Metabolic network reconstruction in Filaria-*Wolbachia* symbiosis

# Aims

Besides new insights on the relationships between worms and endosymbionts, ultimate goal is to deliver panel of new candidate drug targets that lead to macrofilaricidal drugs.

Reconstruction and analysis of combined filaria-Wolbachia metabolic pathways.

* Prelim reconstruction already done on *Loa loa*, which lacks an endosymbiont.
* Similar process will be used here, on *Onchocerca volvulus* and *Brugia malayi*, including their endosymbionts.
* Flux balance analysis will be performed on the networks to identify the contribution of the worm to the fitness of the endosymbionts, and vice versa.
* RNAseq data for different worm life stages will be used in the FBA, providing enzyme expression profiles.
* Will perform *in silico* knockout experiments, and highlight conserved mechanisms of action.

Experimental validation for the essentiality of select enzyme targets.

* 5-10 targets prioritized from previous analysis; should have essential role in adults and not microfilarial stage, be conserved in other filarial worms, and big bonus if specific inhibitors exist.
* Predictions will be validated by RNAi, and if possible, drug inhibition assays.

Contents

[Aims 1](#_Toc476149504)

[Background 3](#_Toc476149505)

[Brugia malayi 3](#_Toc476149506)

[Life cycle 3](#_Toc476149507)

[Prevalence and range 4](#_Toc476149508)

[Lymphatic filariasis 4](#_Toc476149509)

[Genome 4](#_Toc476149510)

[Onchocerca volvulus 5](#_Toc476149511)

[Life cycle 5](#_Toc476149512)

[Prevalence and range 5](#_Toc476149513)

[Onchocerciasis 5](#_Toc476149514)

[Genome 5](#_Toc476149515)

[Metabolic reconstruction 6](#_Toc476149516)

[Wolbachia 6](#_Toc476149517)

[Bm species 7](#_Toc476149518)

[Ov species 7](#_Toc476149519)

[Targeting Wolbachia 7](#_Toc476149520)

[Human filarial infections 7](#_Toc476149521)

[Metabolism 7](#_Toc476149522)

[Energy production 7](#_Toc476149523)

[Redox 8](#_Toc476149524)

[Common irreversible reactions (product) 8](#_Toc476149525)

[Metabolic network reconstruction 8](#_Toc476149526)

[Flux balance analysis 8](#_Toc476149527)

[Compartmentalization 11](#_Toc476149528)

[Incorporating transcriptomics 12](#_Toc476149529)

[Related methods 12](#_Toc476149530)

[Less useful methods 14](#_Toc476149531)

[Implementation notes 14](#_Toc476149532)

[Resources 15](#_Toc476149533)

[References 16](#_Toc476149534)

# Background

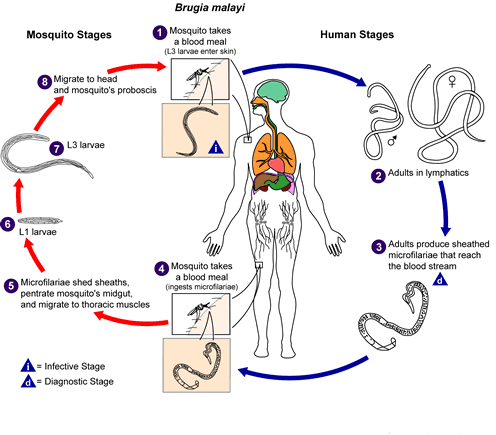
## Brugia malayi

Female adults are 5 cm long, males 2.5 cm.

Only human filarial pathogen that can be maintained in small lab animals.

RNAi works fairly well (Aboobaker and Blaxter, 2003).

### Life cycle



* Microfilariae ingested by mosquito. Exsheathe, molt, penetrate midgut to thoracic muscles, then develop to L1 through L3; process takes 7-21 days.
* L3 migrate to salivary glands then to proboscis, escape onto skin with next bite.
* No reproduction occurs in mosquito.
* L3 actively penetrate skin through bite hole.
* Develop to adults over 6 months in lymphatic system.
* Females release millions of microfilariae per day into the bloodstream.

### Prevalence and range

In 1957, two subspecies were identified.

* Nocturnal periodicity, where microfilariae aren’t detectable in blood during the day, and density peaks between midnight and 2AM.
  + Transmitted by *Mansonia* and some *Anopheles* mosquitos in open swamps and rice growing areas, biting at night. Appear to only infect humans; natural animal infections are rare and lab animals don’t retain infection.
* Nocturnal subperiodicity, where microfilariae are always present in blood, and density peaks between noon and 8PM.
  + Transmitted by *Mansonia* in forest swamps, biting at any time of day. Zoonotic infections are common, and cats, civit cats, dogs, monkeys, slow lorises, and hamsters have been infected in lab.

Hundreds to thousands of mosquito bites likely required before infection, as a bite only transmits a few L3, and a small % ever reach maturity in human. Means visitors are unlikely to ever be infected, only locals.

### Lymphatic filariasis

Affects 120 million people, a third of whom show disfigurement.

Infection and swelling of the lymph system. Appears to take many years to develop, and symptoms include:

* Lymphadenitis, swelling of the lymph nodes.
* Lymphangitis, inflammation of the lymphatic vessels. Abscess and ulcers occur more frequently than with *Wuchereria bancrofti*, and remnants of worms can often be found in the drainage.
* Lymphedema or elephantiasis, is enlargement of the limbs caused by repeated lymphangitis. Unlike *W. bancrofti*, Bm rarely affects the genitals.
* Secondary bacterial infections are common, due to the interference with the immune system.

### Genome

First parasite genome to be sequenced?

TRS strain: 95 Mb, 5 chromosomes, published in (Ghedin *et al.*, 2007).

* N50 93.8 kb, 11,515 proteins found (17.8% of genome); estimated to be ~16,000 proteins in total.
* X/Y sex chromosome.
* 9x coverage, originally 8180 contigs, 15% repeat content, 30.5% GC.

Long-range gene linkage between Bm and *C. elegans* is generally conserved, but local gene order isn’t.

Median of 4 introns per gene (5 in elegans), median exon size is about the same, and median intron length is 3x larger (217 vs 68).

50% of genes have clear orthologs in elegans; 48% average pair-wise identity.

Appear to be missing 9 of 10 enzymes required to synthesize inosine monophosphate from phosphoribosyl pyrophosphate (purine synthesis); the pathway is intact in wBm (Scott and Ghedin, 2009). Similar with heme biosynthesis, and *de novo* riboflavin biosynthesis.

## Onchocerca volvulus

Most pathogenic of the filaria (Cotton *et al.*, 2016).

### Life cycle

### Prevalence and range

Control largely relies on mass drug administration of ivermectin.

* Has successfully eradicated the disease from several countries.
* Can’t be used when *Loa loa* is co-endemic, due to the possibility of irreversible and severe neurological adverse effects, including death.

### Onchocerciasis

Affects 18 million people, of whom 500,000 have visual impairment and 270,000 are blind. Or, 17 million people, 1.2 million with vision impairment or blindness (Hotez *et al.*, 2014).

L1/microfilariae are produced by fertile females within onchocercomata (nodules), which migrate to the skin or other organs. Here they induce inflammatory reactions that are responsible for the pathologies.

### Genome

Very high quality; at chromosome quality (including a sex chromosome) (Cotton *et al.*, 2016).

* Also includes and assembly of its Wolbachia. Version used in the genome paper is a little different from that available on GenBank. Find the modified version at <ftp://ftp.sanger.ac.uk/pub/project/pathogens/Onchocerca/volvulus/>

Also has RNAseq from 8 life stages: nodular microfilariae, skin microfilariae, L2, L3, L3D1, L3D3, adult male, and adult female.

* NCBI BioProject PRJEB2965.

97 Mb nuclear genome, 3 autosomal and 2 sex chromosomes.

* N50 25.5 Mb;
* 94% of the assembled genome is found on 4 huge scaffolds (3 autosomal + X); the Y chromosome is quite small (4.4 Mb; 3.2 Mb of which is shared between X and Y).
* 6 genes missing from CEGMA, 5 of them also missing in filaria; likely only 1 actually missing.
* Moderate synteny to elegans at a very high level; Ov chrm 1 is mostly Ce chrms 1+X, Ov 2 is Ce 3, Ov 3 is Ce 2, and Ov X is Ce 4+5 (with some Ce X).

12,143 protein-coding genes predicted.

* 1,173 (9%) without orthologs.

NOTE: there is a genome assembly for the very closely related *Onchocerca ochengi*, though it’s draft quality (N50 16 kb)

### Metabolic reconstruction

Done in (Cotton *et al.*, 2016).

* 767 reactions (378 distinct enzymes) in Ov, 648 reactions (301 enzymes) in *Loa loa*; share a core set of 648 reactions. 100 of 139 of the additional Ov reactions are provided by Wolbachia.

FBA predicted 71 essential reactions, 112 in *Loa loa* (70 shared).

* Wolbachia appears to provide contributions to fatty acid metabolism, haem synthesis, and nucleotide metabolism. Also, conversion of aspartate to lysine. NAD kinase (EC 2.7.1.23) is predicted as essential to both parasites, but uniquely provided by Wolbachia.

wOv provides an alternate pathway for purine metabolism, it may provide a source of inosine monophosphate. Ov has another enzyme missing in *Loa loa* which could provide a mechanism to selectively target Ov over *Loa loa*; especially testable as inhibitors of purine-nucleoside phosphorylase (EC 2.4.2.1) already exist.

Did not model wOv as a separate compartment with transport reactions between them; would be more accurate, but we may not have the information about the existence of those reactions.

## Wolbachia

First observed by (Mclaren *et al.*, 1975). Present in the vast majority of filarial nematodes. If a population has them, they’re present in every individual. Suggests they are vital.

Found in lateral chords, and located in oocytes, but not male reproductive tract.

Mainly tetracycline and doxycycline, but also rifampicin and azithromycin, show inhibitory effects on filarial development. Drastically reduces *Wolbachia* numbers, removes from some tissues like female reproductive tract.

* No antibiotic effects seen on filarial nematodes lacking endosymbionts, nor from most other antibacterials on Bm.

The endosymbionts are separated into 6 supergroups based on 16S rRNA, the Wolbachia surface protein, and ftsZ phylogenies.

* Four supergroups contain Wolbachia from arthropods, supergroup C has those from Ov and *Dirofilaria immitis*, and supergroup D has those from Bm, W. bancrofti, and Litomosoides sigmodontis.

Closest relatives appear to be the Order Rickettsiales, including *Rickettsia* and others, all parasites of mammals that require arthropod vectors for transmission.

Up to 70% of all insects may harbour Wolbachia.

### Bm species

*Wolbachia pipientis* has a 1 Mb genome (Foster *et al.*, 2005).

Unlike *Rickettsia*, wBm contains genes for synthesis of riboflavin and Flavin adenine dinucleotide. Could be an important source for the nematode, is Bm appears to lack these genes (Foster *et al.*, 2005).

wBm may also provide heme to Bm, as the bacteria has all but 1 gene of the synthetic pathway; that missing reaction may be carried out by an unknown form of the missing gene (protoporphyrinogen oxidase), or may be compensated by another gene as in *E. coli*.

### Ov species

956 kb genome.

Appear to be many cases of HGT to the bacteria, but only 7 larger than 1 kb. Virtually all appear to be fragmented, except for 4 that could be functional. Suggests HGT does not play an important role here.

### Targeting Wolbachia

Has been shown to impede parasite development, fecundity, and survival within humans.

Not currently feasible for mass drug administration.

* 4-6 weeks of daily supervised treatment, and death of adult worms takes 2 years.

## Human filarial infections

Human filarial infections are the leading cause of morbidity in the developing world. This includes lymphatic filariasis (Bm and *W. bancrofti*), river blindness (Ov), and loiasis (*Loa loa*). Currently 150 million people infected, with 1.3 billion at risk in 58 countries.

4 classes of drug available to treat filarial disease: diethylcarbamazine (may interfere with fatty acid metabolism), BZs (microtubules), avermectins (glutamate-gated Cl channels), and oxytetracycines and other antibacterials.

* The anthelmintics target microfilariae or L3s.

Ivermectin is a microfilaricide, and doesn’t affect adult worms; Bm lives for 8 years, Ov for 15.

* Drugs would have to be administered for 30-40 years to eradicate the parasites.

11 countries in Central Africa where Ov is co-endemic with *Loa loa*, targeting the microfilaria can lead to serious adverse events.

* Neurological symptoms or death.

# Metabolism

## Energy production

TCA cycle:

* R00709/R00267: 1 NADH/NADPH. From isocitrate to 2-oxoglutarate.
* R07618: 1 NADH. From 2-oxoglutarate to succinyl-CoA
* R00405: 1ATP. Succinyl-CoA to succinate.
  + R00432 produces GTP instead. Found in elegans, not filaria.
* R02164: 1 ubiquinol (UQH2). From succinate to fumarate.
* R00342: 1 NADH. From malate to oxaloacetate.

Oxidative phosphorylation:

* Complex I oxidizes NADH to NAD, pumps 4 protons across the membrane, and reduces a ubiquinone to ubiquinol in the mitochondrial membrane. The generation of ubiquinol also takes up two protons from inside the mitochrondrion.
* Complex II converts succinate to fumarate and reduces a ubiquinone.
  + *A. suum* possesses a paralog that runs in reverse, allowing anaerobic oxidative phosphorylation with fumarate as the electron acceptor (Shimizu *et al.*, 2012). I believe this works by regenerating ubiquinone, allowing Complex I to pump protons (a reverse of R02164).
* Complex III oxidizes ubiquinol, pumping 2 protons across the membrane, and reduces a cytochrome c.
* Complex IV oxidizes cytochrome c, reduces oxygen to water, and pumps 4 protons.
* Complex V (the ATPase) is driven by the proton gradient and generates ATP from the passage of 4 protons.
* The outer mitochondrial membrane is porous to large and small ions, and freely traversed by uncharged molecules through voltage-dependent anion channels (Kühlbrandt, 2015). So the proton-motive force is generated only across the inner membrane.

## Redox

NADPH and NADH are both important redox cofactors in the cell. NADPH chiefly operates in anabolic reactions, donating its electrons to energy-rich biological molecules. NADH is an intermediate in catabolic reactions such as ATP generation from food.

* The phosphate doesn’t affect the functionality, but allows enzymes to differentiate the molecules, and they are regulated independently.
* The ratio of NAD+ to NADH is kept high, while the ratio of NADP+ to NADPH is kept low. This means there is plenty of NAD+ to act as an oxidizing agent, and plenty of NADPH to act as a reducing agent.

## Common irreversible reactions (product)

In TCA:

* 351 (158), 267 & 709 & 268 (26), 621 (5381)

# Metabolic network reconstruction

Reviewed in (Reed *et al.*, 2006).

In (Gebauer *et al.*, 2016) they generated a metabolic network for *C. elegans* by:

* Used Pathologic (Dale *et al.*, 2010) on the elegans genome, which gave them 218 reactions.
* Then BLAST the elegans genome against metabolic models from yeast and *E. coli*, which increased it to 481.
* They then compared the elegans genome to the human network Recon 2 (Thiele *et al.*, 2013), and accepted reactions only if they were described in WormBase or KEGG; this increased the network to 1914 reactions.

## Flux balance analysis

In general, can have interaction- or constraint-based models; FBA is the latter. It is an approach to analyze a metabolic network, using linear optimization to determine the steady-state reaction flux distribution by maximizing some objective function, commonly ATP production or growth. Described in (Raman and Chandra, 2009).

Widely used approach to study biological networks, calculating the flow of metabolites through an organism’s metabolic network. The steps are outlined below (Orth *et al.*, 2010):

An external file that holds a picture, illustration, etc.
Object name is nihms299330f1.jpg

* S is a m\*n matrix, where m is the number of compounds and n the number of reactions in the network.
* Each row in S describes the productions and uses of a single compound, and each column is one reaction. A -1 indicates that compound is being used, while a 2 would indicate 2 molecules are being produced. The empty cells contain 0.
* The vector v represents the flux through all of the reactions in the network (it has length n).

The FBA model assumes the mass in equals the mass out: . However, any realistic system will have n>m, so there will be no unique solution to the system; any v in the null space of S satisfies it.

* There is information gained in analyzing S, but more detail analyses can be performed if an objective function Z is defined. It could be maximize ATP production, or to maximize the growth rate (the biomass produced). 11 linear and non- objective functions are reviewed in (Schuetz *et al.*, 2007).
* If growth is being modelled, reactions that produce amino acids, nucleotides, lipids, etc are weighted strongly, and the the reaction is scaled so that the flux through it is equal to the exponential growth rate of the organism.
* The vector c is used in this definition, a vector of weights, indicating how much each reaction contributes to the objective function.
* Constraints to the system are incorporated in two ways, as entries in S indicating production and use of metabolites, and as inequalities on the variables during the linear programming step. Reviewed in (Price *et al.*, 2004)
* The linear programming step can be carried out by many different software packages; the COBRA Toolbox is one such, and is a freely available Matlab package.

It has been shown that the network topology is the major factor determining the accuracy of an FBA analysis, while the particular biomass optimization equation plays a lesser role (Puchalka *et al.*, 2008).

Many varied uses of FBA.

* To model aerobic growth of *E. coli*, set oxygen uptake to unrealistically high level, and glucose uptake to something realistic (ex 18.5 mmol gDW-1hr-1). This yielded a growth rate of 1.65 /hr.
* To compare to anaerobic growth, just set oxygen uptake to 0 and rerun. Yields rate of 0.47 /hr.
* Can model the effects of a gene deletion by setting the flux of every reaction associated with that gene to 0. Commonly, a gene is considered essential if the resulting flux is <50% of the wild-type.

Flux variability analysis is a method to min/max every reaction in a network.

* If there are multiple paths for the network to achieve an optimum, this can identify them.
* Identifies reactions with different min and max fluxes under perturbation; analysis of where these are located on the metabolic map will indicate an alternate pathway.

Robustness analysis varies the flux through one reaction, and the impact on the objective function is measured.

* Ex is to set O2 uptake to a realistic number, then vary glucose uptake from 0 through -20 mmol /gDW /hr. Can then plot growth rate (which is just the f value I believe).
* Shows how growth increases rapidly until an inflection point where oxygen uptake becomes the limiting factor; excess glucose is instead shuttled through anaerobic pathways.

Objective functions work well for single cells, but the *a priori* specification introduces bias, which is a problem in multicellular organisms (Basler *et al.*, 2016).

The “shadow price” tells us how the governing constraints on the objective function change, and how the base optimal LP changes (Reznik *et al.*, 2013).

* A negative shadow price (SP) indicates the metabolite has value to the metabolic network. These may be constraining factors for the current model, in that if more was available, more flux would be produced. A reasonable threshold to consider an SP negative might be SP < -0.1.
* A positive SP indicates that the objective function would improve if less of this was made. It may indicate metabolites that are secreted.
* An SP of zero indicates it has no value to the cell. May be secreted?
* There is little to suggest that the correlation between shadow price and the effect on growth is linear.
* The Reznik study found the largest SPs when yeast was under phosphate limitation.
* There are exceptions; mtbs with very negative SPs that exhibit high temporal variation in response to perturbation (growth limiting mtbs should have low variation).
  + One was fructose-1-6-P2. However this resulted in a buildup of PEP in the cell, which allows for fast uptake of glucose when it was again added to the media.
* Mtbs with large negative SPs are usually compounds that the cell needs to control strictly. This is often done by allosteric or post-translational modifications, as they are very fast. An idea for drug target prioritization is to target the mechanisms a cell uses to stabilize some mtb with a highly negative SP.
* So if you’re using a model, and modifying available nutrients, a drop in SP for some metabolite indicates that it has become more important to the cell under those conditions.
* SPs are defined only over a certain range, as dictated by the structure of the feasible solution space. These ranges capture how large a perturbation can e before the genome-scale optimal flux distribution changes sharply (by moving to a new optimum distinct from the old).
* Sometimes a mtb with highly negative SP, if you allow more into the network, will improve the objective fxn. However, other times it does not. This may be due to dependencies; where the objective function really would improve if we would allow more of some other mtb in. The SPs likely wouldn’t be aware of these dependencies, as they are calculated in the dual problem, and the dual only uses the S matrix, the rxn constraints, and the objective function from the primal. The vector of optimal fluxes itself does not appear.

Futile cycles are sets of reactions that result in thermodynamically infeasible fluxes and are a common problem in reconstructions with many transport reactions.

* In (Magnusdottir *et al.*, 2017) they found unfeasibly high export flux of protons out of the cytosol, which resulted in a biologically implausibly high ATP production (average of 933 in the presence and absence of O2).
  + They identified futile cycles by individually constraining each reaction to 0 flux, and computing the flux through the ATP demand reaction.
  + If such a “deleted” reaction lowered the ATP demand flux, it was manually inspected and replaced by an appropriate irreversible reaction. If this change prevented biomass production it was reverted, and they moved on to the next reaction.
  + After curation ATP production was down to 19 and 38 flux in the presence / absence of O2.

## Compartmentalization

A fairly large database was put together to look at the metabolic network in humans: <http://www.metabolicatlas.org/>.

* The Curated Models section contains models for hepatocytes (I believe with compartmentalization), muscle cells, and adipocytes.

The Human Protein Atlas (Uhlen *et al.*, 2010) details the protein expression features of human cells.

* All biological data.
* Could very likely use this to compartmentalize other organisms, at least for common organelles (mitochondria, ER, etc).

The adipocyte model (Mardinoglu *et al.*, 2013) classified enzymes into 8 different compartments: cytosol, nucleus, ER, Golgi, peroxisome, lysosome, mitochondria, and extracellular. It is available from metabolicatlas.org

* Proteins located in the aggresome, centrosome, and cytoskeleton were assigned to the cytoplasm, while those in cell junctions and focal adhesions were assigned to extracellular.

Modeling different tissues as different compartments might require the objective function to be a sum of sub-functions, one per tissue.

* Read (Schuetz *et al.*, 2012); on using multidimensional objective functions.

## Incorporating transcriptomics

GIMME (Becker and Palsson, 2008) does stuff.

TEAM is an extension of GIMME, incorporating temporal expression data.

* Not if also true for GIMME, but here shadow prices have a different interpretation than in standard FBA. Under TEAM, they are defined as the change in the inconcsistency score when the steady state constraint on one metabolite deviates from zero. That means that negative SPs indicate metabolites whose steady-state abundances are decreasing.

## Related methods

Many methods described in (Raman and Chandra, 2009) and (Lewis *et al.*, 2012).

Unbiased methods describe all steady-state flux distributions, including reaction sets that function together without belonging to traditional “pathways”.

* Elementary flux mode analysis (Papin *et al.*, 2004).
  + Elementary modes are a set of vectors derived from the FBA matrix S, where there is a unique set of elementary modes for a given network, each mode has the minimum number of reactions required to exist as a functional unit, and the modes are all routes through a network consistent with property 2.
  + Any possible configuration of flux through a network can be obtained by a linear combination of the elementary modes.
* Extreme pathway analysis (Papin *et al.*, 2004).
  + Share the first 2 properties as elementary modes, along with that the extreme pathways are the systematically independent subset of elementary modes; that is, no extreme pathway can be represented as a nonnegative linear combination of any other extreme pathway. This often means there will be fewer extreme pathways than elementary modes.
  + Any possible configuration of flux through a network can be obtained by a linear combination of the extreme pathways.
* Comparison between EFA and EPA:
  + If the network contains only irreversible reactions, they’re equivalent.
  + A realistic network, like that of a red blood cell, has far more EM (6180) than EP (55).
  + One EP may be used in a huge number of EM (4353). On average, an EM decomposition uses 4 EP.
  + Analysis of a network using EP requires some caution, as it’s possible to cancel reversible fluxes (it would be considered 2 separate reactions in EPA).
* MCMC provide a distribution of feasible fluxes for each reaction.
* These methods detail all possible flux states, but the majority are not used by the cell. Biased methods are good here, as they attempt to only deal with the useful flux states.

MoMA (minimization of metabolic adjustment) is a variant of FBA, and has been shown to better predict lethal gene deletions in *E. coli* compared to FBA (Segrè *et al.*, 2002).

* Relaxes the optimal growth flux for mutants, and instead of finding an optimal growth configuration (maximizing the optimization) it finds the configuration closest to the given “wild-type” configuration.
* The “wild-type” configuration doesn’t have to be from FBA, can be experimentally derived, or anything you like.
* Uses a quadratic optimization function (probably sum of squares; ie distance).

ROOM (regulatory on-off minimization) is another related variant, and performed better in yeast than FBA or MoMA (Shlomi *et al.*, 2005).

* Instead of finding a configuration that is closest to “wild-type” as averaged over the fluxes in the network, it finds the closest configuration as measured by the number of fluxes significantly different.
* Requires a mixed integer linear program solver.

Metabolite essentiality analysis (MEA), which is similar but testing the metabolites directly, was used to good effect to identify a very promising drug compound against a *Vibrio* (Kim *et al.*, 2011).

rFBA (regulatory) is a method of incorporating transcriptional regulation as Boolean operators by Palsson *et al*.

* The metabolism is modeled at steady-state, and the regulatory network is represented by a Boolean function, indicating the expression of each gene.
* Proceeds in iterations; first computes a regulatory net consistent with the FBA steady state (metabolite concentrations may alter expression), then computes a new steady state with FBA.
* In model of *E. coli*, FBA predicted high uptake of both glucose and lactose. rFBA predicted downregulating lactose uptake & processing genes until glucose was gone, followed by their activation, and then the cell using lactose as the carbon source.

SR-FBA is the next step, using a system of linear equations to model the regulatory net instead of Boolean on/off (Shlomi *et al.*, 2007).

* Requires a mixed integer linear program solver.

Energy balance analysis is a method to allow reaction bounds to be dictated by metabolite concentrations (Beard *et al.*, 2002).

* Mixed integer as well as quadratic optimization function.
* Allowed them to explain some previous incorrect essentiality predictions from FBA.

FASIMU is a framework for a few different types of FBA, including a more modern take on EBA (Hoppe *et al.*, 2011).

* Does weighted flux minimization, fitness maximization for partially inhibited enzymes, and concentration-based thermodynamic feasibility constraints.

An approach was made with multi-objective optimization and EBA; designed to work on mammalian liver cells, that change their metabolism based on detected toxins (Nagrath *et al.*, 2007).

Flux coupling finder identifies reactions that are incapable of carrying flux in the current model; likely indicate problems with the network (Burgard *et al.*, 2004). Can also identify hard-coupled reactions; these would be excellent drug targets, as the drug would essentially be hitting 2 targets.

Methods exist to identify alternate optima of a linear programming problem that all produce the same optimal value of the objective function (Lee *et al.*, 2000).

## Less useful methods

OptKnock is a method to identify gene deletions that will help produce a desired metabolite.

* Uses a bi-level optimization, for both biomass and the metabolite of interest.

FBAwMC is a method that incorporates the volume of the cell, and so enzyme availability in a crowded cytoplasm. FBAME does something similar, by restricting flux through specific *E. coli* membrane transporters.

* Worked to good effect in *E. coli*, likely not scalable for me. Also probably requires a lot more biological data.

## Implementation notes

EC predictions in (Cotton *et al.*, 2016) were found by using DETECT v2 (cutoff ILS >=0.9, >=5 positive hits100), BLASTP against SWISSPROT enzymes (cutoff 1E-10), PRIAM (min prob >0.5, profile coverage 70%, check catalytic – TRUE), KAAS, EFICAz, and EC assignments from BRENDA. Reaction assignments to metabolic pathways and pathway hole filling was performed using Pathway Tools V18.

* Priority appears to be BRENDA > DETECT > PRIAM + KAAS agreement.
* Some reactions may be uninformative; if it only has DNA on both sides, remove it. Or just remove metabolites that are unchanged by a reaction.

Model setup

* Directionality of reactions was taken from KEGG. Reversible was indicated by lower bound -1000 & upper bound 1000 mmol (gDW h)-1, while irreversible was indicated by lower/upper of 0/1000.
* Added a non-growth-associated maintenance (NGAM) equation of 5 mmol.
* Glucose uptake set at max of 10 mmol.
* Initially allowed for diffusion of water, O2, CO2, ammonia, diphosphate, phosphate, and ethanol.
* Allowed the transport of all amino acids.
* Model biomass equations taken from previous networks of Toxo and Leishmania.
* Reactions were not compartmentalized. Difficult to assign compartments, and to identify extant transport reactions between them. This assumption is expected to minimize false positives at the expense of increased false negatives (Cotton *et al.*, 2016).

Should have a lower bound on the ATP maintenance reaction; indicates the minimum required to survive. Reflects ATP use of non-growth associated processes like maintenance of electrochemical gradient. For *E. coli* was set to 8.39 mmol gDW-1hr-1. Get this from literature?

If a reaction can be carried out by two different enzymes, the GPR (gene-protein-reaction) will read pqr-2 or pqr-3. So a double-knockout would have to be performed to see the effect of losing the reaction. If multiple enzymes are required for a reaction, the GPR reads pqr-4 and pqr-5. In this case, knocking out either will result in loss of the reaction. Modeling it with FBA involves setting the flux upper and lower boundaries to 0.

Gene knockout studies are useful, and it’s easy to test all possible pairs of knockouts as well, to look for synthetic lethal pairs. No reason to think triples wouldn’t also be doable. Useful?

* Growth ratios are measured by dividing the yield with a knockout by the yield of the unrestrained model.
* A reaction is said to be essential if the growth ratio is less than 0.1.

Can also determine which genes are essential for each biomass precursor. Might be very useful to understand how to grow organism in the lab. Must add a demand reaction to each biomass precursor, then perform gene deletion study for these reactions.

Can identify which genes impact the flexibility of the network the most. Normally will be several (or a few) reactions with variable flux that still lead to the optimal (or very near) objective function. These represent alternate paths through the network that are (nearly) equivalent. To do this, knockout one gene, and then perform FVA on the non-essential genes. This is done when engineering strains to produce products, and they look for reactions with increased flux (through their desired product).

## Resources

Pathway databases:

* KEGG, BioCyc ([www.biocyc.org](http://www.biocyc.org)), Biomodels ([www.biomodels.net](http://www.biomodels.net)), SABIO-RK (analysis; <http://sabio.villa-bosch.de>).

Genome and enzyme information:

* BRENDA ([www.brenda-enzymes.info](http://www.brenda-enzymes.info)), PEDANT (<http://pedant.gsf.de>), Reactome (humans; [www.reactome.org](http://www.reactome.org))

Software analysis:

* COBRA, MetaFluxNet, CellNetAnalyzer, SNA: Stoichiometric network analysis, Yana (also visualization), PathwayAnalyser (FBA, MoMA, deletions), Systems Biology Research Tool (Wagner lab).

FBA/MoMA solvers:

* LINDO, CPLEX, GLPK, OOQP.

# References

Aboobaker,A.A. and Blaxter,M.L. (2003) Use of RNA interference to investigate gene function in the human filarial nematode parasite Brugia malayi. *Mol. Biochem. Parasitol.*, **129**, 41–51.

Basler,G. *et al.* (2016) Control of fluxes in metabolic networks. *Genome Res.*, **26**, 956–968.

Beard,D.A. *et al.* (2002) Energy Balance for Analysis of Complex Metabolic Networks. *Biophys. J.*, **83**, 79–86.

Becker,S.A. and Palsson,B.O. (2008) Context specific metabolic networks are consistent with experiments. *PLoS Comput.*, **4**, e1000082.

Burgard,A.P. *et al.* (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.*, **14**, 301–312.

Cotton,J.A. *et al.* (2016) The genome of Onchocerca volvulus, agent of river blindness. *Nat. Microbiol.*, **2**, 16216.

Dale,J.M. *et al.* (2010) Machine learning methods for metabolic pathway prediction. *BMC Bioinformatics*, **11**, 15.

Foster,J. *et al.* (2005) The Wolbachia genome of Brugia malayi: Endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.*, **3**, 0599–0614.

Gebauer,J. *et al.* (2016) A Genome-Scale Database and Reconstruction of Caenorhabditis elegans Metabolism. *Cell Syst.*, **2**, 312–322.

Ghedin,E. *et al.* (2007) Draft genome of the filarial nematode parasite Brugia malayi. *Science*, **317**, 1756–60.

Hoppe,A. *et al.* (2011) FASIMU: flexible software for flux-balance computation series in large metabolic networks. *BMC Bioinformatics*, **12**, 28.

Hotez,P.J. *et al.* (2014) The Global Burden of Disease Study 2010: Interpretation and Implications for the Neglected Tropical Diseases. *PLoS Negl. Trop. Dis.*, **8**, e2865.

Kim,H.U. *et al.* (2011) Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery. *Mol. Syst. Biol.*, **7**, 460.

Kühlbrandt,W. (2015) Structure and function of mitochondrial membrane protein complexes. *BMC Biol.*, **13**.

Lee,S. *et al.* (2000) Recursive milp model for finding all the alternate optima in lp models for metabolic networks. In, *Computers and Chemical Engineering*., pp. 711–716.

Lewis,N.E. *et al.* (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat. Rev. Microbiol.*, **10**, 291–305.

Magnusdottir,S. *et al.* (2017) Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat. Biotechnol.*, **35**, 81–89.

Mardinoglu,A. *et al.* (2013) Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol. Syst. Biol.*, **9**, 649.

Mclaren,D.J. *et al.* (1975) Micro-organisms in filarial larvae (Nematoda). *Trans. R. Soc. Trop. Med. Hyg.*, **69**, 509–514.

Nagrath,D. *et al.* (2007) Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. *Ann Biomed Eng*, **35**, 863–885.

Orth,J.D. *et al.* (2010) What is Flux Balance Analysis ? *Nat. Biotechnol.*, **28**, 245–248.

Papin,J.A. *et al.* (2004) Comparison of network based pathway analysis methods. *Trends Biotechnol.*, **22**, 400–405.

Price,N.D. *et al.* (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.*, **2**, 886–897.

Puchalka,J. *et al.* (2008) Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput Biol*, **4**, e1000210.

Raman,K. and Chandra,N. (2009) Flux balance analysis of biological systems: Applications and challenges. *Brief. Bioinform.*, **10**, 435–449.

Reed,J.L. *et al.* (2006) Towards multidimensional genome annotation. *Nat. Rev. Genet.*, **7**, 130–141.

Reznik,E. *et al.* (2013) Flux Imbalance Analysis and the Sensitivity of Cellular Growth to Changes in Metabolite Pools. *PLoS Comput. Biol.*, **9**.

Schuetz,R. *et al.* (2012) Multidimensional optimality of microbial metabolism. *Science (80-. ).*, **336**, 601–4.

Schuetz,R. *et al.* (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Mol. Syst. Biol.*, **3**, 119.

Scott,A.L. and Ghedin,E. (2009) The genome of Brugia malayi - All worms are not created equal. *Parasitol. Int.*, **58**, 6–11.

Segrè,D. *et al.* (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 15112–15117.

Shimizu,H. *et al.* (2012) Crystal structure of mitochondrial quinol-fumarate reductase from the parasitic nematode Ascaris suum. *J. Biochem.*, **151**, 589–592.

Shlomi,T. *et al.* (2007) A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol. Syst. Biol.*, **3**, 101.

Shlomi,T. *et al.* (2005) Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 7695–7700.

Thiele,I. *et al.* (2013) A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.*, **31**, 419–25.

Uhlen,M. *et al.* (2010) Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.*, **28**, 1248–1250.