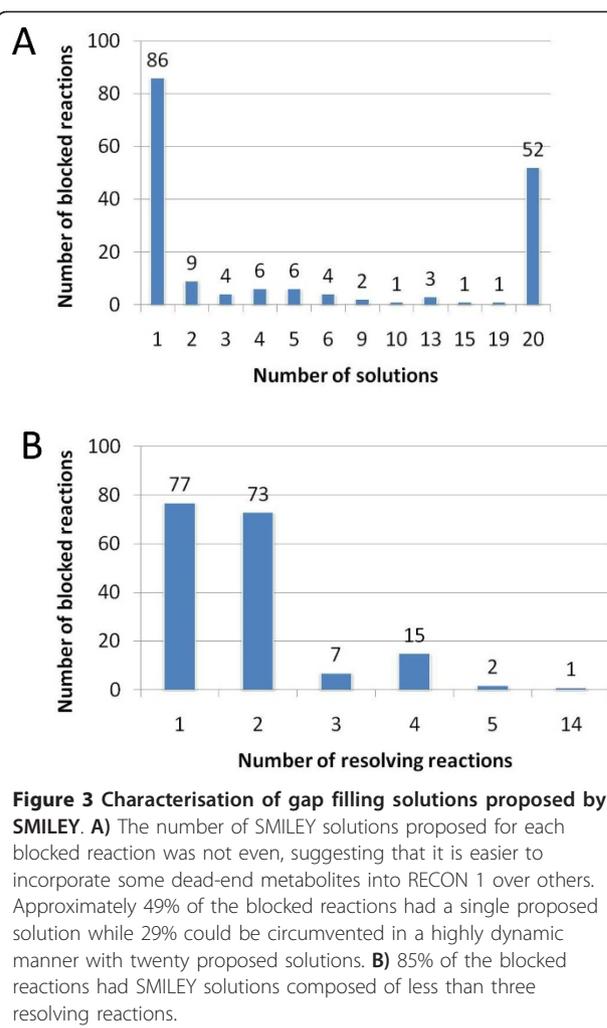
mitochondria due to just two root no-consumption metabolites causing gaps in the biosynthesis of ubiquinone and vitamin D. Similarly, multiple reactions in lipid metabolism were blocked in glycerophospholipid biosynthesis in the cytosol due to the root no-consumption metabolite plasmalogen in the cytosol. In total, 32% of the knowledge gaps investigated were associated with two or more reactions. Finding a solution to the dead-end metabolites in these reaction cascades could therefore result in the connection of multiple blocked reactions back into the metabolic network. The majority of blocked reactions were not, however, part of reaction cascades (Additional file 2).

Knowledge status of reactions causing gaps

We investigated whether the gaps addressed in this study were correlated with a lack of information available for each blocked reaction by assessing their confidence scores. Each reaction in RECON 1 has an assigned confidence score that allows the experimental evidence underlying the reaction to be quickly assessed [11]. Figure 2D shows the distribution of confidence scores of the reactions investigated. Roughly two thirds of the blocked reactions were supported by biochemical or physiological evidence, which is similar to what is observed for all reactions contained in RECON 1 [11]. This distribution implies that the addressed knowledge gaps are not simply due to ill-defined metabolic reactions. Rather, the fate of the participating metabolites within the metabolic network or how they contribute to human metabolism is not known. In order to suggest plausible hypotheses of how this might take place, we investigated whether the blocked reactions could be circumvented or connected back into the RECON 1 using reactions found in the KEGG database [31] in an automated manner by running the SMILEY algorithm [28].

Solutions to a blocked reaction are dependent upon the robustness of its metabolic network

The SMILEY algorithm suggested up to twenty solutions for each blocked reaction. Each solution fell into one of the three category types outlined in Figure 1. We first investigated how trivial it was to bypass the blocked reactions using reactions from the KEGG database (U matrix). Nearly half of the blocked reactions had more than one solution indicating that flux could be restored through these reactions in multiple ways (Figure 3a). Underlying this is that the more robust the non-organism specific reaction environment of the blocked reaction is, the more likely that the blocked reaction will have multiple solutions as there will be more known reactions capable of acting upon the blocked reaction components. For example, a blocked reaction caused by a common, well-characterized metabolite that is a component of



multiple metabolic reactions will have multiple SMILEY solutions while the opposite is true for blocked reactions caused by less well characterized metabolites. The number of solutions obtained for a blocked reaction was independent of the metabolic pathway (Additional file 3). Blocked reactions in RECON 1 are therefore not caused only by dead-end metabolites whose global metabolic role is unknown, rather metabolites and reactions can be well characterized metabolic components of other organisms while their fate is not known in humans. The result complements our earlier observation that knowledge gaps are found in all metabolic pathways, independent of the confidence scores of the reactions making up the particular pathway.

SMILEY solutions involve few resolving reactions

Inspection of the SMILEY output made it clear that assessing each of the 1335 SMILEY solutions represented a time consuming task. Our focus was to assess whether SMILEY could generate biologically relevant

hypotheses of missing reactions in human metabolism rather than produce a detailed list of missing enzyme functionalities. We therefore decided to focus on the SMILEY solutions containing the least number of resolving reactions for each of the 175 blocked reactions hereafter referred to as S1 solutions.

Analysis of S1 solutions showed that flux could be restored through all the blocked metabolites by incorporation of relatively few resolving reactions. Interestingly, 85% of the blocked reactions were circumvented with just one or two resolving reactions (Figure 3b). We observed that this was due to the category of the SMILEY solutions (Figure 4). 74% of the blocked reactions had a category I or III solution, which in the majority of cases involved at most two resolving reactions. The remaining blocked reactions had category II solutions where novel reaction(s) had to be added in order to incorporate the blocked metabolite.

Mindful that we might overlook plausible solutions for blocked reactions having multiple SMILEY solutions by focusing entirely on S1 solutions, we investigated the categories of the alternative solutions. We found that, when available, alternative solutions were either of the same solution category as the S1 solution or were category III transport solutions (Additional file 4). This result indicated that the S1 solutions directly reflect the solution categories available for each blocked reaction although the resolving reactions comprising the solutions were different.

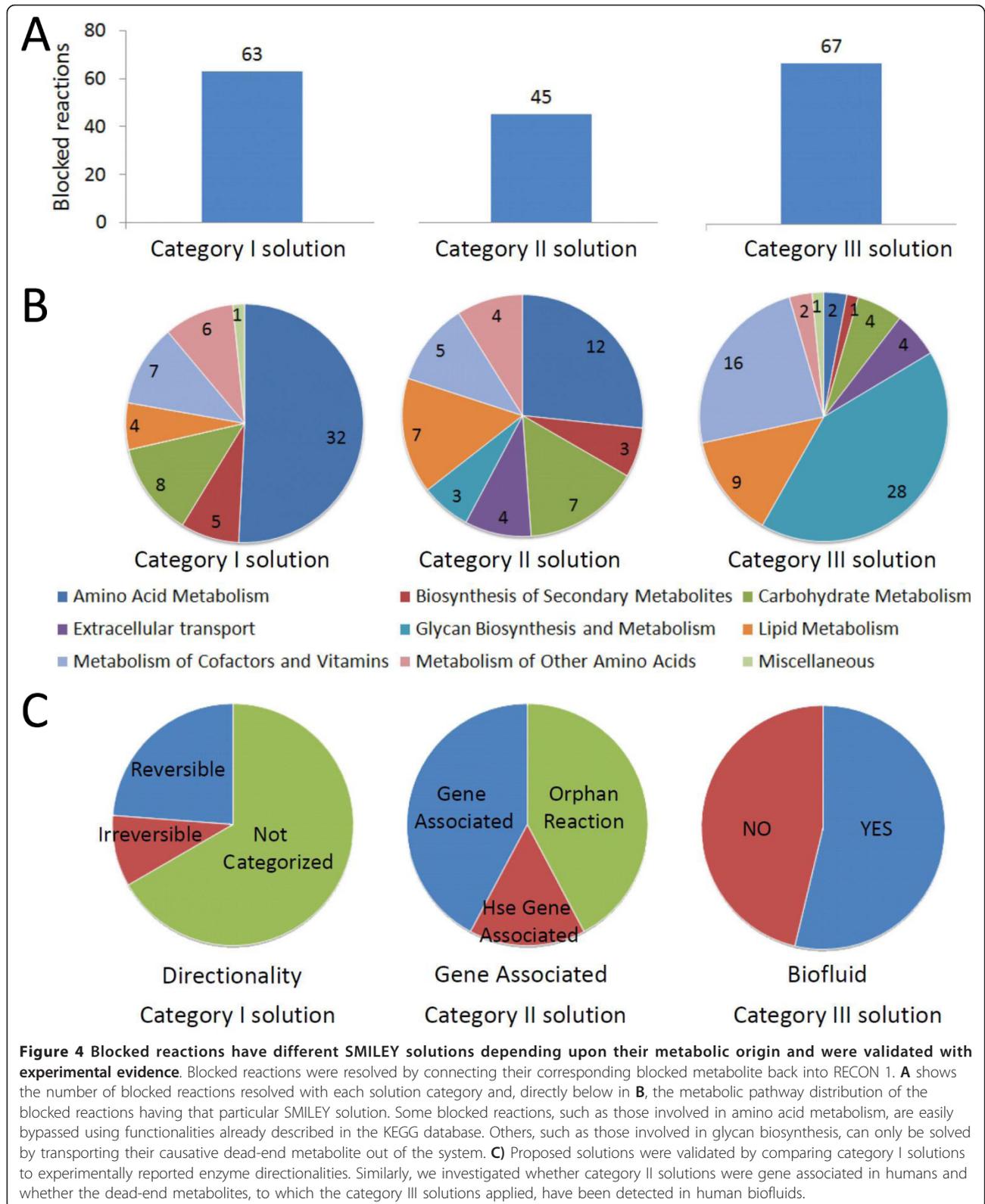
Blocked reactions have different S1 solutions depending on metabolic origin

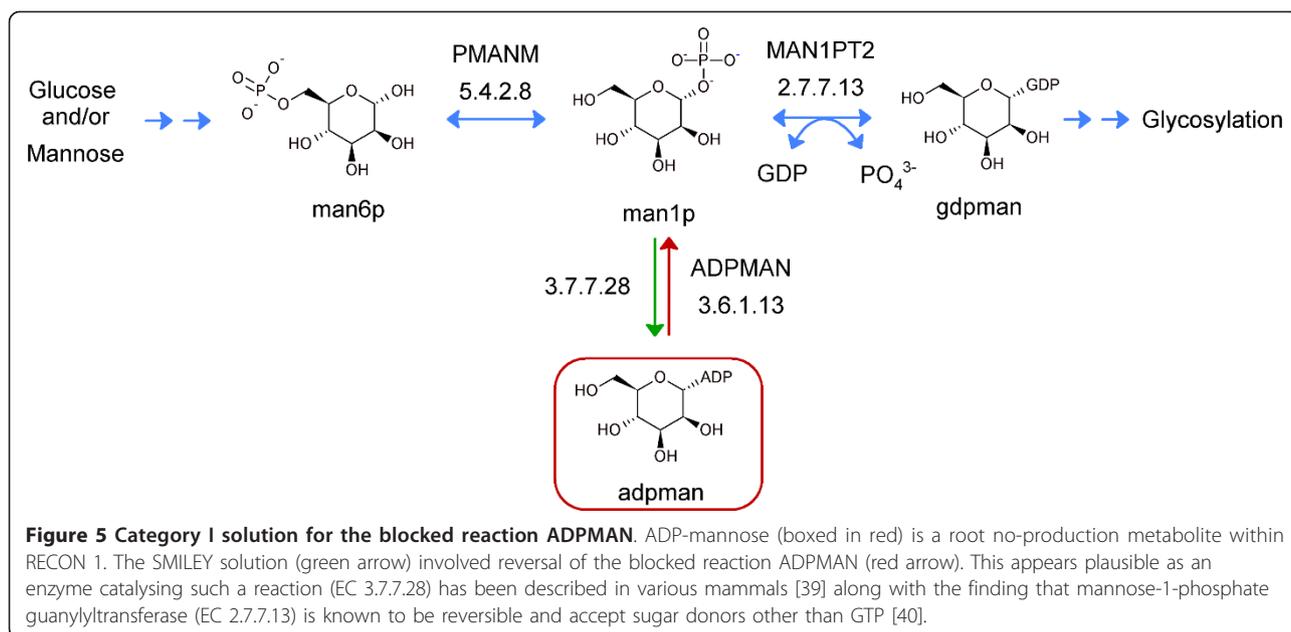
Roughly one third of all the blocked reactions were resolved with category I solutions (Figure 4). Nearly half of these applied to blocked reactions involved in amino acid metabolism (Figure 4B), many of which were part of tryptophan and lysine metabolism in the cytosol. With respect to their metabolic origin, up to 70% of the total number of blocked reactions originating from a particular metabolic pathway had category I solutions (Additional file 5). Notably, none of the blocked reactions involved in glycan biosynthesis and metabolism had category I solutions. Instead, these reactions were primarily resolved by category III transport solutions. These applied to 90% of the blocked reactions involved in glycan biosynthesis and metabolism, more specifically to those in chondroitin sulphate and heparan sulphate degradation in the lysosome and N-glycan biosynthesis in the golgi apparatus. Apart from glycan metabolism, category III solutions applied to multiple blocked reactions involved in the metabolism of cofactors and vitamins and in lipid metabolism. Few blocked reactions within amino acid metabolism and the biosynthesis of secondary metabolites had category III solutions.

Category II solutions were the least common solution. With respect to their metabolic origin, 10-50% of the blocked reactions had this solution type (Additional file 5). Many of these applied to blocked reactions taking part in various metabolic subsystems of amino acid metabolism such as tyrosine metabolism and urea cycle metabolism. Multiple blocked reactions in carbohydrate metabolism and fatty acid metabolism also had category II solutions.

Validation of SMILEY solutions

Category I and III solutions were only generated when no other solution was possible and therefore the blocked reactions that had these solutions contain metabolites whose global metabolic role has not been defined. This is however not a direct indicator of whether or not these solutions are biologically plausible. Category I reversal solutions to blocked reactions were validated by i) quantitative assignment of blocked reaction directionalities based on estimates of their Gibbs free energy changes [33] and ii) a query of experimentally reported reaction directionalities from the Brenda database [34]. Approximately 50% of the blocked reactions were calculated to be reversible while 23% were reported reversible in the Brenda database (Figure 4C and Additional file 6). Only four reactions were reported reversible by both methods which is due to the high number of the blocked reactions having unspecified reaction directionalities in the Brenda database (67%) and the uncertainties associated with reaction Gibbs free energy estimation due to unknown *in vivo* concentrations of the partaking metabolites [35]. These validation methods do not exclude reactions as reversible or not. Ultimately detailed inspection of each gap solution is required. Figure 5 shows the cytosolic reaction ADPMAN, which is catalysed by the nudix hydrolase NUDT5 (EC 3.6.1.13) and blocked in RECON 1 due to the root non-production metabolite ADP-mannose (adpman) [36-38]. The product, mannose-1-phosphate (man1p), is converted to GDP-mannose (gdpman) from either glucose or mannose and is ultimately utilised in N-glycan biosynthesis. A category I solution to ADPMAN was proposed by SMILEY. We found that, although the reversal of NUDT5 activity has not been reported, the conversion of man1p to adpman has been described in rats and cows (EC 2.7.7.28) [39]. Furthermore, the reaction MAN1PT2, catalysed by mannose-1-phosphate guanylyltransferase (EC 2.7.7.13), is known to accept both ATP and ITP in addition to GTP although with reduced activity [40]. This implies that formation of adpman from man1p, which is in effect a reversal of EC 3.6.1.13 activity, is plausible. The results suggest that reversing the directionality of blocked reactions to restore network flux can represent valid biological hypotheses



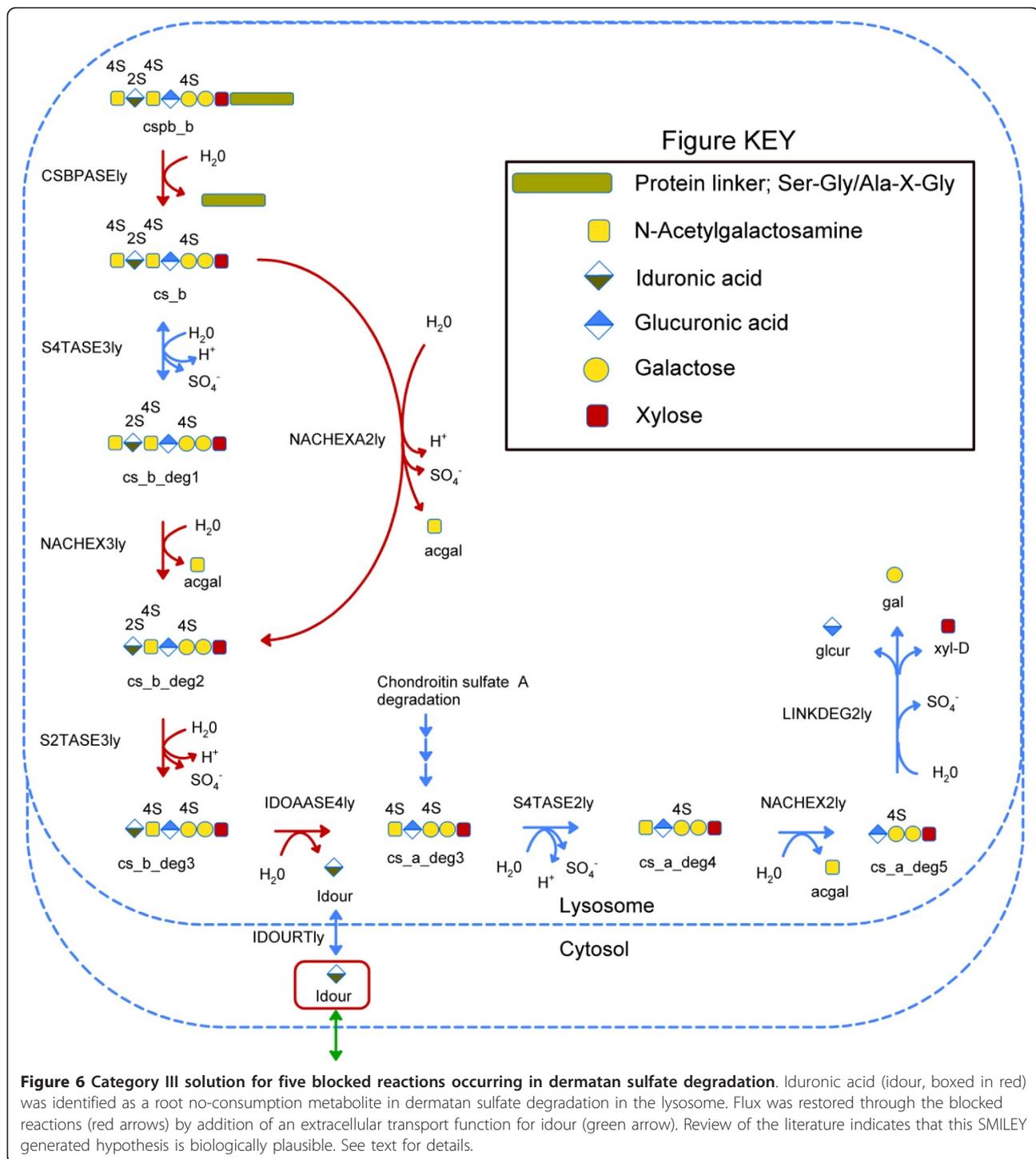


although the biological fate of the dead end metabolite remains unknown.

In order to validate the category III solutions, we investigated whether the metabolites to which the transport solutions applied have been detected in biofluids by querying the Human Metabolome Database [41]. The 67 blocked reactions that had category III S1 transport solutions were part of 24 reaction cascades. We found that 36 of the blocked reactions, part of eight reaction cascades, had blocked metabolites that have been detected in biofluids (Figure 4C). For example, we identified a cascade of five blocked reactions involved in dermatan sulfate degradation in the lysosome. Dermatan sulfate degradation is modelled in RECON 1 by the degradation of a glycosaminoglycan chain of a defined length (Figure 6). Following initial cleavage of the polysaccharide from a serine residue at the reducing end of the core tetrasaccharide linkage by dermatan proteoglycan pretease, degradation occurs through a series of enzymatic steps involving desulfonation by sulfatases and subsequent hydrolysis of the glycosidic bonds connecting the monosaccharide building blocks, iduronic acid and N-galactosamine, by exoglycosidases [42,43]. We found that dermatan sulfate degradation was blocked due to iduronic acid buildup in the cytosol following its cleavage from the dermatan polysaccharide in the reaction IDOAASE4ly (EC 3.2.1.76) and transport out of the lysosome in reaction IDOURtly [44-46]. The SMILEY solution for the blocked reaction IDOAASE4ly proposed transport of iduronic acid out of the cytosol in order to restore flux through the dermatan sulfate subsystem. These results suggest a transport function for

iduronic acid out of the cytosol, which appears plausible as iduronic acid has been detected in urine (M. Fuller, personal communication; [43,47]). Iduronic acid was also identified as a root no-consumption metabolite in heparan sulfate degradation causing twelve blocked reactions. The transport solution for iduronic acid therefore resolves 18 blocked reactions occurring in glycan degradation. We found no reports of cell surface iduronic acid transporters in the literature. Detailed substrate specificity studies for human hexose transporters, however, have not been performed [48].

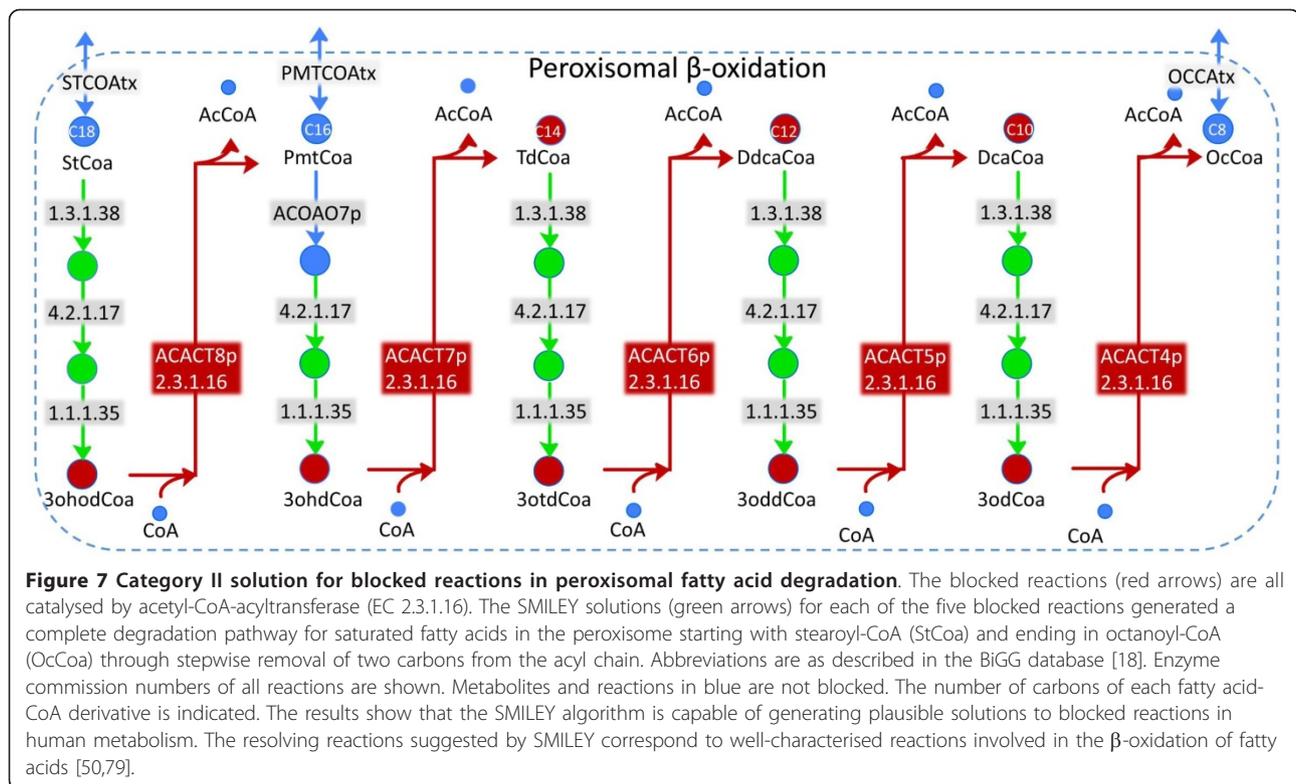
As opposed to category I and III SMILEY solutions, category II solutions involve adding resolving reactions to RECON 1 to restore flux through the dead-end metabolites. Of the 45 blocked reactions with category II solutions we found that seven had solutions composed of resolving reactions associated with human genes. Of those, five were part of peroxisomal beta-oxidation of long chain unsaturated fatty acids (Figure 7). These blocked reactions were all catalysed by 3-ketoacyl coenzyme A thiolase (EC 2.3.1.16), an enzyme responsible for the last step of the four recurring steps of fatty acid β -oxidation [49]. The SMILEY solutions successfully coupled the root no-production metabolites of these reactions, involving 3-oxo fatty acid coenzyme A derivatives, to their corresponding fatty acid in three resolving reactions. For example, the blocked reaction AACT4p that forms octanoyl coenzyme A (ocCoA) from the root no-production metabolite 3-oxodecanoyl-CoA (3odCoA) was coupled to decanoyl-CoA (dcaCoA) through its stepwise oxidation by trans-2-enoyl-CoA reductase (EC 1.3.1.38), enoyl-CoA hydratase (EC 4.2.1.17), and 3-



hydroxyacyl-CoA dehydrogenase (EC 1.1.1.35). Peroxisomal β -oxidation of unsaturated long chain fatty acids has not been completely elucidated in humans. However, it has been shown to generate ocCoA which is then transported to the mitochondria where it can be further oxidised and coupled to ATP production [50] as suggested by SMILEY. Although these results do not

add to current knowledge of metabolism as these gaps result from incomplete modelling in RECON 1 of β -oxidation in the peroxisome, they suggest that SMILEY is capable of generating biologically realistic results to gaps in RECON 1.

The majority of category II solutions were composed of resolving reactions not associated with any known

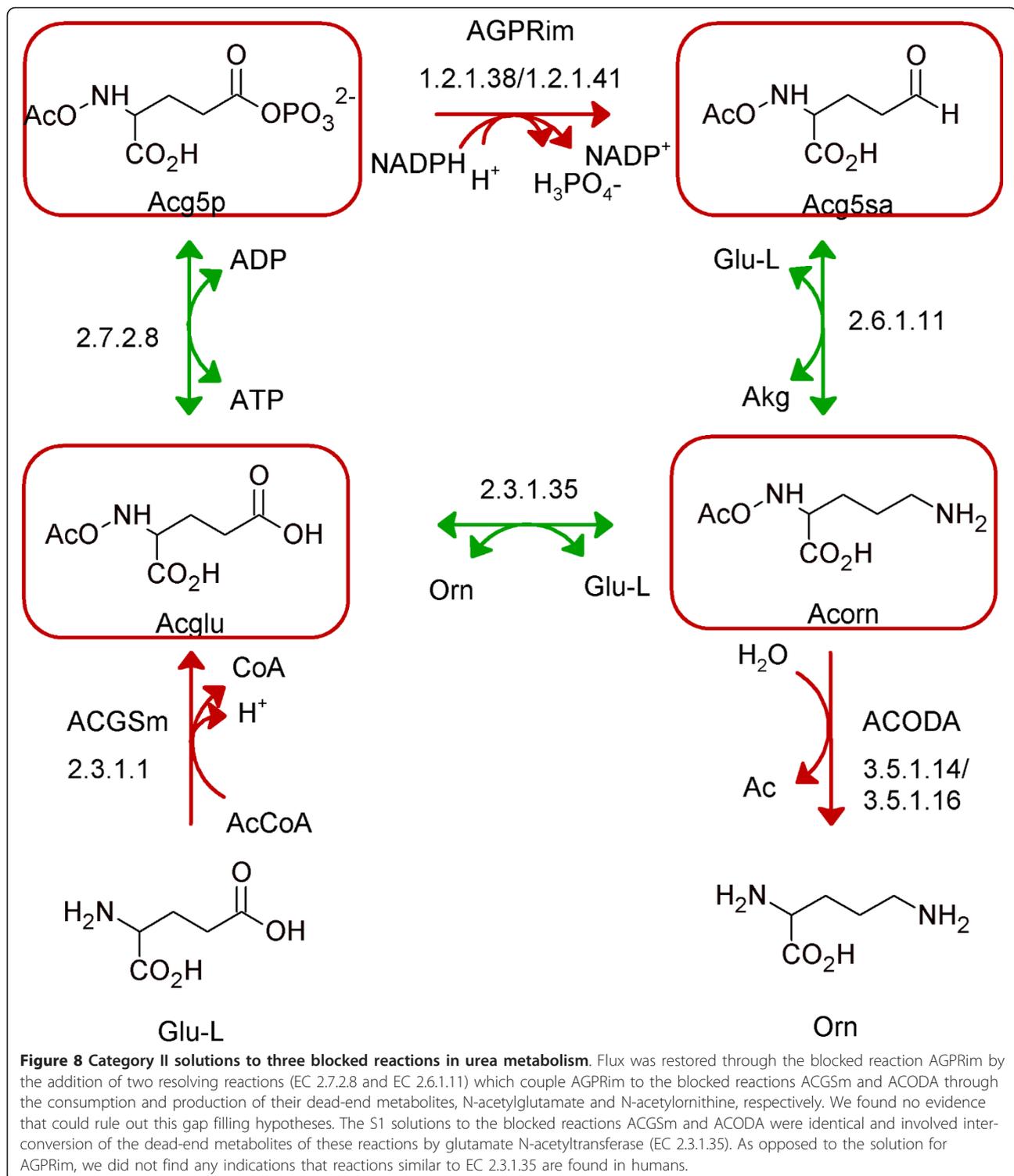


genes (19 solutions) or resolving reactions associated with enzyme activities encoded for by non-human genes (19 solutions) (Figure 4C). These represented the most exciting SMILEY solutions as they have the potential to unearth novel human metabolic enzymes. Figure 8 shows an example of category II solutions to three blocked reactions in urea metabolism. The SMILEY solution to AGPRim (EC 1.2.1.41), involves formation of its root no-production metabolite N-acetylglutamate-5-phosphate (acg5p) from N-acetylglutamate (acglu) and production of N-acetylornithine (acorn) from the root no-consumption metabolite N-acetylglutamate-5-semialdehyde (acg5sa). Acglu and acorn are themselves root no-consumption and root no-production metabolites in the blocked reactions ACGSm (EC 2.3.1.1) and ACODA (EC 3.5.1.16) respectively. Interestingly, the two resolving reactions proposed for AGPRim, catalysed by acetylglutamate kinase (EC 2.7.2.8) and acetylornithine transaminase (EC 2.6.1.11), therefore restore flux through all three blocked reactions. These reactions are part of the arginine biosynthesis pathway in prokaryotes and plants. They have not been associated with human genes, as arginine biosynthesis in mammals has been shown to occur through glutamate rather than N-acetylglutamate [51,52]. Human enzymes with similar functionalities have however been described. For example, human pyrroline-5-carboxylate synthase (P5CS) encodes glutamate-5-semialdehyde dehydrogenase activity (EC

1.2.1.41) and glutamate kinase activity (EC 2.7.2.11). P5CS is therefore a target candidate to catalyse a reaction where acg5sa is produced from acglu. Similarly, while acetylornithine transaminase activity (EC 2.6.1.11) has not been reported in humans, ornithine transaminase activity (EC 2.6.1.13) has [53]. A literature search did not reveal whether human ornithine transaminase is capable of accepting acorn as a substrate. However, ornithine transaminase from *Plasmodium vivax* shows 42-53% sequence similarity to eukaryotic ornithine transaminases and is known to accept both acorn and orn [54]. We suggest that following manual validation the hypotheses generated by SMILEY can guide experimental laboratory research. For selected blocked reactions, this is currently underway in our laboratory.

Discussion

The investigation of large complex systems on a global scale makes it impractical and maybe even impossible to know details about all involved metabolites, genes, and proteins. At the same time, a high level of knowledge about metabolic subsystems and or enzyme activities is necessary in order to come up with hypotheses of particular metabolic fates and novel reactions. In this study we used a systems biology approach to characterise and fill gaps in human metabolism. The key results include i) many dead-end metabolites affect reaction cascades, ii) computationally predicted solutions require thorough



manual curation and biochemical insight, and iii) four biological plausible hypotheses were identified. This work highlights that finding gene candidates for metabolic functions in the human genome is not a trivial issue and the extensive manual effort to curate the

computational predictions of candidate reactions highlight the overall quality and quantity of data included in Recon 1.

We characterised knowledge gaps of human metabolism, represented by blocked reactions and dead-end

metabolites identified in RECON 1 [11], to which solutions, in the form of non-organism specific metabolic reactions, could be found using the computer algorithm SMILEY [28] (Figure 1). We identified 175 blocked reactions, 70% of which had high confidence scores, and observed that while they were unevenly distributed within human metabolic pathways, most were found in the cytosol (Figure 2). Furthermore, we found that they arose due to different dead-end metabolite types that were in some cases responsible for up to 14 blocked reactions. These properties are likely to affect how trivial it will be to address these knowledge gaps experimentally and suggest that the impact of resolving these gaps, both in terms of novel metabolic discovery and their influence on RECON 1, will be different. For example, determining the fate of a metabolite, which results in multiple blocked reactions, will have a different impact on the biological accuracy of RECON 1 than resolving a single blocked reaction. Nevertheless, a single blocked reaction could be of great interest as a candidate target for novel metabolic discovery as its components could represent a drug target and resolving the gap could have unforeseen effects on network robustness, i.e., human metabolism. Also, assaying cytosolic reactions will be more straightforward than determining the function of compartmentalised reactions. Subsequently, which knowledge gaps are chosen for experimental research is ultimately a human decision depending on research goals, biological novelty factors, ease of experimental validation, and underlying evidence of the knowledge gap's validity.

We highlighted four examples of missing knowledge in human metabolism (Figures 5, 6, 7 and 8) that resulted in biologically plausible hypotheses using a combined algorithmic and manual approach. The hypotheses were strengthened with published experimental data. In the case of iduronic acid (Figure 6), a major constituent of glycosamine glycans, we argued for a hypothesis that an extracellular transport reaction needs to be added to RECON 1. Although no direct evidence for such transport could be identified in the human genome, the existence of iduronic acid in human urine (M. Fuller, personal communication) suggests that a transporter may be a biologically plausible solution. Further evidence is that the build-up of iduronic acid in the lysosome has been linked to lysosomal storage disorder caused by defects in the sialic acid lysosomal transporter [55]. Similarly, defects in α -L-iduronidase (3.2.1.76), the exo-glycohydrolase that cleaves iduronic acid off the non-reducing end of dermatan sulfate and heparan sulfate are known to cause a different type of lysosomal storage disorder, called mucopolysaccharidosis I [43]. Despite its apparent involvement in disease, the metabolic fate of iduronic acid is unknown. The present

work highlights knowledge gaps in human metabolic processes, such as the fate of iduronic acid, which in the context of investigating lysosomal storage disorders due to protein deficiencies, have not been relevant but are now required to generate a complete picture of human metabolism. Our gap filling examples showed that algorithms, such as SMILEY, can be used to direct hypotheses of novel functions in human metabolism. Nevertheless, a semi-automated approach was required to assist with the identification of plausible gap filling candidates for experimental verification.

Multiple gap finding and gap filling algorithms exist, including GapFind/GapFill [20] and GrowMatch [29], and the use of alternative algorithms will undoubtedly increase the number of possible hypotheses as they employ different heuristics and data sources (e.g., universal databases). This work does not provide a comprehensive list of possible gap-filling reaction solutions but rather assesses the use of (semi)-automated computational approaches for identifying and completing missing functions in human metabolism on a large-scale. We found that computational tools, such as SMILEY, do not necessarily suggest biologically plausible gap filling hypotheses. The generated hypotheses need to be evaluated in a manual, time-consuming manner, similar to the gap filling process employed during the reconstruction approach [4,26]. The search for novel functions is therefore only semi-automated. Automated algorithms could however be trained, based on manual effort, to prioritize or exclude certain types of solutions. In addition, approaches could be developed that incorporate methods to build hypotheses of genes associated with orphan reactions [56-60], which SMILEY does not directly do.

Identification of genes associated with biological plausible hypotheses as suggested by SMILEY was a major challenge. Relatively few knowledge gaps were resolved using known metabolic functions (Figure 3), and the solutions required detailed literature review such that homology, of what were often prokaryotic genes/proteins to possible human counterparts, could be assessed. In light of our results, we believe that existing automatic gap filling approaches for uncovering gene function will be of limited use for mammals. This limitation arises from a lack of phylogenetic information, which is extensively explored for annotating microbial genomes [58,61]. Although various homology databases exist for mammalian genomes covering up to seventy mammalian species [62], the majority of phenotypic, genetic, and biochemical studies have been performed using mice, and to a lesser extent, human cells. Information derived from these databases therefore originates from few organisms making them less useful for annotation purposes. Furthermore, co-expression analysis is used in

microbes to determine genes with related function [63-65] and could serve as a strategy for gene finding in the human genome. However, analysis of regions of correlated transcription (RCT) in human and mouse identified both related and unrelated genes being co-expressed [66]. The majorities of RCT were not found in both human and mouse, which the authors explained with i) missing definition of homology and/or synteny, ii) no conserved pattern, and/or iii) physiological differences between human and mice. This example highlights the challenges associated with finding novel gene functions in the human genome using established methods from the bacterial world. Novel approaches may include the use of protein-protein interaction data [67-69], tissue-specific information [70,71] and disease information [72] combined with gap filling algorithms. In particular, the latter work [72] observed a high degree of correlation between known co-occurring (co-morbid) diseases in patients and flux-coupling of the reactions that are perturbed in association with each of the disease states. Flux coupled reaction sets [73], or perfectly coupled reaction sets (Co-sets) [74], have been calculated in genome-scale metabolic models. Co-sets are often along linear pathways [74]. Thus, a low co-morbidity of two metabolically linked diseases would indicate a missing link along a Co-set, which would break the flux coupling by creating a pathway split. Similarly, single nucleotide polymorphisms have been mapped onto metabolic networks [75] and may be used for identifying missing functions in human metabolism.

Ubiquitous unknowns e.g. genes with unknown function and orphan enzymes belonging to orthologous families, have been identified as top targets for functional elucidation in terms of biological knowledge payoff as these are ancient in origin and therefore likely to be involved in essential metabolic processes [22,25,76]. We believe that combining a metabolic network approach with knowledge of ubiquitous unknowns could also represent an ideal method for organism specific novel function identification.

Conclusions

The results presented here show that RECON 1 allows the identification of specific metabolic pathways and reactions for which knowledge is lacking; thereby focusing the search for unknown metabolic functions by putting them into context with previously gathered metabolic information. Following identification and characterisation of RECON 1 network gaps, we showed that gap solution hypotheses can be generated automatically and successfully but require detailed, time consuming manual investigation in order to validate their biological plausibility. In this manner we have derived multiple hypotheses, which we intend to use to direct

our knowledge-driven approach towards novel metabolic discoveries.

Methods

Pre-processing to gap identification

RECON 1 was obtained from Duarte et al. [11]. It contains 356 dead-end metabolites causing knowledge gaps in eight cellular compartments. The dead-end metabolites in the compartmentalised model are not unique, as the same metabolite can cause blocked reactions in two or more cellular compartments. In order to address this issue, RECON 1 was decompartmentalised, by placing all intracellular compartment reactions in the cytosol and removing duplicates. Extra-organism located reactions were kept. Therefore, the decompartmentalised network accounts for these two compartments. Subsequently, the number of dead-end metabolites was reduced to 145, as they were unique. Note that this modification affected the following gap analysis in that compartment specific gaps were not considered for gap filling; e.g., a reaction present in mitochondria but missing in cytosol would not result in a gap in the decompartmentalised network. All subsequent analysis was performed using the decompartmentalised version of RECON 1 (Recon_1_decomp).

As a next step, all blocked reactions present in Recon_1_decomp were identified using flux variability analysis (FVA) [27] as reactions unable to carry flux defined by $|V_{max, i}| \leq 10^{-5}$ mmol/g_{dw}/hr and $|V_{min, i}| \leq 10^{-5}$ mmol/g_{dw}/hr for all *i* reactions in the network. All exchange reactions were unconstrained permitting free uptake and secretion of respective metabolites. A total of 285 reactions were identified.

The SMILEY algorithm has been described previously by Reed et al. [28] and implemented in the COBRA toolbox v2.0 (Schellenberger et al, submitted). We downloaded the KEGG [31] Ligand database (as of 1.10.2009), deemed **U**. Furthermore, we constructed a transport matrix, **X**, by defining a transport reaction from cytosol to extra-organism and an exchange reaction for each metabolite occurring Recon_1_decomp (**S**) and **U** (Figure 1). Matrix **U** and **X** served as reaction source for the SMILEY algorithm. Prior to the calculation, metabolites from **S** were matched to **U**. Note that not all metabolites in **S** have a known KEGG ID and that subsequent solutions are sensitive to missing KEGG IDs, meaning that some possible resolving solution may have been missed in our simulation due to this shortcoming. Current work has focused on adding more metabolite identifiers to RECON 1 (Thiele et al, in preparation).

In the next step, each blocked reaction $v_{b, i}$ was chosen as objective function, requiring the SMILEY algorithm to find reactions in **U** and/or **X** to be added to **S**

such that $|v_{b, i}|$ could carry a flux greater than 10^{-5} mmol/g_{dw}/hr. SMILEY is designed to find the shortest possible solution consistent with this requirement [28]. For each blocked reaction, we computed the 20 shortest SMILEY solutions. Note that the number of solutions computed is user defined and in some cases 20 distinct solutions may not exist. For 175 out of 285 blocked reactions at least one SMILEY solution could be found.

Gap analysis and validation of SMILEY output

SMILEY generated a total of 1335 solutions for the 175 blocked reactions. We filtered the solutions such that: i) only category II solutions with the least number of resolving reactions were investigated, and ii) if no category II solution was found, the solution involving the least number of resolving reactions was investigated. In cases where a blocked reaction had multiple solutions with the same number of resolving reactions, a random solution was picked. These criteria generated the S1 SMILEY solution output that is reported in the results.

Biochemical and genetic information concerning each blocked reaction identified by FVA was obtained from the Bigg database <http://bigg.ucsd.edu/>[18]. The reaction specific information allowed blocked reactions to be grouped depending on their metabolic pathway, subsystem, and/or other reaction specific features described within the Bigg database. Similarly, biochemical and genetic information concerning resolving reactions proposed by SMILEY was obtained from the KEGG database <http://www.genome.jp/kegg/>[31], which allowed organisation of the resolving reaction output. For resolving reactions where no human gene or protein information was directly available from Bigg or KEGG, blast homology searches of genes encoding the resolving reaction activity were performed against the human sequence databases on the NCBI website <http://blast.ncbi.nlm.nih.gov/Blast.cgi> and using the STRING database [77]. The localisation of dead-end metabolites in human biofluids was obtained from the human metabolome project <http://www.hmdb.ca/>[41]. Resolving reactions were investigated individually by literature review in order to verify the biological relevance of the proposed SMILEY solutions and generate plausible hypotheses for gap filling of RECON 1. Reaction directionalities were compared with experimentally reported reaction directionalities in the Brenda database [34] and by quantitative assignment of reaction directionality using the von Bertalanffy 1.0 algorithm [33] an extension available freely as part of the openCOBRA project [78]. The details of these calculations will be published in a separate manuscript. Briefly, experimentally determined or computed standard metabolite Gibbs energy transformed to cellular compartmental conditions with respect to *in vivo* pH (pH = 5.5-8.0),

temperature (37°C), ionic strength (0.25 M) and electrical potential (-150 - 30 mV) was used to predict the upper and lower bounds on standard transformed reaction Gibbs energy. The upper and lower bounds are dependent upon *in vivo* metabolite concentration ranges which were set to 10^{-7} - 10^{-2} M. When the transformed reaction Gibbs energy range spans zero the reaction is predicted to be quantitatively reversible. Spontaneous reactions proposed by SMILEY were assessed in a similar manner as enzyme catalysed reactions. All gap filling hypotheses can be found in Additional files 6 and 7. Reaction and sub-cellular compartment abbreviations are as described in the BiGG database.

Additional material

Additional file 1: Metabolic subsystem distribution of the 175 blocked reactions. The distribution of the blocked reactions identified in RECON 1 within metabolic subsystems.

Additional file 2: Dead-end metabolites can cause multiple blocked reactions. In the majority of cases, dead-end metabolites only cause one blocked reaction. When a dead-end metabolite is at the end or beginning of a reaction cascade however, it inhibits flux through all reactions, which are part of the reaction cascade. The figure shows the number of blocked reactions found in the reaction cascades. For example, there is one reaction cascade, which has 14 blocked reactions.

Additional file 3: The SMILEY solution distribution is not dependent upon the metabolic pathway. Blocked reactions can have multiple SMILEY solutions independent of the metabolic pathway, of which the blocked reaction is a part.

Additional file 4: Characterisation of alternative SMILEY solution. The figure shows the solution categories of the alternative SMILEY solutions to blocked reactions, which had a category I (A), category II (B) or a category III (C) S1 solution. A) 30% of the blocked reactions with a category I S1 solution had alternative category III solutions. The remaining 70% had either an alternative category I solution or none at all. B) 61% of the blocked reactions, which had category II S1 solution, had alternative category II or III solutions. C) None of the blocked reactions, which had category, III S1 solutions had alternative category solutions as expected.

Additional file 5: The pathway distribution of blocked reactions and their SMILEY S1 solutions categories. The figure shows the SMILEY S1 solution category as a percentage of the total number of blocked reactions found within a particular metabolic pathway. The number of blocked reactions within each metabolic pathway and their SMILEY solution type is also shown. Some blocked reactions, such as those involved in amino acid metabolism, are easily bypassed using functionalities already described in the KEGG, represented by category I and II SMILEY solutions. Others, such as those involved in glycan biosynthesis, can only be solved by transport of their causative dead-end metabolite out of the system.

Additional file 6: The filtered S1 SMILEY output. The table contains the filtered SMILEY S1 output as described in materials and methods. The identified blocked reaction is given in column A. Columns B-J report properties of the blocked reactions i.e., causative dead end metabolites, BiGG confidence scores, cellular compartment, whether or not the blocked reaction is in a cascade, blocked reaction components, associated genes, the metabolic subsystem and the metabolic pathway the blocked reaction occurs in. Columns K-O report properties of the S1 solution i.e. the solution category, the experimentally reported directionalities and computed directionalities of blocked reactions having category I reversal solutions, the number of resolving reactions composing the solution and the KEGG identifiers of each resolving reaction. The remaining columns report properties of each individual

resolving reaction comprising the SMILEY solution i.e. KEGG identifiers, resolving reaction components, metabolic pathway and subsystem and any associated enzyme commission numbers or known human genes.

Additional file 7: The complete SMILEY output. The table contains the complete SMILEY output with up to twenty solutions for each identified blocked reaction. The identified blocked reaction is given in column A. In columns B-D the properties of the blocked reaction are reported (reaction components, gene and metabolic subsystem). In column E the resolving reactions comprising the complete SMILEY solution for the blocked reaction is reported with KEGG reaction identifiers. In the following columns the properties of the resolving reactions are reported (reaction components, reaction subsystem, enzyme commission number and gene).

Acknowledgements

We thank Jeffrey D. Orth for critical reading of the manuscript, Dr Maria Fuller for her discussion concerning iduronic acid and Hulda Haraldsdottir and Dr. Ronan M.T. Fleming for their inputs on thermodynamic constraints on metabolic reaction directionality. This work was supported by the European Research Council grant proposal no. 232816. IT was supported, in part, by a Marie Curie International Reintegration Grant awarded (N° 249261) within the 7th European Community Framework Program.

Author details

¹Center for Systems Biology, University of Iceland, Sturlugata 8, 101 Reykjavik, Iceland. ²Faculty of Industrial Engineering, Mechanical Engineering & Computer Science, University of Iceland, 101 Reykjavik, Iceland.

Authors' contributions

OR carried out the experiments, analysed the data and wrote the manuscript. IT designed study, carried out experiments, analysed the data and contributed to the writing of the manuscript. BØP contributed to the design and writing of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 29 March 2011 Accepted: 1 October 2011

Published: 1 October 2011

References

1. Reed JL, Famili I, Thiele I, Palsson BO: **Towards multidimensional genome annotation.** *Nat Rev Genet* 2006, **7**:130-141.
2. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ: **Reconstruction of biochemical networks in microorganisms.** *Nat Rev Microbiol* 2009, **7**:129-143.
3. Oberhardt MA, Palsson BO, Papin JA: **Applications of genome-scale metabolic reconstructions.** *Mol Syst Biol* 2009, **5**.
4. Thiele I, Palsson BØ: **A protocol for generating a high-quality genome-scale metabolic reconstruction.** *Nat Protoc* 2010, **5**:93-121.
5. Chang RL, Ghamsari L, Manichaikul A, Hom EFY, Balaji S, Fu W, Shen Y, Hao T, Palsson BØ, Salehi-Ashtiani K, Papin JA: **Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism.** *Mol Syst Biol* 2011, **7**:518.
6. Manichaikul A, Ghamsari L, Hom EFY, Lin C, Murray RR, Chang RL, Balaji S, Hao T, Shen Y, Chavali AK, Thiele I, Yang X, Fan C, Mello E, Hill DE, Vidal M, Salehi-Ashtiani K, Papin JA: **Metabolic network analysis integrated with transcript verification for sequenced genomes.** *Nat Methods* 2009, **6**:589-592.
7. Sheikh K, Förster J, Nielsen LK: **Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of *Mus musculus*.** *Biotechnol Prog* 2005, **21**:112-121.
8. Quek L-E, Nielsen LK: **On the reconstruction of the *Mus musculus* genome-scale metabolic network model.** *Genome Inform* 2008, **21**:89-100.
9. Sigurdsson MI, Jamshidi N, Steingrímsson E, Thiele I, Palsson BO: **A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1.** *BMC Syst Biol* 2010, **4**:140.
10. Seo S, Lewin HA: **Reconstruction of metabolic pathways for the cattle genome.** *BMC Syst Biol* 2009, **3**:33.
11. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BØ: **Global reconstruction of the human metabolic network based on genomic and bibliomic data.** *Proc Natl Acad Sci USA* 2007, **104**:1777-1782.
12. Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I: **The Edinburgh human metabolic network reconstruction and its functional analysis.** *Mol Syst Biol* 2007, **3**:135.
13. Gille C, Bolling C, Hoppe A, Bulik S, Hoffmann S, Hubner K, Karlstadt A, Ganesan R, König M, Rother K, Weidlich M, Behre J, Holzhütter H-G: **HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology.** *Mol Syst Biol* 2010, **6**.
14. Zelezniak A, Pers TH, Soares S, Patti ME, Patil KR: **Metabolic network topology reveals transcriptional regulatory signatures of type 2 diabetes.** *PLoS Comput Biol* 2010, **6**:e1000729.
15. Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E: **Network-based prediction of human tissue-specific metabolism.** *Nat Biotechnol* 2008, **26**:1003-1010.
16. Chang RL, Xie L, Xie L, Bourne PE, Palsson BØ: **Drug off-target effects predicted using structural analysis in the context of a metabolic network model.** *PLoS Comput Biol* 2010, **6**:e1000938.
17. Bordbar A, Lewis NE, Schellenberger J, Palsson BO, Jamshidi N: **Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions.** *Mol Syst Biol* 2010, **6**.
18. Schellenberger J, Park J, Conrad T, Palsson B: **BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions.** *BMC Bioinformatics* 2010, **11**:213.
19. Orth JD, Thiele I, Palsson BØ: **What is flux balance analysis?** *Nat Biotechnol* 2010, **28**:245-248.
20. Satish Kumar V, Dasika MS, Maranas CD: **Optimization based automated curation of metabolic reconstructions.** *BMC Bioinformatics* 2007, **8**:212.
21. Nakahigashi K, Taya Y, Ishii N, Soga T, Hasegawa M, Watanabe H, Takai Y, Honma M, Mori H, Tomita M: **Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism.** *Mol Syst Biol* 2009, **5**:306.
22. Karp PD: **Call for an enzyme genomics initiative.** *Genome Biol* 2004, **5**:401.
23. Lespinet O, Labeledan B: **Orphan enzymes?** *Science* 2005, **307**:42.
24. Chen L, Vitkup D: **Distribution of orphan metabolic activities.** *Trends Biotechnol* 2007, **25**:343-348.
25. Hanson AD, Pribat A, Waller JC, de Crécy-Lagard V: **"Unknown" proteins and "orphan" enzymes: the missing half of the engineering parts list—and how to find it.** *Biochem J* 2010, **425**:1-11.
26. Orth JD, Palsson BØ: **Systematizing the generation of missing metabolic knowledge.** *Biotechnol Bioeng* 2010, **107**:403-412.
27. Mahadevan R, Schilling CH: **The effects of alternate optimal solutions in constraint-based genome-scale metabolic models.** *Metab Eng* 2003, **5**:264-276.
28. Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, Bui OT, Knight EM, Fong SS, Palsson BO: **Systems approach to refining genome annotation.** *Proc Natl Acad Sci USA* 2006, **103**:17480-17484.
29. Kumar VS, Maranas CD: **GrowMatch: an automated method for reconciling in silico/in vivo growth predictions.** *PLoS Comput Biol* 2009, **5**: e1000308.
30. Fong SS, Nanchen A, Palsson BO, Sauer U: **Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes.** *J Biol Chem* 2006, **281**:8024-8033.
31. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M: **KEGG: Kyoto Encyclopedia of Genes and Genomes.** *Nucleic Acids Res* 1999, **27**:29-34.
32. Duarte NC, Herrgård MJ, Palsson BØ: **Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model.** *Genome Res* 2004, **14**:1298-1309.
33. Fleming RMT, Thiele I: **von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models.** *Bioinformatics* 2011, **27**:142-143.
34. Schomburg I, Chang A, Schomburg D: **BRENDA, enzyme data and metabolic information.** *Nucleic Acids Res* 2002, **30**:47-49.
35. Hoppe A, Hoffmann S, Holzhütter H-G: **Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks.** *BMC Syst Biol* 2007, **1**:23.

36. Kim JS, Kim WY, Rho HW, Park JW, Park BH, Han MK, Kim UH, Kim HR: Purification and characterization of adenosine diphosphate ribose pyrophosphatase from human erythrocytes. *Int J Biochem Cell Biol* 1998, **30**:629-638.
37. Gasmil L, Cartwright JL, McLennan AG: Cloning, expression and characterization of YSA1H, a human adenosine 5'-diphosphosugar pyrophosphatase possessing a MutT motif. *Biochem J* 1999, **344**(Pt 2):331-337.
38. Yang H, Slupska MM, Wei YF, Tai JH, Luther WM, Xia YR, Shih DM, Chiang JH, Baikalov C, Fitz-Gibbon S, Phan IT, Conrad A, Miller JH: Cloning and characterization of a new member of the Nudix hydrolases from human and mouse. *J Biol Chem* 2000, **275**:8844-8853.
39. Herman RH: Mannose metabolism. I. *Am J Clin Nutr* 1971, **24**:488-498.
40. Ning B, Elbein AD: Cloning, expression and characterization of the pig liver GDP-mannose pyrophosphorylase. Evidence that GDP-mannose and GDP-Glc pyrophosphorylases are different proteins. *Eur J Biochem* 2000, **267**:6866-6874.
41. Wishart DS, et al: HMDB: the Human Metabolome Database. *Nucleic Acids Res* 2007, **35**:D521-D526.
42. Hoppe W, Rauch U, Kresse H: Degradation of endocytosed dermatan sulfate proteoglycan in human fibroblasts. *J Biol Chem* 1988, **263**:5926-5932.
43. Fuller M, Meikle PJ, Hopwood JJ: Glycosaminoglycan degradation fragments in mucopolysaccharidosis I. *Glycobiology* 2004, **14**:443-450.
44. Scott HS, Anson DS, Orsborn AM, Nelson PV, Clements PR, Morris CP, Hopwood JJ: Human alpha-L-iduronidase: cDNA isolation and expression. *Proc Natl Acad Sci USA* 1991, **88**:9695-9699.
45. Havelaar AC, Mancini GM, Beerens CE, Souren RM, Verheijen FW: Purification of the lysosomal sialic acid transporter. Functional characteristics of a monocarboxylate transporter. *J Biol Chem* 1998, **273**:34568-34574.
46. Mancini GM, de Jonge HR, Galjaard H, Verheijen FW: Characterization of a proton-driven carrier for sialic acid in the lysosomal membrane. Evidence for a group-specific transport system for acidic monosaccharides. *J Biol Chem* 1989, **264**:15247-15254.
47. Toida T, Qiu G, Matsunaga T, Sagehashi Y, Imanari T: Gas Chromatography-Mass Spectrometric Determinations of Iduronic and Glucuronic Acids in Glycosaminoglycans after Reduction of Carboxylic Group Using Sodium Borodeuteride. *Anal Sci* 1992, **8**:799-804.
48. Elbein AD, Freeze H: Glycosylation Precursors. *Essentials of Glycobiology*. 2 edition. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2009.
49. Bout A, Franse MM, Collins J, Blondin L, Tager JM, Benne R: Characterization of the gene encoding human peroxisomal 3-oxoacyl-CoA thiolase (ACAA). No large DNA rearrangement in a thiolase-deficient patient. *Biochim Biophys Acta* 1991, **1090**:43-51.
50. Wanders RJ, Vreken P, Ferdinandusse S, Jansen GA, Waterham HR, van Roermund CW, Van Grunsven EG: Peroxisomal fatty acid alpha- and beta-oxidation in humans: enzymology, peroxisomal metabolite transporters and peroxisomal diseases. *Biochem Soc Trans* 2001, **29**:250-267.
51. Cunin R, Glansdorff N, Piérard A, Stalon V: Biosynthesis and metabolism of arginine in bacteria. *Microbiol Rev* 1986, **50**:314-352.
52. Caldovic L, Tuchman M: N-acetylglutamate and its changing role through evolution. *Biochem J* 2003, **372**:279-290.
53. Ohura T, Kominami E, Tada K, Katunuma N: Crystallization and properties of human liver ornithine aminotransferase. *J Biochem* 1982, **92**:1785-1792.
54. Gafan C, Wilson J, Berger LC, Berger BJ: Characterization of the ornithine aminotransferase from *Plasmodium falciparum*. *Mol Biochem Parasitol* 2001, **118**:1-10.
55. Mancini GM, Beerens CE, Aula PP, Verheijen FW: Sialic acid storage diseases. A multiple lysosomal transport defect for acidic monosaccharides. *J Clin Invest* 1991, **87**:1329-1335.
56. Kharchenko P, Vitkup D, Church GM: Filling gaps in a metabolic network using expression information. *Bioinformatics* 2004, **20**(Suppl 1):i178-185.
57. Kharchenko P, Chen L, Freund Y, Vitkup D, Church GM: Identifying metabolic enzymes with multiple types of association evidence. *BMC Bioinformatics* 2006, **7**:177.
58. Hsiao T-L, Revelles O, Chen L, Sauer U, Vitkup D: Automatic policing of biochemical annotations using genomic correlations. *Nat Chem Biol* 2010, **6**:34-40.
59. Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY: Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. *Nat Biotechnol* 2010, **28**:149-156.
60. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 2010, **28**:977-982.
61. Overbeek R, et al: The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 2005, **33**:5691-5702.
62. Bult CJ, Eppig JT, Kadin JA, Richardson JE, Blake JA: The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res* 2008, **36**:D724-728.
63. Conners SB, Montero CI, Comfort DA, Shockley KR, Johnson MR, Chhabra SR, Kelly RM: An expression-driven approach to the prediction of carbohydrate transport and utilization regulons in the hyperthermophilic bacterium *Thermotoga maritima*. *J Bacteriol* 2005, **187**:7267-7282.
64. Chhabra SR, Shockley KR, Conners SB, Scott KL, Wolfinger RD, Kelly RM: Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. *J Biol Chem* 2003, **278**:7540-7552.
65. Chen L, Vitkup D: Predicting genes for orphan metabolic activities using phylogenetic profiles. *Genome Biol* 2006, **7**:R17.
66. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB: A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA* 2004, **101**:6062-6067.
67. Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes H-W, Ruepp A, Frishman D: The MIPS mammalian protein-protein interaction database. *Bioinformatics* 2005, **21**:832-834.
68. Lynn DJ, Chan C, Naseer M, Yau M, Lo R, Sribrnaia A, Ring G, Que J, Wee K, Winsor GL, Laird MR, Breuer K, Foroushani AK, Brinkman FSL, Hancock REW: Curating the innate immunity interactome. *BMC Syst Biol* 2010, **4**:117.
69. Bossi A, Lehner B: Tissue specificity and the human protein interaction network. *Mol Syst Biol* 2009, **5**:260.
70. Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E: Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* 2008, **26**:1003-1010.
71. Yang R, Su B: Characterization and comparison of the tissue-related modules in human and mouse. *PLoS ONE* 2010, **5**:e11730.
72. Park J, Lee D-S, Christakis NA, Barabási A-L: The impact of cellular networks on disease comorbidity. *Mol Syst Biol* 2009, **5**:262.
73. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD: Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res* 2004, **14**:301-312.
74. Thiele I, Price ND, Vo TD, Palsson BØ: Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet. *J Biol Chem* 2005, **280**:11683-11695.
75. Jamshidi N, Palsson BØ: Systems biology of SNPs. *Mol Syst Biol* 2006, **2**:38.
76. Galperin MY, Koonin EV: "Conserved hypothetical" proteins: prioritization of targets for experimental study. *Nucleic Acids Res* 2004, **32**:5452-5463.
77. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C: STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* 2009, **37**:D412-416.
78. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, Herrgård MJ: Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* 2007, **2**:727-738.
79. Mannaerts GP, van Veldhoven PP: Functions and organization of peroxisomal beta-oxidation. *Ann NY Acad Sci* 1996, **804**:99-115.

doi:10.1186/1752-0509-5-155

Cite this article as: Rolfsson et al: The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions. *BMC Systems Biology* 2011 **5**:155.