Universiteit Leiden
The Netherlands

Metabolomics assisted with stable-isotope labeling: exploring neuronal metabolism related to Parkinson's disease

Huang, L.

Citation

Huang, L. (2024, January 25). *Metabolomics assisted with stable-isotope labeling: exploring neuronal metabolism related to Parkinson's disease*. Retrieved from https://hdl.handle.net/1887/3715034

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3715034

Note: To cite this publication please use the final published version (if applicable).

Chapter 5:

fluxTrAM: Integration of tracer-based metabolomics data into atomically resolved genome-scale metabolic networks for metabolic flux analysis

Based on:

Luojiao Huang[†] , German Preciat[†] , Jesus Alarcon-Gil[†] , Edinson L. Moreno, Agnieszka Wegrzyn, Ines Thiele, Emma Schymanski, Amy Harms, Ronan M.T. Fleming, Thomas Hankemeier

fluxTrAM: Integration of tracer-based metabolomics data into atomically resolved genome-scale metabolic networks for metabolic flux analysis

Manuscript submitted

† Shared first authors

Abstract

Quantitative inference of intracellular reaction rates is essential for characterizing metabolic phenotypes. The classical experimental method for measuring metabolic fluxes makes use of stable-isotope tracing of metabolites through the metabolic network, followed by mass spectrometry analysis. Established ^{13}C-based metabolic flux analysis requires multidisciplinary knowledge in analytical chemistry, cell biology, and mathematical modelling, as well as the use of multiple independent software tools for handling mass spectrometry data and integration with a metabolic model. This approach is usually carried out within a small network rather than at genome-scale. To overcome interdisciplinary barriers and extend flux interpretation towards genome-scale, we developed fluxTrAM, a semi-automated pipeline for processing tracer-based metabolomics data and integrating it with atomically resolved genome-scale metabolic networks to enable flux predictions at genome-scale. fluxTrAM integrates several external software packages with the COBRA Toolbox enabling the generation of metabolite structure and reaction databases for a genome-scale model, processing of labelled mass spectrometry data into standardised mass isotopologue distribution data (MID), and metabolic flux inference. To demonstrate the utility of this pipeline, we generated ^{13}C-labeled metabolomics data on an *in vitro* human induced pluripotent stem cell (iPSC)-derived dopaminergic neuronal culture and processed ^{13}C-labeled MID datasets. In parallel, we generated a chemoinformatic database of standardised metabolite structures and atom-mapped reactions for a genome-scale dopaminergic neuronal metabolic model. MID data were integrated into an atomically resolved metabolic model for flux inference at genome-scale using a novel moiety fluxomics method. Inferred fluxes were compared with the MID data exported to established flux inference software for conventional flux inference with a central metabolic model. Flux vectors inferred by conventional and moiety fluxomic approaches were comparable but not identical. Further analysis are necessary with a genome-scale model that is suitable for both conventional and moiety fluxomic approaches to flux inference. Nevertheless, fluxTrAM enables the automation of labelled liquid chromatography (LC)-mass spectrometry (MS) data processing into MID datasets and atom mapping for any given genome-scale metabolic model. It contributes to the standardization and high throughput of metabolic flux analysis towards a genome-scale.

Introduction

Fluxomics aims for quantification of reaction flux at a genome-scale to comprehensively characterise cellular metabolism. Quantification of reaction flux is challenging as, unlike molecular species abundance, intracellular fluxes cannot be directly measured but must be indirectly inferred from measurements of labelled molecular species [1]. A typical in vitro fluxomic workflow proceeds from experimental design and cell culture implementation, to mass spectrometric data generation and processing, and finally to inference of reaction flux by computational modelling [2]. Therefore, progress in fluxomics is dependent on overcoming challenges in a set of complementary fields, including cell biology, analytical chemistry, and mathematical modelling. Herein, we focus on overcoming challenges in automated data handling, processing of mass spectrometry data and preparing an atom mapped model, which ultimately enables computational inference of metabolic flux.

Metabolomics allows a broad profiling of metabolites that connect diverse biochemical reactions in a metabolic network. Given a chromatographic co-elution of metabolite isotopologue peaks, reliable measurement of stable isotopically labeled metabolites requires a high capacity for analytical technology in molecule separation or peak resolution. Recent developments in mass spectrometry (MS) coupled with separation chromatography (gas chromatography, GC, or liquid chromatography, LC) improve the measurement accuracy of metabolite labelling patterns, as well as increase the coverage of metabolite classes [3,4]. In this way, metabolic flux profiling within a large-scale model becomes possible and can have the advantage of providing a systematic view of metabolism. Along with this, an advanced flux profiling approach at a genome-scale level is required.

By measuring a complex biological sample, high resolution (HR) LC-MS generates a large volume of MS raw data especially with a long analytical run for sufficient chromatographic separation. With the challenges of increased targeted metabolites, co-elution adding to the mass spectral complexity and multiple biological replicates in a tracer experiment, it becomes more difficult to process isotopologue labelling data. Manual peak integration to obtain the mass isotopologue distribution (MID) of each target metabolite becomes increasingly time-consuming and error-prone. To improve data quality and reproducibility, several automated peak detection and extraction packages have been developed for LC-HRMS [5–8], that enable automation of tracer-based metabolomics data processing. For instance, X^{13}CMS provides unbiased retrieval of isotopologue groups from isotopically labelled compounds between experimental conditions. MetExtact focuses on identifying the entire labelled metabolome and annotating unknown metabolites from mixtures of uniformly highly isotope-enriched and native biological samples. mzMatch–ISO offers its functionality in both automated untargeted metabolite annotation and relative quantification

of metabolite isotopologues. For our goal of targeted MID analysis, a pre-evaluation of each tool's suitability for incorporation into a standardised workflow is indispensable.

The advancement of high-throughput experimental technology has increased the demand for computational approaches to integrate complementary sources of omics data [9–11]. One such approach is integration of omics data with genome-scale computational models of biochemical networks, which are reconstructed from experimental data on molecular species and biochemical reactions [12]. Genome-scale models are especially useful for predicting metabolic reaction flux [13]. However, quantitative inference of intracellular metabolic reaction flux by integration of MID data with a genome-scale metabolic model is challenging and typically is done with a subnetwork model [14].

Integration of experimental MID data with metabolic modelling is currently challenging. On the one hand, the input of experimental MID data always needs external pre-processing including correction of naturally occurring isotopes and tracer purity, which involves intense user manipulation. On the other hand, currently available ^{13}C flux analysis programs implement non-linear optimization by minimizing the difference between measured and iteratively simulated MIDs to compute an optimal flux solution within a specified metabolic network [15–18]. Context-specific omics data supported by literature curation as well as modelling expertise are required to generate an organism-specific network model. Furthermore, for flux inference, a metabolic model must be atomically resolved such that the atom mapping from substrate to product atoms is obtained, manually for small networks or automatically for larger models. Automating the MID data processing and atom transitions resolving into a continuous workflow would be beneficial to accelerate metabolic modelling at genome-scale.

A reaction in a genome-scale metabolic network can be represented by a set of atom transitions, each of which connects an atom in a substrate metabolite to this atom in a product metabolite, forming an atom mapping for a reaction. Atom-mappings can be algorithmically predicted at genome-scale [19] and when compared with manually curated biochemical reactions, accurately predict atom mapping for ~90% of human metabolic reactions [20]. Connecting sets of atom mappings for a metabolic network enables one to atomically resolve a metabolic network at genome-scale [13]. Based on this approach, automated workflows for atom mapping have been developed for several species, including human [20] and *Arabidopsis thaliana* metabolism [21]. However, to have atomically balanced metabolic reactions, it is necessary to identify the structures for each metabolite in a genome-scale model. By including external chemoinformatics software such as Open Babel [22], CXCALC [23] and the Reaction Decoder Tool (RDT) [24] it is possible to convert between chemical formats, identify the pH required to balance the number of

hydrogens in the substrates and products, and thereby atom-map balanced metabolic reactions.

Graph theoretical analysis of an atomically resolved metabolic network enables identification of conserved moieties, each of which is a set of atoms that remains invariant with respect to all metabolic transformations in a given network [25]. The existence of conserved moieties means that a biochemical network is a special type of hypergraph [25], in the sense that the underlying stoichiometric matrix has special mathematical properties not shared with a general rectangular matrix corresponding to a general hypergraph. Fundamentally, this special property arises because one can represent a biochemical network at an atomic level [26]. In a companion paper [27], we exploited this property to develop a new mathematical and computational method that linearly relates metabolic reaction flux to the rate at which each labelled or unlabelled conserved moiety transitions between metabolites. Given mass isotopologue distribution data, this *moiety fluxomics* method enables inference of metabolic reaction flux at genome-scale.

Herein, we present fluxTrAM, a novel pipeline for flux inference with tracer-based metabolomics data and atomically resolved genome-scale model. fluxTrAM is an extension to the COBRA Toolbox v3.4 [13] that also takes advantage of existing third-party packages to achieve automated atom mappings and processing of labelled LC-MS data. It generates a chemoinformatic database of standardised and context-specific metabolite structures based on their InChI string and atom-mapped reactions for a given genome-scale metabolic reconstruction. In parallel, wrapper functions interface with third-party packages to enable processing of LC-MS raw data to generate standardised labelling data. This standardised MID data was subsequently used for established approaches to flux inference as well as a novel moiety fluxomics approach to flux inference where MID data was integrated with an atomically resolved metabolic model to infer the internal fluxes at genome-scale [27]. The utility of the pipeline was demonstrated by using stable-isotope labelled LC-MS data from a human dopaminergic neuronal culture, where conventional ^{13}C metabolic flux analysis was compared with fluxes inferred using moiety fluxomics and also genome scale flux predictions that were independent of labelling data. Based on inferred fluxes, the pipeline also enabled prediction of moieties to be considered for labelling when designing a new labelling experiment to further explore dopaminergic neuron metabolism.

Materials and methods

An overview of the methodology is given in **Figure 1** and described in detail below.

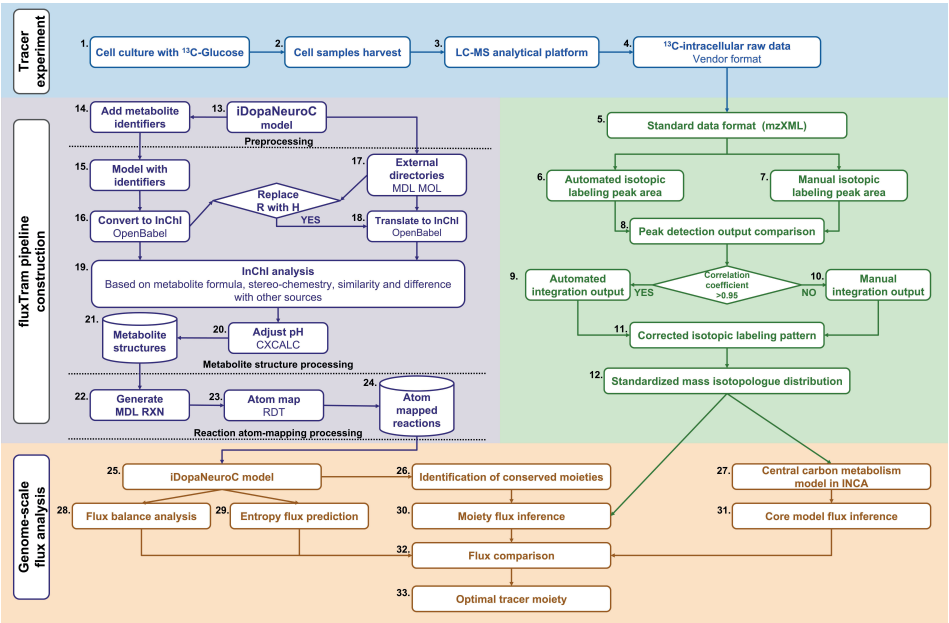

Figure 1. Methodological overview.

In vitro cell culture was fed with fully ^{13}C labelled glucose (1). Labelled cellular samples were collected after certain exposure time (2). A validated LC-MS platform [28] was established to measure labelled intracellular samples (3). Labelled intracellular raw mass spectral data were converted into a standard format (4, 5) and automated peak integration was used to obtain an isotopic labelling pattern (6, 9), unless it showed low correlation (8) to that obtained by manual peak integration (7, 10). Isotopic labelling patterns were converted to mass isotopologue distribution data in a standard format (11, 12). In parallel, atom mapping data was obtained from a genome-scale metabolic network by adding structural chemoinformatic data for each metabolite in the network if required (13-15), or using existing external data (17). The structural chemoinformatic data for each source in the model was used to generate an InChI (16, 18) where any R group in a metabolite structure was replaced by a hydrogen atom. The InChI strings obtained were compared to select the most representative structure for the genome-scale model (19). The number of hydrogen atoms of the highest-scoring metabolite structures was adjusted to match the number of hydrogens at which the metabolite is represented in the metabolic model of dopaminergic neuronal metabolism (iDopaNeuroC) (20), resulting in a database of standardised metabolite structures (21). The metabolic network stoichiometry and metabolite structure database were used to express each reaction in an atomically resolved reaction format (22), which was then atom-mapped (23) to generate an atom-mapped reaction database (24). Atom mapping data was used to identify the conserved moieties in the iDopaNeuroC model (25, 26). The standardised mass isotopologue distributions together with conserved moieties were used as constraints on the

iDopaNeuroC model to infer fluxes using the moiety fluxomic method (30). Fluxes inferred using the moiety fluxomic method were compared with predictions of fluxes without any isotopologue constraints on the iDopaNeuroC using flux balance analysis (FBA, 28) and entropic flux balance analysis [29] (29). Furthermore, the standardised mass isotopologue distribution data was imported established flux inference software (INCA [16,30]) using a central carbon metabolism model (27, 31). We compared inferred fluxes generated from the four different approaches: moiety flux inference, FBA flux prediction, entropy flux prediction and central carbon metabolism flux inference using INCA (32). Optimal genome-scale flux solutions were selected and used to further screen candidate tracer moieties for a new labelling culture experiment (33).

1. Tracer experiment

1.1 In vitro cell culture

Generation of an *in vitro* culture of midbrain-specific dopaminergic neurons followed an established protocol [31,32], with the adaptions described below. This culture method was the same as that used for generation of iDopaNeuroC model [29] a context-specific model of dopaminergic neuronal metabolism.

N2B27 medium preparation

The culture medium, denoted N2B27 medium, was used as the basis to prepare both maintenance and differentiation media. 49.25 mL of culture medium was obtained by mixing 24 mL Neurobasal medium (Invitrogen/Life Technologies), 24 mL of Dulbecco's modified Eagle's medium (DMEM)/F12 medium (Invitrogen/Life Technologies) supplemented with 1% penicillin and streptomycin (Life Technologies), 0.5 mL of 200 mM L-glutamine (Life Technologies), 0.5 mL of B27 supplement without Vitamin A (Life Technologies) and 0.25 mL of N2 supplement (Life Technologies). Glucose-free N2B27 medium was made in the same way, but with the Neurobasal medium replaced with Neurobasal medium with no D-glucose (Invitrogen/Life Technologies), and the DMEM/F12 medium replaced with stable isotope labeling with amino acids in cell culture (SILAC) advanced DMEM/F-12 Flex medium with no glucose (Invitrogen/Life Technologies).

Plate coating

Cell-culture treated 12-well plates (ThermoFisher scientific) were coated with 1% Matrigel (Discovery Labware, Inc., USA, Catalogue number 354277) in 600 µL of DMEM (1X) medium supplemented with knockout serum replacement (ThermoFisher scientific).

Cell seeding and maintenance

At the time of cell seeding, the knockout DMEM (1X) medium from the coating step was removed from each well and the K7 hNESC line was seeded in three replicate wells. The medium to maintain the hNESC in culture, denoted maintenance medium, was based on

N2B27 medium with 0.5 μM PMA (Enzo life sciences), 3 μM CHIR (Axon Medchem) and 150 μM ascorbic acid (Sigma Aldrich). The cell seeding was done by preparing 1.8×10e6 million cells/mL in maintenance medium and adding 300 μL of this preparation together with another 300 μL of maintenance medium to reach 4×10e5 cells per well. The plate was incubated at 37 °C and 5% CO_2 for 48 h.

Neuronal differentiation and maturation

The differentiation medium with PMA was prepared to induce the differentiation of hNESC towards midbrain dopaminergic neurons and consisted of N2B27 medium with 200 μM ascorbic acid, 0.01 ng/μL BDNF (Peprotech), 0.01 ng/μL GDNF (Peprotech), 0.001 ng/μL TGFβ3 (Peprotech), 2.5 μM dbcAMP (Sigma Aldrich) and 1 μM PMA. This medium was completely replaced every 2 days during the next 6 days of culture in the differentiation process. For the maturation of differentiated neurons, PMA is required to be absent from the differentiation medium [31,32]. This differentiation medium without PMA was used from day 9 onwards and complete media replacement was done every 2 days for 2 weeks.

^{13}C-labelled neuron culture and sample collection

^{13}C-labelled differentiation medium without PMA was prepared using glucose-free N2B27 medium, supplemented with 20.4 mM fully carbon labelled U-$^{13}C_6$-glucose (Cambridge Isotope Laboratories, USA), 200 μM ascorbic acid, 0.01 ng/ μL BDNF, 0.01 ng/ μL GDNF, 0.001 ng/ μL TGFβ3, 2.5 μM dbcAMP. For the pilot stable isotopic labelling study set, differentiated neurons were maintained in two groups separately with ^{13}C labelled and unlabelled medium. Three replicate wells of neurons from each group were incubated under 37 °C and 5% CO_2 condition for 4 hours before quenching. For the formal stable isotopic labelling study set, dopaminergic neurons were cultured in labelled medium at 37 °C and 5% CO_2 conditions with different incubation times, from 0min, 5min, 10min, 20min, 40min, 1h, 2h, 3h, 5h, 8h, 12h to 24h. Each condition group was run in triplicates. The spent medium was collected into a 1.5 mL Eppendorf tube. Neurons in each well were immediately quenched by using ice-cold 80% MeOH in water and harvested into a 1.5 mL Eppendorf tube as cell lysate. All samples were fast frozen into liquid nitrogen and stored in the -80 °C freezer until measurement.

1.2 Sample measurement

Labelled cellular samples were measured on a liquid chromatography-mass spectrometry (LC-MS) platform to obtain the mass isotopologue distribution (MID) of target metabolites (*SM4. Table S1*) [28]. The concentrations of U-$^{13}C_6$-glucose in cell and medium were quantified using the same platform. The cell lysate was sonicated, vortexed and then centrifuged at 16000 relative centrifugal force (rcf), at 4 ℃ for 10 min. Cell pellets were collected to measure the protein content using a bicinchoninic acid (BCA) assay (Thermo

Fisher Scientific Inc, USA). Supernatants were transferred into clean 1.5 mL Eppendorf tubes and evaporated to dryness in a Labcono SpeedVac (MO, United State). Each sample was reconstituted with 60 μL ice cold methanol/water (80%/20%; v/v). 50 μL of the reconstitution volume was collected and transferred into a new Eppendorf tube as a cell supernatant sample. Next, 50 μL (5 μL) of each cell supernatant (medium) sample was treated with liquid-liquid extraction by adding 40 μL of ice-cold methanol/water (80%/20%; v/v), 45 μL of ice cold milliQ water and 65 μL of ice-cold chloroform, followed with mixing and vortexing for 5 min and centrifuging at 16000 rcf 4 ℃ for 10 min. 130 μL of the aqueous phase was transferred into a new Eppendorf tube and extracted again by adding 25 μL of ice-cold methanol/water (50%/50%; v/v) and 65 μL of ice-cold chloroform, followed with mixing and vortexing for 5 min and centrifuging at 16000 rcf, 4 ℃ for 10 min. 140 μL of the upper aqueous phase was collected and taken to dryness. The residue was reconstituted with 50 μL of methanol/water (50%/50%; v/v), vortex-mixed, and centrifuged to remove debris. The supernatants were finally transferred to vials for LC-MS analysis.

LC-MS analysis was performed on a SCIEX tripleTOF 5600 MS system (SCIEX, USA) coupled to a Waters Acquity UPLC Class II (Waters, USA) equipped with a SeQuant® ZIC®-cHILIC HPLC column (2.1 mm x 100 mm, 3.0 μm, Merck, Germany). The LC-MS method was previously reported [28]. Mobile phase A was 90% acetonitrile, 10% water with 5 mM ammonium formate, and mobile phase B was 10% acetonitrile, 90% water with 5 mM ammonium formate. LC elution followed a nonlinear gradient that starts from 0% B (0-2 min) to 40% B (20 min), ends up with 0% B (20.1-35 min). The flow rate was 0.25 mL/min and the injection volume was 3 μ L. The MS detection was set for a scan range of 50-900 *m/z*. Acquired raw LC-MS data files were stored in vendor-specific formats (.wiff and .wiff.scan).

2. Tracer-based metabolomics data processing

A pipeline for processing of tracer-based metabolomic data was developed, which takes raw mass spectrometry data as input and ultimately generates mass isotopologue distribution data in standard format as output (**Figure 2**). This pipeline is implemented in MATLAB (Mathworks Inc, USA), but calls several established external software tools for (i) LC-MS raw data conversion, (ii) Peak detection and extraction, (iii) Isotopologue peak correction, (iv) Isotopologue peak summary.

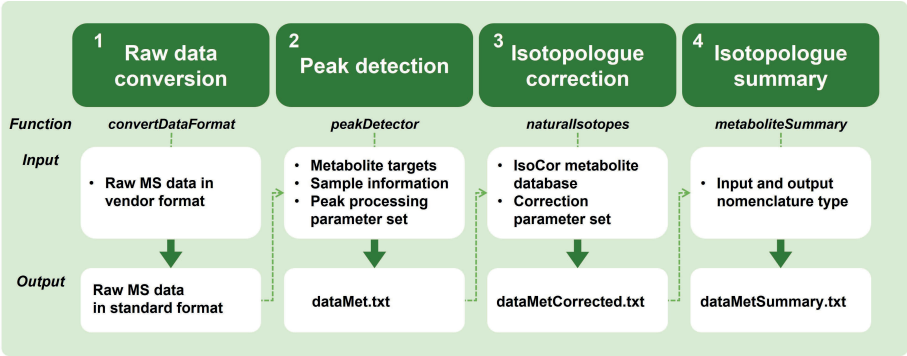

Figure 2. The schematic representation of tracer-based metabolomics data processing pipeline. In step 1, 'Raw data conversion', raw LC-MS data was processed by the function '*convertDataFormat*' into MS data in standard format. In step 2, 'Peak detection', standardized MS data used additional inputs for running function '*peakDetector*' to generate a result table of metabolite integration in a text file 'dataMet.txt'. At step 3, 'Isotopologue correction', the dataMet.txt file, the isoCor metabolite database, and the correction parameter set were used as inputs for running the function '*naturalIsotopes*' to remove the isotopic impurity of tarcer and the naturally occurring isotopic abundance. This created a 'dataMetCorrected.txt' file with a corrected integration table. In step 4, "Isotopologue summary', the dataMetCorrected.txt file was used as an input together with the entered nonmenclature types for running function '*metaboliteSummary*', which finally generated a 'dataMetSummary.txt' file with an integration table including interconverted metabolite nonmenclatures and summarized MID results shown in mean and standardized deviation.

LC-MS raw data conversion

A wrapper function (***convertDataFormat, MATLAB***) was created based on ProteoWizard msconvert tool [33] for the pipeline to run 'msconvert.exe' externally and automatically but still within a MATLAB environment. It is used to convert vendor-specific formats to standard formats (mzML, mzXML, MGF, MS2/CMS2/BMS2, mzIdentML). The specific output format for fluxTrAM is centroid mode in mzXML with zlib compression.

Peak detection and extraction

To select one automated peak extraction software package, two candidate packages (mzMatch-ISO [7] and X^{13}CMS [6]) were compared with two manual peak picking software packages (ElMAVEN [34], Skyline [35]). The two manual peak integration software packages (Skyline and ElMAVEN) utilised a common target metabolite list with exact molecular formula and expected retention time for peak detection. A threshold of 10 ppm was applied so that interferences from ions with close masses could be eliminated during peak extraction. In a first step of applying automated packages, XCMS [5] was used for the initial peak detection and alignment. The optimal values for the parameter settings

used in XCMS, mzMatch-ISO and X^{13}CMS analysis were tested in R (*as supplementary files*: X^{13}CMS_src.R, mzmatch_install.R). Five types of metabolite classes from the pilot stable isotopic labelling experiment were analysed, which included three amino acids (alanine, aspartate, glutamate), three organic acids (phosphoenolpyruvate, ketoglutarate, fumarate), one sugar phosphate (fructose 1,2-biphosphate), one tripeptide (glutathione), and one nucleotide (ATP). Pearson correlation was used to evaluate the normalised integrated peak area in automated and manual ways, and finally select one package for the automated data processing.

A wrapper function (***peakDetector, MATLAB)*** was created based on (XCMS [5], mzMatch-ISO [7]) packages for the pipeline to execute R packages still within the MATLAB environment. Converted LC-MS data was first analysed using XCMS to extract peaks. All detected peak features were then processed by mzMatch-ISO for aligning, noise filtering, gap-filling peaks, and later forming into a combined PeakML file containing all samples. Based on the input of a targeted metabolite list, automated profiling could be performed to extract relevant isotopologue peaks.

Isotopologue peak correction

For the correction of incomplete labeling of the isotopic tracer and naturally isotopic abundance in metabolites, a wrapper function (***naturalIsotopes, MATLAB***) was created based on IsoCor [36] for the pipeline to excute the Python module within a MATLAB environment. Relying on the experimental inputs, such as MS resolution, isotopic tracer type, isotopic purity, the extracted isotopologue peaks for each metabolite were corrected to only preserve the isotopic abundance resulting from the isotopic tracer. The final output was the corrected MID data of each metabolite in each sample.

Isotopologue peak summary

The function ***metaboliteSummary*** was created to summarize the MID data from each individual sample into the average MID and corresponding standard deviation. In addition, it also recognises and converts between different possible metabolite nomenclatures (such as Virtual Metabolic Human [37] (VMH) Abbreviation or Full Name, the Kyoto Encyclopaedia of Genes and Genomes database [38] (KEGG), the Human Metabolome Database [39] (HMDB), the International Union of Pure and Applied Chemistry (IUPAC), the International Chemical Identifier (InChI) String and InChlKey [40], the Simplified Molecular-Input Line-Entry System (SMILES) [41]) through the VMH database [37].

3. Atomically resolving genome-scale models

A parallel novel software pipeline was developed to atomically resolve any given genome-scale metabolic model, where the result is a chemoinformatic database of metabolite structures and atom mapped reactions. This pipeline is implemented in MATLAB, but calls

several optional external software tools, such as Open Babel [22] and CXCALC [23] for chemoinformatic data processing and the RDT [24] to atom map metabolic reactions (Java SE Development Kit is required). Installation of these external software tools is optional but recommended to obtain metabolites with higher confidence score. An overview of the methodology to generate an atomically resolved genome-scale model is given in **Figure 1** (13-24) and further described below.

3.1 Metabolite structures

Metabolites were represented with database identifiers, e.g., Virtual Metabolic Human [37] (VMH), PubChem [42], the Kyoto Encyclopaedia of Genes and Genomes database [38] (KEGG), Chemical Entities of Biological Interest [43] (ChEBI) or the Human Metabolome Database [39] (HMDB). In addition to database identifiers, the metabolite structures were also represented using several formats including metabolite chemical tables (MDL MOL) that list all of the atoms in a molecule, as well as their coordinates and bonds [44]; the Simplified Molecular-Input Line-Entry System (SMILES), which uses a string of ASCII characters to describe the structure of a molecule [41]; or the International Chemical Identifier (InChI) developed by the IUPAC, which provides a standard representation for encoding molecular structures using multiple layers [40] as illustrated in *Figure S2*.

For a given metabolite, comparing different metabolic databases, one may obtain substantial structural diversity due to the inclusion of different isomers in the different sources. The fluxTrAM pipeline uses all of the aforementioned databases to obtain and ultimately select a single InChI for each metabolite. InChI was chosen due to its standard structure for encoding molecular information and it is a database independent representation [40]. fluxTrAM analyses all the InChI strings obtained from each source database assigning a score based on the criteria shown in **Table 1** to identify the metabolic structure that most closely resembles the metabolite described in the genome-scale reconstruction. The InChI for each metabolite is scored considering the chemical formula excluding hydrogen atoms, similarity with other databases, stereochemistry and charge in order to avoid loss of the predictive capacity of computer models, due to propagation of an inconsistent structures.

Table 1. InChI based comparison

Each InChI layer contains information that can be used to identify the metabolite structure that is most similar to the metabolite in the genome-scale reconstruction.

Concept	Score	Description
Chemical formula	0 or 10	The chemical formula indicated in the genome-scale model is compared with that obtained from the InChI. This feature is given more weight in order to keep the metabolite in the genome-scale model, where all reactions are stoichiometrically consistent and mass balanced. Hydrogen atoms were ignored in this comparison since they can be modified based on the charge (**Figure S2**).

Charge	0 or 1	The charge indicated in the genome-scale model is compared with the charge obtained from the source.
Stereochemical information	0 or 1	Indicates whether the InChI contains stereochemical information or not.
Standard	0 or 1	Indicates whether the InChI is standardised or not.
Similarity with other databases	0 or 1	The number of sources where the InChI strings are identical, divided by the total number of sources.
Main layer similarity	0 or 1	The number of sources where the main layers are identical, divided by the total number of sources.
InChI with more layers	0 or 1	InChI with more layers.

Letters in a formula that do not correspond to a chemical element, typically described with an R, represent any group of atoms attached to the rest of the molecule for example an alkyl group or hydrogen atom. To compare metabolite structures with R groups, these atoms are modified by replacing each non-chemical atom with hydrogen only for the InChI comparison (**Figure 3.1**). If the number of hydrogen atoms in the metabolite structure obtained from a database and the metabolite obtained from the genome-scale model are different, software (CXCALC [23]) is used to adjust the charge and number of hydrogen atoms of a metabolite based on the pH (**Figure 3.2**). To ensure database consistency, the molecular structures of the metabolites in the database were represented as either molecular graphs or hydrogen-suppressed molecular graphs (**Figure 3.3**). Finally, the results were presented as molecular structures in five different formats (metabolite chemical tables (MDL MOL) [44], the Simplified Molecular-Input Line-Entry System (SMILES) [41], the International Chemical Identifier (InChI), the InChIkey [40] and an image (JPEG file) containing a graphical representation of the metabolite structure (when CXCALC is installed).

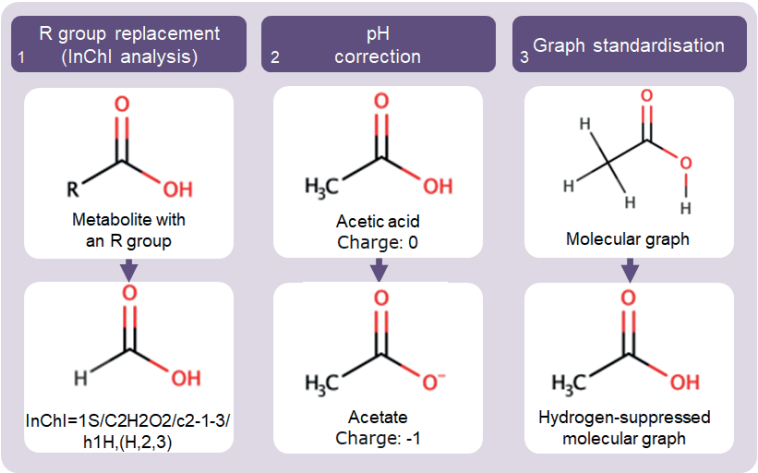

Figure 3. Example molecular graphs.

Metabolite structure represented in different forms, as well as how it is processed by the pipeline. (1) If an unspecified alky group attached to the carboxyl group is represented by a non-chemical atom (R), the R is replaced by a hydrogen to form a molecule that can be converted into an InChI. (2) The pH of the acetic acid molecule is changed in order to obtain the charge described in the model. (3) In order to achieve implicit hydrogen standardization, a molecular graph representing a metabolite's structural formula is converted to a hydrogen-suppressed molecular graph, which is a molecular graph with the hydrogen vertices removed.

3.2 Atomically resolved reactions

A reaction database was generated for all reactions where the corresponding metabolite structures were available, as shown in the example using a metabolic model, Recon3D [45] (**Figure 4**).

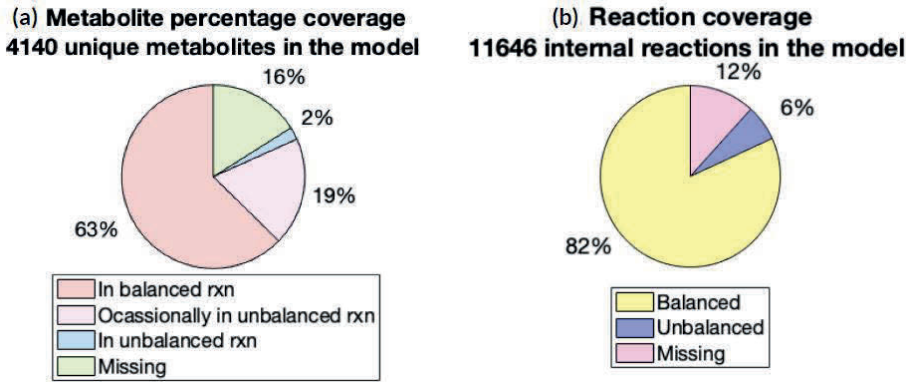

Figure 4. Coverage of metabolite structures and atom mapped reactions in the model case of Recon3D.

The fluxTrAM pipeline was used to collect 84% of the metabolite structures in Recon3D, while only 2% correspond to the metabolite structures that are always present in unbalanced reactions (a). The metabolite structures were used to atom-map 88% of the reactions. 12% of the reactions are missing because at least one metabolite structure is missing, and 6% are atomically unbalanced (b).

The obtained database of metabolite structures and reaction stoichiometriesfrom the genome-scale model was then used to generate chemical tables for each reaction (MDL RXN format) and was atom mapped using the Reaction Decoder Tool algorithm [24], interfaced with the COBRA Toolbox [13]. Atom mapped metabolic reactions were represented in reaction chemical tables (MDL RXN [44]) and reaction SMILES, while unmapped metabolic reactions were represented in International chemical identifiers for reactions [46] (RInChI; see *Figure S2*). Additionally, the reactions were atom mapped to represent their reaction mechanism. To demonstrate the general utility of this software, we generated atomically resolved reaction databases for iDopaNeuroC model, as well as

several other species and community metabolic models, such as *E. coli* core [47], gut microbial organisms (AGORA2 [48] and Recon3D [45]).

4. Inference of metabolic fluxes

4.1 Isotopomer Network Compartmental Analysis of central carbon metabolism

To perform conventional ^{13}C metabolic flux analysis, a stoichiometric model of central carbon metabolism which included glycolysis, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle was extracted from the iDopaNeuroC model. First, this model was edited in a MATLAB-based ^{13}C metabolic flux inference software package, isotopomer network compartmental analysis (INCA, [16]), describing reaction stoichiometries and carbon transitions between substrate and products (*SM4. Table S2*). Then standardised MID data was imported into INCA to infer a steady-state metabolic flux distribution in a central carbon metabolic model. INCA employs a type of decomposition of an isotopomer network to simulate MIDs as a function of metabolic flux [49], then metabolic flux is inferred by minimising the lack-of-fit between experimentally measured and computationally simulated MIDs using a Levenberg-Marquardt optimisation algorithm. Flux inference was repeated 50 times from random initial metabolic flux values to search for a global minimum fit. To assess goodness-of-fit, we subjected flux results to a chi-square statistical test and calculated 95% confidence intervals for each estimated flux value by evaluating the sensitivity of the sum of squared residuals to parameter variations.

4.2 Inference of metabolic flux at genome-scale with moiety fluxomics

Previously [26], we provided a mathematical definition of a conserved moiety and demonstrated how identification of the complete set of conserved moieties for a given metabolic network leads to a novel *moiety graph decomposition* of a stoichiometric matrix. We applied moiety graph decomposition to obtain the complete set of conserved moieties for a genome-scale model from human stem cell-derived, midbrain-specific, dopaminergic neurons *in vitro*, specfically iDopaNeuroC as described elsewhere [29]. In a companion paper, we mathematically and computationally demonstrate how moiety graph decomposition enables the development of a computationally efficient constraint-based modelling method to infer metabolic fluxes from isotope labelling data for a using complete metabolic network at genome-scale [27]. We applied this moiety fluxomic method to infer metabolic reaction fluxes at genome-scale using the aforementioned standardised MID data and a genome-scale model of dopaminergic neuronal metabolism.

4.3 Prediction of metabolic flux at genome-scale

Two additional methods, neither of which used any tracer-based metabolomic data, were used to predict metabolic reaction flux in the genome-scale model of dopaminergic neuronal metabolism: (i) Flux Balance Analysis (FBA) [50], with maximisation of ATP consumption

(VMHID: ATPM) to simulate a high demand for energy of dopaminergic neurons [51], (ii) Entropic Flux Balance Analysis (EFBA) maximisation of the entropy of forward and reverse fluxes, a least biased flux prediction given the available data [52], which also ensures that predicted flux satisfies energy conservation and the second law of thermodynamics [29]. Visualisation of all relevant flux distribution maps was performed in the software VANTED (Visualization and Analysis of Networks containing Experimental Data), v2.1.0 [53].

Results

We present fluxTrAM, a semi-automated pipeline for quantitative flux inference at genome-scale, including modules for processing of raw tracer-based mass spectrometry data into standardised mass isotopologue distribution and for generation of a chemoinformatic database of standardised metabolite structures and atom-mapped reactions at genome-scale. We demonstrated the performance of these two modules using tracer-based metabolomics data from iPSC-derived midbrain-specific dopaminergic neurons and a corresponding neuronal metabolic model, respectively. Next, we utilised the intermediate results of the pipeline for conventional ^{13}C flux analysis with a central carbon metabolism model, and for moiety fluxomics analysis at genome-scale.

1. Tracer experiment

Intracellular ^{13}C-glucose increased rapidly after 5 minutes and began to stabilise within 5 hours (**Figure 5**). The neuronal culture system reached a metabolic steady state within 24 hours uptake of ^{13}C-glucose.

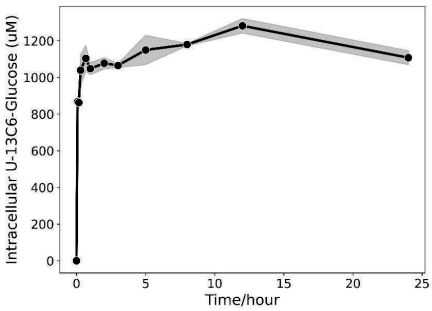

Figure 5. Progression toward metabolic steady state in differentiated dopaminergic neurons by uptake of U-$^{13}C_6$-glucose.

The intracellular level of U-$^{13}C_6$-glucose measured at 0min, 5min, 10min, 20min, 40min, 1h, 2h, 3h, 5h, 8h, 12h and 24h, constant level reached at 24h.

2. Tracer-based metabolomics data processing

2.1 Comparison for peak detection and extraction packages

Mass isotopologue peak integration using two automated software packages (mzMatch-ISO [7] and X^{13}CMS [6]) and two manual peak picking software packages (ElMAVEN [34], Skyline [35]) were compared and the resulting isotopologue fractions for a set of 9 test metabolites are shown in **Figure 6**. The Pearson correlation of the isotopologue fractions between automated and manual is shown in **Table 2**. The test metabolites were representative of clear ion chromatographic peaks varied in peak width, intensity, and retention time. Given a list of well-established retention times and mass-to-charge ratios (*m/z*) for either respective metabolites monoisotopic peak in ElMAVEN or their mass isotopologues as well in Skyline, manual peak picking software still requires manual inspection across the integration results of every sample. Finally, ElMAVEN and Skyline showed comparable integration results for all test metabolites. The two automated packages differed in their peak integration performance. X^{13}CMS returned isotopologue fractions that were less consistent with the integration results from the other packages. For example, no peak was extracted for phosphoenolpyruvate and compared to manual peak integration, there were low Pearson correlation coefficients for isotopologue fractions of aspartate and fructose 1,6-biphosphate. In contrast, for each of the test metabolites, mzMatch-ISO returned a similar isotopologue fraction to both of the manual software packages, which overall showed a high Pearson correlation coefficient above 0.95. Their respective statistical significance was below 0.002 (*SM4. Table S3*).

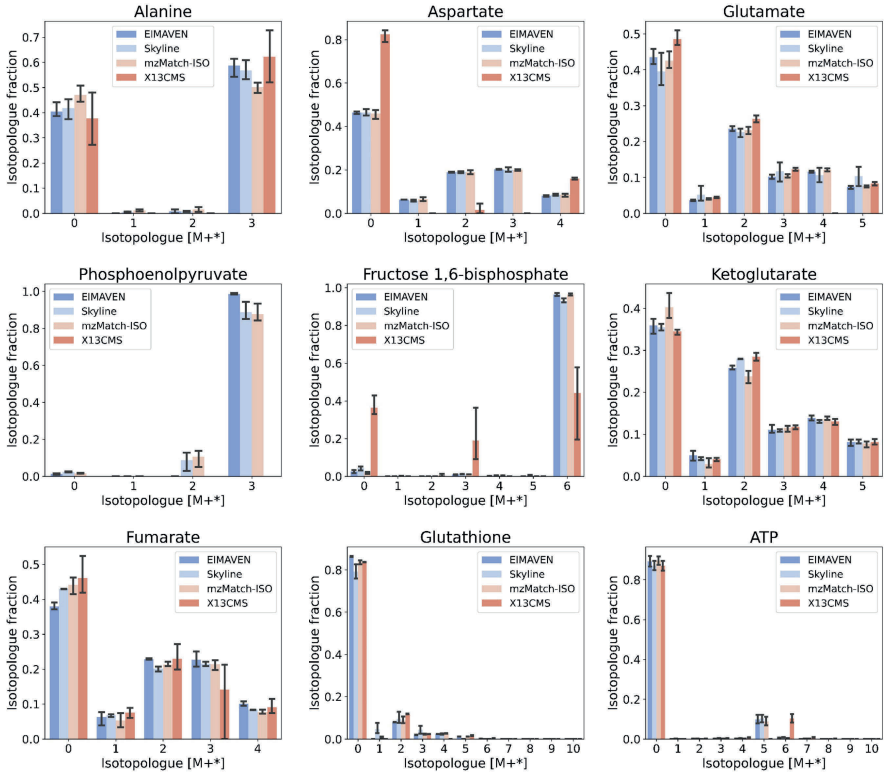

Figure 6. Comparison between the manual and automated peak detection and extraction results. The resulting metabolite isotopologue fractions of manual peak picking (ElMAVEN, Skyline) were compared to automated isotopologue peak integration (X^{13}CMS, mzMatch-ISO) for the tested nine metabolites.

Table 2. Correlation analysis of integration results from automated packages and manual software.

Pearson correlation coefficients between automatically and manually collected isotopologue fractions for each of the tested nine metabolites. ND = not detected.

Metabolite	**mzMatch-ISO** *vs* **Skyline**	**mzMatch-ISO** *vs* **ElMAVEN**	**X^{13}CMS** *vs* **Skyline**	**X^{13}CMS** *vs* **ElMAVEN**
Alanine	0.979	0.969	0.971	0.984
Aspartate	0.997	0.998	0.872	0.87
Glutamate	0.984	1	0.959	0.954
Phosphoenolpyruvate	0.99	0.99	ND	ND
Fructose 1,6-bisphosphate	1	1	0.645	0.637
Ketoglutarate	0.975	0.989	0.998	0.989
Fumarate	0.995	0.985	0.899	0.852
Glutathione	0.996	1	0.995	0.998
ATP	1	1	0.985	0.985

2.2 Peak detection and extraction evaluation

To test the peak detection and extraction performance of the incorporated mzMatch-ISO [7] package within our pipeline, we next analysed mass isotopologue distribution of essential amino acids and non-essential amino acids in neurons from the pilot study set. The integration output of 14 amino acids was compared between groups of neurons treated with labelled glucose or unlabelled glucose, the result for nine amino acids is shown in **Figure 7**. For each amino acid, the statistical differences in the isotopic incorporation between labelled and unlabelled group were evaluated based on an external t-test examination (*SM4. Table S4*). Essential amino acids can only be taken up from diet and cannot be synthesised de novo. Non-essential amino acids can be synthesised, mainly from glucose. The isotopologue fractions of six essential amino acids showed a naturally labelled isotopologue pattern in labelled neurons and had no significant difference from unlabelled neurons, including methionine, threonine, leucine, isoleucine, phenylalanine, tryptophan. In contrast, five nonessential amino acids including alanine, serine, aspartate, asparagine, glutamate showed various labelling patterns in labelled neurons. Moreover, the calculated MIDs in *Table S4* showed high consistency across the three labelled or unlabelled sample replicates. Conditionally essential amino acids, proline, tyrosine, and glutamine, showed different preferences for synthesis from glucose in neurons. A significant labelling pattern from U-$^{13}C_6$-glucose could be observed in proline. Meanwhile, glutamine and tyrosine showed little isotopic incorporation (**Figure 7**). The calculated MIDs were consistent among the labelled or unlabelled replicates.

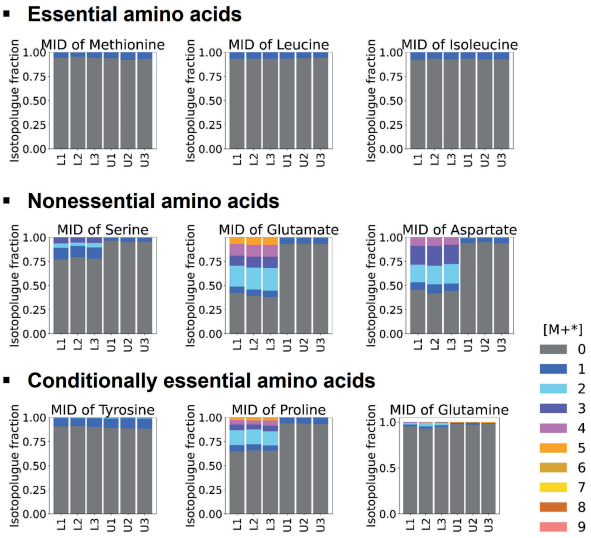

Figure 7. The mass isotopologue distribution (MID) of nine amino acids in the labelled neurons (n=3) and unlabelled neurons (n=3).

The labelled isotopomer pattern of methionine, leucine and isoleucine (essential amino acids), serine, glutamine, aspartate (nonessential amino acids) and tyrosine, proline, glutamine (conditionally essential amino acids) shown in both labelled and unlabelled neurons. Labelled neuron samples, abbreviated in L1, L2, L3: Three replicate samples in the group of cultured neurons with isotopic tracer. Unlabelled neuron samples, abbreviated in U1, U2, U3: Three replicate samples in the group of cultured neurons with no isotopic tracer. The color represents the different detected mass isotopologues as defined by the legend.

2.3 Data processing evaluation for a formal study set

To test the pipeline performance for processing a larger sample set, we further analysed the dynamic changes in mass isotopologue distribution of central carbon metabolites for neurons over a time course of 24 hours. A total of 36 samples for 10 metabolites (52 isotopologues) were processed automatically by fluxTram. The MID results can be found in *SM4. Table S5* and visualised in **Figure 8**. The labelling fractions of metabolites increased immediately after the initial time point. The glycolytic intermediates, phosphoenolpyruvate, pyruvate and lactate attained isotopic equilibrium within 5 hours. They showed high abundances of fully ^{13}C labelled states (M+3). The tricarboxylic acid cycle, including ketoglutarate, succinate, fumarate, malate, citrate and glutamate, took nearly 24 hours to reach an isotopic steady state. Each of them showed mass isotopologue distribution in multiple ^{13}C labelled states, ranging from one ^{13}C labelled to fully ^{13}C labelled.

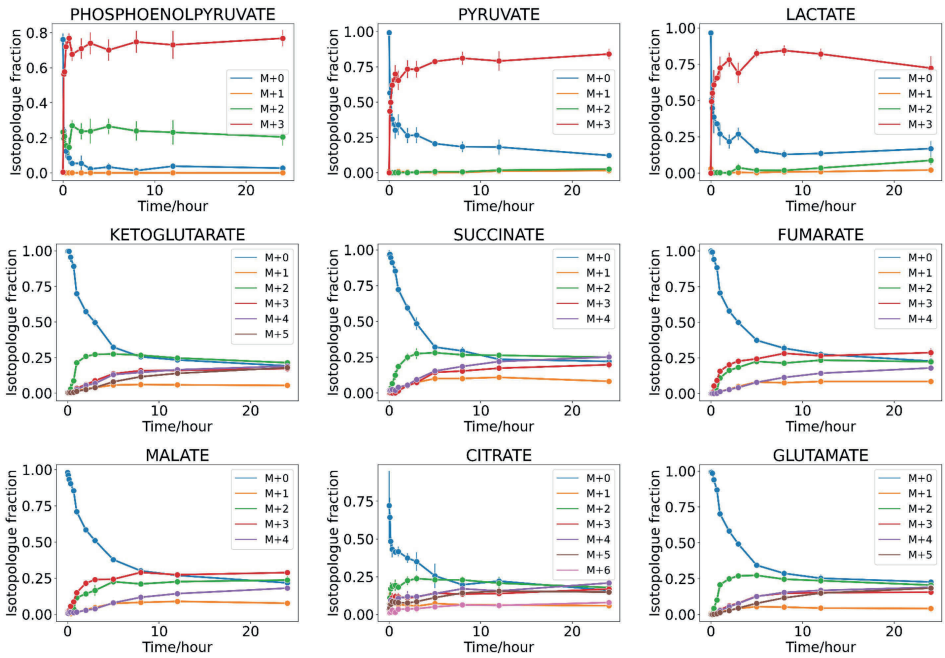

Figure 8. Evaluation of isotopic steady state in differentiated dopaminergic neurons by uptake of U-$^{13}C_6$-glucose.

Mass isotopologue distributions as a function of time for intracellular metabolites within 24h of feeding isotopically labelled substrate, U-$^{13}C_6$-glucose.

3. Databases of metabolite structures and atom mapped-reactions

3.1 Metabolite structures

Figure 9 shows a comparison of the identifiers in the iDopaNeuroC model (See **Table 1**). This comparison only reflects the specificity of the metabolite identifiers in the iDopaNeuroC model, not the quality of the databases consulted. The VMH database [37] contains the molecular structures that most closely resemble the metabolites described in the iDopaNeuroC model, providing the largest number of atomically balanced reactions. The more specific a metabolite structure is, the more likely it is to obtain a matched structure for a genome-scale metabolic network, as evidenced by **Figure 9** G. The metabolite formula in the iDopaNeuroC model matched well with the formula from ChEBI [43], HMDB [39], KEGG [38], and PubChem [42] databases, however, the charge is inconsistent with the model, so they received a lower score in the InChI string analysis (**Figure 9** A, B, D and E). On the other hand, the InChIs and SMILES identifiers in the model have high compatibility in the formula and charge, but they lack stereochemical information, reducing their specificity (**Figure 9** C and F).

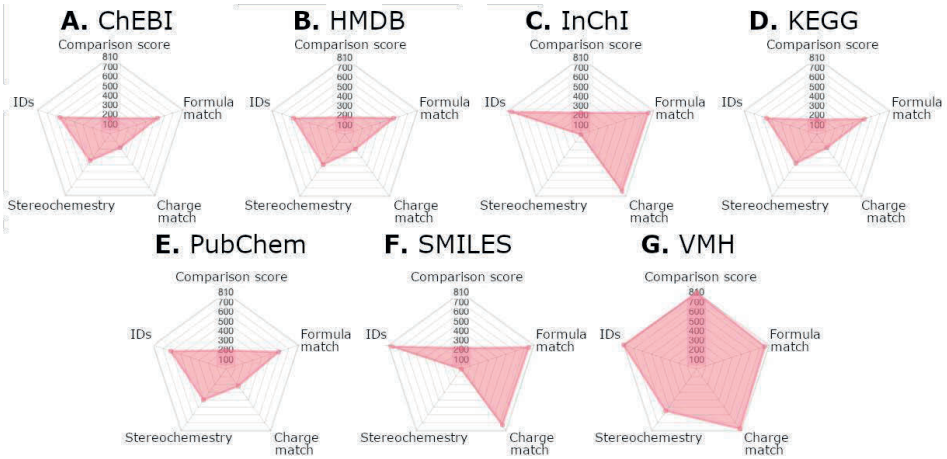

Figure 9. Comparison of metabolite structure identifiers in the iDopaNeuroC model.

Each of the metabolite identifiers evaluated in this work was compared using various criteria, including 1) IDs: the number of identifiers compared; 2) Comparison score: the number of times they received the best score; 3) Formula match: the number of times the chemical formula matched the metabolite formula in the iDopaNeuroC model (excluding the number of hydrogen atoms); 4) Charge match: the charge consistency with the model; and 5) Stereochemistry: the number of times stereochemical information was included.

3.2 Atomically resolved reactions

The genome-scale model of dopaminergic neuronal metabolism, iDopaNeuroC [29], was atomically resolved. We obtained almost all (810/818) structures corresponding to the metabolites in the iDopaNeuroC model. Of these, 677 metabolites were always present in atomically balanced reactions, 116 metabolites were occasionally present in atomically unbalanced reactions, and 15 metabolites were always present in atomically unbalanced reactions. Two molecular structures were not used, because at least one molecular structure was required to complete the reaction in which they were involved. Almost all (2,254/2,262) internal metabolic reactions in the iDopaNeuroC model could be atom-mapped, of which 2,155 were atomically balanced and 99 were unbalanced. 18 metabolic reactions could not be written, because at least one molecular structure was missing from the metabolite database. A list of the problematic reactions, along with an explanation of why each reaction was problematic, is provided in the supplementary information. **Table 3** provides a summary of the atomically resolved coverage of the iDopaNeuroC model as well as several additional species and community metabolic models.

Table 3. Atomically resolved coverage of metabolism in different models.

	iDopaNeuroC [29]	***E.coli* core** [47]	**AGORA2** [54]	**Recon3D** [45]
Metabolites in the database	810	53	1894	3534
Metabolites in balanced reactions	677	53	1510	2594
Metabolites occasionally in unbalanced reactions	116	0	260	789
Metabolites always in unbalanced reactions	15	0	66	92
Metabolites not used	2	0	58	59
Missing metabolites	8	1	1697	665
Reactions in the database	2254	71	4803	10279
Atom-mapped reactions	2254	71	4623	10279
Balanced reactions	2155	71	4583	9533
Unbalanced reactions	99	0	220	746
Missing reactions	18	3	1697	1367

4. Inference of metabolic fluxes

4.1 ^{13}C metabolic flux analysis of iPSC-derived midbrain-specific dopaminergic neurons

Steady-state flux inference was performed on the formal sample set collected at a 24-h time point on the basis of a stoichiometric model created in INCA covering 36 reactions (54 fluxes) from central carbon metabolism. As experimental constraints, the INCA flux analysis utilised uptake rates (glucose, glutamate), secretion rates (lactate), and ten standardised metabolite MIDs (glutamate, glutamine, ketoglutarate, succinate, fumarate, malate, citrate, lactate, pyruvate, and phosphoenolpyruvate). The final stationary ^{13}C flux distribution results in **Figure 10** showed glucose served as a primary nutrient to neurons compared to glutamine. Remarkably high glycolytic activity and relatively low metabolic activity in the tricarboxylic acid cycle were observed in neurons, whereas the pentose phosphate pathway showed almost no metabolic activity.

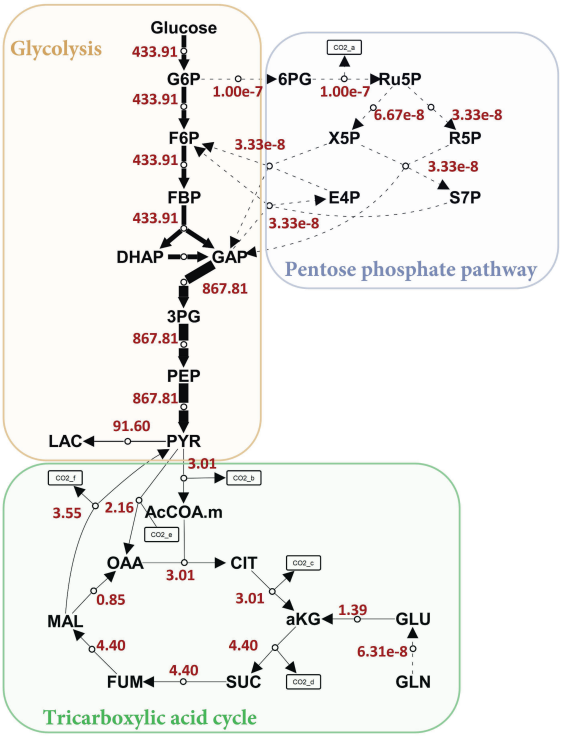

Figure 10. ^{13}C metabolic flux analysis of differentiated dopaminergic neurons within central carbon metabolism network.

A steady-state flux distribution map covering glycolysis, tricarboxylic acid cycle, and pentose phosphate pathway for iPSC-derived midbrain-specific dopaminergic neurons. G6P: Glucose 6-phosphate; F6P: Fructose 6-phosphate; FBP: Fructose 1,6-bisphosphate; DHAP: Dihydroxyacetone phosphate; GAP: or G3P, Glyceraldehyde 3-phosphate; 3PG: 3-Phosphoglycerate; PEP: Phosphoenolpyruvate; PYR: Pyruvate; LAC: Lactate; AcCOA.m: Acetyl coenzyme A in mitochondria; OAA: Oxaloacetate; CIT: Citrate; aKG: alpha-Ketoglutarate; MAL: Malate; FUM: Fumarate; SUC: Succinate; GLN: Glutamine; GLU: Glutamate; 6PG: 6-Phosphogluconate; Ru5P: Ribulose 5-phosphate; R5P: Ribose 5-phosphate; X5P: Xylulose 5-phosphate; E4P: Erythrose 4-phosphate; S7P: Sedoheptulose 7-phosphate.

4.2 Genome-scale flux analysis

Flux balance analysis [50] and entropic flux balance analysis [29] predicted reaction fluxes in the iDopaNeuroC model without any constraints from isotope labelling data. The standardised MIDs (glutamate, glutamine, ketoglutarate, succinate, fumarate, malate, citrate, lactate, pyruvate, phosphoenolpyruvate) were used as new constraints to the iDopaNeuroC

model, and helped to generate a third metabolic flux solution with a newly developed moiety solver algorithm [27]. Three genome-scale flux solutions were compared to INCA flux within the same central carbon metabolism network in order to evaluate their flux rationality.

Regarding an accurate mapping of reaction flux at the genome scale to the corresponding reaction in the core model, we adhered to the rule of combining steady-state flux, as depicted in **Figure 11.A**. The self-defined central carbon metabolism model in INCA only describes the major carbon transitions from glucose without indicating other reaction substances and cofactors. However, metabolic reactions from the iDopaNeuroC model cover multiple reactions related to the same carbon transition and with diverse reaction substances or cofactor conversions involved. For instance, phosphoenolpyruvate can use different diphosphate-nucleotides as reactants and convert into pyruvate. In this sense, we merged the net fluxes of parallel reactions that share the same carbon transitions from the iDopaNeuroC model together. In another case, a simplified reaction was described in the INCA model to represent a series of reactions in the iDopaNeuroC model. For instance, the conversion from 3-phosphoglycerate into phosphoenolpyruvate is via the intermediate metabolite 2-phosphoglycerate. Therefore, the minimum flux of consecutive reactions was used to represent the simplified reaction flux for the iDopaNeuroC model. In a third case, if the intermediate metabolite receives net fluxes both from substrates A and C, the final flux representing conversion from A to C is considered to be zero. Based on the calculated flux distributions in the same core model (*SM4. Table S6*), a Spearman correlation analysis showed that the entropy and moiety flux solution and the INCA model flux solution had a relatively good correlation, however, the FBA flux solution had a poor correlation to any other flux solution (**Figure 11.B**).

In terms of the reaction directionality of the nutrient exchange, the entropy, moiety and INCA model flux solutions all showed the uptake of glucose, glutamate, glutamine, and the secretion of lactate and pyruvate in neurons, whereas the FBA flux solution showed the uptake of lactate and pyruvate and no uptake of glutamine. A comparison of internal reactions was then only focused on entropy, moiety, and INCA model flux solutions. Within the subsystem of glycolysis/gluconeogenesis, all relevant reactions share the same directionality between the three flux solutions. The flux value of the reaction converting pyruvate to lactate seen in the entropy and moiety flux solution fell within the INCA flux range. Several reactions of the pentose phosphate pathway showed in opposite directions to INCA model flux solutions, especially for moiety flux solution. They all, however, had low flux compared to glycolytic reactions. Generally, pyruvate can be converted into acetylCoA via pyruvate dehydrogenase (PDH) and join the tricarboxylic acid cycle. Besides, pyruvate

anaplerosis can covert pyruvate directly into oxaloacetate via pyruvate carboxylase (PC) to compensate for metabolite loss from tricarboxylic acid cycle due to biomass production [55]. Interestingly, the reaction flux of PDH and PC observed in the moiety flux solution both fell within the INCA flux range. Similar results can be also found in a few other reactions. Regardless of this, three reactions of the tricarboxylic acid cycle showed in opposite directions to INCA model flux solutions, as shown in **Figure 11.C**. A final steady-state flux distribution map based on entropy, moiety and INCA model flux solutions can be seen in **Figure 12**.

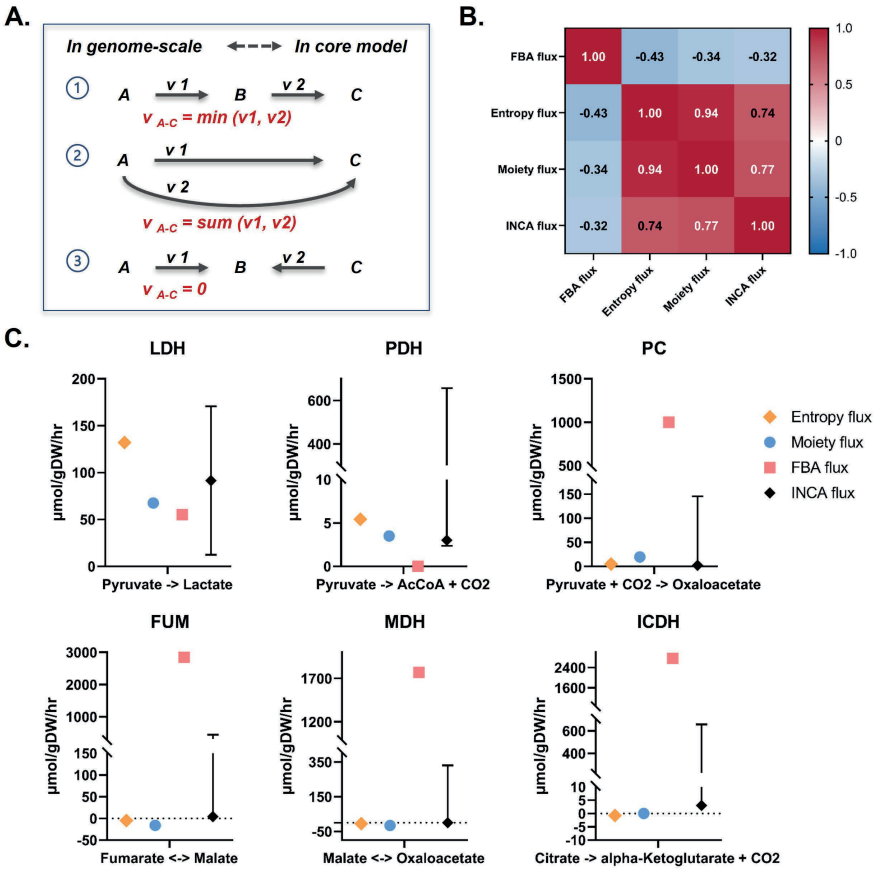

Figure 11. Flux comparison between FBA flux, entropy flux, moiety flux, and INCA model flux within the same central carbon metabolism network for iPSC-derived midbrain-specific dopaminergic neurons.

A. Illustration on mapping reaction fluxes at the genome scale to the corresponding reaction flux in the core model. B. Spearman correlation analysis between FBA flux, entropy flux, moiety flux, and INCA flux within the same network. C. Comparison of individual reaction fluxes from glycolysis and tricarboxylic acid cycle.

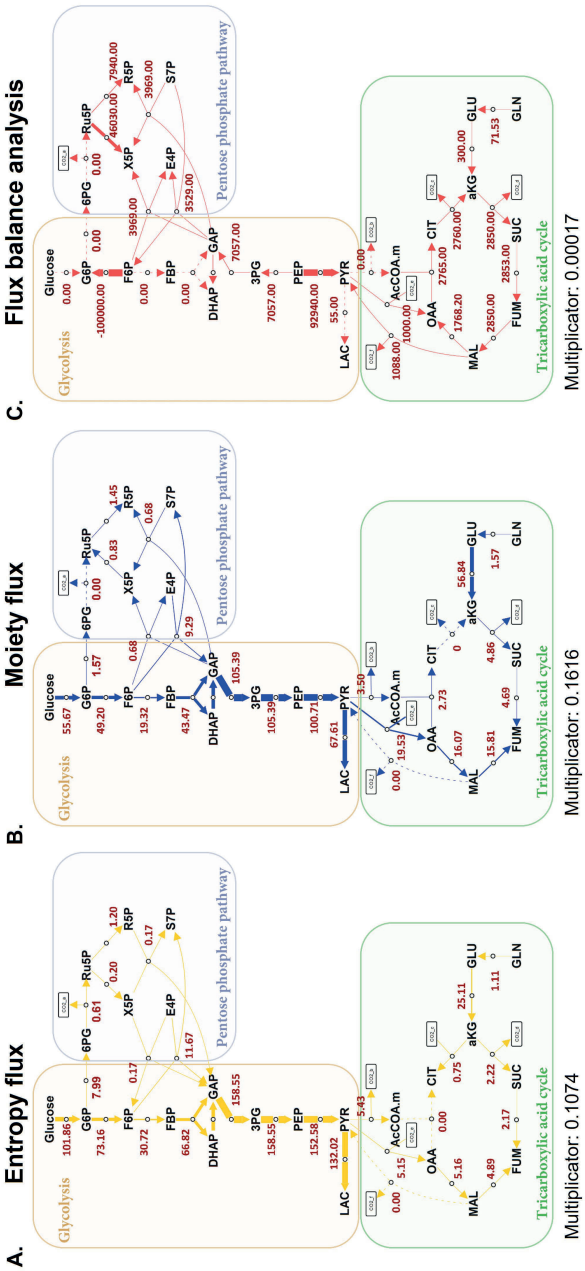

Figure 12. Genome-scale metabolic flux analysis of differentiated dopaminergic neurons projected within central carbon metabolism network.

A steady-state flux distribution map over glycolysis, tricarboxylic acid cycle, and pentose phosphate pathway for iPSC-derived midbrain-specific dopaminergic neurons based on entropy (A, shown in yellow), moiety (B, shown in blue), FBA (C, shown in red) flux solutions.

Besides, the top 15 active pathways within the genome-scale network based on entropic flux balance analysis were summarised in **Figure 13**. In comparison to the results of entropy flux, moiety flux solutions exhibited comparable average flux results in most of the subsystems. Higher flux activity can be found in the subsystems of fructose and mannose metabolism, oxidative phosphorylation, glutamate metabolism, and citric acid cycle, whereas lower flux activity was observed in the subsystem of glycolysis/gluconeogenesis.

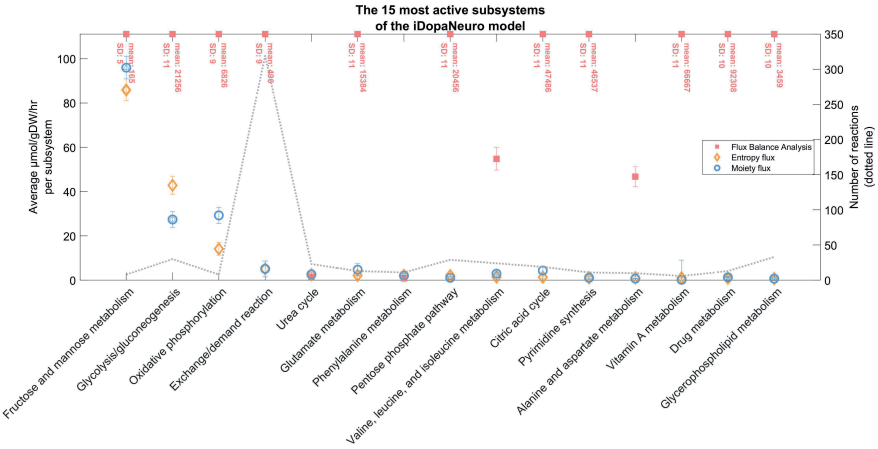

Figure 13. Subsystem activity.

Summary of top 15 active pathway subsystems in the iDopaNeuroC model based on the objective function of entropic flux balance analysis, and the correalted results with FBA flux (red), and moiety flux (blue) analysis.

Regarding the phenotypes for carbon source and energy production, entropy and moiety flux shared some similarities and showed a big difference from FBA flux. Notably, with moiety flux, the reaction of ATP synthase in mitochondria produced the most ATP, whereas phosphoglycerate kinase in the cytoplasm produced the most ATP with entropy flux. Both methods relied on glucose as their primary carbon source. (**Table 4**)

Table 4. Modelling phenotypes.
Comparison of energetic phenotypes obtained using various genome-scale modelling approaches.

	FBA flux	Entropy flux	Moiety flux
	Carbon sources		
Metabolite#1	Bicarbonate	D-Glucose	D-Glucose
Flux	-4268 µmol/gDW/hr	-427.44 umol/gDW/hr	-427.44 umol/gDW/hr
Formula	CHO_3	$C_6H_{12}O_6$	$C_6H_{12}O_6$
Metabolite#2	D-Glucose	L-Arginine	L-Arginine
Flux	-440.40 umol/gDW/hr	-29.35 umol/gDW/hr	-27.77 umol/gDW/hr
Formula	$C_6H_{12}O_6$	$C_6H_{15}N_4O_2$	$C_6H_{15}N_4O_2$
Metabolite#3	L-Threonine	Glycine	Glycine
Flux	-117.30 umol/gDW/hr	-25.81 umol/gDW/hr	-27.06 umol/gDW/hr
Formula	$C_4H_9NO_3$	$C_2H_5NO_2$	$C_2H_5NO_2$
	ATP production		
1st ATP source	Nucleotide interconversion	Phosphoglycerate Kinase	ATP synthase
Flux	$-1e10^5$ umol/gDW/hr	-158.55 umol/gDW/hr	110.88 umol/gDW/hr
2nd ATP source	Succinate- Coenzyme A Ligase	ATP synthase	Phosphoglycerate Kinase
Flux	$-1e10^5$ umol/gDW/hr	52.23 umol/gDW/hr	-105.392 umol/gDW/hr
3rd ATP source	ATP synthase	Pyruvate Kinase	Pyruvate Kinase
Flux	33370 umol/gDW/hr	48.15 µmol/gDW/hr	-34.184 µmol/gDW/hr

5. Identification of conserved moieties and new tracer design

A list of 10 conserved moieties within the atomically resolved iDopaNeuroC model was identified (*SM4. Table S7*). Each conserved moiety corresponded to its own subnetwork. To predict the best conserved moieties to label in a future tracer-based metabolomic experiment, we first compared the coverage of internal reactions in every subnetwork. Conserved moiety #59 had the highest coverage of carrying internal reactions flux. Moreover, four of moiety-relevant subsystems were confirmed as top-15 active subsystems in entropy flux-based solution, and two of moiety-relevant subsystems were confirmed as top-15 active subsystems in moiety flux-based solution. Conserved moieties #5 and #56 showed about 50% internal reaction coverage. Nevertheless, they had no active subsystems hits and conserved moieties #56 was not found within any uptaken metabolite in the whole subnetwork. Overall, the conserved moiety #59 showed the most promising candidate because of its subnetwork activity. To further trace that conserved moiety within network through experiments, a partially $^{13}C_4$, D, $^{15}N_2$, ^{18}O-labelled thymidine can be designed as a new culture tracer (**Figure 14**).

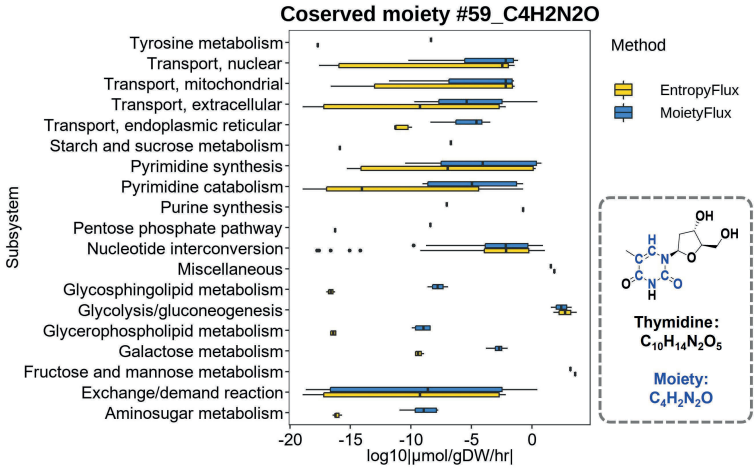

Figure 14. Conserved moiety #59_$C_4H_2N_2O$

Overall reaction flux plot is classified by metabolic subsystem based on entropy flux and moiety flux predictions. Inside the dotted line illustrates a determined tracer that can be utilised in a new tracer-based metabolomic experiment.

Discussion

This work describes a semi-automated pipeline which streamlines and standardises the procedure for preparing an atomically resolved metabolic model and processing tracer-based metabolomics data to perform metabolic flux analysis at genome-scale.

Tracer-based metabolimics data processing

One part of fluxTrAM pipeline is focused on the standardised and high-throughput processing for stable isotope labelling data and ^{13}C metabolic flux analysis. The accurate mass isotopologue distribution of the target metabolites depends on good peak detection and extraction. Based on our evaluation, mzMatch-ISO showed higher accuracy than $X^{13}CMS$ in collecting targeted isotopologue information. X13CMS is a package mainly designed for untargeted metabolomics [6], and mzMatch is a package for both untargeted and targeted metabolomics, mzMatch-ISO especially showed its unique advantage in isotopic labelling data analysis with a list of target analytes. The function of ***peakDetector*** based on the mzMatch-ISO package presented good consistency in the isotopologue faction analysis for essential amino acids between the labelled and unlabelled groups while showing significant differences for non-essential amino acids. The automated connection between different packages guaranteed that no additional data format manipulation occurs, leading to a consecutive data processing workflow. This was demonstrated in the analysis of a time-course dataset from the formal study set. In addition, the pipeline is flexible in that it allows self-defined parameter settings, and the output result for each function step

can be freely exported. It is currently only applicable to processing LC-MS datasets due to the restriction of selected automated peak extraction package. When appropriate packages are developed by the metabolomics community, future development can be extended to LC-MS/MS and GC-MS datasets.

In the application of ^{13}C metabolic flux analysis for iPSC-derived dopaminergic neurons, reaction flux was obtained within a central carbon metabolism network model in INCA software. The experimental flux range within 95% confidence interval is valuable in reflecting the basic energy metabolism regulation on a healthy neuronal model. The net reaction flux map showed a very active glycolysis activity compared to the tricarboxylic acid (TCA) cycle and the pentose phosphate pathway. However, this model was built based on many assumptions to obtain a successful fit. Only a few nutrient transports (glucose, glutamine, glutamate) were described in the INCA model compared to 320 exchange reactions in the iDopaNeuroC model. Oxidative sources of unlabelled carbon such as amino acids and fatty acids in the TCA cycle were not considered in this model. The output carbon flux was limited to the secretion of lactate and pyruvate. Besides, metabolites were assigned in one single pool without a compartmental difference. Although the experimental flux solution was based on limited network size and many pre-defined assumptions, it still offers a flux solution that can indicate a feasible cellular activity with statistically approved goodness-of-fit and biological significance.

Metabolite structure and atom mapping databases

Obtaining metabolite identifiers is critical for expanding the database's coverage. Similarly, having a wide range of sources increases the probability of obtaining the metabolites that best represent those described in the genome-scale model. An InChI represents various metabolite structures depending on the level of detail included in it (see *Figure S1*). The main layer consists of layers 1 (standardised or not), 2 (chemical formula), 3 (atoms connectivity), and 4 (hydrogen atom connectivity), e.g., InChI=1S/C3H7NO2/c1-2(4)3(5)6/h2H,4H2,1H3,(H,5,6), which represents the alanine molecule; layer 5 adds the charges, e.g., adding p-1, representing the alaninate molecule; and finally, layer 6 indicates the stereochemistry, e.g., adding t2-/m0/s, representing the L-alaninate molecule. As the specificity of the metabolites increases, so does the number of atomically balanced reactions in the database. A high quality atomically resolved genome-scale metabolic model may have multiple applications, some of which were explored in this work. Among the applications are the estimation of fluxes in the metabolic network in combination with MID data [27], identifying conserved moieties in the metabolic network to optimise tracer-based experimental design [25,26], to calculate the potential energy in molecules [56] or to

identify the amount of enthalpy change associated with a metabolic reaction via bond enthalpies (See supplementary information).

Metabolite structures may be obtained in a variety of ways, including drawing based on the literature using chemoinformatic software [23] or obtained from metabolic databases either manually or computationally given various different identifiers. Different approaches exist to differentiate isomers in order to obtain a consistent representation of a metabolite structure from a biological system, such as the CAS Registry Number, a unique numerical identifier that is independent of any chemical nomenclature system and is widely used to identify chemical substances with a maximum capacity of 1 e9 identifiers [57]; or an InChI, representing a chemical structure in a standardised manner, allowing the generation of a unique identifier for a chemical structure [40], while other chemoinformatic formats such as SMILES [41] and chemical tables [44], may use different annotations to represent the same molecule. Nevertheless, assigning the correct identifier for a metabolite in a genome-scale model is a difficult task due to the various isomers that a molecule may have, and more work should be done in this area to avoid mismatches. This is the case of the metabolite nicotinamide adenine dinucleotide, which is represented in the iDopaNeuroC model [29] with the identifier CHEBI:15846 from the ChEBI database, whereas the consistent identifier based on the number of hydrogen atoms should be CHEBI:57540. Tools for mapping identifiers between different biological databases [58], such as BridgeDB [59], the chemical translation service [58,60] and MetaboAnnotator [61], can be useful to assigning metabolite identifiers in accordance with the genome-scale model.

Integrative metabolic flux analysis and new experimental tracer design

FBA was used to predict the steady-state flux distribution assuming maximization of neuronal ATP production. Entropic flux balance analysis does not maximise the flux of a single reaction but rather assumes the objective function is to maximise the entropy of every forward and reverse internal reaction flux and also predicts a thermodynamically feasible flux vector. In addition, the moiety fluxomics predicts a new genome-scale flux solution in the atomically resolved iDopaNeuroC model by integrating the labelled MID data as extra constraints. The latter two approaches outperformed the FBA method as evaluated by their higher correlations with the flux inference using conventional metabolic flux analysis with INCA while using a central carbon metabolism network model. Moiety flux inference [27]shares some similarities with the entropic flux balance analysis prediction in terms of metabolic phenotype. For instance, glucose was predicted as one of the primary carbon sources for dopaminergic neuron. Nevertheless, it also shows some inconsistencies with entropy modelling solution. Moiety flux predicted a very active ATP production via ATP synthase (four protons for one ATP) in mitochondria, while entropy flux predicted a major

ATP production from cytoplasm. While maintaining a comparable ATP production, incorporating labelled MID data reformed some active subsystems of the neuronal metabolism, shown as more active oxidative phosphorylation, glutamate metabolism, and citric acid cycle, and less active glycolysis/gluconeogenesis.

Tracer choice plays an important role in determining the precision of network flux estimation [62,63]. We applied U-$^{13}C_6$-glucose as the tracer for iPSC-derived dopaminergic neurons, which is not an ideal tracer to elucidate flux through the pentose phosphate pathway. Although glucose is a major nutrient used for energy production, neurons can utilise multiple other nutrient sources as well. To improve the accuracy of genome-scale flux prediction, a validation experiment with a new tracer is further required. Identification of conserved moieties has strong potential for use in design of tracer-based metabolomic experiments [64]. By isotopically labelling any single atom in a conserved moiety, one can use the predict the reachable set reactions that could contain that isotopic label.

In our study, the results of the genome-scale flux prediction helped to determine optimal tracers by evaluating flux activity of candidate conserved moieties on intracellular metabolism. This will facilitate future study of particularly significant metabolic pathways in more detail. On the other hand, it also contributes to the systematic interaction between genome-scale modelling and experimental validation towards a more efficient level. The conserved moiety proposed in this study, present in the metabolite thymidine, was found in multiple subsystems making it a potential candidate for the design of novel labelled molecules for a tracer-based metabolomics experiments. Ikuno reported an increased level of thymidine in the plasma of prodromal Parkinson's disease mouse model, which can cause mitochondrial disorders characterized by mitochondrial DNA (mtDNA) depletion [65]. In Anandagopu *et al.* [66], the distribution of thymidine was found to be decreased in PINK1 mRNA *homo sapiens* sequences, a mutation associated with mitochondrial dysfunction and oxidative stress. Not all of the conserved moieties in the iDopaNeuroC were identified. This was hampered by the absence of metabolite structures for some reactants, e.g., lipids with R groups in the structure, as they precluded the atomic resolution of all reactions in the iDopaNeuroC model. However, it was possible to atom map the majority of internal reactions, which permitted the identification of the majority of conserved moieties [25,26] in the iDopaNeuroC model.

Conclusions

Unlike a snapshot of metabolic phenotypes reflected by metabolite concentration, metabolic fluxes offer a functional perspective on cellular metabolism. Current methods for flux inference are often restricted to use with small-scale metabolic networks and are rarely applicable at a genome-scale level. To standardise and streamline the implementation of

metabolic flux analysis at genome-scale, we developed a semi-automated pipeline, fluxTrAM, that enables the integration of tracer-based metabolomics data with atomically resolved metabolic networks for metabolic flux inference. The first module of fluxTrAM implemented automated processing of tracer-based LC-MS raw data into standardised MID data. The second module atomically resolves any given genome-scale metabolic model, and results in the creation of a chemoinformatic database of standardized and context-specific metabolite structures and atom-mapped reactions. We demonstrated the utility of this pipeline by integrating experimental mass isotopomer distribution data with an atomically resolved dopaminergic neuronal model using a novel moiety fluxomics method that enables flux inference at genome-scale. The fluxes inferred by moiety fluxomics showed good correlation with fluxes inferred using conventional ^{13}C flux solution obtained based on a central carbon metabolism network. To conclude on the relative merits of these approaches, further analyses are necessary with a genome-scale model that is suitable for both conventional and moiety fluxomic approaches to flux inference. Genome-scale flux inference has the potential to broaden the analysis of metabolic flux beyond central carbon metabolism and combined with the conserved moiety analysis, enables design of labelled tracer for new labelling experiments.

Data Availability:

The fluxTrAM pipeline can be found in the COBRA Toolbox v3.4 [13] (https://github.com/opencobra/COBRA.tutorials/tree/master/analysis/tracerMetabolomics2atomResolvedModels). The metabolite structures and atom-mapped reactions generated by the fluxTrAM pipeline from the genome-scale models iDopaNeuroC [29], E.coli core [47], AGORA2 [54] and Recon3D [45] can be found on GitHub (https://github.com/opencobra/COBRA.papers/2023_tracerMetabolomics2atomResolvedModels) and in the VMH database [37].

Scripts for running flux comparison analysis can be found on GitHub (https://github.com/opencobra/COBRA.papers/2023_tracerMetabolomics2atomResolvedModels)

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement number 668738, the China scholarship council [No.201806210057], the Dutch National Institutes of Health (ZonMw) TKI-LSH Neuromet project (LSHM18092) and the Dutch Research Council (NWO) 'Investment Grant NWO Large' program, for the 'Building the infrastructure for Exposome research: Exposome-Scan' project (No. 175.2019.032).

References

1. Basler G, Fernie AR, Nikoloski Z. Advances in metabolic flux analysis toward genome-scale profiling of higher organisms. Bioscience Reports. 2018;38: BSR20170224. doi:10.1042/BSR20170224
2. Antoniewicz MR. A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications. Metabolic Engineering. 2021;63: 2–12. doi:10.1016/j.ymben.2020.11.002
3. Jiang L, Shestov AA, Swain P, Yang C, Parker SJ, Wang QA, et al. Reductive carboxylation supports redox homeostasis during anchorage-independent growth. Nature. 2016;532: 255–258. doi:10.1038/nature17393
4. McAtee Pereira AG, Walther JL, Hollenbach M, Young JD. 13C Flux Analysis Reveals that Rebalancing Medium Amino Acid Composition can Reduce Ammonia Production while Preserving Central Carbon Metabolism of CHO Cell Cultures. Biotechnology Journal. 2018;13: 1700518. doi:10.1002/biot.201700518
5. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Anal Chem. 2006;78: 779–787. doi:10.1021/ac051437y
6. Huang X, Chen Y-J, Cho K, Nikolskiy I, Crawford PA, Patti GJ. X13CMS: Global Tracking of Isotopic Labels in Untargeted Metabolomics. Anal Chem. 2014;86: 1632–1639. doi:10.1021/ac403384n
7. Chokkathukalam A, Jankevics A, Creek DJ, Achcar F, Barrett MP, Breitling R. mzMatch–ISO: an R tool for the annotation and relative quantification of isotope-labelled mass spectrometry data. Bioinformatics. 2013;29: 281–283. doi:10.1093/bioinformatics/bts674
8. Bueschl C, Kluger B, Neumann NKN, Doppler M, Maschietto V, Thallinger GG, et al. MetExtract II: A Software Suite for Stable Isotope-Assisted Untargeted Metabolomics. Anal Chem. 2017;89: 9518–9526. doi:10.1021/acs.analchem.7b02518
9. Zierer J, Menni C, Kastenmüller G, Spector TD. Integration of 'omics' data in aging research: from biomarkers to systems biology. Aging Cell. 2015;14: 933–944. doi:10.1111/acel.12386
10. Cisek K, Krochmal M, Klein J, Mischak H. The application of multi-omics and systems biology to identify therapeutic targets in chronic kidney disease. Nephrology Dialysis Transplantation. 2016;31: 2003–2011. doi:10.1093/ndt/gfv364
11. Das T, Andrieux G, Ahmed M, Chakraborty S. Integration of Online Omics-Data Resources for Cancer Research. Frontiers in Genetics. 2020;11. Available: https://www.frontiersin.org/articles/10.3389/fgene.2020.578345
12. Palsson BØ. Systems Biology: Constraint-based Reconstruction and Analysis. Cambridge University Press; 2015.
13. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc. 2019;14: 639–702. doi:10.1038/s41596-018-0098-2
14. Pey J, Rubio A, Theodoropoulos C, Cascante M, Planes FJ. Integrating tracer-based metabolomics data and metabolic fluxes in a linear fashion via Elementary Carbon Modes. Metabolic Engineering. 2012;14: 344–353. doi:10.1016/j.ymben.2012.03.011
15. Weitzel M, Nöh K, Dalman T, Niedenführ S, Stute B, Wiechert W. 13CFLUX2—high-performance software suite for 13C-metabolic flux analysis. Bioinformatics. 2013;29: 143–145. doi:10.1093/bioinformatics/bts646
16. Young JD. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. Bioinformatics. 2014;30: 1333–1335. doi:10.1093/bioinformatics/btu015
17. Quek L-E, Wittmann C, Nielsen LK, Krömer JO. OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. Microbial Cell Factories. 2009;8: 25. doi:10.1186/1475-2859-8-25
18. Foguet C, Jayaraman A, Marin S, Selivanov VA, Moreno P, Messeguer R, et al. p13CMFA: Parsimonious 13C metabolic flux analysis. PLOS Computational Biology. 2019;15: e1007310. doi:10.1371/journal.pcbi.1007310
19. Kumar A, Maranas CD. CLCA: Maximum Common Molecular Substructure Queries within the MetRxn Database. J Chem Inf Model. 2014;54: 3417–3438. doi:10.1021/ci5003922
20. Preciat Gonzalez GA, El Assal LRP, Noronha A, Thiele I, Haraldsdóttir HS, Fleming RMT. Comparative evaluation of atom mapping algorithms for balanced metabolic reactions: application to Recon 3D. Journal of Cheminformatics. 2017;9: 39. doi:10.1186/s13321-017-0223-1
21. Huß S, Judd RS, Koper K, Maeda HA, Nikoloski Z. An automated workflow that generates atom mappings for large-scale metabolic models and its application to Arabidopsis thaliana. The Plant Journal. 2022;111: 1486–1500. doi:10.1111/tpj.15903
22. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. Journal of Cheminformatics. 2011;3: 33. doi:10.1186/1758-2946-3-33
23. Standardizer Transform | Chemaxon Docs. [cited 15 Jan 2023]. Available: https://docs.chemaxon.com/display/docs/standardizer-transform.md
24. Rahman SA, Torrance G, Baldacci L, Martínez Cuesta S, Fenninger F, Gopal N, et al. Reaction Decoder Tool (RDT): extracting features from chemical reactions. Bioinformatics. 2016;32: 2065–2066. doi:10.1093/bioinformatics/btw096
25. Haraldsdóttir HS, Fleming RMT. Identification of Conserved Moieties in Metabolic Networks by Graph Theoretical Analysis of Atom Transition Networks. PLOS Computational Biology. 2016;12: e1004999. doi:10.1371/journal.pcbi.1004999
26. Ghaderi S, Haraldsdóttir HS, Ahookhosh M, Arreckx S, Fleming RMT. Structural conserved moiety splitting of a stoichiometric matrix. Journal of Theoretical Biology. 2020;499: 110276. doi:10.1016/j.jtbi.2020.110276

27. Fleming RMT, Haraldsdottir HS, Preciat G, Huang L, Thiele I, Harms A, et al. Conserved moiety fluxomics.
28. Hosseinkhani F, Huang L, Dubbelman A-C, Guled F, Harms AC, Hankemeier T. Systematic Evaluation of HILIC Stationary Phases for Global Metabolomics of Human Plasma. Metabolites. 2022;12: 165. doi:10.3390/metabo12020165
29. Preciat G, Moreno EL, Wegrzyn AB, Willacey CCW, Modamio J, Monteiro FL, et al. Mechanistic model-driven exometabolomic characterisation of human dopaminergic neuronal metabolism. bioRxiv; 2021. p. 2021.06.30.450562. doi:10.1101/2021.06.30.450562
30. Rahim M, Ragavan M, Deja S, Merritt ME, Burgess SC, Young JD. INCA 2.0: A tool for integrated, dynamic modeling of NMR- and MS-based isotopomer measurements and rigorous metabolic flux analysis. Metabolic Engineering. 2022;69: 275–285. doi:10.1016/j.ymben.2021.12.009
31. Reinhardt P, Glatza M, Hemmer K, Tsytsyura Y, Thiel CS, Höing S, et al. Derivation and expansion using only small molecules of human neural progenitors for neurodegenerative disease modeling. PLoS One. 2013;8: e59252. doi:10.1371/journal.pone.0059252
32. Moreno EL, Hachi S, Hemmer K, Trietsch SJ, Baumuratov AS, Hankemeier T, et al. Differentiation of neuroepithelial stem cells into functional dopaminergic neurons in 3D microfluidic cell culture. Lab Chip. 2015;15: 2419–2428. doi:10.1039/C5LC00180C
33. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A Cross-platform Toolkit for Mass Spectrometry and Proteomics. Nat Biotechnol. 2012;30: 918–920. doi:10.1038/nbt.2377
34. Agrawal S, Kumar S, Sehgal R, George S, Gupta R, Poddar S, et al. El-MAVEN: A Fast, Robust, and User-Friendly Mass Spectrometry Data Processing Engine for Metabolomics. In: D'Alessandro A, editor. High-Throughput Metabolomics: Methods and Protocols. New York, NY: Springer; 2019. pp. 301–321. doi:10.1007/978-1-4939-9236-2_19
35. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 2010;26: 966–968. doi:10.1093/bioinformatics/btq054
36. Millard P, Delépine B, Guionnet M, Heuillet M, Bellvert F, Létisse F. IsoCor: isotope correction for high-resolution MS labeling experiments. Bioinformatics. 2019;35: 4484–4487. doi:10.1093/bioinformatics/btz209
37. Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, Preciat G, et al. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Res. 2019;47: D614–D624. doi:10.1093/nar/gky992
38. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research. 2016;44: D457–D462. doi:10.1093/nar/gkv1070
39. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res. 2018;46: D608–D617. doi:10.1093/nar/gkx1089
40. Heller SR, McNaught A, Pletnev I, Stein S, Tchekhovskoi D. InChI, the IUPAC International Chemical Identifier. Journal of Cheminformatics. 2015;7: 23. doi:10.1186/s13321-015-0068-4
41. Anderson E, Veith G, Weininger D. SMILES: a line notation and computerized interpreter for chemical structures. 1987. Available: https://www.semanticscholar.org/paper/SMILES%3A-a-line-notation-and-computerized-for-Anderson-Veith/ae92190b303a24d6df28df571dd1631ad95e8af3
42. Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, et al. PubChem Substance and Compound databases. Nucleic Acids Res. 2016;44: D1202–D1213. doi:10.1093/nar/gkv951
43. Hastings J, de Matos P, Dekker A, Ennis M, Harsha B, Kale N, et al. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. Nucleic Acids Res. 2013;41: D456-463. doi:10.1093/nar/gks1146
44. Dalby A, Nourse JG, Hounshell WD, Gushurst AKI, Grier DL, Leland BA, et al. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. J Chem Inf Comput Sci. 1992;32: 244–255. doi:10.1021/ci00007a012
45. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nat Biotechnol. 2018;36: 272–281. doi:10.1038/nbt.4072
46. Grethe G, Blanke G, Kraut H, Goodman JM. International chemical identifier for reactions (RInChI). Journal of Cheminformatics. 2018;10: 22. doi:10.1186/s13321-018-0277-8
47. Orth JD, Fleming RMT, Palsson BØ. Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide. EcoSal Plus. 2010;4. doi:10.1128/ecosalplus.10.2.1
48. Heinken A, Hertel J, Acharya G, Ravcheev DA, Nyga M, Okpala OE, et al. Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. Nat Biotechnol. 2023; 1–12. doi:10.1038/s41587-022-01628-0
49. Young JD, Walther JL, Antoniewicz MR, Yoo H, Stephanopoulos G. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. Biotechnol Bioeng. 2008;99: 686–699. doi:10.1002/bit.21632
50. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nat Biotechnol. 2010;28: 245–248. doi:10.1038/nbt.1614
51. Pissadaki EK, Bolam JP. The energy cost of action potential propagation in dopamine neurons: clues to susceptibility in Parkinson's disease. Front Comput Neurosci. 2013;7: 13. doi:10.3389/fncom.2013.00013

52. Fleming RMT, Maes CM, Saunders MA, Ye Y, Palsson BØ. A variational principle for computing nonequilibrium fluxes and potentials in genome-scale biochemical networks. Journal of Theoretical Biology. 2012;292: 71–77. doi:10.1016/j.jtbi.2011.09.029
53. Rohn H, Junker A, Hartmann A, Grafahrend-Belau E, Treutler H, Klapperstück M, et al. VANTED v2: a framework for systems biology applications. BMC Systems Biology. 2012;6: 139. doi:10.1186/1752-0509-6-139
54. Heinken A, Acharya G, Ravcheev DA, Hertel J, Nyga M, Okpala OE, et al. AGORA2: Large scale reconstruction of the microbiome highlights wide-spread drug-metabolising capacities. bioRxiv; 2020. p. 2020.11.09.375451. doi:10.1101/2020.11.09.375451
55. Buescher JM, Antoniewicz MR, Boros LG, Burgess SC, Brunengraber H, Clish CB, et al. A roadmap for interpreting 13C metabolite labeling patterns from cells. Curr Opin Biotechnol. 2015;34: 189–201. doi:10.1016/j.copbio.2015.02.003
56. Wang L, Upadhyay V, Maranas CD. dGPredictor: Automated fragmentation method for metabolic reaction free energy prediction and de novo pathway design. PLOS Computational Biology. 2021;17: e1009448. doi:10.1371/journal.pcbi.1009448
57. Jacobs A, Williams D, Hickey K, Patrick N, Williams AJ, Chalk S, et al. CAS Common Chemistry in 2021: Expanding Access to Trusted Chemical Information for the Scientific Community. J Chem Inf Model. 2022;62: 2737–2743. doi:10.1021/acs.jcim.2c00268
58. Haraldsdóttir HS, Thiele I, Fleming RM. Comparative evaluation of open source software for mapping between metabolite identifiers in metabolic network reconstructions: application to Recon 2. J Cheminform. 2014;6: 2. doi:10.1186/1758-2946-6-2
59. van Iersel MP, Pico AR, Kelder T, Gao J, Ho I, Hanspers K, et al. The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services. BMC Bioinformatics. 2010;11: 5. doi:10.1186/1471-2105-11-5
60. Wohlgemuth G, Haldiya PK, Willighagen E, Kind T, Fiehn O. The Chemical Translation Service—a web-based tool to improve standardization of metabolomic reports. Bioinformatics. 2010;26: 2647–2648. doi:10.1093/bioinformatics/btq476
61. Thiele I, Preciat G, Fleming RMT. MetaboAnnotator: an efficient toolbox to annotate metabolites in genome-scale metabolic reconstructions. Bioinformatics. 2022;38: 4831–4832. doi:10.1093/bioinformatics/btac596
62. Metallo CM, Walther JL, Stephanopoulos G. Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells. J Biotechnol. 2009;144: 167–174. doi:10.1016/j.jbiotec.2009.07.010
63. Antoniewicz MR. 13C metabolic flux analysis: optimal design of isotopic labeling experiments. Current Opinion in Biotechnology. 2013;24: 1116–1121. doi:10.1016/j.copbio.2013.02.003
64. Wiechert W. 13C Metabolic Flux Analysis. Metabolic Engineering. 2001;3: 195–206. doi:10.1006/mben.2001.0187
65. Ikuno M, Yamakado H, Amano I, Hatanaka Y, Uemura N, Matsuzawa S-I, et al. Mitochondrial dysfunction in a mouse model of prodromal Parkinson's disease: A metabolomic analysis. Neurosci Lett. 2021;765: 136267. doi:10.1016/j.neulet.2021.136267
66. Anandagopu P, Banu S, Li J. Low thymine content in PINK1 mRNAs and insights into Parkinson's disease. Bioinformation. 2010;4: 452–455. doi:10.6026/97320630004452

Supplementary Materials

1. Inconsistencies in atom mapping data

Not all of the metabolites from the HMDB database were collected, resulting in a low collection of metabolites. This was due to connection problems with the database, as shown by the The HyperText Transfer Protocol (HTTP) status 503, which indicates that the server is not ready to handle the request.

Inconsistent metabolites or reactions can cause unbalanced reactions. With the iDopaNeuroC model, 15 metabolites were always present in at least one of the 100 unbalanced reactions; of these metabolites, 9 have non-chemical atoms that are used to represent a variable metabolite pool or to highlight the reacting zone. Additionally, there are inconsistencies between the information from the sources and the information in the iDopaNeuroC model in 11 metabolites, such as metabolite 9E-elaidic acid, where the InChI with the highest score has a different chemical formula than the formula described in the iDopaNeuroC model, C18H34O2 in all the sources and in the model $C_{39}H_{64}N_7O_{17}P_3S$. Furthermore, because the metabolites were obtained from open databases, the model's identifiers in some cases represent a molecule with no stereochemistry or charge, as demonstrated by the metabolite L-arginine, from which the VMH alone obtained the highest score. L-arginine is represented in on the VMH [31] as arg_L, in KEGG [39] as C00062 in PubChem [43] as 6322 , in HMDB [40] as HMDB00517 and in ChEBI [44] as CHEBI:16467.

1.1 Inconsistencies caused by R groups

The R groups can cause inconsistencies because they represent a substructure that, because they are not defined, can result in unbalanced reactions, such as reaction sphingomyelin phosphodiesterase (VMH ID: HMR_0795).

Reaction	VMH ID	Formula
Glucosylceramidase	GBA	h2o[c] + gluside_hs[c] -> glc_D[c] + crm_hs[c]
Glucosylceramidase (lysosome)	GBAl	h2o[l] + gluside_hs[l] -> crm_hs[l] + glc_D[l]
S-Adenosyl-L-Methionine: Phosphatidylethanolamine N-Methyltransferase	HMR_0653	amet[c] + pe_hs[c] -> 2 h[c] + ahcys[c] + M02686[c]
Ceramide glucosyltransferase	HMR_0761	crm_hs[c] + udpg[c] -> h[c] + udp[c] + gluside_hs[c]
Sphingomyelin phosphodiesterase	HMR_0795	h2o[c] + sphmyln_hs[c] -> h[c] + crm_hs[c] + cholp[c]
Phosphatidic acid phosphatase	PPAP	h2o[c] + pa_hs[c] -> pi[c] + dag_hs[c]
Phosphatidylserine decarboxylase	PSDm_hs	h[m] + ps_hs[m] -> co2[m] + pe_hs[m]
Phosphatidylserine synthase	PSSA1_hs	ser_L[c] + pchol_hs[c] <=> chol[c] + ps_hs[c]
Phosphatidylserine synthase	PSSA2_hs	ser_L[c] + pe_hs[c] <=> etha[c] + ps_hs[c]
Psdm_Hsc	PSDm_hsc	h[c] + ps_hs[c] -> co2[c] + pe_hs[c]
Sphingomyelin synthase	SMS	pchol_hs[c] + crm_hs[c] -> dag_hs[c] + sphmyln_hs[c]

1.2 Inconsistencies caused by protons

Inconsistencies with protons are caused by metabolite pH, which changes the number of hydrogens, resulting in unbalanced reactions, such as reaction N-Acyl-Aliphatic-L-Amino Acid Amidohydrolase (VMH ID: RE2642C)

Reaction	VMH ID	Formula
Cardiolipin synthase	CLS_hs	cdpdag_hs[c] + pglyc_hs[c] -> h[c] + cmp[c] + clpn_hs[c]
Glucosylceramidase	GBA	h2o[c] + gluside_hs[c] -> glc_D[c] + crm_hs[c]

Glucosylceramidase (lysosome)	GBAl	h2o[l] + gluside_hs[l] -> crm_hs[l] + glc_D[l]
S-adenosyl-L-methionine: phosphatidyl-N-dimethylethanolamine N-methyltransferase	HMR_0657	amet[c] + M02758[c] -> ahcys[c] + pchol_hs[c]
Ceramide glucosyltransferase	HMR_0761	crm_hs[c] + udpg[c] -> h[c] + udp[c] + gluside_hs[c]
Metabolism of LeuSerTrp (formation/degradation)	LEUSERTRPr	2 h2o[c] + leusertrp[c] <=> ser_L[c] + leu_L[c] + trp_L[c]
Lysophospholipase	LPASE	h2o[c] + lpchol_hs[c] -> h[c] + Rtotal[c] + g3pc[c]
Nicotinate D-ribonucleoside kinase	NICRNS	atp[c] + nicrns[c] -> h[c] + adp[c] + nicrnt[c]
Nucleotide phosphatase	NP1	h[c] + nac[c] + r1p[c] -> pi[c] + nicrns[c]
Phospholipase A2	PLA2_2	h2o[c] + pchol_hs[c] -> h[c] + Rtotal2[c] + lpchol_hs[c]
Phospholipase A2 (extracellular)	PLA2_2e	h2o[e] + pchol_hs[e] -> h[e] + lpchol_hs[e] + Rtotal2[e]
RE1266C	RE1266C	o2[c] + 4mop[c] -> h[c] + co2[c] + CE2028[c]
Aminoacid N-acetyltransferase	RE2031M	accoa[m] + ala_L[m] <=> h[m] + coa[m] + CE1554[m]
N-acyl-aliphatic-L-amino acid amidohydrolase	RE2642C	h2o[c] + CE1554[c] <=> ac[c] + ala_L[c]

1.3 Missing reactions

Finally, three of the missing reactions have stoichiometry that cannot be represented with integers, resulting in the failure to generate an MDL RXN reaction. Additionally, not all metabolites were present in the degradation (VMH id: HDL_HSDEG) and formation (VMH id: HDL_HSSYN) of HDL due to the limited information available about metabolites in the iDopaNeuroC model.

Reaction	VMH ID	Formula
Cytochrome C oxidase; mitochondrial complex IV	CYOOm3	o2[m] + 7.92 h[m] + 4 focytC[m] -> 1.96 h2o[m] + 4 h[c] + 4 ficytC[m] + 0.02 o2s[m]
Degradation of HDL	HDL_HSDEG	h2o[e] + hdl_hs[e] -> 2 chsterol[e] + 2 pchol_hs[e] + Rtotal[e] + Rtotal2[e] + Rtotal3[e] + glyc[e] + HC00004[e] + HC00006[e] + HC00007[e] + HC00008[e] + HC00009[e]
Formation of HDL	HDL_HSSYN	2 chsterol[e] + 2 pchol_hs[e] + tag_hs[e] + HC00004[e] + HC00006[e] + HC00007[e] + HC00008[e] + HC00009[e] -> hdl_hs[e]
HMR_0017	HMR_0017	M02909[e] -> 0.125 octa[e] + 0.125 dca[e] + 0.125 but[e] + 0.125 C01601[e] + 0.125 M02108[e] + 0.125 M03117[e] + 0.125 M03134[e] + 0.125 caproic[e]
Phosphate transport via Na+ symporter	PIt8	1.5 na1[e] + pi[e] <=> pi[c] + 1.5 na1[c]

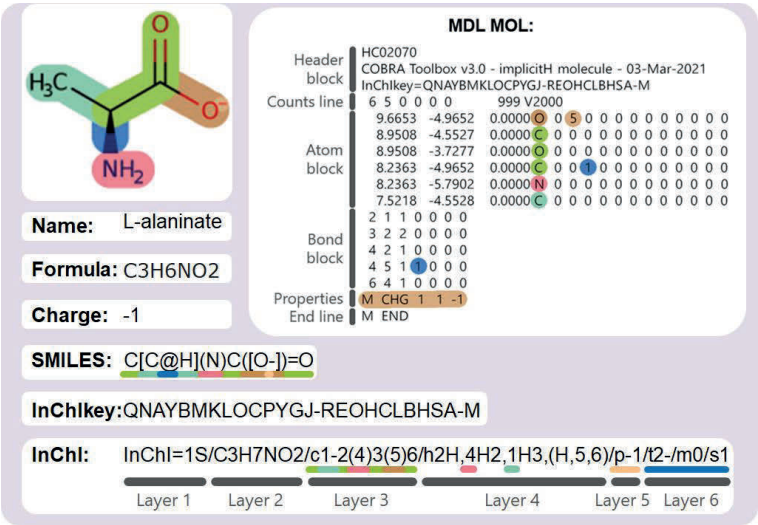

Figure S1. Comparison of metabolite structure formats.

An L-alaninate molecule represented by a hydrogen-suppressed molecular graph (implicit hydrogens). The main branch of the molecule can be seen in green; the additional branches can be seen in brown, pink and turquoise. The stereochemistry of the molecule is highlighted in blue, the double bond in dark green and the charges in light brown. The same colours are used to indicate where this information is represented in the different metabolite strucure formats. The InChI is divided into layers, each of which begins with a lowercase letter, except for Layers 1 and 2. Layer 1 indicates if the InChI is standardised. Layer 2 shows the chemical formula in a neutral state. Layer 3 indicates the connectivity between the atoms (ignoring hydrogen atoms). Layer 4 demonstrates the connectivity of the hydrogen atom. Layer 5 indicates the charge of the molecule. Layer 6 shows the stereochemistry. Additional layers can be added, but they cannot be represented with a standard InChI.

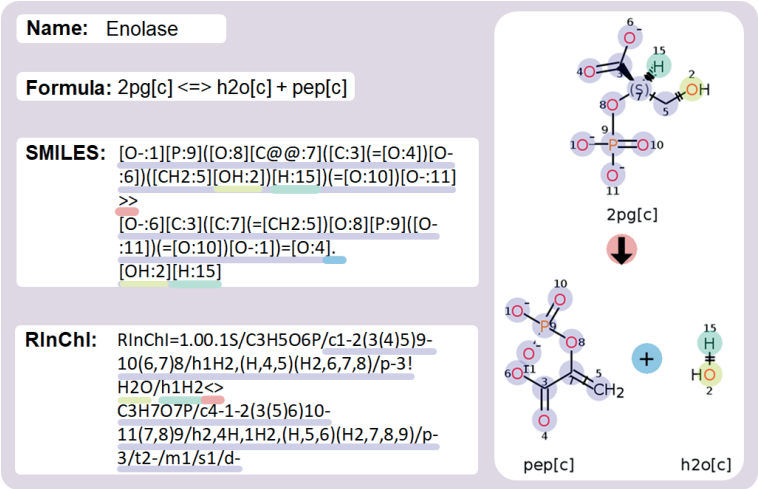

Figure S2. Chemoinformatic formats for chemical reactions.

Atom-mapped representation of enolase reaction in different chemoinformatic formats. The larger moiety in the reaction is shown in blue, the hydroxide moiety in yellow, the hydrogen atom in green, the sum sign in dark blue and the reaction sign in red.

Table S1. Metabolites targeted in the mass isotopologue distribution analysis by fluxTrAM

KEGG id	Name	Formula	LC-MS retention time
C00064	Glutamine	C5H10N2O3	9.952
C00025	Glutamate	C5H9NO4	9.206
C00074	Phosphoenolpyruvate	C3H5O6P	10.57
C00022	Pyruvate	C3H4O3	3.911
C00186	Lactate	C3H6O3	5.587
C00026	Ketoglutarate	C5H6O5	7.272
C00042	Succinate	C4H6O4	6.18
C00122	Fumarate	C4H4O4	8.049
C00149	Malate	C4H6O5	7.979
C00158	Citrate	C6H8O7	12.54

Table S2. List of reactions and atom transitions for the central carbon metabolism model created in INCA

ID	Equation
GLCt1r	glc_D.e (abcdef) -> glc_D (abcdef)
HEX1	glc_D (abcdef) -> g6p (abcdef)
PGI	g6p (abcdef) -> f6p (abcdef)
PFK	f6p (abcdef) -> fdp (abcdef)
FBA	fdp (abcdef) -> dhap (cba) + g3p (def)
TPI	dhap (abc) <-> g3p (abc)
GAPD/PGK	g3p (abc) <-> 3pg (abc)
PGM/ENO	3pg (abc) <-> pep (abc)
PYK	pep (abc) <-> pyr (abc)
LDH_L	pyr (abc) <-> lac_L (abc)
L_LACt2r	lac_L (abc) -> lac_L.e (abc)
G6PDH2r/PGL	g6p (abcdef) -> 6pgc (abcdef)
GND	6pgc (abcdef) -> ru5p_D (bcdef) + co2 (a)
RPI	ru5p_D (abcde) <-> r5p (abcde)
RPE	ru5p_D (abcde) <-> xu5p_D (abcde)
TALA	s7p (abcdefg) + g3p (hij) <-> e4p (defg) + f6p (abchij)
TKT2	xu5p_D (abcde) + e4p (fghi) <-> g3p (cde) + f6p (abfghi)
TKT1	xu5p_D (abcde) + r5p (fghij) <-> g3p (hij) + s7p (fgabcde)
PYRt2m/PDHm	pyr (abc) -> accoa (bc) + co2 (a)
CSm	accoa (ab) + oaa (cdef) <-> cit (fedbac)
ACONTm/ICDHxm	cit (abcdef) <-> akg (abcde) + co2 (f)
AKGDm/SUCOAS1m	akg (abcde) <-> succ (bcde) + co2 (a)
SUCD1m	succ (abcd) <-> fum (abcd)
FUMm	fum (abcd) <-> mal_L (abcd)
MDHm	mal_L (abcd) <-> oaa (abcd)
PCm	pyr (abc) + co2 (d) -> oaa (abcd)
GLUDxm	glu_L (abcde) <-> akg (abcde)
GLUVESSEC	glu_L.e (abcde) -> glu_L (abcde)
GLNtm/GLUNm	gln_L (abcde) <-> glu_L (abcde)
In_gln_L	gln_L.in (abcde) -> gln_L (abcde)
Out_Pyr	pyr (abc) -> pyr.out (abc)
ME	mal_L (abcd) -> pyr (bcd) + co2 (a)
In_CO2	co2.in (a) -> co2 (a)
Out_CO2	co2 (a) -> co2.out (a)
Out_gln_L	gln_L (abcde) -> gln_L.out (abcde)
In_Pyr	pyr.in (abc) -> pyr (abc)

Table S3. Correlation analysis of integration results from automated packages and manual software

Pearson correlation p values between automatically and manually collected isotopologue fractions for each of the tested nine metabolites. ND = not detected.

Metabolite	**mzMatch-ISO** *vs* **Skyline**	**mzMatch-ISO** *vs* & **ElMAVEN**	**X^{13}CMS** *vs* **& Skyline**	**X^{13}CMS** *vs* **& ElMAVEN**
Phosphoenolpyruvate	7.22E-10	6.77E-10	ND	ND
Fumarate	3.06E-14	3.18E-11	5.19E-06	5.54E-05
Ketoglutarate	6.29E-12	1.13E-14	8.45E-21	9.25E-15
Alanine	3.02E-08	2.09E-07	1.44E-07	8.31E-09
Aspartate	6.77E-16	1.02E-16	2.23E-05	2.48E-05
Glutamate	2.10E-13	8.95E-29	3.29E-10	8.17E-10
Fructose 1,6-bisphosphate	3.98E-31	1.33E-41	1.60E-03	1.91E-03
Glutathione	5.99E-34	9.49E-51	5.44E-33	1.89E-39
ATP	1.25E-53	1.61E-58	3.51E-25	2.08E-25

Table S4. The t-test examination over the automatically processed MID results of amino acids between labelled and unlabelled samples

class	metabolite	isotopologue	Sample_L1	Sample_L2	Sample_L3	Standard deviation in Sample_L	Sample_U1	Sample_U2	Sample_U3	Standard deviation in Sample_U	statistic	p value
Conditionally essential amino acid	Proline	0	0.65	0.66	0.66	0.01	0.94	0.93	0.93	0.00	-67.20	**<0.00001***
		1	0.07	0.06	0.05	0.01	0.06	0.07	0.07	0.00	-1.17	0.30582
		2	0.15	0.16	0.15	0.00	0.00	0.00	0.00	0.00	59.87	**<0.00001***
		3	0.06	0.05	0.06	0.01	0.00	0.00	0.00	0.00	16.68	**0.00008***
		4	0.05	0.04	0.05	0.01	0.00	0.00	0.00	0.00	15.56	**0.00010***
		5	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	18.09	**0.00005***
	Tyrosine	0	0.91	0.91	0.90	0.01	0.89	0.89	0.88	0.00	5.44	**0.00553***
		1	0.09	0.08	0.09	0.01	0.10	0.10	0.11	0.00	-4.34	**0.01222***
		2	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.00	-1.28	0.26843
		3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-4.28	**0.01287***
		4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.97	**0.00768***
		5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-10.51	**0.00046***
		6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.37390
		7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.61	**0.00993***
		8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.37390
		9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
	Glutamine	0	0.95	0.93	0.94	0.01	0.98	0.97	0.98	0.01	-4.23	**0.01342***
		1	0.02	0.02	0.02	0.00	0.01	0.02	0.01	0.01	1.75	0.15481
		2	0.02	0.04	0.03	0.01	0.00	0.00	0.00	0.00	5.52	**0.00527***
		3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.45	**0.00107***
		4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	6.22	**0.00340***
		5	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	-9.65	**0.00065***
Essential amino acid	Methionine	0	0.94	0.95	0.94	0.01	0.94	0.92	0.93	0.01	2.61	0.05955
		1	0.06	0.05	0.06	0.01	0.06	0.08	0.07	0.01	-2.79	**0.04921***
		2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.44	0.67992
		3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.37390
		4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.92	0.12718
	Threonine	0	0.96	0.95	0.96	0.00	0.95	0.96	0.95	0.01	0.89	0.42255
		1	0.04	0.04	0.04	0.00	0.05	0.04	0.04	0.01	-0.80	0.46735
		2	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	-1.22	0.29111
		3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.98	**0.00760***
		4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
	Leucine	0	0.94	0.93	0.93	0.00	0.93	0.94	0.94	0.00	-1.86	0.13635
		1	0.06	0.06	0.07	0.00	0.06	0.06	0.06	0.00	1.85	0.13858
		2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.76	0.49062
		3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
	Isoleucine	0	0.92	0.93	0.93	0.01	0.93	0.93	0.93	0.00	-0.21	0.84734
		1	0.07	0.06	0.07	0.01	0.07	0.07	0.07	0.00	-0.14	0.89837
		2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.61	**0.02246***
		3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
	Phenylalanine	0	0.89	0.90	0.90	0.00	0.90	0.90	0.93	0.02	-1.52	0.20249
		1	0.10	0.09	0.10	0.00	0.09	0.09	0.07	0.02	1.32	0.25742
		2	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	1.45	0.22057

		3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.05	0.35351
		6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.37390
		8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.86	0.43674
		9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
	Tryptophan	0	0.93	0.89	0.89	0.02	0.91	0.91	0.87	0.02	0.10	0.92863
		1	0.07	0.11	0.11	0.03	0.08	0.09	0.12	0.02	-0.17	0.87263
		2	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.22	0.28787
		3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
		10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.76	0.49057
		11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		
Nonessential amino acid	Alanine	0	0.45	0.43	0.49	0.03	0.97	0.92	0.94	0.02	-21.10	**0.00003***
		1	0.03	0.03	0.02	0.00	0.03	0.08	0.05	0.03	-1.68	0.16747
		2	0.03	0.04	0.02	0.01	0.00	0.00	0.00	0.00	5.33	**0.00596***
		3	0.49	0.50	0.46	0.02	0.00	0.00	0.00	0.00	40.92	**<0.00001***
		0	0.45	0.43	0.49	0.03	0.97	0.92	0.94	0.02	-21.10	**0.00003***
		1	0.03	0.03	0.02	0.00	0.03	0.08	0.05	0.03	-1.68	0.16747
		2	0.03	0.04	0.02	0.01	0.00	0.00	0.00	0.00	5.33	**0.00596***
		3	0.49	0.50	0.46	0.02	0.00	0.00	0.00	0.00	40.92	**<0.00001***
	Serine	0	0.77	0.80	0.78	0.02	0.96	0.95	0.95	0.01	-18.11	**0.00005***
		1	0.13	0.11	0.12	0.01	0.03	0.04	0.04	0.01	16.12	**0.00009***
		2	0.04	0.03	0.05	0.01	0.00	0.00	0.01	0.00	11.15	**0.00037***
		3	0.06	0.05	0.06	0.01	0.00	0.00	0.00	0.00	20.31	**0.00003***
	Aspartate	0	0.45	0.41	0.44	0.02	0.94	0.95	0.94	0.01	-41.50	**<0.00001***
		1	0.08	0.10	0.08	0.01	0.05	0.04	0.05	0.01	4.75	**0.00895***
		2	0.18	0.19	0.20	0.01	0.01	0.00	0.00	0.00	36.83	**<0.00001***
		3	0.19	0.20	0.20	0.00	0.00	0.00	0.00	0.00	80.64	**<0.00001***
		4	0.09	0.09	0.08	0.01	0.00	0.00	0.00	0.00	21.86	**0.00003***
	Asparagine	0	0.87	0.88	0.86	0.01	0.99	1.00	0.99	0.00	-22.49	**0.00002***
		1	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	2.36	0.07771
		2	0.08	0.08	0.09	0.00	0.01	0.00	0.01	0.00	38.45	**<0.00001***
		3	0.04	0.03	0.04	0.01	0.00	0.00	0.00	0.00	8.97	**0.00085***
		4	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	4.18	**0.01394***
	Glutamate	0	0.42	0.39	0.38	0.02	0.93	0.93	0.93	0.00	-41.84	**<0.00001***
		1	0.06	0.07	0.07	0.00	0.07	0.06	0.07	0.00	0.40	0.71246
		2	0.22	0.23	0.24	0.01	0.00	0.00	0.00	0.00	39.95	**<0.00001***
		3	0.11	0.11	0.12	0.00	0.00	0.00	0.00	0.00	39.81	**<0.00001***
		4	0.12	0.13	0.12	0.00	0.00	0.00	0.00	0.00	48.27	**<0.00001***
		5	0.07	0.08	0.08	0.00	0.00	0.00	0.00	0.00	43.22	**<0.00001***

Table S5. The MID results of central carbon metabolism for neurons monitored over a time course of 24h

			^{13}C	0 min		5 min		10 min		20 min		40 min		1 h		2 h		3 h		5 h		8 h		12 h		24 h		
met_ID	met_Name	Formula	isotope	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	reconMap3
C00064	GLUTAMINE	C5H10N2O3	0	0.989	0.005	0.991	0.004	0.991	0.003	0.992	0.005	0.987	0.001	0.960	0.005	0.933	0.006	0.903	0.010	0.867	0.002	0.838	0.008	0.829	0.002	0.712	0.005	gln_L
C00064	GLUTAMINE	C5H10N2O3	1	0.011	0.006	0.009	0.004	0.009	0.003	0.007	0.004	0.008	0.000	0.010	0.004	0.009	0.002	0.018	0.005	0.021	0.002	0.020	0.003	0.022	0.001	0.030	0.004	gln_L
C00064	GLUTAMINE	C5H10N2O3	2	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.004	0.001	0.020	0.001	0.038	0.003	0.048	0.003	0.057	0.001	0.061	0.003	0.062	0.003	0.090	0.003	gln_L
C00064	GLUTAMINE	C5H10N2O3	3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.004	0.001	0.008	0.001	0.013	0.002	0.022	0.004	0.031	0.002	0.033	0.001	0.060	0.002	gln_L
C00064	GLUTAMINE	C5H10N2O3	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.001	0.007	0.001	0.012	0.002	0.020	0.001	0.029	0.006	0.031	0.001	0.057	0.003	gln_L
C00064	GLUTAMINE	C5H10N2O3	5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.000	0.005	0.001	0.007	0.001	0.013	0.000	0.021	0.001	0.023	0.002	0.051	0.002	gln_L
C00025	GLUTAMATE	C5H9NO4	0	0.993	0.004	0.992	0.004	0.985	0.001	0.939	0.002	0.869	0.006	0.702	0.012	0.581	0.004	0.492	0.016	0.343	0.009	0.286	0.007	0.252	0.021	0.226	0.004	glu_L
C00025	GLUTAMATE	C5H9NO4	1	0.007	0.003	0.005	0.004	0.006	0.001	0.009	0.002	0.007	0.000	0.020	0.004	0.029	0.002	0.041	0.004	0.054	0.006	0.051	0.001	0.044	0.003	0.041	0.005	glu_L
C00025	GLUTAMATE	C5H9NO4	2	0.000	0.000	0.003	0.001	0.008	0.002	0.042	0.004	0.100	0.003	0.207	0.011	0.249	0.002	0.269	0.006	0.272	0.002	0.246	0.007	0.235	0.007	0.205	0.006	glu_L
C00025	GLUTAMATE	C5H9NO4	3	0.000	0.000	0.000	0.000	0.001	0.000	0.005	0.001	0.010	0.001	0.032	0.003	0.061	0.004	0.078	0.002	0.128	0.005	0.147	0.005	0.152	0.009	0.155	0.010	glu_L
C00025	GLUTAMATE	C5H9NO4	4	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.001	0.009	0.001	0.026	0.000	0.054	0.003	0.076	0.008	0.125	0.004	0.154	0.003	0.168	0.009	0.189	0.014	glu_L
C00025	GLUTAMATE	C5H9NO4	5	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.004	0.001	0.013	0.000	0.027	0.001	0.044	0.001	0.077	0.005	0.116	0.001	0.149	0.008	0.184	0.009	glu_L
C00074	PHOSPHOENOLPYRUVATE	C3H5O6P	0	0.762	0.030	0.239	0.021	0.215	0.044	0.124	0.036	0.085	0.032	0.054	0.016	0.054	0.041	0.022	0.008	0.034	0.019	0.013	0.007	0.038	0.010	0.027	0.004	pep
C00074	PHOSPHOENOLPYRUVATE	C3H5O6P	1	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	pep
C00074	PHOSPHOENOLPYRUVATE	C3H5O6P	2	0.233	0.036	0.198	0.052	0.208	0.039	0.155	0.022	0.146	0.053	0.269	0.029	0.237	0.043	0.238	0.067	0.265	0.041	0.239	0.053	0.231	0.067	0.205	0.048	pep
C00074	PHOSPHOENOLPYRUVATE	C3H5O6P	3	0.005	0.008	0.564	0.044	0.577	0.005	0.720	0.030	0.769	0.025	0.676	0.036	0.709	0.057	0.740	0.060	0.701	0.059	0.748	0.060	0.731	0.076	0.768	0.044	pep
C00022	PYRUVATE	C3H4O3	0	0.992	0.014	0.565	0.036	0.502	0.094	0.379	0.048	0.300	0.058	0.338	0.072	0.262	0.051	0.265	0.054	0.205	0.019	0.183	0.035	0.181	0.050	0.121	0.018	pyr
C00022	PYRUVATE	C3H4O3	1	0.008	0.014	0.000	0.000	0.000	0.000	0.001	0.002	0.000	0.000	0.008	0.007	0.005	0.008	0.000	0.000	0.000	0.000	0.001	0.002	0.011	0.002	0.014	0.005	pyr
C00022	PYRUVATE	C3H4O3	2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.003	0.005	0.007	0.006	0.006	0.010	0.017	0.015	0.025	0.012	pyr
C00022	PYRUVATE	C3H4O3	3	0.000	0.000	0.435	0.036	0.498	0.094	0.620	0.047	0.699	0.059	0.654	0.066	0.733	0.059	0.732	0.056	0.787	0.017	0.811	0.043	0.791	0.066	0.840	0.034	pyr
C00186	LACTATE	C3H6O3	0	0.966	0.016	0.505	0.084	0.448	0.064	0.386	0.108	0.340	0.005	0.270	0.073	0.217	0.046	0.269	0.040	0.154	0.014	0.128	0.023	0.136	0.022	0.169	0.047	lac_L
C00186	LACTATE	C3H6O3	1	0.030	0.013	0.001	0.002	0.002	0.003	0.003	0.006	0.002	0.001	0.003	0.003	0.001	0.002	0.005	0.005	0.003	0.003	0.009	0.002	0.010	0.001	0.021	0.006	lac_L
C00186	LACTATE	C3H6O3	2	0.005	0.003	0.001	0.003	0.000	0.000	0.003	0.004	0.004	0.006	0.003	0.005	0.001	0.002	0.037	0.024	0.019	0.009	0.019	0.013	0.035	0.013	0.087	0.029	lac_L
C00186	LACTATE	C3H6O3	3	0.000	0.000	0.493	0.083	0.551	0.062	0.608	0.099	0.655	0.009	0.724	0.075	0.781	0.046	0.689	0.067	0.825	0.024	0.844	0.032	0.820	0.035	0.723	0.080	lac_L
C00026	KETOGLUTARATE	C5H6O5	0	0.997	0.001	0.998	0.001	0.996	0.002	0.955	0.030	0.891	0.020	0.698	0.008	0.573	0.008	0.496	0.023	0.323	0.007	0.255	0.009	0.233	0.007	0.192	0.002	akg
C00026	KETOGLUTARATE	C5H6O5	1	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.009	0.003	0.001	0.016	0.006	0.031	0.006	0.039	0.003	0.057	0.002	0.060	0.004	0.058	0.004	0.054	0.004	akg
C00026	KETOGLUTARATE	C5H6O5	2	0.000	0.000	0.000	0.000	0.000	0.000	0.029	0.033	0.087	0.018	0.213	0.009	0.258	0.008	0.272	0.012	0.275	0.006	0.266	0.004	0.247	0.014	0.214	0.014	akg
C00026	KETOGLUTARATE	C5H6O5	3	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.010	0.002	0.034	0.004	0.059	0.003	0.087	0.003	0.136	0.001	0.158	0.004	0.160	0.003	0.172	0.006	akg
C00026	KETOGLUTARATE	C5H6O5	4	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.004	0.002	0.025	0.001	0.054	0.004	0.068	0.003	0.130	0.006	0.146	0.003	0.164	0.005	0.190	0.007	akg
C00026	KETOGLUTARATE	C5H6O5	5	0.003	0.000	0.002	0.001	0.003	0.002	0.004	0.002	0.005	0.001	0.012	0.000	0.025	0.001	0.039	0.002	0.080	0.002	0.115	0.005	0.139	0.005	0.179	0.004	akg
C00042	SUCCINATE	C4H6O4	0	0.967	0.025	0.969	0.005	0.946	0.020	0.911	0.019	0.852	0.028	0.723	0.022	0.594	0.021	0.484	0.040	0.321	0.019	0.294	0.023	0.235	0.019	0.221	0.022	succ
C00042	SUCCINATE	C4H6O4	1	0.008	0.013	0.006	0.001	0.000	0.000	0.000	0.000	0.004	0.005	0.038	0.008	0.050	0.008	0.074	0.016	0.101	0.015	0.101	0.010	0.109	0.010	0.082	0.001	succ
C00042	SUCCINATE	C4H6O4	2	0.007	0.006	0.007	0.005	0.033	0.013	0.066	0.010	0.124	0.028	0.185	0.019	0.250	0.022	0.275	0.034	0.281	0.014	0.266	0.014	0.263	0.012	0.249	0.022	succ
C00042	SUCCINATE	C4H6O4	3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.014	0.002	0.050	0.010	0.073	0.007	0.144	0.010	0.153	0.007	0.173	0.013	0.197	0.022	succ
C00042	SUCCINATE	C4H6O4	4	0.018	0.008	0.018	0.003	0.022	0.015	0.024	0.015	0.019	0.004	0.040	0.007	0.055	0.006	0.094	0.010	0.154	0.008	0.185	0.010	0.220	0.025	0.251	0.034	succ
C00122	FUMARATE	C4H4O4	0	1.000	0.000	0.992	0.008	0.992	0.009	0.941	0.004	0.884	0.004	0.706	0.017	0.579	0.006	0.500	0.009	0.373	0.006	0.318	0.023	0.276	0.014	0.227	0.008	fum
C00122	FUMARATE	C4H4O4	1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.006	0.028	0.007	0.049	0.008	0.080	0.002	0.075	0.005	0.084	0.009	0.084	0.008	fum
C00122	FUMARATE	C4H4O4	2	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.007	0.022	0.003	0.109	0.019	0.162	0.005	0.183	0.018	0.225	0.013	0.213	0.004	0.234	0.013	0.223	0.021	fum
C00122	FUMARATE	C4H4O4	3	0.000	0.000	0.008	0.008	0.008	0.009	0.053	0.010	0.092	0.007	0.156	0.019	0.203	0.011	0.227	0.002	0.243	0.007	0.282	0.005	0.264	0.017	0.287	0.027	fum
C00122	FUMARATE	C4H4O4	4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.003	0.013	0.003	0.028	0.007	0.041	0.004	0.078	0.003	0.113	0.019	0.142	0.014	0.179	0.011	fum
C00149	MALATE	C4H6O5	0	0.980	0.006	0.961	0.017	0.933	0.011	0.903	0.004	0.854	0.005	0.710	0.021	0.584	0.001	0.510	0.018	0.378	0.012	0.301	0.010	0.269	0.017	0.217	0.013	mal_L
C00149	MALATE	C4H6O5	1	0.002	0.003	0.003	0.003	0.015	0.004	0.005	0.003	0.011	0.007	0.012	0.003	0.032	0.004	0.047	0.002	0.075	0.003	0.082	0.012	0.089	0.006	0.077	0.008	mal_L
C00149	MALATE	C4H6O5	2	0.004	0.004	0.008	0.007	0.012	0.003	0.022	0.004	0.037	0.008	0.114	0.009	0.142	0.005	0.165	0.034	0.225	0.003	0.210	0.007	0.225	0.015	0.237	0.026	mal_L
C00149	MALATE	C4H6O5	3	0.007	0.001	0.020	0.003	0.025	0.004	0.055	0.004	0.086	0.004	0.150	0.015	0.214	0.002	0.240	0.017	0.243	0.010	0.290	0.011	0.273	0.004	0.288	0.017	mal_L
C00149	MALATE	C4H6O5	4	0.007	0.002	0.007	0.004	0.015	0.004	0.014	0.004	0.011	0.001	0.015	0.003	0.028	0.002	0.038	0.003	0.079	0.005	0.118	0.010	0.143	0.006	0.181	0.012	mal_L
C00158	CITRATE	C6H8O7	0	0.721	0.225	0.643	0.127	0.484	0.049	0.432	0.052	0.414	0.016	0.416	0.030	0.374	0.029	0.350	0.059	0.255	0.078	0.197	0.030	0.221	0.022	0.159	0.011	cit
C00158	CITRATE	C6H8O7	1	0.072	0.067	0.076	0.027	0.082	0.019	0.071	0.015	0.092	0.022	0.065	0.024	0.062	0.005	0.055	0.037	0.073	0.022	0.064	0.014	0.063	0.009	0.058	0.014	cit
C00158	CITRATE	C6H8O7	2	0.063	0.059	0.110	0.049	0.170	0.040	0.186	0.020	0.193	0.036	0.181	0.019	0.224	0.025	0.237	0.020	0.230	0.023	0.229	0.020	0.207	0.040	0.179	0.034	cit
C00158	CITRATE	C6H8O7	3	0.047	0.032	0.064	0.027	0.085	0.032	0.110	0.029	0.118	0.021	0.117	0.023	0.110	0.016	0.115	0.016	0.142	0.007	0.135	0.006	0.139	0.010	0.168	0.011	cit
C00158	CITRATE	C6H8O7	4	0.048	0.038	0.052	0.019	0.095	0.020	0.086	0.014	0.088	0.028	0.106	0.020	0.119	0.035	0.118	0.010	0.139	0.020	0.170	0.006	0.156	0.015	0.209	0.030	cit
C00158	CITRATE	C6H8O7	5	0.037	0.025	0.045	0.020	0.068	0.012	0.086	0.025	0.081	0.008	0.079	0.007	0.076	0.007	0.085	0.030	0.111	0.021	0.142	0.017	0.155	0.011	0.148	0.014	cit
C00158	CITRATE	C6H8O7	6	0.011	0.011	0.011	0.008	0.016	0.011	0.029	0.012	0.014	0.006	0.035	0.016	0.036	0.008	0.039	0.007	0.050	0.019	0.063	0.017	0.060	0.021	0.079	0.017	cit

Table S6. Comparison of four sets of flux distributions in the core model

RXN_ID	RXN_NAME	RXN_Formula	subSystems	INCA model Equation	EntropyFlux	MoietyFlux	FBAFlux	INCAmodel	StdErr	LB	UB
R1: Merge			**Transport, extracellular**	**Glucose.e -> Glucose**	**427.4398**	**427.4396**	**440.0000**	**433.91**	**0.19**	**421.25**	**446.56**
GLCt1r	Glucose Transport (Uniport)	glc_D[e] <=> glc_D[c]			358.0923	357.2980	100000.0000				
GLCt4	Glucose Transport via Sodium Symport	na1[e] + glc_D[e] <=> na1[c] + glc_D[c]			69.3475	70.1416	-99560.0000				
R2: Reversed			**Transport, extracellular**	**Pyr -> Pyr.e**	**6.9377**	**5.1566**	**-100000.0000**	**774.60**	**0.47**	**660.74**	**896.96**
PYRt2r	Transport of Pyruvate Reversible via Proton Symport	h[e] + pyr[e] <=> h[c] + pyr[c]		Pyr -> Pyr.out	-6.9377	-5.1566	100000.0000	852.29	0.23	764.61	940.10
				Pyr.in -> Pyr				77.69	0.23	43.14	103.86
R3: GLNt4	**Transport of L-Glutamine via Sodium Symport, Reversible**	**na1[e] + gln_L[e] -> na1[c] + gln_L[c]**	**Transport, extracellular**	**Gln.e -> Gln**	**1.5147**	**2.1135**	**0.0000**	**6.31E-08**	**0.00**	**-1484.30**	**1228.80**
				Gln.in -> Gln				1.63E-07	2.72E-08	0.00	1228.80
				Gln -> Gln.out				1.00E-07	0.00	0.00	1484.30
R4: Merge			**Transport, extracellular**	**Glu.e -> Glu**	**2.2764**	**2.7179**	**100009.2040**	**1.39**	**0.06**	**1.24**	**1.55**
GLUt6	Glutamate Transport via Na, H Symport And K Antiport	h[e] + 3 na1[e] + k[c] + glu_L[e] -> h[c] + 3 na1[c] + glu_L[c] + k[e]			0.0000	-0.0001	100000.0000				
GLUVESSEC	L-Glutamate Secretion via Secretory Vesicle (ATP Driven)	h2o[c] + atp[c] + glu_L[c] -> h[c] + adp[c] + pi[c] + glu_L[e]			0.0000	0.0000	0.0000				
GLUPROASCT1	Glutamate Intake by System Asct-1 Transporter	na1[e] + glu_L[e] + pro_L[c] -> na1[c] + glu_L[c] + pro_L[e]			2.2764	2.7180	9.2040				
R5: Reversed			**Transport, extracellular**	**Lac -> Lac.e**	**84.5875**	**35.4093**	**-100000.0000**	**91.60**	**0.44**	**12.39**	**170.81**
L_LACt2r	L-Lactate Reversible Transport via Proton Symport	h[e] + lac_L[e] <=> h[c] + lac_L[c]			-84.5875	-35.4093	100000.0000				
R6: Merge			**Glycolysis/gluconeogenesis**	**Glucose -> G6P**	**101.8607**	**55.6727**	**0.0001**	**433.91**	**0.19**	**421.25**	**446.56**
HEX1	Hexokinase (D-Glucose:ATP)	atp[c] + glc_D[c] -> h[c] + adp[c] + g6p[c]			47.0870	22.6522	0.0001				
CBPPer	Carbamoyl Phosphate Phosphotransferase, Endoplasmic Reticulum	h[r] + glc_D[r] + cbp[r] -> co2[r] + nh4[r] + g6p[r]			0.0000	0.0004	0.0000				
r0354	Hexokinase	itp[c] + glc_D[c] -> h[c] + idp[c] + g6p[c]			4.6387	3.3390	0.0000				
r0355	Hexokinase	datp[c] + glc_D[c] -> h[c] + dadp[c] + g6p[c]			50.1350	29.6811	0.0000				
R7: PGI	**Glucose-6-Phosphate Isomerase**	**g6p[c] <=> f6p[c]**	**Glycolysis/gluconeogenesis**	**G6P -> F6P**	**73.1642**	**49.1976**	**-100000.0000**	**433.91**	**0.70**	**421.25**	**446.56**
R8: PFK	**Phosphofructokinase**	**atp[c] + f6p[c] -> h[c] + adp[c] + fdp[c]**	**Glycolysis/gluconeogenesis**	**F6P -> FBP**	**30.7213**	**19.3229**	**0.0000**	**433.91**	**0.19**	**421.25**	**446.56**
R9: FBA	**Fructose-Bisphosphate Aldolase**	**fdp[c] <=> dhap[c] + g3p[c]**	**Glycolysis/gluconeogenesis**	**FBP -> DHAP + GAP**	**66.8249**	**43.4680**	**0.0000**	**433.91**	**0.19**	**421.25**	**446.56**
R10: TPI	**Triose-Phosphate Isomerase**	**dhap[c] <=> g3p[c]**	**Glycolysis/gluconeogenesis**	**DHAP <-> GAP**	**78.6574**	**52.0826**	**-7498.0000**	**433.91**	**0.19**	**421.25**	**446.56**
R11: Merge			**Glycolysis/gluconeogenesis**	**GAP <-> 3PG**	**158.5497**	**105.3870**	**-7057.0000**	**867.81**	**0.25**	**842.51**	**893.12**
GAPD	Glyceraldehyde-3-Phosphate Dehydrogenase	pi[c] + nad[c] + g3p[c] <=> h[c] + nadh[c] + 13dpg[c]			158.5497	105.3870	-7057.0000				
PGK	Phosphoglycerate Kinase	atp[c] + 3pg[c] <=> adp[c] + 13dpg[c]			-158.5497	-105.3920	7057.0000				
ACYP	Acylphosphatase	h2o[c] + 13dpg[c] -> h[c] + pi[c] + 3pg[c]			0.0000	-0.0005	0.0000				
R12: Merge			**Glycolysis/gluconeogenesis**	**3PG <-> PEP**	**158.5497**	**105.3930**	**-7057.0000**	**867.81**	**0.25**	**842.51**	**893.12**
PGM	Phosphoglycerate Mutase	2pg[c] <=> 3pg[c]			-158.5497	-105.3930	7057.0000				
ENO	Enolase	2pg[c] <=> h2o[c] + pep[c]			158.5497	105.3940	-7057.0000				
R13: Merge			**Glycolysis/gluconeogenesis**	**PEP <-> Pyr**	**152.5825**	**100.7097**	**92940.0000**	**867.81**	**0.25**	**842.51**	**893.12**
PYK	Pyruvate Kinase	h[c] + adp[c] + pep[c] -> atp[c] + pyr[c]			48.1542	34.1840	0.0000				
r0413	DATP:Pyruvate O2-Phosphotransferase	h[c] + pep[c] + dgdp[c] -> pyr[c] + dgtp[c]			5.5925	4.0385	0.0000				
r0280	DATP:Pyruvate O2-Phosphotransferase	h[c] + pep[c] + dadp[c] -> pyr[c] + datp[c]			45.2266	26.0918	0.0000				
r0122	GTP:Pyruvate O2-Phosphotransferase	h[c] + pep[c] + gdp[c] -> pyr[c] + gtp[c]			6.6746	4.1498	0.0000				
r0153	CTP:Pyruvate O2-Phosphotransferase	h[c] + pep[c] + cdp[c] -> pyr[c] + ctp[c]			5.9584	4.8780	92940.0000				
r0165	UTP:Pyruvate O2-Phosphotransferase	h[c] + udp[c] + pep[c] -> pyr[c] + utp[c]			40.9762	27.3676	0.0000				
R14: Merge				**Pyr <-> Lac**	**132.0235**	**67.6056**	**55.0000**	**91.60**	**0.44**	**12.39**	**170.81**
LDH_Lm	L-Lactate Dehydrogenase	nad[m] + lac_L[m] <=> h[m] + nadh[m] + pyr[m]	Pyruvate metabolism		0.1083	-0.0032	1205.0000				
LDH_L	L-Lactate Dehydrogenase	nad[c] + lac_L[c] <=> h[c] + pyr[c] + nadh[c]	Glycolysis/gluconeogenesis		-131.1714	-66.3706	-100000.0000				
r0173	(S)-Lactate:NAD+ Oxidoreductase	nad[x] + lac_L[x] <=> h[x] + nadh[x] + pyr[x]	Glycolysis/gluconeogenesis		-0.9604	-1.2318	98740.0000				
R15: Merge			**Pentose phosphate pathway**	**G6P -> 6PG**	**7.9932**	**1.5662**	**0.0000**	**1.00E-07**	**0.83**	**0.00**	**446.49**
G6PDH1rer	Glucose-6-Phosphate Dehydrogenase, Endoplasmic Reticulum	nad[r] + g6p[r] <=> h[r] + nadh[r] + 6pgl[r]			0.0067	0.7647	-0.0386				
G6PDH2rer	Glucose 6-Phosphate Dehydrogenase, Endoplasmic Reticulum	nadp[r] + g6p[r] <=> h[r] + nadph[r] + 6pgl[r]			0.2160	0.3525	0.0386				
G6PDH2c	Glucose 6-Phosphate Dehydrogenase	3 nadp[c] + 3 g6p[c] -> 3 h[c] + 3 nadph[c] + 3 6pgl[c]			7.7705	0.4489	0.0000				
PGLer	6-Phosphogluconolactonase, Endoplasmic Reticulum	h2o[r] + 6pgl[r] -> h[r] + 6pgc[r]			0.2227	1.1183	0.0000				
PGLc	6-Phosphogluconolactonase	3 h2o[c] + 3 6pgl[c] -> 3 h[c] + 3 6pgc[c]			7.4804	0.4068	0.0000				
PGL	6-Phosphogluconolactonase	h2o[c] + 6pgl[c] -> h[c] + 6pgc[c]			0.8702	0.1302	0.0000				
R16: Merge			**Pentose phosphate pathway**	**6PG -> Ru5P + CO2**	**0.6090**	**1.95E-05**	**0.00E+00**	**1.00E-07**	**0.83**	**0.00**	**446.49**
GNDer	Phosphogluconate Dehydrogenase, Endoplasmic Reticulum	nadp[r] + 6pgc[r] -> nadph[r] + co2[r] + ru5p_D[r]			0.0000	0.0000	0.0000				

GNDc	Phosphogluconate Dehydrogenase	3 nadp[c] + 3 6pgc[c] -> 3 nadph[c] + 3 co2[c] + 3 ru5p_D[c]			0.4625	0.0000	0.0000				
GND	Phosphogluconate Dehydrogenase	nadp[c] + 6pgc[c] -> nadph[c] + co2[c] + ru5p_D[c]			0.1464	0.0000	0.0000				
R17: Merge			**Pentose phosphate pathway**	**Ru5P <-> R5P**	**1.2029**	**1.4473**	**7940.0000**	**3.33E-08**	**0.28**	**-2.17E-19**	**148.83**
r0249	D-Ribose-5-Phosphate Ketol-Isomerase	r5p[r] <=> ru5p_D[r]			-0.2660	-0.3116	-100000.0000				
RPI	Ribose-5-Phosphate Isomerase	r5p[c] <=> ru5p_D[c]			-0.9369	-1.1357	92060.0000				
R18: Merge			**Pentose phosphate pathway**	**Ru5P <-> X5P**	**0.1986**	**-0.8334**	**46030.0000**	**6.67E-08**	**0.55**	**0.00**	**297.66**
RPEc	Ribulose 5-Phosphate 3-Epimerase	2 ru5p_D[c] <=> 2 xu5p_D[c]			0.1326	-0.5669	-53970.0000				
RPE	Ribulose 5-Phosphate 3-Epimerase	ru5p_D[c] <=> xu5p_D[c]			0.0660	-0.2665	100000.0000				
R19: TALA	**Transaldolase**	**g3p[c] + s7p[c] <=> f6p[c] + e4p[c]**	**Pentose phosphate pathway**	**S7P + GAP <-> E4P + F6P**	**-11.6670**	**-9.2856**	**3529.0000**	**3.33E-08**	**0.28**	**0.00**	**148.83**
R20: TKT2	**Transketolase**	**xu5p_D[c] + e4p[c] <=> g3p[c] + f6p[c]**	**Pentose phosphate pathway**	**X5P + E4P <-> GAP + F6P**	**0.1656**	**-0.6825**	**-3969.0000**	**3.33E-08**	**0.28**	**1.08E-19**	**148.83**
R21: TKT1	**Transketolase**	**xu5p_D[c] + r5p[c] <=> g3p[c] + s7p[c]**	**Pentose phosphate pathway**	**X5P + R5P <-> GAP + S7P**	**0.1656**	**-0.6801**	**-3969.0000**	**3.33E-08**	**0.28**	**0.00**	**148.83**
R22: PDHm	**Pyruvate Dehydrogenase**	**nad[m] + coa[m] + pyr[m] -> nadh[m] + co2[m] + accoa[m]**	**Glycolysis/gluconeogenesis**	**Pyr -> AcCoA.m + CO2**	**5.4261**	**3.4981**	**0.0001**	**3.01**	**0.13**	**2.37**	**657.25**
R23: Merge			**Citric acid cycle**	**AcCoA.m + OAA <-> Cit**	**0.0000**	**2.7340**	**2765.0000**	**3.01**	**0.13**	**2.37**	**657.25**
ACITL	ATP-Citrate Lyase	atp[c] + coa[c] + cit[c] -> adp[c] + pi[c] + accoa[c] + oaa[c]			0.0000	-0.0003	0.0000				
CSm	Citrate Synthase	h2o[m] + accoa[m] + oaa[m] -> h[m] + coa[m] + cit[m]			0.0000	2.7337	2765.0000				
R24: Merge			**Citric acid cycle**	**Cit <-> aKG + CO2**	**-0.7505**	**0.0000**	**2760.0000**	**3.01**	**0.13**	**2.37**	**657.25**
ACONT	Aconitase	cit[c] <=> icit[c]			-0.2777	1.7068	100000.0000				
ACONTm	Aconitate Hydratase	cit[m] <=> icit[m]			-0.4727	0.0831	2764.0000				
ICDHxm	Isocitrate Dehydrogenase (NAD+)	nad[m] + icit[m] -> akg[m] + nadh[m] + co2[m]			0.0000	-0.0007	100000.0000				
ICDHy	Isocitrate Dehydrogenase (NADP)	nadp[c] + icit[c] -> nadph[c] + akg[c] + co2[c]			0.0000	0.0000	0.0000				
ICDHyp	Isocitrate Dehydrogenase (NADP+)	nadp[x] + icit[x] -> akg[x] + co2[x] + nadph[x]			0.0000	0.0000	0.0000				
ICDHyrm	Isocitrate Dehydrogenase (NADP+)	nadp[m] + icit[m] <=> nadph[m] + akg[m] + co2[m]			-3.6789	-1.2666	-97240.0000				
R25: Merge			**Citric acid cycle**	**aKG <-> Suc + CO2**	**2.2193**	**4.8601**	**2850.0000**	**4.40**	**0.16**	**1.70**	**438.58**
AKGDm	2-Oxoglutarate Dehydrogenase	akg[m] + nad[m] + coa[m] -> nadh[m] + co2[m] + succoa[m]			2.2193	4.8601	2851.0000				
SUCOAS1m	Succinate- Coenzyme A Ligase (GDP-Forming)	coa[m] + gtp[m] + succ[m] <=> pi[m] + gdp[m] + succoa[m]			-1.5895	-3.4023	97150.0000				
SUCOASm	Succinate- Coenzyme A Ligase (ADP-Forming)	coa[m] + atp[m] + succ[m] <=> adp[m] + pi[m] + succoa[m]			-0.6298	-1.4952	-100000.0000				
R26: Merge			**Citric acid cycle**	**Suc <-> Fum**	**2.1684**	**4.6900**	**2853.0000**	**4.40**	**0.16**	**1.70**	**438.58**
SUCD1m	Succinate Dehydrogenase	fad[m] + succ[m] <=> fadh2[m] + fum[m]			0.6600	0.7903	2853.0000				
r0509	Succinate:Ubiquinone Oxidoreductase	q10[m] + succ[m] -> q10h2[m] + fum[m]			1.5084	3.8997	0.0000				
R27: Merge			**Citric acid cycle**	**Fum <-> Mal**	**-4.8871**	**-15.8098**	**2850.0000**	**4.40**	**0.16**	**1.70**	**438.58**
FUM	Fumarase	h2o[c] + fum[c] <=> mal_L[c]			-5.6093	-29.8207	-97150.0000				
FUMm	Fumarase, Mitochondrial	h2o[m] + fum[m] <=> mal_L[m]			0.7221	14.0109	100000.0000				
R28: Merge			**Citric acid cycle**	**Mal <-> OAA**	**-5.1599**	**-16.0664**	**1768.2000**	**0.85**	**0.22**	**-0.73**	**330.08**
MDH	Malate Dehydrogenase	nad[c] + mal_L[c] <=> h[c] + nadh[c] + oaa[c]			-2.5304	-2.0983	508.2000				
MDHx	Malate Dehydrogenase, Peroxisomal	nad[x] + mal_L[x] <=> h[x] + nadh[x] + oaa[x]			0.4263	0.7782	-98740.0000				
MDHm	Malate Dehydrogenase, Mitochondrial	nad[m] + mal_L[m] <=> h[m] + nadh[m] + oaa[m]			-3.0558	-14.7463	100000.0000				
R29: PCm	**Pyruvate Carboxylase**	**pyr[m] + atp[m] + hco3[m] -> h[m] + adp[m] + pi[m] + oaa[m]**	**Pyruvate metabolism**	**Pyr + CO2 -> OAA**	**5.1495**	**19.5274**	**1000.0000**	**2.16**	**0.12**	**1.44**	**145.56**
R30: Merge			**Glutamate metabolism**	**Glu <-> aKG**	**25.1130**	**56.8432**	**300.0000**	**1.39**	**0.06**	**-270.67**	**19.59**
GLUDxm	Glutamate Dehydrogenase (NAD), Mitochondrial	h2o[m] + nad[m] + glu_L[m] <=> h[m] + akg[m] + nadh[m] + nh4[m]			14.1898	54.5894	100000.0000				
GLUDym	Glutamate Dehydrogenase (NADP), Mitochondrial	h2o[m] + nadp[m] + glu_L[m] <=> h[m] + nadph[m] + akg[m] + nh4[m]			10.9232	2.2538	-99700.0000				
R31: Merge			**Glutamate metabolism**	**Gln <-> Glu**	**1.1065**	**1.5719**	**71.5300**	**6.31E-08**	**2.72E-08**	**-81.82**	**18.20**
GLUNm	Glutaminase, Mitochondrial	h2o[m] + gln_L[m] -> glu_L[m] + nh4[m]			1.1065	1.5714	71.5300				
HMR_9802	Glutaminase	h2o[c] + gln_L[c] -> nh4[c] + glu_L[c]			0.0000	0.0005	0.0000				
R32: Merge			**Pyruvate metabolism**	**Mal -> Pyr + CO2**	**0.0000**	**0.0013**	**1088.0000**	**3.55**	**0.14**	**0.00**	**48.97**
ME1m	Malic Enzyme (NAD), Mitochondrial	nad[m] + mal_L[m] -> nadh[m] + co2[m] + pyr[m]			0.0000	0.0013	0.0000				
ME2	Malic Enzyme (NADP)	nadp[c] + mal_L[c] -> pyr[c] + nadph[c] + co2[c]			0.0000	0.0000	1088.0000				
ME2m	Malic Enzyme (NADP), Mitochondrial	nadp[m] + mal_L[m] -> nadph[m] + co2[m] + pyr[m]			0.0000	0.0000	0.0000				

Table S7. A list of 10 conserved moieties identified within the atomically resolved iDopaNeuroC model

Index	Metabolites	Rxns	Moiety_formula	Mass	Minimal_mass_metabolite	Name	Formula	Mass_fraction	Internal reaction coverage	internal subsystem hits in entropy flux-based top 15 active subsystems	internal subsystem hits in moiety flux-based top 15 active subsystems	Uptaken metabolite
28	59	228	'S'	31.9720707	'HC00250[e]'	'Hydrosulfide'	'HS'	0.96944123	40.80%	1	1	Cystine
56	22	60	'C5HA2O5'	140.982398	'HC02056[c]'	'1,2-Diacylglycerol-Ld-Pc-Pool'	'C5H6O5R2'	0.965490531	54.80%	0	0	\
5	27	105	'C19H7N7O4'	397.055952	'fol[c]'	'Folate'	'C19H18N7O6'	0.902129546	55.60%	0	0	Folate
59	60	294	C4H2N2O'	94.0167127	'ura[c]'	'Uracil'	'C4H4N2O2'	0.839230542	75.60%	4	2	Thymidine
27	44	125	'C5H3N2'	91.0296231	'im4act[c]'	'Imidazole-4-Acetaldehyde'	'C5H6N2O'	0.82718098	30.10%	0	0	Histidine
93	31	114	'C6H7O'	95.0496898	'4mop[c]'	'4-Methyl-2-Oxopentanoate'	'C6H9O3'	0.736504322	22.60%	1	1	Leucine
90	30	110	'C6H6O'	94.0418648	'3mop[c]'	'3-Methyl-2-Oxopentanoate'	'C6H9O3'	0.728695065	22.60%	1	1	Isoleucine
29	39	130	'C9H6N'	128.050024	'srtn[c]'	'Serotonin'	'C10H13N2O'	0.723026586	27.00%	0	0	Tryptophan
25	37	148	'S'	31.9720707	'3mtp[c]'	'3-Methyl-Thio-Propionate'	'C4H8OS'	0.307336327	44.30%	1	1	methionine
44	26	103	'CHN'	27.010899	'crtn[c]'	'Creatinine'	'C4H7N3O'	0.238909951	28.60%	1	1	Arginine