

1 **MitoCore: A curated constraint-based model for simulating human**  
2 **central metabolism**

3  
4 **Anthony C. Smith<sup>1</sup>, Filmon Eyassu<sup>1</sup>, Jean-Pierre Mazat<sup>2,3</sup> and Alan J. Robinson<sup>1\*</sup>**

5  
6 <sup>1</sup> MRC Mitochondrial Biology Unit, University of Cambridge, Cambridge Biomedical Campus, Hills  
7 Road, Cambridge, CB2 0XY, UK

8 <sup>2</sup> Institute of Biochemistry and Genetics of the Cell, CNRS-UMR5095, 1 Rue Camille Saint Saëns, 33077  
9 Bordeaux cedex, France

10 <sup>3</sup> University Bordeaux Segalen, 146 rue Léo Saignat, 33076 Bordeaux cedex, France

11  
12 \* Corresponding author

13  
14 [AJR: alan.robinson@mrc-mbu.cam.ac.uk](mailto:alan.robinson@mrc-mbu.cam.ac.uk)

15

16

17

18

19

1 **Abstract**

2

3 **Background:** The complexity of metabolic networks can make the origin and impact of profound  
4 changes in central metabolism occurring during disease difficult to understand. Computer simulations  
5 can help unravel this complexity, and progress has advanced in genome-scale metabolic models.  
6 However, many current models produce unrealistic results when challenged to simulate abnormal  
7 metabolism as they include incorrect specification and localization of reactions and transport steps,  
8 incorrect reaction parameters, and confounding of prosthetic groups and free metabolites in  
9 reactions. Other common drawbacks are due to their scale, such as being difficult to parameterise  
10 and simulation results being hard to interpret. Therefore, it remains important to develop smaller,  
11 manually curated models to represent central metabolism accurately.

12

13 **Results:** We present MitoCore, a manually curated constraint-based computer model of human  
14 metabolism that incorporates the complexity of central metabolism and simulates this metabolism  
15 successfully under normal and abnormal conditions, including hypoxia and mitochondrial diseases.  
16 MitoCore describes 324 metabolic reactions, 83 transport steps between mitochondrion and cytosol,  
17 and 74 metabolite inputs and outputs through the plasma membrane, to produce a model of  
18 manageable scale for easy data interpretation. Its key innovations include accurate partitioning of  
19 metabolism between cytosol and mitochondrial matrix; correct modelling of connecting transport  
20 steps; proper differentiation of prosthetic groups and free co-factors in reactions; and a new  
21 representation of the respiratory chain and the proton motive force. MitoCore's default parameters  
22 simulate normal cardiomyocyte metabolism, and to improve usability and allow comparison with  
23 other models and types of analysis, its reactions and metabolites have extensive annotation, and  
24 cross-reference identifiers from Virtual Metabolic Human database and KEGG. These innovations—  
25 including over 100 reactions absent or modified from Recon 2—are essential to model central  
26 metabolism accurately.

27

28 **Conclusion:** We anticipate MitoCore as a research tool for scientists, from experimentalists looking to  
29 interpret data and generate further hypotheses, to experienced modellers predicting the  
30 consequences of disease or using computationally intensive methods that are infeasible with larger  
31 models, as well as a teaching tool for those new to modelling and needing a small manageable model  
32 on which to learn and experiment.

33

34 **Keywords:** Constraint-based model, metabolic network, flux balance analysis, central metabolism,  
35 mitochondria; mitochondrial metabolism

## 1 **Background**

2 Human central metabolism is a large and complex system under sensitive homeostatic control, and its  
3 disturbance is causative or associated with many diseases and responses to toxins. However, it is  
4 often difficult to relate more than a handful of these changes to their underlying origin or their down-  
5 stream impact, due to the highly connected nature of the reactions of central metabolism. Computer  
6 models are widely accepted in many fields as a tool to incorporate complexity and simulate changes,  
7 allowing predictions to be made and providing a unifying framework to interpret empirical data,  
8 especially from large, noisy and incomplete data sets. Yet modelling is treated with scepticism by  
9 many biomedical researchers despite their potential broad utility (1). Simple models of enzyme  
10 kinetics (using the assumptions of Henri-Michaelis-Menten kinetics (2)) are familiar to biomedical  
11 scientists, but are impractical for simulations of central metabolism due to every reaction needing  
12 parameterisation, alongside the computational expense of solving the large set of differential  
13 equations. However, constraint-based models of metabolism used in conjunction with methods such  
14 as flux balance analysis (3) are particularly useful for simulating metabolic changes in large metabolic  
15 networks as they can incorporate flexibility, do not require any kinetic parameters and are  
16 computationally inexpensive. Many genome-scale constraint-based models (4-10), representing  
17 known enzymes encoded in the human genome, have covered central metabolism and used to  
18 successfully model some diseases (11,12). But these models do not simulate the realistic production  
19 rate of ATP (with one recent exception (10)), a crucial element of modelling central metabolism.  
20 Furthermore, the interpretation of simulation results from thousands of reactions is difficult  
21 (especially for new-comers), and erroneous “short-circuits” and energy-generating cycles commonly  
22 occur without extensive manual curation (13). In addition, attempts to simulate disease can result in  
23 the prediction of physiologically improbable reaction fluxes. These are caused by several common  
24 problems including: incorrect parameters for directionality constraints, the assignment of reactions to  
25 the wrong cellular compartments, or inaccurate representations of pathways, enzymes, transport  
26 steps, prosthetic groups and metabolites. These errors can introduce unrealistic bypasses and  
27 shuttles that appear to compensate for a disease state. For example, proton-coupled mitochondrial  
28 transporters running in reverse and thus pumping protons that contribute to ATP generation, and the  
29 confounding of free co-factors with prosthetic groups, especially the FAD of mitochondrial succinate  
30 dehydrogenase and electron-transferring flavoprotein (ETF), leading to unrealistic fluxes of electrons  
31 between isolated complexes. These problems are common in genome-scale models that include an  
32 initial auto-generation of the reaction network from databases that can include incomplete or  
33 incorrect annotation. These issues are particularly acute for modelling mitochondrial metabolism and  
34 metabolite transport, as all the current genome-scale models neglect the electrical gradient  
35 component ( $\Delta\Psi$ ) of the proton motive force (PMF) and the correct proton cost of making ATP by ATP  
36 synthase in animals (14). It is also sometimes questionable whether the enormous size and  
37 complexity of these genome-scale models benefits simulations of subsystems of cellular metabolism,

1 such as central metabolism. Furthermore, genome-scale models make some techniques  
2 computationally infeasible due to their scale, such as elementary mode analysis (15).

3

4 By using smaller curated models validated against data from normal and disease metabolism these  
5 problems can be avoided. A more focussed and carefully defined model also allows the user to be  
6 confident in each reaction and more clearly elucidate the behaviour of the system, including any  
7 short-comings, and interpret the results more readily. We previously applied this approach for our  
8 *iAS253* model of the mitochondrion, which we used to simulate metabolic diseases of the  
9 tricarboxylic acid cycle (16). This model was then used as a basis to simulate other disorders including  
10 hypoxia during cardiac ischemia (17), fumarate hydratase deficiency (18) and common diseases of the  
11 mitochondrial electron transport chain (19). These simulation results were used to generate detailed  
12 mechanistic hypotheses for data interpretation and to design further experiments. However, we  
13 recognised that this model could be improved upon by constructing a new model that encompassed  
14 more of central metabolism and explicitly modelled physiochemical features such as the  
15 mitochondrial proton motive force. In particular, ease of use would be improved by providing  
16 extensive annotation of reactions and their parameters.

17

18 Here we present MitoCore, a new constraint-based model of central metabolism that addresses these  
19 issues and comprehensively expands upon and refines our previous mitochondrial models. This model  
20 has been designed to be easy-to-use, includes extensive annotation, has default parameters to  
21 simulate human cardiomyocyte metabolism, and is encoded in the widely used SBML format (20). We  
22 anticipate the model will be of great use to those wishing to interpret empirical data by comparing it  
23 to simulations of central metabolism and thus investigate predictive models of disease and  
24 toxicology.

25

## 1 **Results**

### 2 *Building the MitoCore model of human central metabolism*

3 MitoCore is a constraint-based model of central metabolism with two compartments; one  
4 representing the cytosol, outer mitochondrial membrane, inter-membrane space and cytosolic side of  
5 the inner mitochondrial membrane, and the other the mitochondrial side of the inner membrane and  
6 the mitochondrial matrix. It models 324 reactions of which 157 have been assigned to the  
7 mitochondrial compartment, and 167 assigned to the cytosolic compartment. 83 mitochondrial  
8 transport steps connect the two compartments, of which 30 are modelled on the known transport  
9 mechanisms of characterised transport proteins of the inner mitochondrial membrane, whereas the  
10 other 53 represent known transport capabilities of the membrane, such as diffusion of small  
11 metabolites. 74 transport steps at the cytosolic boundary represent the import and export of  
12 metabolites across the plasma membrane, such as oxygen, carbon dioxide, glucose, fatty acids and  
13 amino acids. For convenience MitoCore includes four 'pseudo' reactions that summarise aspects of its  
14 biological activity and can be used by flux balance analysis as objective functions: ATP hydrolysis  
15 (representing cellular ATP demand), and the biosynthesis of heme, lipids and amino acids  
16 (representing the anabolism of biomolecules for cellular growth).

17

18 To create the MitoCore model of central metabolism, we built a list of candidate reactions to include  
19 by considering human reactions in the KEGG (21), HumanCyc (22) and BRENDA (23) databases that  
20 use any metabolites involved in central metabolism, and assigned each reaction to the appropriate  
21 cellular compartment(s) by assessing the localisation evidence collated in the MitoMiner database  
22 (24) for its catalysing protein. Reaction directionality was assigned by taking the consensus from  
23 annotation in metabolic databases, estimates of Gibbs free energy (25,26), and general rules of  
24 irreversibility (27). For each reaction extensive additional annotation was recorded including the  
25 original KEGG identifiers, EC number, description, gene mappings (both HUGO gene symbol and  
26 Ensembl identifiers), and evidence for the gene's expression in heart and the protein's mitochondrial  
27 localisation.

28

29 To enable comparison of results from MitoCore to those of the popular genome-scale models such as  
30 Recon 2 (5), MitoCore re-used identifiers for metabolites and reactions present in the Virtual  
31 Metabolic Human database (<https://vmh.uni.lu>) where possible. However, it was necessary to create  
32 105 new reactions for MitoCore (additional file 6) that were either absent from the Virtual Metabolic  
33 Human database (such as transport steps or compartment specific versions), were inaccurately  
34 described (such as specifying prosthetic FAD as a free co-factor), or were needed to represent new  
35 features (such as the proton motive force). New reaction identifiers were appended with the suffix  
36 'MitoCore'. Pathways represented in MitoCore include glycolysis, pentose phosphate pathway, TCA  
37 cycle, electron transport chain, synthesis and oxidation of fatty acids, ketone body and amino acid  
38 degradation and cover all parts of central metabolism involved directly or indirectly with ATP

1 production. Finally, MitoCore was extensively tested to ensure that it contained no erroneous energy-  
2 generating cycles, it was capable of simulating disorders such as ischemia and mitochondrial diseases,  
3 and that each reaction was capable of having a flux (depending on the constraints placed on the  
4 cytosolic boundary transport steps) and so ensure it contained no reaction dead-ends. MitoCore was  
5 encoded in SBML (20) (additional file 1) and a companion annotation Excel spreadsheet produced  
6 (additional file 2).

7

### 8 *Simulating cardiomyocyte metabolism using MitoCore and flux balance analysis*

9 MitoCore's default reactions and parameters are optimised for cardiomyocytes and use the  
10 metabolites available to healthy hearts of glucose, fatty acids, ketone bodies and amino acids  
11 (references listed in additional file 2). To demonstrate that MitoCore was capable of producing  
12 physiological relevant results using these parameters, we simulated cardiomyocyte metabolism by  
13 using flux balance analysis (3). Flux balance analysis calculates the optimum rate of turnover, or flux,  
14 of metabolite through each reaction of a network, given a particular objective. To reflect the primary  
15 role of central metabolism in cardiomyocytes we set the simulation's objective as maximum ATP  
16 production (by maximising flux through the pseudo reaction of ATP hydrolysis) and calculated the  
17 optimum reaction fluxes through central metabolism. The resultant reaction fluxes simulated core  
18 metabolism correctly with activity of all respiratory complexes and the TCA cycle (Fig 1, additional file  
19 2). As metabolic fuels were provided in slight excess, the availability of oxygen limited the overall  
20 fluxes. Simulated ATP production was 100.9  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry weight. Sources of acetyl-CoA  
21 for the TCA cycle were fatty acid degradation (55.0%), glucose oxidation (26.4%), lactate oxidation  
22 (8.4%), ketone body degradation (6.1%), amino acid degradation (3.8%) and glycerol oxidation (0.3%).  
23 The amino acids degraded and used to produce ATP were histidine, isoleucine, leucine, lysine,  
24 threonine, valine, arginine, aspartate, cysteine, glycine, proline, serine, asparagine, and alanine.  
25 Ammonia, produced as a by-product of amino acid degradation, was exported from the system.

26

### 27 *Metabolite degradation and ATP yields using MitoCore*

28 To demonstrate how different metabolites are degraded appropriately in the MitoCore model and  
29 the effect of the implementation of the proton motive force on ATP production, we performed a  
30 series of simulations using particular 'fuel' metabolites in isolation. We simulated the oxidation and  
31 degradation of glucose, lactate, hexadecanoic acid, hydroxybutanoate, acetoacetate and 20 different  
32 amino acids in separate simulations. In each simulation, one fuel was allowed a maximum uptake of 1  
33  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry weight while all others were set to zero, then the objective function was  
34 set to maximise ATP hydrolysis, and the optimum reaction fluxes in the metabolic network were  
35 calculated by using flux balance analysis (additional file 3). The simulations showed unsurprisingly that  
36 fatty acids were the most energy rich fuels to metabolise (ATP production of 112  $\mu\text{mol}/\text{minute}/\text{gram}$   
37 of dry weight), followed by the amino acids tryptophan (43  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry weight),  
38 isoleucine (38  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry weight), leucine (37  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry weight) and

1 phenylalanine (36  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry weight) and then glucose (33  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry  
2 weight).

3

#### 4 *The effect of proton leak on the electron transport chain in MitoCore*

5 To show the MitoCore model can simulate experimentally induced conditions that are directly  
6 affected by the proton motive force, we performed a series of simulations representing an increasing  
7 proton leak across the mitochondrial inner membrane. We introduced a flux through the transport  
8 step that represents the leak of matrix protons to the cytosol through the uncoupling protein 2  
9 (UCP2, Reaction ID: HtmB\_MitoCore) in cardiomyocytes (28), while using the objective function of  
10 maximum ATP hydrolysis during flux balance analysis. The optimum reaction fluxes (additional file 4)  
11 were identical to those under default constraints with the exception of the steps involved in  
12 mitochondrial ATP synthesis. The flux through ATP synthase was progressively reduced as the proton  
13 leak increased until it reversed under very high leak, using ATP synthesized in the Krebs cycle. At the  
14 highest level of proton leak the ATP/ADP carrier reversed to provide additional ATP from glycolysis  
15 (Reaction ID: ATPtmB\_MitoCore). Similarly the phosphate carrier also reversed at the highest proton  
16 leak value.

17

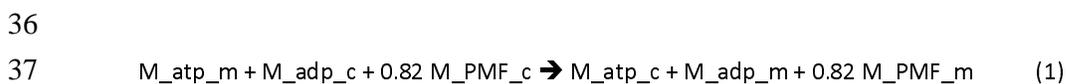
18

## 1 Discussion

2 Here we present MitoCore, a curated, constraint-based model of human central metabolism  
3 (additional file 1) designed as a predictive model of metabolism in disease and toxicology, and for use  
4 by a wide range of researchers. It covers all major pathways involved in central metabolism using 407  
5 reactions and mitochondrial transport steps, and 74 transport steps over the plasma membrane. To  
6 increase the metabolic flexibility of MitoCore, we included a large number of reactions that were not  
7 assigned to classical pathways, but could have potentially important roles in supporting central  
8 metabolism. MitoCore was parameterised and annotated for cardiomyocyte metabolism, which is  
9 useful for many types of analyses as the cardiomyocyte can metabolise a wide range of substrates  
10 and has reactions common to many other cell types, as well as representing the metabolism of an  
11 organ of utmost importance in human health, disease and toxicology. This allows the simulation  
12 results to be generalisable, without having features that are particularly cell specific, such as those  
13 found only in hepatocytes. However, we have included reactions that are switched off by default in  
14 the cardiac model, but can be activated to represent the abilities of other cell types, e.g.  
15 gluconeogenesis, ketogenesis,  $\beta$ -alanine synthesis and folate degradation. Thus, the model allows  
16 biologically relevant flux distributions to be generated 'out of the box' without altering the model,  
17 while allowing for easy modification to represent other cell types.

18  
19 MitoCore has several unique features. The first is the accurate partitioning of metabolism between  
20 the mitochondrion and the cytosol by using extensive localisation data and annotation. This  
21 partitioning has an important impact on model behaviour as the limited transport steps into the  
22 mitochondrial matrix result in dramatic differences in metabolite availability in the matrix compared  
23 to the cytosol. Therefore, it is crucial that reactions are assigned to the correct compartment. We  
24 achieved this by manually evaluating the localisation of each reaction's catalysing protein by using the  
25 mitochondrial localisation evidence in the MitoMiner database (24). MitoMiner collates GFP tagging,  
26 large-scale mass-spectrometry mitochondrial proteomics studies and mitochondrial targeting  
27 sequence predictions, with detailed annotation from the Gene Ontology and metabolic pathway data  
28 from KEGG. MitoMiner also contains homology information allowing localisation evidence to be  
29 shared amongst species. For reactions catalysed by enzymes with a large amount of evidence for  
30 mitochondrial localisation but lacking specific evidence for being in the mitochondrial matrix or matrix  
31 side of the inner membrane, we applied the principle of metabolite availability—a reaction can only  
32 be present in a compartment if all its substrates are available and its products can be used by  
33 reactions within the same compartment (16). A summary of this localisation evidence is provided in  
34 the mitochondrial evidence section of the supplementary annotation file (additional file 2, 'Reaction  
35 & Fluxes' worksheet) and consists of confidence scores from the MitoCarta 2 (29) inventory of genes  
36 that encode mitochondrial proteins. This dataset is derived by using machine learning to evaluate  
37 mitochondrial localisation data from training on collections of characterised mitochondrial and non-  
38 mitochondrial genes.

1  
2 The partitioning of metabolism between the mitochondrion and the cytosol logically led us  
3 reconsider the proton motive force (PMF) and the role of protons crossing the inner mitochondrial  
4 membrane as part of oxidative phosphorylation, as it produces the majority of cellular ATP. The PMF  
5 is achieved by complexes I, III and IV pumping matrix protons across the mitochondrial inner  
6 membrane and into the intermembrane space. This proton-pumping creates a proton motive force  
7 (PMF) across the membrane that has two components: a proton gradient ( $\Delta\text{pH}$ ) coupled with an  
8 electrical membrane potential ( $\Delta\Psi$ ). The energy for proton-pumping comes from the transfer of  
9 electrons down the respiratory chain from NADH and ubiquinone to oxygen, to form water.  
10 Additional electrons are passed into the respiratory chain from the TCA cycle by complex II, and from  
11 the degradation of fatty acids and amino acids by the electron-transfer flavoprotein (ETF). ATP  
12 synthase uses the PMF to power ATP synthesis from ADP and phosphate by channelling protons back  
13 across the membrane. It is thus clear that it is necessary to distinguish as Peter Mitchell did (30), the  
14 protons involved in chemical reactions taking place in an isolated compartment (“scalar protons”),  
15 from the protons crossing between compartments (“vectorial protons”). To better model oxidative  
16 phosphorylation, we devised a new representation of the PMF and mitochondrial respiratory chain—  
17 the second unique feature of MitoCore. MitoCore’s representation of the respiratory chain differs in  
18 many key aspects to other metabolic models, due to how it models vectorial protons and accounts for  
19 both components of the PMF. MitoCore represents the PMF as a metabolite that is co-transported in  
20 steps that transport charged metabolites or protons across the inner mitochondrial membrane, such  
21 as the reactions of the respiratory complexes, and in proton-coupled and electrogenic transport  
22 steps. As MitoCore’s representation of the PMF incorporates both  $\Delta\Psi$  and  $\Delta\text{pH}$ , we reflect their  
23 relative contributions to the overall PMF by co-transporting 0.82 PMF metabolites for transport steps  
24 that affect  $\Delta\Psi$  and 0.18 PMF metabolites for transport steps that affect  $\Delta\text{pH}$ . These values are the  
25 average of published figures of the relative contributions of  $\Delta\text{pH}$  and  $\Delta\Psi$  by several authors (see  
26 additional file 5 for details and references). Therefore, the reactions that represent complexes I, III  
27 and IV of the respiratory chain move PMF metabolites that correspond to the number of protons they  
28 pump from the matrix to the cytosol, as a proton in this case affects both  $\Delta\Psi$  and  $\Delta\text{pH}$ . ATP synthase  
29 needs to transport 2.7 PMF metabolites back to the matrix to synthesise one molecule of ATP (as it  
30 uses 2.7 protons per molecule of ATP (14)). Transport steps between the two compartments are  
31 modelled in the same way; for example the mitochondrial ATP/ADP carrier 1 (SLC25A4) requires 0.82  
32 PMF to be co-transported with each imported  $\text{ADP}^{3-}$  nucleotide to reflect the charge difference  
33 between  $\text{ATP}^{4-}$  and  $\text{ADP}^{3-}$  that affects only  $\Delta\Psi$  (equation 1), whereas the proton-coupled phosphate  
34 carrier (SLC25A3) imports 0.18 PMF as overall transport is electro-neutral and so only affects  $\Delta\text{pH}$   
35 (equation 2):





3  
4 This separate modelling of vectorial protons using an additional new PMF metabolite enabled the  
5 impact of electrogenic transporters on the PMF to be accounted for in flux balance analysis  
6 simulations for the first time and prevented simulation artefacts where (scalar) protons generated or  
7 removed from other parts of metabolism allow unrealistically high ATP production, as can be the case  
8 in other models. We have removed FADH from reactions (as specified in databases such as KEGG)  
9 where it is treated as a redox cofactor (such as for complex II and the ETF), and instead directly  
10 coupled the reactions to the reduction of ubiquinone to ubiquinol. This keeps separate the electrons  
11 entering the respiratory chain from different sources, which can otherwise become connected via a  
12 shared FADH metabolite, and is particularly relevant under perturbed conditions where the  
13 erroneous connection of free co-factor and bound prosthetic FADH can cause unrealistic bypasses to  
14 occur. This problem is endemic in large-scale models that auto-generate reaction networks directly  
15 from metabolic databases without manual curation of reactions. We model the generation of reactive  
16 oxygen species (ROS) at 0.001% of the flux through complex I, to reflect the primary site of ROS  
17 generation from the respiratory chain and reduce the efficiency of proton pumping as electrons  
18 escape (31,32). Further reactions convert ROS to water at the expense of NADPH.

19  
20 The third unique feature of MitoCore is the modelling of the transport steps that connect the  
21 cytosolic and mitochondrial compartments. We included four different transport categories. First  
22 were transport steps based on the characterised mitochondrial transport proteins. Many carriers can  
23 transport a range of related substrates (although with different affinities) and each metabolite  
24 combination was modelled as a separate step including counter exchange and proton-coupling. In  
25 some cases proton-coupled transporters were represented by two reactions to model separately the  
26 forward and reverse directions, and a proton (plus co-transported PMF metabolite) only used for  
27 movement down the proton gradient. This prevented transporters being used in reverse to pump  
28 protons artificially (and so transferring PMF) and thereby contribute to ATP production by the ATP  
29 synthase. The second category was for metabolites whose transporters are unidentified. We chose to  
30 model these steps as uniporters that were not proton coupled. If the metabolite was charged and  
31 moving into the matrix we assumed this would impact  $\Delta\Psi$  and co-transported 0.82 PMF metabolite  
32 accordingly. The third category was for metabolites that can diffuse across the inner mitochondrial  
33 membrane—including oxygen, carbon dioxide and water—and these were modelled as reversible  
34 uniport transport steps. Finally we modelled the insertion of lipids into the mitochondrial inner  
35 membrane via the flippases, which was coupled to ATP hydrolysis.

36  
37 Other improvements to MitoCore included modelling the protein complexes as one reaction rather  
38 than a series of linked reactions, as can often be found in metabolic pathway databases and other

1 models. This is particularly important for gene knockout simulations where a gene deficiency will  
2 knock out a protein complex rather than just one of the constituent reactions. To facilitate gene-  
3 based analyses (such as knock out studies) we provide reaction-to-gene mappings with both gene  
4 symbols and Ensembl identifiers. To summarise the behaviour of MitoCore, we also defined four  
5 'pseudo' reactions for biosynthesis of biomolecules (amino acids, lipids, DNA) required by cells for  
6 growth and ATP hydrolysis to reflect the energy demand of the cell. These pseudo reactions are  
7 designed to be used as objective functions during flux balance analysis.

8

9 A weakness of many metabolic models is the lack of provenance about their components. For  
10 example recording why reactions have been included, why directionality constraints have been set, or  
11 the origin of reaction parameters. Thus for MitoCore we created a supplementary annotation  
12 spreadsheet (additional file 2) to record this provenance. For example, the directionality evidence  
13 section describes why a constraint has been set; a manually evaluated consensus of information from  
14 the KEGG (21), HumanCyc (22) and BRENDA (23) databases, general rules of irreversibility (27), large  
15  $\Delta G$  values from eQuilibrator (26) or estimated using a group contribution method (25), and  
16 information from the literature. In cases for which reaction directionality was unclear, it was kept  
17 reversible. To address why a reaction has been included in our cardiomyocyte model, we included a  
18 heart expression section consisting of RNAseq and immunochemistry expression levels of genes taken  
19 from the Human Protein Atlas (version 14) (9). The spreadsheet also includes gene mappings,  
20 identifiers from Recon 2 and KEGG, mitochondrial localisation evidence (as described above) and  
21 baseline reaction fluxes when the objective function was maximum ATP production under normal  
22 conditions. The spreadsheet also serves as a useful template to map flux distributions onto, as  
23 reaction fluxes can be grouped in an intuitive way directly against useful supplementary information.  
24 When combined with the small size of the model, simulations can be generated extremely quickly and  
25 then easily interpreted.

26

27 To demonstrate that MitoCore functions realistically, we simulated cardiomyocyte metabolism using  
28 the default parameters with flux balance analysis (3) and the objective of maximum ATP production.  
29 The reaction fluxes showed central metabolism was modelled correctly, with the largest fluxes  
30 through the TCA cycle and respiratory chain. Numerous fuel sources (fatty acids, glucose, lactate,  
31 ketone bodies and amino acids) were imported, degraded and entered the TCA cycle at several  
32 different points (Figure 1). The sources of acetyl-CoA for the TCA cycle were similar to the  
33 experimental measurements of well-perfused heart (33)—experiments report 60-90% of acetyl-CoA  
34 derives from fatty acids compared to 55% in the MitoCore simulation, whereas glycolysis (including  
35 lactate oxidation) accounts for 10-40% of acetyl-CoA compared to 34.8% in the MitoCore simulation.  
36 These results support the biological relevance and parameter choice of the model.

37

1 To show the importance of explicitly modelling the PMF, we simulated the maximum ATP production  
2 achievable using 1  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry weight of common metabolic fuels in isolation  
3 (effectively calculating their ATP yields) including for glucose and the fatty acid hexadecanoic acid  
4 (additional file 3). In MitoCore each glucose produced 33 ATP in comparison to 32 calculated  
5 theoretically (34) and 32 in Recon 2.2 (10). For the fatty acid hexadecanoic acid 112 ATP was  
6 produced compared to 108 theoretically (34) and 107 in Recon 2.2. A significant difference between  
7 the models is the number of protons required by ATP synthase to produce one molecule of ATP—  
8 MitoCore uses 2.7 (based on the structure of the ATP synthase (14)) whereas the theoretical  
9 calculations use 2.5 and Recon 2.2 uses 4.0. Due to also considering additional factors that affect ATP  
10 production—including the impact of all the transport steps on the PMF as well as ROS production and  
11 removal—we believe our figure is likely to be more accurate than these other calculations. When  
12 comparing maximum ATP production using default parameters to the previous *iAS253* model (16),  
13 ATP production was notably lower (101 vs 140  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry weight) demonstrating that  
14 the separation between the PMF and protons is an important improvement. These simulations also  
15 show that the model in each case uses canonical degradation pathways, further increasing confidence  
16 that the model is capable of producing realistic results using a wide range of metabolites. The results  
17 from the proton leak simulations (additional file 4) show the use of metabolites to represent the PMF  
18 enables the model to replicate experimental observations under perturbed conditions that would  
19 otherwise be impossible, such as the reversal of ATP synthase, the ATP/ADP carrier and the  
20 phosphate carrier, which is well-known behaviour in absence of respiratory chain activity, for instance  
21 in  $p^0$  cells to maintain a proton motive force (35,36).

22  
23

1 **Conclusions:**

2 MitoCore is a constraint-based model of human central metabolism, provided with default  
3 parameters that provide physiologically realistic reaction fluxes for cardiomyocytes. The model has  
4 several innovations including a new representation of the respiratory chain and proton motive force,  
5 and partitioning of reactions to subcellular compartments based on the latest localisation evidence.  
6 Each of the 402 reactions was manually evaluated for directionality, expression of its gene in heart  
7 and subcellular localisation of its protein resulting in an accurate depiction of central metabolism. To  
8 allow MitoCore to be easily used and to make the results directly comparable with other models and  
9 compatible with other types of analyses, we have used identifiers from the Virtual Metabolic Human  
10 database for both reactions and metabolites where possible, and recorded KEGG identifiers in the  
11 annotation. To help ease of use, we also provide an annotation spreadsheet that provides gene  
12 mappings, localisation and heart expression evidence, and notes on parameter choice. MitoCore is  
13 provided in SBML format to be compatible with a wide range of software. We hope MitoCore will be  
14 of use as a research tool to a wide range of biomedical scientists and students—from experienced  
15 modellers interested in central metabolism or using computationally intensive methods that are  
16 infeasible on a genome-scale model, to those new to modelling who would like begin by using a small  
17 manageable model, with application as a predictive model of disease.

## 1 **Methods**

### 2 **Identifying reactions of central metabolism to include in MitoCore**

3 An updated version of the *iAS253* mitochondrial model was the starting point for the MitoCore model  
4 (19). The model was expanded by searching for human reactions in KEGG (21), HumanCyc (22) and  
5 BRENDA (23) that were missing from the *iAS253* model and could impact central metabolism. Each  
6 reaction was reassessed for subcellular localisation and directionality (see below), and to ensure the  
7 reactions were present in most tissues including cardiac, tissue expression of the genes and proteins  
8 was verified by using the Human Protein Atlas (9).

9

### 10 **Partitioning reactions between the cytosol and mitochondrion**

11 To partition the reactions into either cytosol or mitochondrion, each enzyme was manually evaluated  
12 using the mitochondrial localisation evidence in the MitoMiner database (24). All available  
13 experimental localisation evidence, mitochondrial targeting sequence predictions and annotation  
14 were considered, including from homologs from mouse, rat and yeasts. For reactions with strong  
15 evidence for mitochondrial localisation, but where matrix localisation is unclear, we used the principal  
16 of metabolite availability (16). Reactions residing in the mitochondrial matrix or matrix side of the  
17 inner membrane were assigned to the mitochondrial compartment. Transport steps were created to  
18 connect the cytosol and matrix based on the transport properties of the membrane (active transport,  
19 diffusion, etc.). Each reaction was cross-referenced with the Virtual Metabolic Human database  
20 (<https://vmh.uni.lu>) and used its identifiers for reactions and metabolites where possible.

21

### 22 **Assigning reaction directionality**

23 Reaction directionality was manually evaluated for each reaction in MitoCore. The KEGG (21),  
24 HumanCyc (22) and BRENDA (23) databases were consulted and general rules of irreversibility were  
25 taken into account (such as most reactions consume ATP rather than produce it and carbon dioxide is  
26 normally produced not consumed) (27). We also considered the  $\Delta G$  values for reactions, both  
27 calculated by eQuilibrator (26) and estimated using the group contribution method (25), and large  
28 changes noted. Finally we consulted the literature if reaction directionality was conflicted or unclear.  
29 If support for irreversibility was poor or a consensus could not be found, the reaction was assigned  
30 reversible. The information used to make each assessment was recorded in the directionality  
31 evidence column of the annotation spreadsheet (additional file 2). Further refinement of reaction  
32 directionality was used to eliminate loops that could produce metabolites such as ATP and NADH for  
33 'free', and to prevent the interconversion of NADH and NADPH unless experimentally verified.

34

### 35 **Defining reactions**

36 To improve modelling of the proton motive force (PMF) across the mitochondrial inner membrane,  
37 we introduced a pseudo-metabolite to model the effect on the PMF of proton and electrogenic  
38 transport steps across the inner mitochondrial membrane. Reactions representing respiratory

1 complexes as well as proton-coupled transport steps were rewritten to use this new species. To  
2 prevent unrealistic bypasses between free and complex-bound prosthetic flavin adenine  
3 dinucleotides FADH and FAD<sup>+</sup> were removed from all reactions and replaced with ubiquinone and  
4 ubiquinol. Many reactions were rewritten to represent better the underlying biology, such as the  
5 generation of ROS from complex I and the proton-pumping stoichiometry of the respiratory  
6 complexes. In some cases the subcellular localisation was changed by using mitochondrial metabolite  
7 species, or new reactions were written using existing metabolite species.

8

9 In total 105 MitoCore reactions were different to or absent from Recon 2 (additional file 6). To  
10 highlight where there are differences between MitoCore reactions and corresponding reactions in the  
11 Virtual Metabolic Human database, the MitoCore reaction identifiers use the suffix 'MitoCore' and  
12 the same Recon 2 identifier if one exists.

13

#### 14 **Refining the MitoCore model**

15 Finally, the model was extensively tested to ensure all reactions were capable of having flux,  
16 erroneous energy-generating cycles were removed, the model behaved physiologically under normal  
17 and perturbed conditions, and all the objective functions were feasible. The model was encoded in  
18 SBML v2.1 and its validity checked with the SBML online validator (37).

19

#### 20 **Simulating cardiomyocyte metabolism**

21 We simulated metabolism in cardiomyocytes with flux balance analysis (FBA) by using MitoCore's  
22 default parameters that have been experimentally recorded for heart tissue. FBA has been described  
23 extensively elsewhere, but can summarised as calculating the reaction turnover, or fluxes (flows) of  
24 metabolites through a network of biochemical reactions assuming a pseudo steady state (3). The  
25 fluxes through the network are constrained by the stoichiometry and directionality of the reactions as  
26 well as flux capacity and cytosol boundary uptake ranges. Cytosol boundary transport steps model the  
27 import and export of metabolites to the cell, but the overall rate of production and consumption of  
28 metabolites is assumed to be zero, hence a pseudo steady state. A metabolic objective function is  
29 chosen for a simulation, and FBA is used to calculate an optimal set of reaction fluxes that maximises  
30 this function. For this simulation we use maximum ATP production as the objective function because  
31 energy generation is one of the primary purposes of central metabolism in cardiomyocytes.

32

33 For simulations of ATP yield, all cytosol boundary uptake fluxes for metabolites that could be  
34 degraded to produce ATP were set to zero, oxygen was increased to 50  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry  
35 weight (so that limited oxygen availability would not affect the results), while other cytosol boundary  
36 conditions were unaltered. The uptake flux of each metabolite of interest was then increased to 1  
37  $\mu\text{mol}/\text{minute}/\text{gram}$  of dry weight.

38

1 For simulations of proton leak, the lower bound of the reaction representing the gene UCP2 (Reaction  
2 ID: HtmB\_MitoCore) was increased over a series of simulations, thus forcing a minimum flux through  
3 the reaction.

4

5 All FBA simulations were performed using MATLAB (Math Works, Inc, Natick, MA) and the COBRA  
6 Toolbox (38), with the linear programming solver GLPK (<http://www.gnu.org/software/glpk>).

7

#### 8 **Availability of data and material**

9 All data generated or analysed during this study are included in this published article and its  
10 supplementary information files. In addition, the model and annotation file will be available at the  
11 MRC Mitochondrial Biology Unit website (<http://www.mrc-mbu.cam.ac.uk/mitocore/>).

12

#### 13 **Competing interests**

14 The authors declare that they have no competing interests.

15

#### 16 **Funding**

17 ACS, FE and AJR were supported by the Medical Research Council, UK. JPM was supported by the Plan  
18 cancer 2014-2019 No BIO 2014 06 and the French Association against Myopathies.

19

#### 20 **Author's contributions**

21 ACS and AJR conceived the study. ACS and JPM devised the methodology. ACS created the model.

22 ACS, FE, JPM, and AJR participated in the curation and testing of the model.

23 All authors contributed to, read and approved the final manuscript.

24

25

#### 26 **Acknowledgements**

27 We wish to thank Lukasz Zielinski and Alexander Smith for their input and testing of the model, and  
28 Edmund Kunji for discussions on the transport of metabolites across the mitochondrial membrane.

29

#### 30 **Figure Legend**

31 Fig 1. Summary of the major active pathways of central metabolism in the flux balance analysis  
32 simulation using the MitoCore model with its default parameters and the objective function of  
33 maximum ATP production. Values of all fluxes are reported in additional file 1.

34

#### 35 **Supporting information captions**

36 Additional file 1. MitoCore model encoded in the SBML format. (XML 795KB)

- 1 Additional file 2. Companion annotation spreadsheet recording evidence and provenance for  
2 reactions and parameters used in MitoCore and fluxes using default parameters for cardiomyocytes.  
3 (XLSX 247 KB)  
4 Additional file 3. Flux distributions around the MitoCore model for maximum ATP production with  
5 different metabolic fuels. (XLSX 263 KB)  
6 Additional file 4. Flux distributions around the MitoCore model for maximum ATP production with  
7 varying proton leak through reaction representing UCP2. (XLSX 230 KB)  
8 Additional file 5. Experimental measurements of the relative contributions of  $\Delta\text{pH}$  and  $\Delta\Psi$  to the  
9 proton motive force. (XLSX 12 KB)  
10 Additional file 6. Table of 105 new or altered reactions from those in the Virtual Metabolic Human  
11 database (Recon 2). (XLSX 85 KB)  
12

13 **References:**

- 14 1. Kirk PDW, Babbie AC, Stumpf MPH. Systems biology (un)certainties. *Science*. 2015 Oct  
15 23;350(6259):386–8.  
16 2. Cornish-Bowden A, Mazat J-P, Nicolas S. Victor Henri: 111 years of his equation. *Biochimie*.  
17 2014 Dec;107 Pt B:161–6.  
18 3. Orth JD, Thiele I, Palsson BO. What is flux balance analysis? *Nat Biotechnol*. 2010 Mar  
19 1;28(3):245–8.  
20 4. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, et al. Global reconstruction of the  
21 human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA*.  
22 2007 Feb 6;104(6):1777–82.  
23 5. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven  
24 global reconstruction of human metabolism. *Nature*. 2013 Mar 3;31(5):419–25.  
25 6. Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, et al. The Edinburgh human  
26 metabolic network reconstruction and its functional analysis. *Mol Syst Biol*. 2007 Jan 1;3:135.  
27 7. Mardinoglu A, Agren R, Kampf C, Asplund A, Nookaew I, Jacobson P, et al. Integration of  
28 clinical data with a genome-scale metabolic model of the human adipocyte. *Mol Syst Biol*.  
29 2013;9:649.  
30 8. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlén M, Nielsen J. Genome-scale metabolic  
31 modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver  
32 disease. *Nat Comms*. 2014;5:3083.  
33 9. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics.  
34 Tissue-based map of the human proteome. *Science*. 2015 Jan 23;347(6220):1260419–9.  
35 10. Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, et al. Recon 2.2: from  
36 reconstruction to model of human metabolism. *Metabolomics*. 2016;12(7):109.  
37 11. Yizhak K, Chaneton B, Gottlieb E, Ruppin E. Modeling cancer metabolism on a genome scale.  
38 *Mol Syst Biol*. EMBO Press; 2015 Jun 1;11(6):817–7.  
39 12. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and

- 1 associated cellular functions. *Nat Rev Genet.* 2014 Feb;15(2):107–20.
- 2 13. Fritzscheier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ. Erroneous energy-generating  
3 cycles in published genome scale metabolic networks: Identification and removal. *PLoS*  
4 *Comput Biol.* Public Library of Science; 2017 Apr;13(4):e1005494.
- 5 14. Watt IN, Montgomery MG, Runswick MJ, Leslie AGW, Walker JE. Bioenergetic cost of making  
6 an adenosine triphosphate molecule in animal mitochondria. *Proc Natl Acad Sci USA.* 2010  
7 Sep 28;107(39):16823–7.
- 8 15. Trinh CT, Wlaschin A, Sreenc F. Elementary mode analysis: a useful metabolic pathway analysis  
9 tool for characterizing cellular metabolism. *Appl Microbiol Biotechnol.* 2009 Jan;81(5):813–26.
- 10 16. Smith AC, Robinson AJ. A metabolic model of the mitochondrion and its use in modelling  
11 diseases of the tricarboxylic acid cycle. *BMC Systems Biology.* 2011 Jun 29;5(1):102.
- 12 17. Chouchani ET, Pell VR, Gaude E, Aksentijević D, Sundier SY, Robb EL, et al. Ischaemic  
13 accumulation of succinate controls reperfusion injury through mitochondrial ROS. *Nature.*  
14 2014 Nov 20;515(7527):431–5.
- 15 18. Ashrafian H, Czibik G, Bellahcene M, Aksentijević D, Smith AC, Mitchell SJ, et al. Fumarate is  
16 cardioprotective via activation of the Nrf2 antioxidant pathway. *Cell Metabolism.* 2012 Mar  
17 7;15(3):361–71.
- 18 19. Zieliński ŁP, Smith AC, Smith AG, Robinson AJ. Metabolic flexibility of mitochondrial  
19 respiratory chain disorders predicted by computer modelling. *Mitochondrion.* 2016 Sep  
20 30;31:45–55.
- 21 20. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, et al. The systems biology  
22 markup language (SBML): a medium for representation and exchange of biochemical network  
23 models. *Bioinformatics.* 2003 Mar 1;19(4):524–31.
- 24 21. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for  
25 gene and protein annotation. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D457–62.
- 26 22. Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PD. Computational prediction  
27 of human metabolic pathways from the complete human genome. *Genome Biol.* 2005 Jan  
28 1;6(1):R2.
- 29 23. Chang A, Schomburg I, Placzek S, Jeske L, Ulbrich M, Xiao M, et al. BRENDA in 2015: exciting  
30 developments in its 25th year of existence. *Nucleic Acids Res.* 2015 Jan;43(Database  
31 issue):D439–46.
- 32 24. Smith AC, Robinson AJ. MitoMiner v3.1, an update on the mitochondrial proteomics  
33 database. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D1258–61.
- 34 25. Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V. Group contribution method for  
35 thermodynamic analysis of complex metabolic networks. *Biophys J.* 2008 Aug 1;95(3):1487–  
36 99.
- 37 26. Flamholz A, Noor E, Bar-Even A, Milo R. eQuilibrator—the biochemical thermodynamics  
38 calculator. *Nucleic Acids Res.* 2012 Jan;40(Database issue):D770–5.
- 39 27. Ma H, Zeng A-P. Reconstruction of metabolic networks from genome data and analysis of  
40 their global structure for various organisms. *Bioinformatics.* 2003 Jan 22;19(2):270–7.
- 41 28. Laskowski KR, Russell RR. Uncoupling proteins in heart failure. *Curr Heart Fail Rep.* 2008

- 1 Jun;5(2):75–9.
- 2 29. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian  
3 mitochondrial proteins. *Nucleic Acids Res.* 2016 Jan 4;44(D1):D1251–7.
- 4 30. Mitchell P. Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic  
5 type of mechanism. *Nature.* 1961 Jul 8;191:144–8.
- 6 31. Cochemé HM, Murphy MP. Complex I is the major site of mitochondrial superoxide  
7 production by paraquat. *J Biol Chem.* 2008 Jan 25;283(4):1786–98.
- 8 32. Murphy MP. How mitochondria produce reactive oxygen species. *Biochem J.* 2009 Jan  
9 1;417(1):1–13.
- 10 33. Stanley WC, Recchia FA, Lopaschuk GD. Myocardial substrate metabolism in the normal and  
11 failing heart. *Physiol Rev.* 2005 Jul 1;85(3):1093–129.
- 12 34. Nelson DL, Cox MM, Lehninger AL. *Lehninger Principles of Biochemistry.* New York, W. H.  
13 Freeman; 2013.
- 14 35. Dupont C-H, Mazat J-P, Guerin B. The role of adenine nucleotide translocation in the  
15 energization of the inner membrane of mitochondria isolated from  $\Delta+$  and  $\Delta_0$  strains of  
16 *Saccharomyces cerevisiae*. *Biochem Biophys Res Commun.* 1985 Nov;132(3):1116–23.
- 17 36. Buchet K, Godinot C. Functional F1-ATPase essential in maintaining growth and membrane  
18 potential of human mitochondrial DNA-depleted rho degrees cells. *Journal of Biological  
19 Chemistry.* 1998 Sep 4;273(36):22983–9.
- 20 37. Bornstein BJ, Keating SM, Jouraku A, Hucka M. LibSBML: an API Library for SBML.  
21 *Bioinformatics.* 2008 Mar 13;24(6):880–1.
- 22 38. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative  
23 prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0.  
24 *Nature Protocols.* 2011 Sep;6(9):1290–307.

25

