THE UNIVERSITY *of* EDINBURGH

The physics of biological evolution

Steven James Court

Doctor of Philosophy
The University of Edinburgh
2014

Lay summary

Many factors influence biological evolution, including mutation, gene transfer, habitat structure and the complicated ecological interactions that exist between species and the environment. In this thesis we use physics approaches to investigate some of these factors.

In the first part, we look at evolution on the level of metabolism; the set of chemical reactions powering life. Despite the diversity of life, much of the metabolism of these different organisms is the same – certain chemical pathways are utilized by all forms of life. This important observation can have two explanations. Firstly, it is possible that after billions of years of evolution, biology has discovered the most optimal way to perform certain functions. The alternative explanation is that these shared features appeared early in the history of life and are now used by all organisms by virtue of this fact alone; they are a frozen-in accident. This latter scenario means that, in principle, "better" options could exist. We investigate this question by choosing a highly conserved chemical pathway and writing a computer program to explore all the possible alternatives. We find that many alternatives can be generated by our program but that, in the present study, the real pathway appears to be optimal under realistic cellular conditions. Our approach also allows novel chemical pathways to be discovered, pathways which may already be operating in biology but are yet to be characterized experimentally. Our approach can thus be used to study the evolution of metabolism, as well as offering direction for future metabolic experiments.

In the second part of this thesis we introduce a novel statistical physics model which captures the dynamics of an infectious process in a population. Crucially, this model incorporates both horizontal transmission (infection from another individual) and vertical transmission (from parent to offspring). This is important because evolution strongly depends on both of these modes of transmission; in the microbial world many organisms can freely exchange genetic material with other members of the community. This is believed to have been extremely important for the earliest organisms, but is also currently of huge concern since it facilitates the rapid spread of antibiotic resistance among bacteria.

Abstract

Part I: A remarkable feature of life on Earth is that despite the apparent observed diversity, the underlying chemistry that powers it is highly conserved. From the level of the nucleobases, through the amino acids and proteins they encode, to the metabolic pathways of chemical reactions catalyzed by these proteins, biology often utilizes identical solutions in vastly disparate organisms. This universality is intriguing as it raises the question of whether these recurring features exist because they represent some truly optimal solution to a given problem in biology, or whether they simply exist by chance, having arisen very early in life's history. In this project we consider the universality of metabolism – the set of chemical reactions providing the energy and building blocks for cells to grow and divide. We develop an algorithm to construct the complete network of all possible biochemically feasible compounds and reactions, including many that could have been utilized by life but never were. Using this network we investigate the most highly conserved piece of metabolism in all of biology, the trunk pathway of glycolysis. We design a method which allows a comparison between the large number of alternatives to this pathway and which takes into account both thermodynamic and biophysical constraints, finding evidence that the existing version of this pathway produces optimal metabolic fluxes under physiologically relevant intracellular conditions. We then extend our method to include an evolutionary simulation so as to more fully explore the biochemical space.

Part II: Studies of population dynamics have a long history and have been used to understand the properties of complex networks of ecological interactions, extinction events, biological diversity and the transmission of infectious disease. One aspect of these models that is known to be of great importance, but one which nonetheless is often neglected, is spatial structure. Various classes of models have been proposed with each allowing different insights into the role space plays. Here we use a lattice-based approach. Motivated by gene transfer and parasite dynamics, we extend the well-studied contact process of statistical physics to include multiple levels. Doing so generates a simple model which captures in a general way the most important features of such biological systems: spatial structure and the inclusion of both vertical as well as horizontal transmission. We show that spatial structure can produce a qualitatively new effect: a coupling between the dynamics of the infection and of the underlying host population, even when the infection does not affect the fitness of the host. Extending the model to an arbitrary number of levels, we find a transition between regimes where both a finite and infinite number of parasite levels are sustainable, and conjecture that this transition is related to the roughening transition of related surface growth models.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Parts of this work have been published in [1, 2].

(*Steven James Court, 2014*)

Acknowledgements

Undertaking this PhD has been an important personal experience for me, and one which would not have been possible without the support and guidance that I received.

I first want to thank my supervisor, Rosalind Allen, whose technical and moral support was often needed. The enthusiasm you show for your work, and your ability to jump into such a wide range of scientific research has been inspiring. Your patience, understanding and constant encouragement has been truly appreciated.

I also wish to thank both of my second supervisors, Richard Blythe and Bartek Waclaw, who provided essential input to the work presented in this dissertation, were always welcoming and available for helpful discussions and who have also provided inspiration through their dedication to their work.

Friends are vitally important. To my office mates in 1511 – Guilia, Matt, Alys and Fred – thanks for the company throughout these years, as well as to all my fellow PhD students who I've come to know. Thanks to my flatmates – Sam C, Sam B, Chris, Ryan and Øyvind – who kept it fun and put up with the times of stress. Hat tip to the "XM-ers" and GAHA for the love.

Thanks to all my family; to my Mum and Dad for the absolute confidence and pride that only parents can have.

I greatfully acknowledge The Carnegie Trust, whose Carnegie-Caledonian PhD Scholarship provided the financial support that allowed this thesis to be completed.

Finally I thank a number of people who will never know how much of a role they played; the chance encounters at various summer schools and conferences. It is surprising how much of an impact that a single lecture or conversation with a passionate professor or PhD student can have, and how long-lasting this can be.

Thank you all.

Contents

List of Figures

List of Tables

Chapter 1

Overview

1.1 Many factors influence biological evolution

The process of evolution has resulted in the variety of life observed today. Evolution arises from just three simple principles: replication, variation (mutation), and competition (differences in fitness). From the *last universal ancestor*, existing around 3.8 billion years ago, life has been subject to random mutations under the continuous pressure of natural selection, with the organisms best adapted to a given environment being most likely to survive and reproduce. Many factors are important in understanding this evolution, from the fundamental molecular processes underpinning reproduction, mutation and gene aquisition, to the dynamics arising from the complex ecological interactions both between species and with their physical environment. Hence the study of evolutionary biology necessarily encompasses ideas from a range of disciplines.

In fact, it is becoming increasingly difficult to distinguish between traditionally distinct scientific disciplines. A large amount of research now comes from groups composed of scientists from a variety of backgrounds; those who may still identify themselves as mathematicians, physicists or computer scientists working alongside those with a formal biological training. The fields of biophysics, systems biology and bioinformatics are in rapid expansion. The result has been a sharing of knowledge and ideas, and the increased use of modelling, analytical and experimental tools originally developed in the physical sciences. From the mapping of our DNA through the human genome project, the application of statistical physics methods in evolutionary genetics and the studying of emergent phenomena in complex systems at the molecular and ecological

levels, such collaborations have led to great leaps in the understanding of our origins.

1.2 Outline of thesis

This thesis is split into two parts in which I will consider two different aspects of evolution. The first part is concerned with evolution at the most fundamental level, investigating the evolutionary principles governing metabolism – the set of biochemical reactions that powers every living cell. Many of the metabolic pathways essential in energy production and biosynthesis appear in a vast number of totally disparate organisms. This raises fundamental evolutionary questions. Do these highly conserved pathways exist because they represent the only solution to life, or do they represent one possibility from many? If one can envisage alternative biochemical pathways, is the observed universality the result of a long evolutionary optimization, or the result of a chance event in life's history?

In this thesis we attempt to address such questions by considering perhaps the most highly conserved metabolic pathway – the trunk pathway of lower glycolysis. By generating a network of all relevant biochemical compounds and reactions, including many that are not known to exist in nature, we enumerate all potential alternatives to this pathway and find that, in principle, a great number of alternatives exist. We then introduce a method for comparing these pathways based on the metabolic flux that they can sustain and show that, given realistic thermodynamic and biophysical constraints, the real pathways perform best under typical physiological conditions.

In the second part we present a new statistical physics model for the dynamics of a parasitic infection within a host population. Parasitic infections are ubiquitous, and play crucial roles in the dynamics and evolution of ecosystems. Such infections can be transmitted either horizontally (from other infected individuals) or vertically (passing from parent to offspring). These two processes can have very different consequences, but surprisingly few models incorporate both. Those which do, neglect spatial structure, which is known to be a crucial factor in population dynamics models. Our statistical physics model is the first to incorporate both horizontal and vertical transmission in a spatially structured population in a way that is generic enough to reveal important new physical phenomena.

The general interplay between these two modes of transmission is also of critical importance in evolutionary theory. Although evolution is often thought of as descent with modification (i.e. vertical transfer of genes), in recent decades the extent of

horizontal gene transfer (HGT) has been realised. HGT allows genes to be transferred rapidly, with genes being acquired either through direct transmission from other organisms, or being salvaged from the environment. This process is believed to have played a significant role in evolution during the early stages of life, and is still of huge consequence today as it enables the rapid spread of traits, such as antibiotic resistance, throughout microbial communities. The outline of this thesis is as follows.

Part I. Chapter 2 provides an introduction to metabolism, covering the basic biological and mathematical theory, and gives an overview of the remarkable universality and possible optimality of metabolism. Chapter 3 reviews the necessary biochemical and thermodynamic knowledge that will enable us to generate the complete network of biochemical compounds and reactions, with Chapter 4 detailing the generation of this network. In Chapter 5 we use this network to enumerate all feasible biochemical alternatives to the trunk pathway, and introduce the methodology for comparing their metabolic fluxes, showing that the real pathway outperforms the alternatives. Chapter 6 discusses alternative flux calculations and introduces a program that we are writing to perform evolutionary simulations using our network.

Part II. Chapter 7 reviews the importance of spatial models in biology and in particular, lattice-based models. The utility of such lattice models in physics is discussed, and the canonical model of an infectious process, the contact process (CP), is introduced. The CP's role in studies of non-equilibrium systems, phase transitions and disordered systems is also reviewed. In Chapter 8 we introduce our stacked contact processes extension to the original CP and analyze the steady-state behaviour of the two-level system before generalizing to multi-level systems. Chapter 9 concludes the thesis.

Part I

Chapter 2

Introduction to metabolism

2.1 Metabolism and life

2.1.1 What is metabolism?

Metabolism is the set of chemical reactions that drives life, producing energy from food and using chemical building blocks to synthesize new biological material. These reactions form biochemical pathways that can be broadly divided into two categories: catabolism and anabolism. Catabolic processes take in large molecules and break them down into smaller ones, biochemically capturing the energy released in the process. Anabolic pathways on the other hand consume this energy and use it to build new cellular material out of smaller "precursor" moleclues. Figure 2.1 gives a schematic view of intermediary metabolism and illustrates the modularity of the metabolic network: each colour corresponds to a set of reactions involved in a specific aspect of metabolism. For example the blue region in the centre corresponds to carbohydrate metabolism, while the reactions in the pink region in the upper right corner correspond to cofactor and vitamin metabolism.

2.1.2 ATP: the cell's energy currency

The way in which a cell stores the energy derived from catabolism is through the formation of ATP (adenosine triphosphate), the cell's energy currency [3]. ATP can be viewed as the fully charged state of a biological battery, and contains a chain of three linked phosphate groups (Figure 2.2). The *phosphoanhydride* bonds between

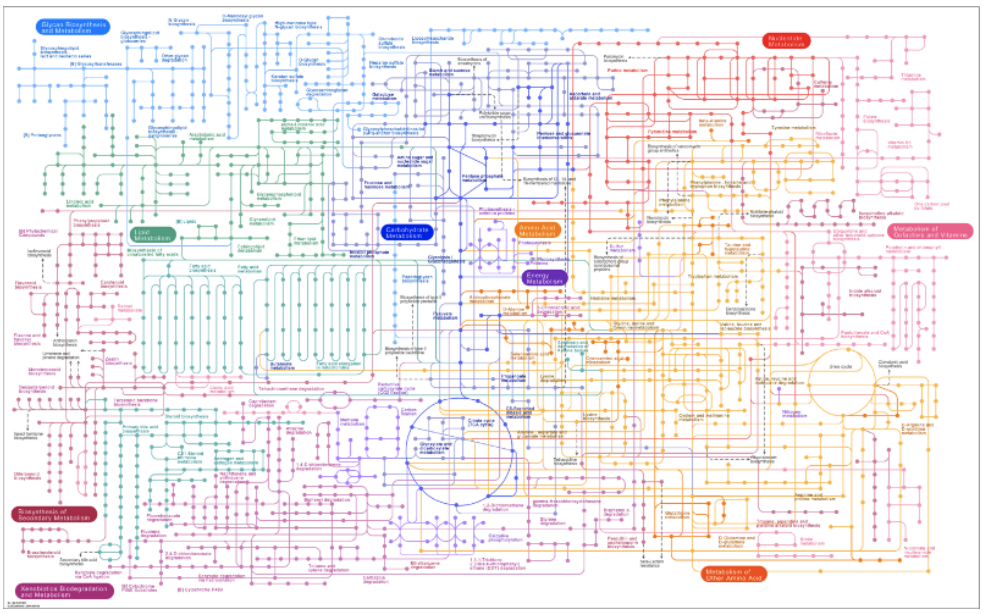


Figure 2.1 Schematic chart of intermediary metabolism, taken from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [4]. The complicated network of interactions has a modular structure and can be viewd as the sum of simpler modules, each performing a distinct function. Different colours represent different functional modules.

the phosphate groups are often referred to as high-energy bonds; upon hydrolysis, the cleavage of a chemical bond through the addition of water, the terminal phosphate group is lost to form ADP (adenosine diphosphate) and the energy released in this process can be enzymatically coupled to drive thermodynamically unfavourable reactions. Likewise, energetically favourable reactions can be coupled to the generation of ATP from ADP, allowing the cell to harvest and store biochemical energy. Although other nucleoside triphosphates such as GTP, CTP and UTP are utilized for a small number of metabolic tasks, the cycling of ATP-ADP is by far the most important mode of energy exchange in biology.

2.1.3 Redox reactions

Redox reactions form a crucial class of reactions in metabolism. A redox reaction is a coupled oxidation and reduction, involving a change in the oxidation state of a pair of compounds. This change occurs through the transfer of electrons, with oxidation reactions corresponding to a loss of electrons and reduction reactions corresponding to a gain in electrons. In biochemistry, nicotinamide adenine dinucleotide (NAD^+), shown in Figure 2.3 is an important molecule involved in the transfer of electrons. NAD^+ is an electron acceptor and is thus an oxidizing agent: as it gains electrons it is reduced to NADH, and the molecule from which the electrons have been donated is simultaneously

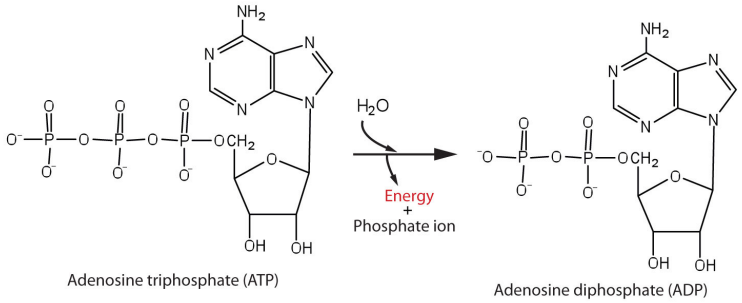


Figure 2.2 ATP hydrolysis can be used to drive thermodynamically unfavourable reactions. The structure of an ATP molecule consists of the nucleobase adenine, connected to the sugar ribose, with three phosphate groups attached. Upon hydrolysis, the terminal phosphate group is lost to form ADP and a phosphate ion, with a large amount of energy being released in the process.

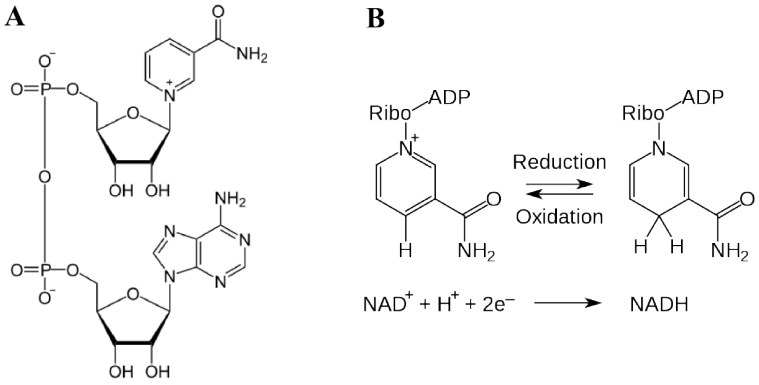


Figure 2.3 NAD^+ is an important electron carrier in metabolism. (A) The chemical structure of NAD^+ in its oxidized state. (B) The reduction of NAD^+ to NADH. A general reaction coupled to this might be $RH_2 + NAD^+ \rightarrow NADH + H^+ + R$, where R is the reactant, which is oxidized through the removal of two hydrogen atoms.

oxidized.

Reductants, such as NADH, are produced during the oxidation of reduced compounds including sugars and fatty acids, and the energy stored in them can be coupled to energy production in aerobic organisms; NADH plays an important role in cellular respiration, in a process known as oxidative phosphorylation [3]. In this process, the energy released from the flow of electrons from NADH to some terminal electron acceptor, such as oxygen, can be used to drive the production of ATP. A number of electron carriers are utilized in biology, with $NADP^+$ and FAD being other important examples. These carriers differ in the roles that they play in metabolism, as well as in the number of electrons that they may transfer [3].

2.1.4 Enzymes

In the cell, each reaction is carried out by an enzyme – a protein molecule which acts as a biological catalyst, being recovered unaltered after the reaction. By lowering the activation energy, an enzyme can speed up the rate of the reaction by a factor of between 10^6 and 10^{18} [5]. The cell's use of enzymes to catalyze its reactions not only facilitates faster kinetics. It also enables thermodynamically unfavourable reactions to be driven by chemically coupling them to energy-supplying reactions, such as the consumption of ATP already discussed. Many enzymes in biology have evolved to be highly specific, catalyzing only one particular reaction. The evolution of this specialization allowed for increases in the reaction rates, as well as for a strong degree of control over the functioning of metabolism.

2.1.5 Metabolic control

Since many cells utilize multiple food sources and exist in a constantly varying environment, it is important to be able to control which biochemical reactions are operating at any one time. Enzymes are also hugely important in this process; cells mainly control their metabolism by either controlling the amount of a given enzyme in the cell, or by controlling its activity, as depicted in Figure 2.4. The most obvious method of control is arguably that of altering the amount of the enzymes themselves, Figure 2.4A. Through transcriptional regulation, the expression of genes encoding the enzymes of a certain pathway can be up or down regulated, or perhaps be switched off entirely. These genetic circuits offer a range of possible dynamics to the cell, but do not offer a rapid response to sudden changes – it takes time for enzyme synthesis and degradation to have an effect on the enzyme levels.

The most rapid and flexible form of enzymatic regulation is allosteric regulation, Figure 2.4B, resulting in a change in the activity of the enzyme. Upon binding to an enzyme, certain effector molecules can induce a conformational change in its structure. This structural change can increase or decrease the binding strength of the enzyme to its substrate, leading to either allosteric activation or inhibition. Inhibition can also occur if a molecule binds directly to the active sites, acting as a physical barrier to the substrate of the reaction. These effectors can be metabolites from different modules of metabolism, or can be metabolites produced in the same pathway. For instance, if the end product of a metabolic pathway could allosterically inhibit the initial step, this would provide an effective way to slow down the process when product levels built up

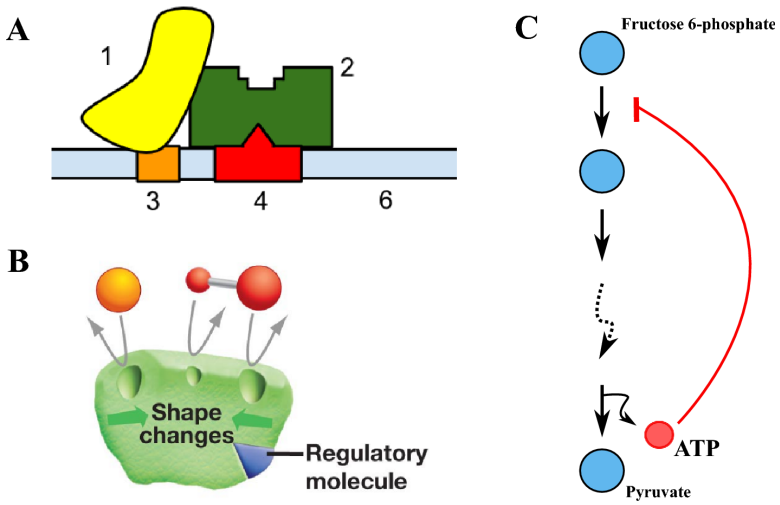


Figure 2.4 Forms of metabolic control. (A) Transcriptional regulation controls the amount of enzyme present. Here, some repressor molecule (2), prevents the RNA polymerase (1) from binding to the DNA, preventing the enzyme encoded by the downstream genes (6) from being produced. (B) Schematic of allosteric regulation (inhibition). Some regulatory molecule binds to the enzyme, inducing a conformational change that inhibits the binding of the enzyme's primary substrates. The regulatory molecule can also serve to increase this binding (allosteric activation). (C) A simplified view of glycolysis. Glycolysis produces ATP, which allosterically inhibits an earlier reaction and hence reduces the rate of glycolysis when ATP is abundant.

past a certain level. This is actually the case in the metabolic pathway of glycolysis, where one of the products, ATP, allosterically inhibits an earlier reaction, slowing down the rate of the pathway when ATP is abundant. This is shown schematically in Figure 2.4C. These processes have evolved to form a coherent and astonishingly complex set of regulatory mechanisms, controlling the cell's biochemical reactions in an ever-changing environment.

2.1.6 Why study metabolism?

Metabolism represents the driving force behind life and, as we will see in Section 2.3, a significant number of metabolic pathways exists conserved in all organisms. Studying extant metabolism might thus help us to better understand the evolution and origin of life itself. One scenario for the origin of life is in fact a "metabolism first" one, in which life is proposed to have originated from self-replicating metabolic reactions [6]. If this is the case, studying the most highly conserved components of metabolism could provide information on the conditions under which it arose.

A solid understanding of metabolism is not just of fundamental scientific interest, but leads to a vast range of applications in industry and medicine. Many human diseases

are essentially metabolic in nature; genetic mutations can lead to an altered behaviour of the proteins in our cells, disrupting or altering enzyme function. The various forms of cancer are themselves the result of genetic mutations, and a large body of research has been aimed at understanding the altered metabolisms that arise in these cells [7]. The ultimate outcome of such work being to take advantage of this knowledge in developing treatments that specifically target these modified cancer cells [8]. Even in the case of bacterial or fungal infections, it is ultimately an understanding of their specific metabolisms that is the key to developing effective treatments.

The benefit of being able to control and modify metabolism has led to a host of industrial applications. With advances in both theoretical and experimental techniques, it has become possible to genetically engineer organisms for the optimization of desired processes, and to perform entirely new functions by incorporating enzymes from a range of different organisms. Examples include the production of medically important chemicals, such as a genetically modified yeast that produces a compound used in the treatment of malaria [9]; or the production of biofuels from modified photosynthetic organisms such as algae [10]. This rapidly evolving science is not just restricted to microorganisms. Plants are being engineered that produce herbicides and pesticides, that are more resistant to drought and frost, and that have enhanced nutritional benefits, such as the now famous Golden Rice which was developed to combat vitamin A deficiency in certain parts of the world [11].

The introduction of novel genes into an organism, one that has been shaped by evolution over millions of years, can have unpredictable and dramatic consequences. If the full potential of these technologies are to be realised, and for this to be done safely, a fundamental grasp of the biochemical, thermodynamic and physical processes that underpin metabolism must be obtained. “Good engineers must know the principles that make their machines tick” [12].

2.1.7 What physicists have to offer

The boundaries between scientific disciplines are becoming increasingly blurred. With such a huge amount of biological data being produced, efficient techniques are required to collect these and incorporate them into the constantly evolving models, and much of this work has come from those who might still identify themselves as physicists, mathematicians or computer scientists. Many of the modelling tools, analytical and computational, developed in the study of idealized physical systems have found a role in the vastly more complicated systems of biology. Modelling a network of metabolic

reactions for instance can results in a complex set of nonlinear interactions and feedback loops, and can require sophisticated optimization and parameter estimation techniques. Of particular interest for us, tools from the field of network theory have provided an understanding of the structure of metabolism and helped elucidate potential design principles behind it.

2.2 Graphical representation of metabolism

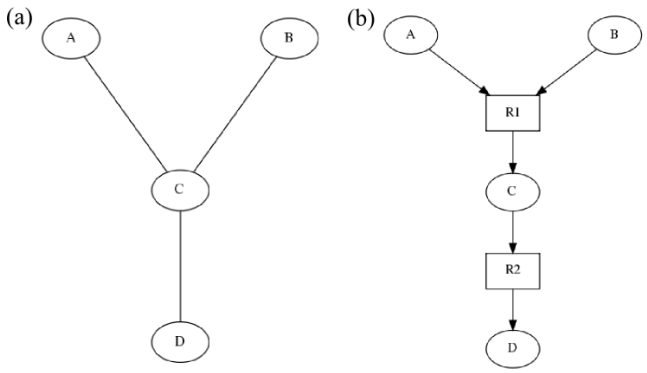


Figure 2.5 Two possible graphical representations of the 2-reaction scheme $A + B \rightarrow C$ and $C \rightarrow D$. An undirected unipartite network is used in (a) connecting all substrates of a reaction to all products, while (b) shows a directed bipartite network where metabolite and reaction nodes are denoted by ovals and boxes respectively.

I will be using the term metabolic *network* throughout this thesis, so it is worth taking a moment to clarify what is meant by this. Mathematically, a network or graph $G(\{E\}, \{V\})$ is an object formed from a set of vertices (or nodes) $\{V\}$ connected by a set of edges $\{E\}$. Since metabolism is a set of biochemical reactions, each of which can be viewed as connecting its chemical input molecules (reactants) to its output molecules (products), a graphical representation is an obvious choice. Various authors have used different types of graphs for representing metabolism, depending on the properties being investigated. Broadly these can be divided into directed or undirected, unipartite or bipartite networks, depending on whether the edges have an associated directionality and the number of node types present in the network. Figure 2.5 illustrates this for a network consisting of just two reactions, showing an undirected unipartite representation alongside a directed bipartite representation. In the unipartite network there is only one type of node, corresponding to metabolites say, and a graphical representation connects all metabolites which participate in the same reaction, shown in Figure 2.5(a). The bipartite representation, Figure 2.5(b), contains two different node types, corresponding to reactions and metabolites. This allows a greater level of detail to be captured, since in the unipartite projection it is never clear

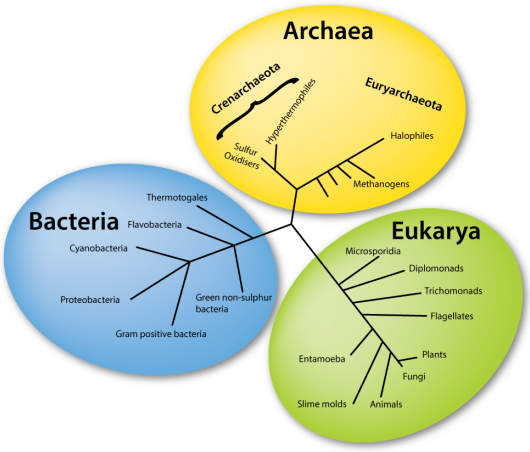


Figure 2.6 The tree of life is split into three domains. Bacteria and archaea are prokaryotes. Prokaryotes are predominantly unicellular organisms and are characterized by their lack of a membrane-bound nucleus. Eukaryotes hold their DNA in such a nucleus and exist in both uni- and multicellular forms. All plants, animals and fungus belong to the class of eukaryotes.

which metabolites take part in any given reaction.

Since a key property of a graph is that it consists of edges connecting two nodes, any multi-species chemical reactions, of the form $A + B \rightarrow C + D$ for example, technically involves multiple nodes and cannot then be fully captured in the classical graphical representation. For instance, analyzing the structure of the bipartite representation in Figure 2.5(b) may obscure the fact that compound D cannot be reached from A alone, but requires the simultaneous presence of compound B. To fully account for such features requires the use of hypergraphs [13], in which an "edge" is generalized so as to include any number of vertices. By using these graphical network representations to describe metabolism, it has become possible to employ the many tools developed in graph theory in analyzing the structural design and functionality of metabolism. We will encounter several such examples later in this thesis.

2.3 The universal nature of metabolism

Life on Earth is characterized by both its diversity and universality. The tree of life is divided into the three domains shown in Figure 2.6; eukaryotes, bacteria and archaea. Both within and among these domains there is a wide range of body shape and biological function, yet despite this huge diversity of life, many key aspects of the underlying chemistry exist conserved across all organisms. An obvious example is the DNA-based machinery for storing genetic information. DNA (deoxyribonucleic acid), is built from

the same four nucleobases in all organisms; guanine, adenine, thymine, and cytosine (G, A, T and C). Triplets of these nucelobases, *codons*, encode the specific amino acids required for life. Despite there being more than 500 naturally occurring amino acids [14], only 22 of these are used in protein building, with only 20 of these being encoded by any genome. These amino acids are then linked into chains to form proteins, which go on to play a structural role in the cell or to function as enzymes. These enzymes produce another layer of universality, with a number of highly conserved enzymes existing at the core of metabolism [15]. In the context of metabolism, such observations of universality raise a number of fundamental questions. For instance, could it be the case that the enzyme mechanisms used by life are the only solution to a basic biochemical problem? – or do other potential mechanisms exist? If so, is the universality we see today the optimal result from billions of years of natural selection, or has historical contingency played a role, whereby some early chance event has "frozen in" a potentially sub-optimal solution as extra layers of complexity have been built on top of it?

The universality of life is particularly striking at the level of the metabolic chart, Figure 2.1, with large portions of this chart appearing in totally disparate oganisms; from the bacteria found at the bottom of the ocean to the red blood cells of the largest land mammals. In Section 2.5 we will discuss some of these highly conserved biochemical pathways.

2.3.1 Biology utilizes a small subset of biochemistry

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database [4] is an important online resource that contains information on the metabolic reactions of nearly 3000 organisms. In a recent study Barve and Wagner [16] collected all well-defined compounds and reactions from the KEGG database and were left with a "universal metabolism", from which the metabolic network of all 3000 organisms can be constructed, that included 5030 metabolites and 5906 reactions. Bearing in mind that the metabolic network of the bacterium *E. coli* contains on the order of 1000 metabolites and 2000 reactions [17], these numbers suggest that a large amount of the biochemistry is being shared by these different organisms. These numbers appear even more modest when one considers the vast number of biochemical possibilities that nature could have used: in a study aimed at generating a small-molecule chemical universe of compounds of potential interest in drug development, Fink et al. [18] generated a database of nearly 14 million compounds from first principles. In Chapter 4 we will perform a similar process and find that tens of thousands of biochemically feasible compounds

are produced even when restricting to the simplest of chemical structures, supporting the notion that biology only utilizes a small fraction of the available biochemistry.

2.3.2 Metabolic networks share similar topology

It is not only at the level of the individual reactions and pathways that one can observe the universality of metabolism. This is also apparent in the actual structure of metabolic networks, where it has been shown that the metabolic networks of organisms from all three domains of life share similar topology. In 2000, Jeong et al. [19] presented the first systematic study of this. By representing the metabolism of 43 organisms as unipartite graphs (in which the nodes represented metabolites connected by edges if they appeared in the same reaction) they studied such network properties as the degree distribution, mean path length and network diameter. The purpose of this study was to look for design principles in these networks by comparing their structure to that predicted by the theory of random networks developed by Erdös-Réyni [20]. One feature of random networks is that their degree distribution is described by a Poisson distribution. The degree of a node is simply the number of edges that connect to it, while the degree distribution $P(k)$ of the network gives the probability that a randomly chosen node will have degree k. Interestingly, Jeong et al. found that all 43 metabolic networks exhibited a power-law degree distribution, $P(k) \sim k^{-\gamma}$, with an exponent of about $\gamma = 2.2$ in every organism. This so-called "scale-free" topology implies the existence of a small number of highly connected hub nodes which participate in a large number of reactions, leading to a small average path length and "small world" structure [21]. It has been proposed that such a structure is biologically beneficial as it allows the rapid propagation of signals through the network and thus fast response times. Scale-free networks are also known to have the property of robustness with respect to random deletions of nodes or edges [22].

This scale-free structure arises naturally from a simple model of network growth known as preferential attachment [23], where new nodes are connected preferentially to more highly connected nodes in the existing network. In the context of metabolic networks, this at least seems biologically reasonable. New metabolic reactions are likely to appear by duplication of existing enzymes, implying that these reactions will tend to involve already highly-connected metabolites. The study of Jeong et al. sparked a flurry of research, with subsequent work also finding commonalities in the hierarchical organization and modularity of the metabolic networks [24]. It is important to note that the work of Jeong et al. led to some debate about whether the nature of the mapping

between metabolic network and graphical representation might alter the statistics of the analysis, or even their biological interpretation [13, 25–27]. Nevertheless, it is striking that these similarities exist over such a wide range of organisms.

2.4 Thermodynamics of metabolism

A major constraint on the evolution of metabolism comes from the laws of thermodynamics. These laws dictate whether a given reaction will occur spontaneously in nature, based on its inherent free energy change and on the concentrations of its reactants and products. The change in the Gibbs free energy associated with a reaction determines the net direction of its biochemical flux. For a system at constant temperature and pressure, equilibrium is characterized by a minimization of the Gibbs free energy. The second law of thermodynamics thus requires that all biochemical reactions flow in the direction of decreasing Gibbs free energy.

The Gibbs free energy change for the reaction $\mathrm{S}_1 + \mathrm{S}_2 + ... \rightarrow \mathrm{P}_1 + \mathrm{P}_2 + ...$, converting substrates $\{S_i\}$ into products $\{P_j\}$, is written

$$\Delta_r G = \Delta_r G^0 + RT \ln \Gamma, \tag{2.1}$$

where $\Gamma = \prod_j [P_j] / \prod_i [S_i]$ is the reaction quotient – the ratio of product to substrate concentrations – and $\Delta_r G^0$ indicates the free energy change at standard conditions, where all metabolite concentrations are 1 M. Note that in biochemistry it is usual to use this unit of "molar concentration", where 1 molar (1 M) is equivalent to 1 mole per litre (1 mol/L). The value $R = 8.314$ J mol^{-1} K^{-1} is the gas constant and T is the temperature in Kelvin. In this thesis I will talk of biological *standard conditions*, which is just a convention that defines an arbitrary set of aqueous conditions close to those found in many living cells: $T = 298$ K, pH = 7.0 and ionic strength $I = 0.2$ M.

For a reaction to proceed spontaneously, the second law of thermodynamics states that its $\Delta_r G$ must be negative. If $\Delta_r G$ is positive, then it is the reverse reaction that is thermodynamically favoured. For a reaction with a positive $\Delta_r G^0$, a negative $\Delta_r G$ might be attainable by increasing the ratio of substrates to products, i.e. decreasing Γ. Shifts in the metabolite concentrations can alter the free energy by a significant amount: if the concentration of one of the products in a reaction is increased by a factor of 10, then the free energy change will be shifted by $RT \ln(10) = 5.71$ kJ/mol at $T = 298$ K. Altering the concentrations of multiple compounds can potentially shift $\Delta_r G$ by tens of kJ/mol. Note an important outcome of this: the reversibility of a

reaction also somewhat depends on the reaction stoichiometry [28]. Assuming that all compounds have some physiological upper and lower limit on their concentrations, the reaction quotient Γ can attain a much larger range of values when there are a greater number of metabolites involved in the reaction. This means that the standard free energy of a reaction $\Delta_r G^0$ is only of limited use in determining the directionality of a reaction under a particular set of physiological conditions.

For a reaction which is at thermodynamic equilibrium the net flux is zero, corresponding to $\Delta_r G = 0$. Substituting this into Eq. (2.1) we see that the reaction quotient, which at equilibrium we denote as the *equilibrium constant* q, becomes

$$q = \Gamma_{eq} = e^{-\Delta_r G^0 / RT}. \tag{2.2}$$

This is an important quantity that will feature in many of the calculations in this thesis.

2.5 Central carbon metabolism

The majority of a cell is composed of molecules made of carbon atoms, including carbohydrates, lipids, amino acids and nucleic acids. Central carbon metabolism is a term used to describe the set of metabolic reactions that are vital in both energy production and the generation of precursor molecules that the cell can use to build more complex cellular components. Central carbon metabolism (Figure 2.7) consists of the most highly conserved pathways in metabolism, including glycolysis, the tricarboxylic acid cycle, and the pentose phosphate pathway. As such, it is believed to be the component of metabolism most ancient in origin; it is expected that much of central carbon metabolism was shared by even the earliest forms of life [29–31].

Because the pathways contained in Figure 2.7 are central to all forms of life, and will be discussed again in this thesis, I will briefly introduce them here. Glycolysis (discussed in detail in Section 2.6) is vital in producing ATP under both aerobic and anaerobic conditions. Glycolysis feeds into the tricarboxylic acid (TCA) cycle, which is used to produce energy in all aerobic organisms [3]. The TCA cycle produces NADH (indicated as electrons in the figure) which are used to generate ATP through oxidative phosphorylation. Under anaerobic conditions this respiration cannot occur, and instead the product of glycolysis, pyruvate, is further metabolized in a process known as fermentation (Section 2.6.4). In addition to these roles, each of these pathways produces precursor metabolites that feed into multiple biosynthetic pathways. These are shown in the red boxes in Figure 2.7. The pentose phosphate pathway can run in

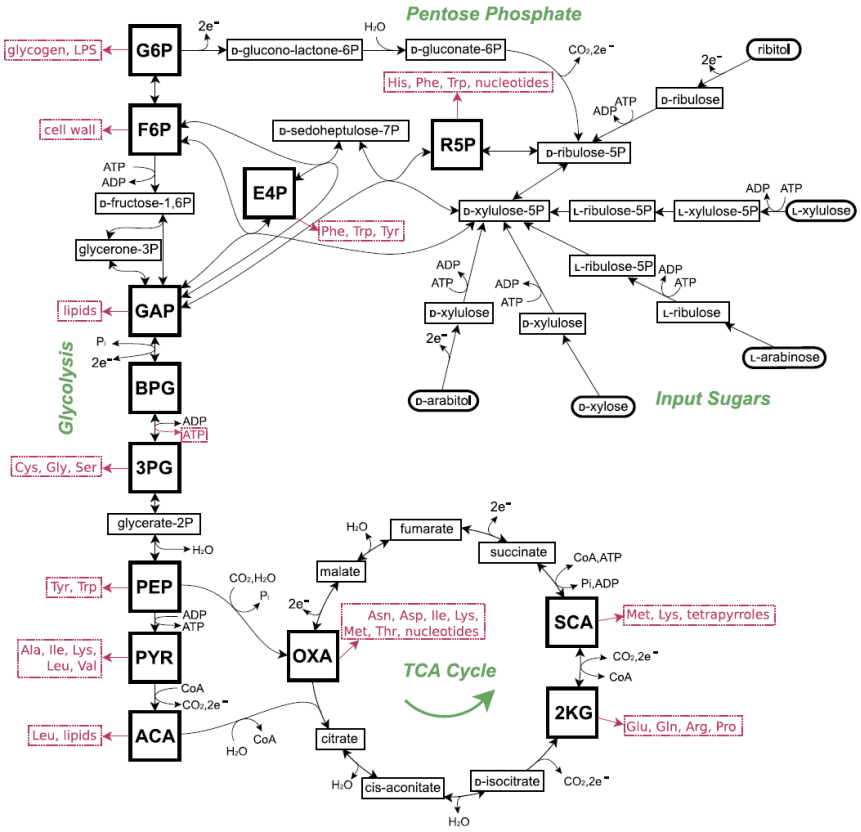


Figure 2.7 Central carbon metabolism, essential for energy production and biosynthesis, consists of the most highly conserved metabolic pathways. Shown is glycolysis, the tricarboxylic acid cycle and the pentose phosphate pathway. Figure taken from [29].

parallel to glycolysis, and has primarily an anabolic rather than a catabolic role. It is used to produce NADPH (the phosphorylated version of NADH) important in fatty acid synthesis, as well as precursors for the synthesis of nucleic and amino acids [3].

Central carbon metabolism forms the core of Morowitz's shell-structure view of metabolism [32]. The substrates of this metabolic core are comprised solely of carbon, hydrogen and oxygen with phosphorous entering only in the way of phosphate groups. Higher level shells see the incorporation of nitrogen and sulfur, followed by the production of the cofactor molecules (ATP, NAD^+, etc.). The hypothesis put forward by Morowitz is that metabolism recapitulates biogenesis [33]; that is the order of biogenesis follows this shell structure. Studying the core then, that may be traced back to the last universal common ancestor, provides a chance to illuminate any processes that might have directed its emergence and subsequent evolution.

2.6 Glycolysis and the trunk pathway

The main focus of the next chapters will be specifically on glycolysis (and the closely related gluconeogenesis) and so here I will discuss this pathway in some detail. Glycolysis is present in some form in every living organism, and serves a dual function in both providing cellular energy in the form of ATP and in producing important precursor metabolites. It exists predominantly as either the Embden-Meyerhof-Parnas (EMP) pathway found in eukaryotes and many bacteria, or the Entner-Doudoroff (ED) pathway which is found in some bacteria and archaea [34]. Both forms are shown in Figure 2.8, and some rarer variations are discussed in detail in Section 5.10.

2.6.1 Embden-Meyerhof-Parnas pathway

The canonical EMP glycolytic pathway (Figure 2.8) transforms one glucose molecule into two pyruvate molecules, producing two ATPs and one NADH in the process. This pathway is almost linear and can be divided into two parts: an "upper" chain of reactions involving 6-carbon molecules, which connects glucose to glyceraldehyde-3-phosphate (G3P), and a lower chain of reactions involving 3-carbon molecules, which connects G3P to pyruvate. This lower section of glycolysis is often called the trunk pathway, and represents one of the most highly conserved sections of metabolism. It is shared by a number of pathways, including the EMP, ED and pentose phosphate pathway and is central to the metabolism of all organisms [30, 31, 35, 36].

Upper-half of EMP

Step (1) of the EMP pathway involves the phoshphorylation of glucose to glucose-6-phosphate through a coupling with ATP consumption. This is followed in step (2) by an isomerization reaction to fructose-6-phosphate and then in step (3) by a second phosphorylation via ATP consumption, forming fructose 1,6-bisphosphate. The upper half of the EMP pathway has thus consumed 2 ATP molecules. Step (4) involves an aldol cleavage reaction that splits fructose 1,6-bisphosphate into the two 3-carbon molecules dihydroxyacetone-phosphate (DHAP) and glyceraldehyde 3-phosphate. Note that both products of this cleavage reaction are phosphorylated; this is going to enable the production of ATP in the lower half of the pathway. In step (5) DHAP is isomerized to G3P, so that both compounds follow the same lower trunk pathway.

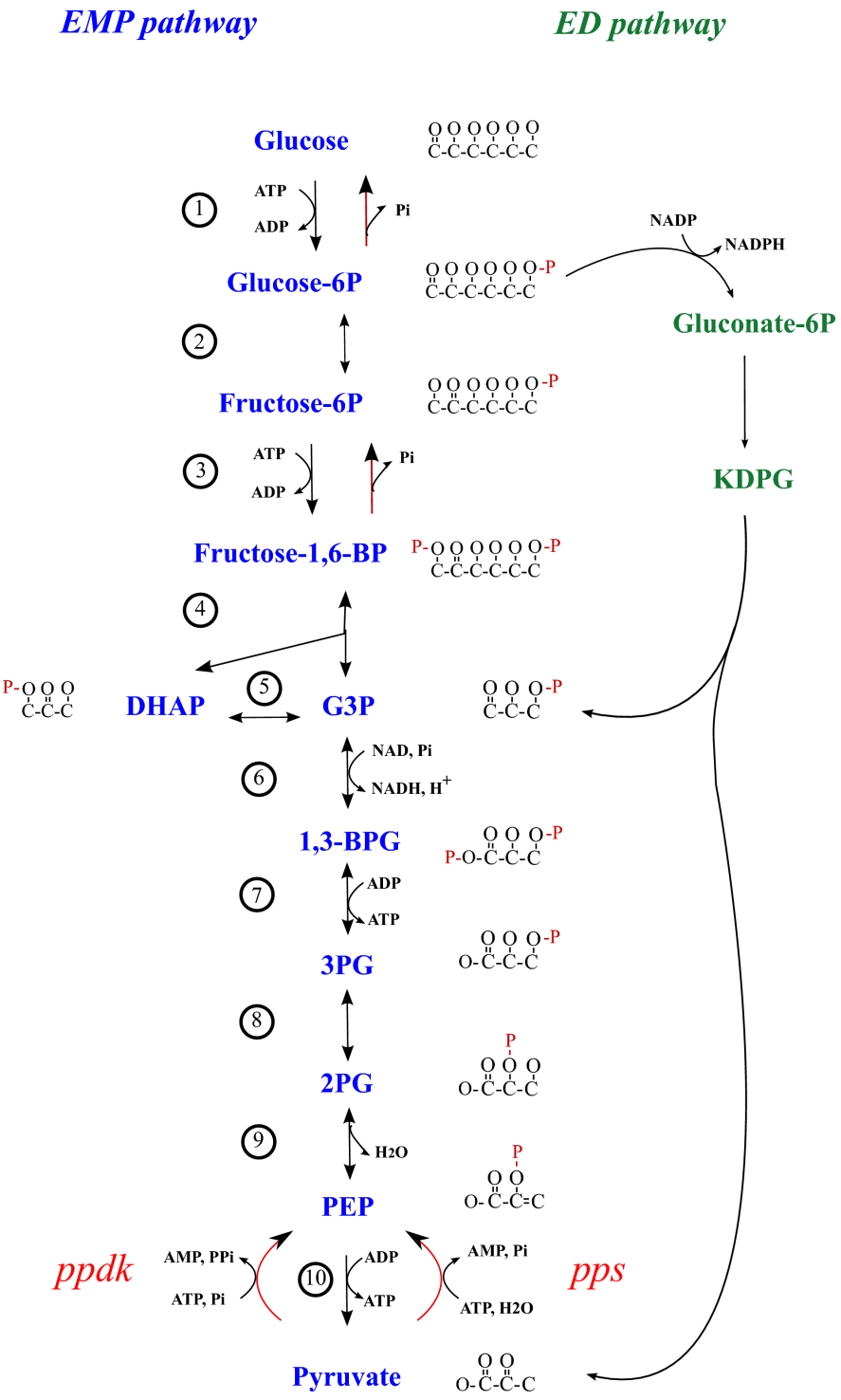


Figure 2.8 The EMP and ED glycolytic pathways. Red arrows show gluconeogenic reactions. Chemical structures excluding hydrogen are shown for the EMP pathway, where the red P represents the phosphate group PO_3^{2-} attached to a metabolite. Abbreviations: DHAP, dihydroxyacetone phosphate; G3P, glyceraldehyde-3-phosphate; 1,3-BPG, 1,3-bisphosphoglycerate; 3PG, 3-phospho-glycerate; 2PG, 2-phospho-glycerate; PEP, phosphoenolpyruvate. Pi and PPi are orthophosphate and pyrophosphate, respectively.

Trunk pathway

The first reaction in the trunk pathway, step (6), sees G3P being simultaneously oxidized (coupling the reduction of NAD^+ to NADH) and phosphorylated through the incorporation of an orthophosphate, yielding 1,3-bisphosphoglycerate (1,3-BPG). In step (7) this acquired phosphate group is then immediately used to generate an ATP molecule, and the resulting 3-phosphoglycerate (3PG) then undergoes an isomerization reaction to 2-phosphoglycerate (2PG) in step (8). The penultimate step (9) of glycolysis is the dehydration reaction of 2PG to form phosphoenolpyruvate (PEP), which then uses its high phosphoryl transfer potential to generate another ATP molecule as it is converted to pyruvate in the final step (10). Since there are two G3P molecules passing through this lower reaction sequence and each of them produces two ATP molecules, the initial investment of two ATPs in the upper half is recovered and the EMP pathway results in the net production of two ATP molecules.

2.6.2 Entner-Doudoroff path

The ED pathway shares the same set of reactions in the lower section, differing only in its upper half (Figure 2.8). We have seen that in the EMP pathway glucose is phosphorylated twice, allowing both G3P and DHAP to produce ATP through substrate-level phosphorylation in the lower reactions. In the ED pathway, glucose is only phosphorylated once before being oxidized to 2-keto-3-deoxy-6-phosphogluconate (KDPG). Since this compound only contains one phosphoryl group, only one of its products upon cleavage will too. KDPG is split into pyruvate and G3P, with G3P following the shared trunk pathway. The effect of this is that the ED pathway produces a net yield of just 1 ATP compared to the 2 produced in the EMP path.

2.6.3 Phosphorylation of intermediates

The phosphorylation of the intermediates does not only facilitate the production of ATP but also alters the membrane permeability of a molecule. The hydrophobic lipid membrane of a cell is largely impermeable to charged molecules, which is important since by having to actively transport ions through the membrane (using proteins called ion pumps) the cell can control the salt concentrations and cellular pH. Uncharged molecules can however passively diffuse through this membrane at a much higher rate. In fact, the small uncharged molecules glyceraldehyde and dihydroxyacetone (the unphosphorylated version of G3P and DHAP), have such a high membrane permeability

that a significant portion of these metabolites can leak out of a cell within seconds [12]. The phosphate moiety can prevent this; the electrostatic charge of the phosphate group serves to retain these molecules within the cell. For example, glucose is prevented from diffusing through the membrane as a result of the initial phosphorylation reaction in step (1) of Figure 2.8. Notice that all of the compounds in Figure 2.8 contain a phosphate group, with the exception of pyruvate. Here, the charge induced by the dissociation of the carboxyl group ($COOH \rightarrow COO^-$) is enough to prevent the loss of this compound. In fact, all core metabolites contain one or both of these groups [37].

2.6.4 The fate of pyruvate

Depending on the environment, the pyruvate produced in glycolysis can follow two courses. An organism can produce energy from sugar in two ways: fermentation or respiration. If terminal electron acceptors are present, an organism can pass pyruvate into the tricarboxylic acid (TCA) cycle for respiration, which generates NADH molecules that can be used in turn to generate a net production of $\sim$30 ATPs through the process of oxidative phosphorylation [3].

In an environment in which no terminal electron acceptors are present, the organism cannot utilize any form of respiration and glycolysis is the sole provider of the cell's ATP needs, producing just 2 ATP molecules from glucose in a process known as fermentation. In fermentation, pyruvate is further metabolized to a range of compounds (dependent on the organism) including lactate, acetate or ethanol. Whatever the end product of fermentation, the redox state of the cell must be maintained. An excess of reducing agents (NADH) is prevented in all scenarios by recovering the NAD^+ molecule consumed in the first reaction of the trunk pathway (e.g. by coupling NADH oxidization to the reduction of pyruvate to lactate). We will discuss fermentation further in Section 5.12.

2.6.5 Gluconeogenesis

As well as breaking down complex molecules to make energy, cells also need to be able to build complex molecules from smaller parts. Gluconeogenesis is essentially the reverse process of glycolysis. It consumes ATP to form glucose, which in some organisms can be stored for times of starvation, and produces the same precursor molecules as in glycolysis. From Figure 2.8 we can see that glycolysis and gluconeogenesis share many of the same reactions, but running in reverse, with the exception of three reactions

(reactions 1, 3, and 10). These reactions are all highly exergonic ($\Delta_r G \ll 0$) in the glycolytic direction, and are thus thermodynamically unfavourable in the reverse direction. The reactions shown in red in Figure 2.8 demonstrate how these reactions are bypassed in gluconeogenesis. The two reactions (1 and 3) in the upper half of glycolysis both involve the consumption of an ATP molecule and both have relatively large $\Delta_r G$ values of about -20 kJ mol^{-1}. To overcome this barrier by altering the concentrations of the metabolites alone would require a change in their ratio by a factor of close to 3000. Instead, for gluconeogenesis to be thermodynamically feasible without hugely altering the state of the cell, both reactions are replaced by a hydrolysis reaction, where the phosphate group is lost through the addition of water and the costly ATP molecule is not produced.

Focusing specifically on the trunk pathway, just one reaction differs – the exergonic conversion of PEP to pyruvate coupled to the phosphorylation of ADP to ATP which has a standard free energy change of $\Delta_r G^0 \sim -29$ kJ mol^{-1}. In gluconeogenesis, prokaryotic cells overcome this with the hydrolysis of ATP to AMP, either with the release of an orthophosphate (Pi) in the "phosphoenolpyruvate synthase" (pps) route, or via the consumption of an orthophosphate and release of pyrophosphate (PPi) in the "pyruvate, phosphate dikinase" (ppdk) route. Both of these are shown in Figure 2.8 and will be referred to throughout this thesis as the "pps" and "ppdk" pathways.

2.6.6 Evolutionary origins

Although the glycolytic-gluconeogenic pathway is central to the metabolism of all organisms, our understanding of its origins remains speculative. It is known that while the upper part of this pathway exists in several distinct variants, the lower trunk pathway is ubiquitous and contains enzymes which are distributed across all three domains of life [38]. Sequence analysis of these enzymes indicates that they were all present within a common ancestor prior to any significant phylogenetic divergence [30].

Hyperthermophiles are organisms that have optimal growth temperatures above 80^{o}C and are found in hot springs and deep-sea hydrothermal vents. Hyperthermophilic archaea and bacteria are believed to be among the most ancient and slowly evolving organisms on the planet [38] and so studying the core pathways of these organisms might shed light on the earliest forms of life. It has been shown that although a small number of glycolytic trunk pathway variants exist (discussed in Section 5.10), all organism, including the hyperthermophilic archaea, carry out gluconeogenesis via the reversed EMP pathway [31]. This has led to the hypothesis that the original function of

this pathway was an anabolic one, possibly hinting at an anabolic origin of life. In this thesis we will explore this hypothesis by investigating all biochemical alternatives to this trunk pathway, and show that the routes available for gluconeogenesis are necessarily more constrained by the biochemistry.

2.7 Modelling metabolism

Over the decades a huge effort has been made to construct mathematical models of metabolism. One motivation behind these efforts is simply to understand the fundamental principles that power life. These models can also guide experimentation, suggesting ways in which to improve the efficiency of industrial processes such as fermentation or fuel production, and offer a way to explore the often counterintuitive outcomes that result from modifiying a system as highly complex as a metabolic network. The complex structure and nonlinear dynamics of metabolic networks can prevent even the simplest systems from being understood intuitively and so a variety of theoretical methods have been studied. Because these models will form a key part of my work, I here describe them in some detail.

The modelling of metabolism can broadly be divided into two categories: kinetic and structural (flux balance) models. Traditionally, the analysis of metabolic networks involved the study of kinetic models to understand the behaviour of a pathway. Kinetic models describe the dynamical changes in concentrations of the metabolites in the pathway, and represent the network of reactions as a system of time-dependent ordinary differential equations (ODEs). Since the number of parameters in any single enzyme mechanism can be large, the experimental effort to determine their values is also enormous and as will be discussed, the process is fraught with difficulties. As a result, this approach was limited to single pathways or a small number of interacting paths. Nevertheless, these pioneering studies enabled the fundamental properties underpinning metabolism to be uncovered, shedding light on numerous dynamical phenomena including oscillations, allosteric feedback and flux control which arise in metabolic pathways [39].

In recent years, rapid DNA sequencing has made it possible to catalogue the metabolic enzymes of many different organisms. This sparked interest in looking at metabolic networks on the scale of the whole organism, genome-scale models, which can contain on the order of hundreds, or thousands of reactions. It is difficult, and perhaps currently infeasible, to construct kinetic models for such large scale networks, so researchers have

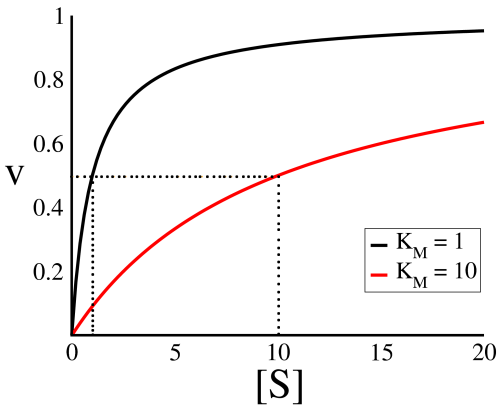


Figure 2.9 Michaelis-Menten kinetics of Eq. (2.8) for two different values of K_M. In both cases $V_m = 1$ and the dotted lines show $[S] = K_M$ at which $v = \frac{1}{2}V_m$.

resorted to structural models, which use an approach known as flux balance analysis. Flux balance analysis depends predominantly on the network topology and leaves out enzyme kinetics altogether. Despite the obvious lack of biochemical realism, these models have proven extremely useful in predicting the distribution of fluxes in different environments and the effect of gene knockouts. Since I will not make use of such models in this thesis, I will not discuss them here. A brief outline can be found in Appendix A.1.4 and many good reviews exist [40, 41].

Both kinetic and structural models provide vital contributions to the field. Although it might seem that genome-scale networks would not be amenable to a kinetic study, there has been an increased focus towards this goal [42–44]. Through the use of generic enzyme rate equations in place of the more detailed mechanistic ones, and along with sophisticated techniques for parameter estimation, large-scale dynamic models have been constructed with some success in recreating experimentally observed features of metabolism.

2.7.1 Kinetic modelling

Here, I discuss the basic forms of the rate equations most frequently used in kinetic models and that form the basis of the research described in this thesis.

Michaelis-Menten kinetics

Consider perhaps the simplest representation of a metabolic reaction: the irreversible transformation of a substrate S into product P, catalyzed by an enzyme E:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_{\text{cat}}} E + P \, . \tag{2.3}$$

The values k_i correspond to the rate constants of the individual steps, with k_{cat} being the enzyme turnover number; the maximum number of substrate molecules an enzyme can convert per unit time. Assuming mass-action kinetics, the system of equations governing the reaction is:

$$\frac{d[S]}{dt} = -k_1[E][S] + k_{-1}[ES] \, , \tag{2.4}$$

$$\frac{d[P]}{dt} = k_{\text{cat}}[ES] \, , \tag{2.5}$$

$$\frac{d[ES]}{dt} = k_1[E][S] - k_{-1}[ES] - k_{\text{cat}}[ES] \, , \tag{2.6}$$

$$\frac{d[E]}{dt} = -k_1[E][S] + k_{-1}[ES] + k_{\text{cat}}[ES] \, . \tag{2.7}$$

To date, no analytical solution exists to the above set of equations. Nonetheless it is possible to approximate these to obtain a useful form describing an irreversible reaction. In 1913 Michaelis and Menten [45] showed that if we assume that the substrate and substrate-enzyme complex are in rapid equilibrium, meaning that $k_1, k_{-1} \gg k_{\text{cat}}$, then the canonical rate equation for such an enzyme catalyzed reaction takes the form:

$$v = \frac{d[P]}{dt} = \frac{V_m[S]}{K_M + [S]}, \tag{2.8}$$

where $V_m = k_{\text{cat}}[E]_t$ is the maximal rate of the reaction, with $[E]_t = [E] + [ES]$ being the total enzyme concentration (free plus substrate-bound forms). $[S]$ is the substrate concentration and K_M is the Michaelis-Menten constant or dissociation constant of substrate S, corresponding to the concentration of substrate at which the reaction rate is half its maximal value (Figure 2.9). An alternative derivation by Briggs and Haldane [46] invoked the assumption that the concentration of the enzyme-substrate complex $[ES]$ is constant in time, but note that both approaches ultimately result in the same form of Eq. (2.8). For this reason the above rate is often used without acknowledging the underlying assumptions of the model. Both derivations are given in Appendix A.1.

The Michaelis-Menten rate, Eq. (2.8), is widely used in metabolic modelling because it provides a simple form that captures some essential features of biological reactions; an

initially linear increase in rate at low substrate concentrations, approaching a limiting maximal rate upon substrate-saturation of the enzyme (Figure 2.9). The Michaelis-Menten rate is governed by just two parameters, V_m and K_M, and it is relatively straightforward for these to be measured in *in vitro* biochemical experiments.

Despite the simplicity of the Michaelis-Menten form, it is only applicable to irreversible reactions. Many reactions in biology are required to be able to proceed in both directions and indeed all enzyme catalyzed reactions are technically reversible. It is thus useful to extend the simple irreversible model to the case of reversible reactions. From the mechanism

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k^-_{\text{cat}}}{\overset{k^+_{\text{cat}}}{\rightleftharpoons}} EP \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P, \tag{2.9}$$

the corresponding reaction rate

$$v = \frac{d[P]}{dt} = \frac{V_f}{K_S} \frac{1}{1 + \frac{[S]}{K_S} + \frac{[P]}{K_P}} \left([S] - \frac{[P]}{q} \right), \tag{2.10}$$

can be derived, where $q = [P]_{eq}/[S]_{eq} = e^{\frac{-\Delta G^0}{RT}}$ is the equilibrium constant of the reaction and $V_f = k^+_{\text{cat}}[E]_t$ is the maximal rate in the forward direction. K_S and K_P are the dissociation constants of the substrate and product respectively. A full derivation of Eq. (2.10) can also be found in Appendix A.1. Notice that with the product concentration now appearing in the denominator, the reaction rate decreases at high product concentrations also. This is easily seen by rearranging Eq. (2.10) following [47]:

$$v = k^+_{\text{cat}}[E]_t \times \left(\frac{[S]/K_S}{1 + [S]/K_S + [P]/K_P} \right) \times (1 - e^{\Delta G/RT}) \equiv V_f \times \kappa \times \gamma \,, \tag{2.11}$$

where $\kappa = ([S]/K_S)/(1 + [S]/K_S + [P]/K_P)$ relates to the level of substrate saturation and $\gamma = (1 - e^{\Delta G/RT})$ is a measure of the thermodynamic driving force[1]. This form clearly shows that the inhibition of the reaction at high product concentrations arises from both the enzyme saturation (the amount of free enzyme decreases as a function of P) as well as the reaction thermodynamics: increasing the concentration of the product will simultaneously lower the thermodynamic favourability of the reaction by increasing ΔG, and decrease the rate through the saturation term. This form allows us to distinguish the importance of the three contributions V_f, κ and γ, denoted the capacity, fractional saturation and thermodynamic terms respectively. Depending on

[1]Note that the free energy appearing in this expression is no longer the standard free energy. Using $\Delta G = \Delta G^0 + RT \ln \frac{P}{S}$, allows $\frac{P/S}{q} = \frac{P/S}{e^{-\Delta G^0/RT}}$ to be replaced by $e^{\Delta G/RT}$.

the parameters chosen, the rate of the reaction may be dominated by thermodynamics, saturation, or both [47].

More complicated enzyme mechanics

In many cases, even more complex kinetic equations are needed to accurately describe metabolic reactions.

Multiple substrates and products For example, most biological reactions actually involve multiple substrate and product metabolites, such as the simple reaction $A+X \rightleftharpoons B+Y$ that arises any time a unimolecular reaction is coupled to the conversion of a cofactor molecule, such as ATP. This leads to far more possibilities for the rate law, depending on how the substrates and products bind and leave the enzyme. To illustrate this, the most common mechanisms that occur for the above two-substrate reaction are *compulsory-order*, in which a certain substrate must bind to the enzyme before the other, *random-order* in which metabolites bind and unbind in any order, and *double-displacement* reactions where one substrate binds and modifies the enzyme before unbinding and allowing the second substrate to bind. The different mechanisms lead to slightly different rate laws. To first approximation, each of these rates is a special case of a general rate equation [39] but, depending on the parameters chosen, they can lead to notably different reaction rates [42].

Allosteric regulation and cooperativity Many metabolic reactions are affected not just by the substrate and product concentrations but also by direct regulation of the enzyme activity via the binding of other molecules called reaction modifiers. These modifiers can be activators or inhibitors, and provide an important way for the cell to control metabolism. There are a number of proposed mechanisms through which this regulation can occur, leading to a number of models for enzyme inhibition and allosteric regulation. Since this thesis will not be concerned with metabolic regulation, we will not discuss them here, but the interested reader can refer to [5, 39]. Briefly though, the various modifications can affect the form of the reaction rate, through modifications to V_f or K_M that depend on the concentrations of the reaction modifiers, and can lead to a deviation from the hyperbolic rate equation of the simple Michaelis-Menten dynamics (Figure 2.9).

With most reactions involving multiple substrates, and many being regulated by multiple metabolites, the exact form for the rate of a specific enzyme is an experimentally

challenging and time consuming task. As a result, generic rate equations have been proposed.

Generic rate equations

To accurately model the full dynamics of a metabolic system, the precise mechanism and parameters for every reaction would need to be known as input, along with the parameters associated with every regulatory molecule. Measuring such values is a huge experimental undertaking. Typically, the purified enzyme has to be first extracted from the organism before a multitude of experiments are conducted in the presence of various effector metabolites, at a range of concentrations, so as to determine all the necessary kinetic parameters. Some reactions involve multiple regulatory interactions: in one model of glycolysis in yeast [48], the phosphofructokinase enzyme catalyzing the transformation of fructose-6-phosphate to fructose-1,6-bisphosphate involves a total of 15 parameters. In addition to the sheer number of parameters, there is also the problem that the *in vitro* conditions under which their values are being determined do not match the natural *in vivo* cellular environment, leading to discrepancies between experiment and model [48, 49]. For example, the temperature, pH and salt concentrations under which a reaction proceeds all affect its rate. Inside the natural cellular environment there is also the potential for an enormous number of enzyme-ligand and enzyme-enzyme interactions, however an unbiased screen for all effectors is unfeasible and so only those which are a priori suspected to affect enzyme activity will be studied. As a result, for the majority of metabolic pathways the enzyme characterization is incomplete. Motivated by this issue, there has been an effort to develop generalized rate laws which can substitute for the detailed mechanistic ones. These should capture the full range of possible dynamics while making simplifying assumptions about the actual reaction mechanisms, which are usually unknown, in order to minimize the number of kinetic parameters that are required. The use of such generalized rate equations has become more common in recent years and are essential if one is to systematically turn metabolic networks into kinetic models [44].

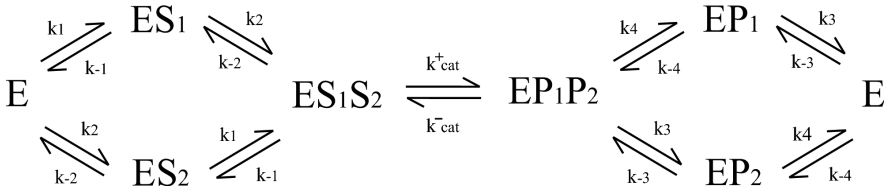


Figure 2.10 A bi-bi molecular reaction with random order binding mechanism.

Here I briefly review one form of generalized rate equation, which will be used later in this thesis. The basis for this approach is the *common modular* rate law, a special case of the family of modular rate laws proposed in [43]. The common modular rate is derived from a simple random-order enzyme mechanism that extends the reversible Michaelis-Menten equation to any number of substrates and products. Figure 2.10 shows the case of a reaction with two substrates and two products, a bi-bi reaction. For a reaction converting i substrate metabolites into j product metabolites, the common modular rate law it is expressed as

$$v(\mathbf{S},\mathbf{P}) = [E]_t \frac{k^+_{\text{cat}} \prod_i \widetilde{S}_i - k^-_{\text{cat}} \prod_j \widetilde{P_j}}{\prod_i (1+\widetilde{S}_i) + \prod_j (1+\widetilde{P_j}) - 1} \tag{2.12}$$

$$= [E]_t k^+_{\text{cat}} \left(\frac{\prod_i \widetilde{S}_i}{\prod_i (1+\widetilde{S}_i) + \prod_j (1+\widetilde{P_j}) - 1} \right) \left(1 - \frac{\Gamma}{q} \right) , \tag{2.13}$$

where the normalized reactant constant $\widetilde{S}_i = [S_i]/K_{S_i}$ is the metabolite concentration divided by its reactant constant. A derivation can be found in Appendix A.1. Mechanistically this rate corresponds to the random-ordering binding shown in Figure 2.10. For the reaction to occur, first all substrates must bind to the enzyme (in any order) and after the conversion the products may dissociate from the enzyme in any order, with simultaneous substrate-product binding disallowed.

Liebermeister [42, 43] describes other rate laws similar to that of Eq. (2.12). These differ from (2.12) in the form of their denominator and correspond to slight variations in the enzyme mechanism. For instance, the *simultaneous binding* modular rate law has a denominator equal to $\prod_i (1+\widetilde{S}_i) \prod_j (1+\widetilde{P_j})$. This form contains extra terms compared to the common modular rate law, and corresponds to the case in which simultaneous binding of substrate and product metabolites is allowed, resulting in a smaller reaction rate than in Eq. (2.12). Rate equations of this type can also be easily modified to incorporate allosteric regulation [42, 43]. Such generalized rate laws can be applied to metabolic networks in an automated fashion, and parameter fitting procedures can be used to obtain the parameter values that best agree with the experimental data [50]. In this thesis we will use the common modular rate law to model glycolysis and will compare its predictions to those of simpler rate laws.

Having introduced the basics of metabolism and its modelling, we will now consider in what ways the metabolic networks observed in nature can be understood in terms of evolutionary design and optimality.

2.8 Evolution acts to optimize metabolism

The aim of this project is ultimately to develop an approach that is able to analyse and compare all biochemically feasible alternatives to a metabolic pathway. It is currently unclear whether existing pathways have been selected for from a large or small set of biochemical possibilities when realistic biochemical, biophysical and thermodynamic constraints are applied. If there are other solutions, are the existing pathways a result of historical contingency during early evolution, or are they instead optimal solutions to the problem of energy and biomass production?

Since the central metabolic pathways are highly conserved throughout all forms of life, it is likely that they have been subjected to billions of years of natural selection. As such, it is not unreasonable to view their development as an evolutionary optimization problem, and their current form as the solution to this problem. Indeed, much work has been based on this assumption and has offered new insights into the pressures that directed evolution. Some of these studies have approached questions of optimality by constructing simplified artificial metabolic networks [51–53], while many have limited their scope to biochemical compounds and reactions that are known to exist in living organisms [16, 54, 55]. Here I give an overview of this work and in the subsequent chapters we will aim to improve on this by examining the full biochemical space available to a pathway.

2.8.1 Defining the evolutionary goal

The first task in any evolutionary study is to identify the goal function being optimized by evolution. This is not always a simple task [56]. On the level of the organism, it must ultimately be the case that the organism that reproduces fastest is more likely to outgrow its competitors. This will however depend on many factors, such as the organism's ability to cope with a fluctuating environment [57], varying nutrient sources [16] and direct species-species interactions [58]. At the level of metabolic pathways, each section of metabolism is performing its own distinct function but must work in unison to bring about the optimal strategy for cell growth and division. This brings complications since it is possible that a naive optimization of one pathway, with respect to one specific objective, could negatively affect the functioning of another pathway, resulting in a net detrimental effect to the organism. Such trade-offs have been studied extensively in biology, and there can be a need for multi-objective optimization [56, 59], in which several competing goals must be satisfied simultaneously.

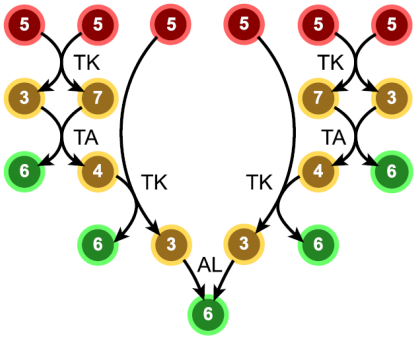


Figure 2.11 The solution to the pentose phosphate game, converting six pentoses to five hexoses in the smallest number of steps. The numbers correspond to the number of carbon atoms in the molecule. TK and TA correspond to the two enzymes, transketolase and transaldolase, which transfer 2 and 3 carbon groups respectively. Figure taken from [29].

It is also true that the present function of a metabolic pathway may not be what it originally evolved for, known as an exaptation in evolutionary theory [16]. This creates the possibility that certain traits are not optimal at all, but are the result of a frozen in evolutionary accident. Distinguishing between the roles of natural selection and chance in the evolutionary history may not always be feasible. Despite the complexity of this area, the procedure of examining individual parts of metabolism and their possible optimization has led to great insights into the evolution and functioning of metabolism.

2.8.2 Structural simplicity and enzyme cost

One of the earliest studies in this direction was conducted in 1985 by Meléndez-Hevia and Isodoro, who looked at the structural design of the pentose phosphate pathway (PPP) [53]. The PPP is part of central carbon metabolism and can run in conjunction with glycolysis, but its main function is anabolic rather than catabolic. The net effect of the PPP is to transform six pentoses (5-carbon sugars) into five hexoses (6-carbon sugars). To study the structure of the PPP, the authors abstracted the problem of converting these sugar molecules into a game in which the rules represented the behaviour of the relevant enzymes: the sugar molecules may transfer either two or three carbons from one to another. By analyzing all combinations of these rules, it was discovered that the natural PPP corresponds to the conversion that involved the smallest possible number of biochemical steps; with the principle emerging being one of structural simplicity. The solution found in [53] is shown in Figure 2.11.

Another key study, in 2010, revisited the principle of structural simplicity in metabolism, with Noor et al. [29] applying the same idea to the entirety of central carbon metabolism

(CCM). This study concluded that, given a defined set of essential precursor metabolites that must be produced, the existing pathways of CCM are the shortest possible biochemical routes connecting these precursors. To show this, the authors examined every pair of metabolites and looked for biochemically feasible alternative pathways between them. The study was not restricted to only those metabolites and reactions found in any database. Instead, each enzyme mechanism known to operate in central carbon metabolism was assigned a generalized molecular operation. Then by applying all transformations that could act on some molecule, and iteratively doing the same on each of the products of that process, they were able to check for biochemical shortcuts between every pair of metabolites present in CCM. An important finding from this study was that there are often multiple minimal-length paths between pairs of precursors. Since this work did not include any discussion of thermodynamics, considering only path length, no attempt was made to further investigate how these pathways differed. Distinguishing between these alternative minimal-length paths will be one of the main lines of investigation in the following chapters.

Both of these studies suggest an evolutionary drive towards simplicity in design, and such a drive is consistent with reasonable biological arguments. For instance it is known that proteins are costly to produce and that cells making excess enzymes suffer from a reduced growth rate [60]. An argument can also be made from the point of view of robustness – by minimizing the number of enzymatic reactions, there is a smaller chance that deleterious mutations will disrupt a given pathway.

2.8.3 Flux and thermodynamic design

Another key concept in our understanding of the design of metabolism is flux. The flux through a pathway is simply the rate of turnover of the molecules in a pathway, or can be thought of as the flow of mass through the pathway. An increased flux will lead to an increased production of any compounds in the pathway, such as an increased rate of ATP production, or cell components. Hence flux is an important quantity, and in many cases it is reasonable to assume that an increased flux for certain processes could confer some fitness advantage.

In particular, Heinrich et al. [52] have asked whether a maximization in the flux of ATP production could explain the form of the glycolytic pathway. By studying the kinetics of generic linear reaction chains, they showed that by arranging the pathway so as to incorporate highly exergonic steps ($\Delta_r G \ll 0$) at the start of the pathway (such as ATP consuming reactions) and placing the endergonic processes (such as ATP production)

at the end of the chain, the flux of the pathway can be maximized. Interestingly, this principle is consistent with the form of real glycolysis, wherein the upper half is the ATP-consuming phase and the lower half is the ATP-producing phase. As we will see later in the thesis however, the distinction between ATP-consuming and ATP-producing as exergonic and endergonic respectively, is not so clear in the context of real biochemical pathways.

In an extension to this work, Ebenhöh [51] considers the stoichiometry of both ATP and NADH producing pathways, corresponding to glycolysis and the TCA cycle coupled to oxidative phosphorylation. The model abstracts a set of generic reactions that act on molecular skeletons, including phosphorylation through the use of both orthophosphates and ATP, along with NADH production and consumption reactions. The goal of the study is to use an evolutionary simulation to explore the vast space of reaction arrangements, characterizing the different pathways by their total ATP production rate. Similar results are obtained from before with respect to the ordering of ATP consumption and production reactions, showing in addition that stoichiometries with a high ATP production rate usually have their NADH producing reactions near the end, corresponding to the TCA cycle. The study also highlights the possibility of alternative pathway designs that are able to produce ATP with only a small reduction in rate, demonstrating that there may indeed be multiple viable options available to life.

2.9 Aims of the following chapters

The majority of studies on the evolution of metabolism have either made use of simplified artificial metabolic networks, or have limited their scope to biochemical compounds and reactions known to exist in living organisms. Both of these approaches have drawbacks: the former does not capture real biochemistry, while the latter is limited to metabolites and reactions found in well-studied organisms. The aim of this project is improve on this; ultimately, we aim to develop an approach that is able to generate and compare all biochemically feasible alternatives to a metabolic pathway. The hope is that by studying what does not exist in nature, but that plausibly could exist, it will be possible to gain some insight into the present form of metabolism. Did natural selection or chance play the dominant role?

Before enumerating and comparing all pathways, we will first have to generate a network of all possible biochemical compounds and reactions, and this process is the subject

of Chapter 4. Chapter 5 will then employ the network generated to investigate the trunk pathway of glycolysis and gluconeogenesis, the most conserved set of reactions in metabolism. We will introduce a method for comparing all possible pathways, and show that the real trunk pathway has the potential to carry a higher flux than any of the alternatives. In Chapter 6 we extend the approach to non-linear pathways and imbed this into an evolutionary simulation to more fully explore the biochemical space. Before constructing our biochemical network I will first introduce the chemistry and thermodynamics that will be needed to do so.

Chapter 3

Biochemical reactions and thermodynamics

3.1 Introduction

In the previous chapter we have already encountered the most important aspects of metabolism, including some basic biochemistry and thermodynamics. In this chapter I will expand on some concepts in chemistry that will be relevant for us, and discuss the thermodynamics of biochemical reactions in greater detail. All of this will be required so that in Chapter 4 we are able to generate a complete network of metabolites and reactions, along with the associated thermodynamics, from basic biochemical principles.

3.2 Reactions of biochemistry

We have encountered a number of reaction types in the previous chapter, including isomerizations, redox reactions and phosphoryl transfers from ATP. Since I will refer to these and others repeatedly throughout the thesis I will here summarize the most important terms and concepts.

There are 6 broad classes of reactions that operate in biochemistry, and each is catalyzed by an enzyme whose name corresponds to the type of reaction it facilitates: oxidoreductases, transferases, hydrolases, lyases, isomerases and ligases. It is worthwhile clarifying what each of these means.

- Oxidoreductases are enzymes that catalyze the transfer of electrons between molecules – oxidation and reduction reactions.
- Transferases facilitate the transfer of molecular subgroups between molecules. For example, the transfer of the phosphate group from ATP to glucose is carried out by a type of transferase.
- Hydrolases catalyze hydrolysis reactions, the breaking of a bond through the addition of water.
- Lyases involve the breaking of a chemical bond through some process other than hydrolysis. They are unique in that they require only one substrate for the reaction to proceed in one direction, but two substrates in the other direction.
- Isomerases carry out isomerization reactions; reactions in which the substrate and product are isomers, both having the same chemical formula but a different structure.
- Ligases catalyze the joining of two molecules through the formation of a new chemical bond.

3.2.1 EC classification

In biology there exists an enzyme classification scheme that assigns a unique identifier to every enzyme known to exist in nature [61, 62]. This Enzyme Commission (EC) number consists of four separate numbers that provide increasingly detailed information about the type of reaction being catalyzed. The first number corresponds to the broadest level of classification, and takes a value from 1 to 6, corresponding to the six reaction types as ordered above. Taking a specific example, the enzyme EC 5.4.2.12, the first number 5 tells us that this enzyme is an isomerase. The subclass EC 5.4 corresponds to intramolecular transferases and EC 5.4.2 is specific to phosphotransferases (the movement of a phosphate group within the molecule). The fourth and final number indicates the precise substrates and products of the reaction; EC 5.4.2.12 is the enzyme catalyzing the isomerization of 3-phosphoglycerate to 2-phosphoglycerate, found in the trunk pathway of glycolysis, and involves the movement of the phosphate group to an adjacent carbon atom. A full breakdown of every subclass can be found at [62]. We will make use of this classification scheme in the following chapters.

3.3 Thermodynamics of biochemical reactions

3.3.1 Free energy of formation

In Section 2.4 we discussed the change in Gibb's free energy associated with a reaction, $\Delta_r G$. Another important concept is the free energy of formation of a compound, $\Delta_f G$. This value corresponds to the energy released when one mole of the substance is formed from its constituent elements in their standard state and if known, can be used to calculate the free energy change associated with a biochemical reaction. Consider a generic reaction containing i substrates $\{S_i\}$ and j products $\{P_j\}$ with arbitrary stoichiometric coefficients $\{m_i\}$ and $\{n_j\}$:

$$m_1 S_1 + m_2 S_2 + ... \rightarrow n_1 P_1 + n_2 P_2 + ... \tag{3.1}$$

If the free energies of formation were known for each compound in the above reaction, then under standard conditions, indicated by the superscript 0, the free energy change associated with the reaction could be calculated as

$$\Delta_r G^0 = \sum_j n_j \cdot \Delta_f G^0_{P_j} - \sum_i m_i \cdot \Delta_f G^0_{S_i}. \tag{3.2}$$

Note that these *standard* conditions correspond to $P = 1$ atm, $T = 298$ K, pH $= -\log_{10}[\mathrm{H}^+] = 0$ and all compounds are assumed to be present in an aqueous solution at a concentration of 1M. In reality, when the concentrations take a range of values not equal to 1 M, the actual free energy of the reaction can be found from

$$\Delta_r G = \Delta_r G^0 + RT \ln \frac{[P_1]^{n_1}[P_2]^{n_2}...}{[S_1]^{m_1}[S_2]^{m_2}...} \ . \tag{3.3}$$

3.3.2 Correction in free energy due to pH

In biology, no cell exists at the *standard* pH value of zero. The form of Eq. (3.3) makes it clear that the thermodynamics of any reaction involving protons will be affected by the pH. For example, consider a reaction in which n protons (H^+) are produced:

$$\text{substrates} \rightarrow \text{products} + nH^+ \ . \tag{3.4}$$

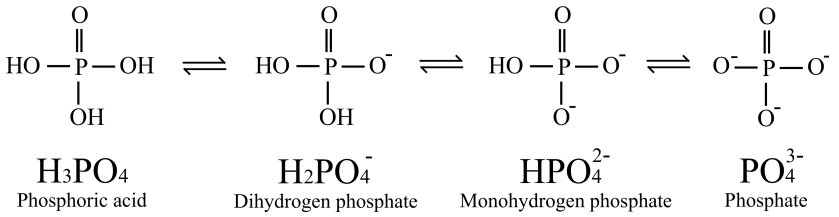


Figure 3.1 Orthophosphate exists in solution as an equilibirum mixture of its four distinct protonation states. The fully protonated form is shown on the left and the fully unprotonated form on the right. Under normal biological conditions it is present primarly in the forms $H_2PO_4^-$ and HPO_4^{2-}.

If all other compounds exist at the standard 1 M concentrations, then the free energy change at some non-zero pH is simply

$$\Delta_r G^{\mathrm{pH}} = \Delta_r G^0 + nRT\ln[H^+] = \Delta_r G^0 + nRT\ln 10^{-\mathrm{pH}} \,. \tag{3.5}$$

3.3.3 Free energy of formation of a biochemical reactant

Equation (3.2) allows the free energy of a reaction to be calculated if the free energy of formation of each compound is known. For this to be true however, requires that the compounds exist in the form that they appear in the reaction equation. Compounds in solution exist as a mixture of various forms, at different levels of protonation. For example, orthophosphate Pi does not solely exist as the usually quoted $H_2PO_4^-$, but rather as a mixture of this and the forms PO_4^{3-}, HPO_4^{2-} and H_3PO_4 as illustrated in Figure 3.1. Carbon dioxide in solution actually exists as a mixture of the dissociated carbonate forms H_2CO_3, HCO_3^- and CO_3^{2-}, and ATP also exists as a mixture: ATP^{4-}, $HATP^{3-}$, etc. Following Alberty [63] these mixtures are called "biochemical reactants" and are denoted by enclosing them in parentheses, e.g. (Pi) represents the equilibrium mixture of the different protonated forms listed above.

Consider a biochemical *reactant* (A) that exists at equilibrium as a mixture of different *species* $\{A_i\}$. If each species has a standard free energy of formation $\Delta_f G_i^0$, then at thermodynamic equilibrium the proportion r_i of species i is given by the Boltzmann distribution:

$$r_i = \frac{\exp(-\Delta_f G_i^0/RT)}{\sum_j \exp(-\Delta_f G_j^0/RT)}. \tag{3.6}$$

The standard free energy of formation of the mixture is [64]:

$$\Delta_f G^0_{(A)} = \sum_i r_i \Delta_f G_i^0 + RT \sum_i r_i \ln r_i \tag{3.7}$$

which is the mole-fraction-weighted average of the mixture (first term) plus the Gibbs energy of mixing (second term). This second term is equivalent to $-T\Delta S$ where $\Delta S = -R\sum_i r_i \ln r_i$ is the entropy of mixing [64]. Substituting Eqn. (3.6) into Eqn. (3.7), leads to an expression for the standard free energy of formation of a biochemical reactant (A):

$$\Delta_f G^0_{(A)} = -RT\ln\left[\sum_i \exp\left(-\Delta_f G^0_i/RT\right)\right]. \tag{3.8}$$

The most important case in the present work is when the species $\{A_i\}$ are the dissociated forms of an acid H_nA. In the general case, a molecule can pass between several states of dissociation

$$\begin{aligned} &H_nA \longleftrightarrow H^+ + H_{n-1}A^- \\ &H_{n-1}A^- \longleftrightarrow H^+ + H_{n-2}A^{2-} \\ &... \\ &HA^{(n-1)-} \longleftrightarrow H^+ + A^{n-} \end{aligned}$$

so that in aqueous solution the acid exists as an equilibrium mixture of all these states. This means that Eq. (3.8) can be used to calculate the standard free energy of formation of the biochemical reactant, (H_nA), at pH=0 and I=0, where the free energy of the proton is defined to be zero:

$$\Delta_f G^0_{(H_nA)} = -RT\ln\left[e^{-\Delta_f G^0_{H_nA}/RT} + e^{-\Delta_f G^0_{H_{n-1}A^-}/RT} + ... + e^{-\Delta_f G^0_{A^{n-}}/RT}\right]. \tag{3.9}$$

If we then wish to calculate this free energy at the more biologically realistic value of pH=7, Eqn. (3.5) can be substituted in place of the $\Delta_f G^0$ values in Eq. (3.9) to give

$$\Delta_f G^{\mathrm{pH}}_{(H_nA)} = -RT\ln\left[\sum_{m=0}^{n} \exp\left(\frac{-\Delta_f G^0_{H_{n-m}A^{m-}}}{RT} - m\ln 10^{-\mathrm{pH}}\right)\right]. \tag{3.10}$$

3.3.4 Ionic strength

Another important property of a solution is what is known as its ionic strength,

$$I = \frac{1}{2}\sum_k c_k z_k^2 \tag{3.11}$$

where c_k and z_k are the molar concentration and charge number of ion k respectively. The ions created when ionic compounds are dissolved in water can affect the physical properties of the solution, and importantly for us, affect the free energy change

associated with a reaction occurring in the solution. When species have electric charges, their thermodynamic properties are corrected to account for the ionic strength of the solution as described by the extended Debye-Hückel theory [63, 65]. The Mathematica notebook of Alberty [66] incorporates the effect of a non-zero ionic strength through a modified version of Eqn. (3.10) [63]:

$$\Delta_f G^{\mathrm{pH,I}}_{(A)} = -RT\ln\left[\sum_i \exp\left(\frac{-\Delta_f G^0_i}{RT} + N_H(i)\ln 10^{-\mathrm{pH}} - \frac{2.91482(Z_i^2 - N_H(i))I^{\frac{1}{2}}}{1+1.6I^{\frac{1}{2}}}\right)\right], \tag{3.12}$$

where Z_i is the charge of species i and $N_H(i)$ the total number of hydrogen atoms. Equation (3.12) allows the free energy of formation of a biochemical *reactant* $\Delta_f G^{\mathrm{pH,I}}_{(A)}$ in an aqueous solution of specified pH and ionic strength to be calculated if the free energies of formation of the individual *species* $\Delta_f G^0_i$ are known. We will make use of Eq. (3.12) later in this thesis.

3.3.5 Experimental thermodynamic data

With enormous experimental effort, the standard free energies of formation $\Delta_f G^0_i$ of every species of 199 biochemical reactants common in biochemistry have been determined and are available in the Mathematica notebook of Alberty [66, 67]. With this data, using Eq. (3.12) it is possible to calculate the free energies of formation of these 199 biochemical reactants as a function of pH and ionic strength I.

3.3.6 Free energy of biochemical reactions

Now that we know how to calculate the free energy of formation of a biochemical reactant, calculating the free energy change of the reaction in aqueous solution is simply through the application of Eq. (3.2). As a specific example, consider the hydrolysis of ATP: (ATP) + H_2O → (ADP) + (Pi). From Eq. (3.2), this has a free energy change of

$$\Delta_r G^{\mathrm{pH,I}} = \Delta_f G^{\mathrm{pH,I}}_{(\mathrm{Pi})} + \Delta_f G^{\mathrm{pH,I}}_{(\mathrm{ADP})} - \Delta_f G^{\mathrm{pH,I}}_{(\mathrm{ATP})} - \Delta_f G^{\mathrm{pH,I}}_{\mathrm{H_2O}}, \tag{3.13}$$

where each of these energies of formation can be calculated from Eq.(3.12) from either [65] or the Mathematica notebook [66]. What this equation corresponds to is the transformation of 1 mole of ATP, which is actually an equilibirum mixture of different species, and 1 mole of water, into 1 mole of ADP and orthophosphate Pi which both also exist as mixtures of species. This is true irrespective of the actual reaction occurring – the exact reaction mechanism does not even need to be known to calculate the free

energy change. For example, if we assume that the actual mechanism of the above reaction involving biochemical reactants is $ATP^{4-} + H_2O \rightarrow ADP^{3-} + H_2PO_4^-$; that is only ATP^{4-} participates in hydrolysis; then Eq. (3.13) still applies. This is easily demonstrated in the schematic below:

$$\begin{array}{rcl} (\mathrm{ATP}) + \mathrm{H_2O} & \equiv & \mathrm{ATP^{4-}} + \quad \mathrm{H_2O} \quad + \quad xyz \\ & & \updownarrow \\ & & \mathrm{ADP^{3-}} + \quad \mathrm{H_2PO_4^-} \quad + \quad xyz \quad \equiv \quad (\mathrm{ADP}) + (\mathrm{Pi}) \end{array}$$

where xyz represents some number of protons or other ions present in the biochemical reactant. Since these are the same on both sides of the reaction they cancel out and do not appear in any calculation of the energy. This also means that in the actual reaction mechanism, we obviously do not need to worry about any other ions or molecules involved, as long as they appear on both sides of the reaction equation.

In this section I have presented the thermodynamics required to calculate the free energy change associated with a reaction under aqueous conditions, in which the biochemical *reactants* exist as an equilibrium mixture of different *species*. This requires the free energies of formation of the reactants to be known, which requires data on the the free energies of formation of each species. Alberty's thermodynamic data [66, 67] provides a way to calculate this for many important biochemical reactants. In the next section I will introduce the group contribution method. This provides a way to estimate $\Delta_f G$ for compounds for which experimental data does not exist. In Section 4.5.1 we will return to this in more detail, when we use it to calculate the free energy of formation for a large set of compounds generated by our algorithm, many of which do not exist in any database.

3.3.7 Group contribution method

The available thermodynamic data for metabolism is far from complete. For example, $\Delta_r G$ values of only 5% of the biochemical reactions contained in the KEGG database have been determined experimentally [28], and only about 8% of those in the metabolic network of *E.coli.* are known [68]. It would thus be advantageous to be able to estimate these unknown values from the available data. The group contribution method can be used to this end [28, 68–70].

The basic idea behind these group contribution methods is that any molecule can be split into a number of functional groups, with each group making a specific contribution

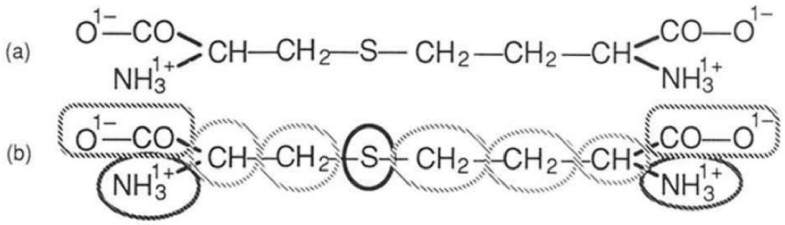


Figure 3.2 The group contribution method assumes that molecules can be broken down into molecular groups and that each of these groups contributes independently to the free energy of formation of that molecule. Shown is the structure of cystathionine and its corresponding break down into chemical groups. Figure taken from [70].

to the total $\Delta_f G$ of the molecule. This is illustrated in Figure 3.2. By identifying the free energy contribution from each of a small set of molecular subgroups, it will then be possible to estimate the $\Delta_f G$ for each of the molecules in the large set that can be built from these groups.

The use of group contribution methods has been around for some time [71], however this early work was designed for compounds existing in the ideal gaseous state and so is not applicable to biochemical reactants in solution. The work of Mavrovouniotis [69, 70] was the first to extend the group contribution method to reactions in aqueous solution. Here, a set of biochemical groups are defined that contribute independently to the free energy of formation of a molecule.[1] In this original work, biochemical compounds do not exist as mixtures of various protonated forms (i.e. as biochemical reactants) but are assumed to exist solely in their dominant state as determined in aqueous solution at pH=7 and T=298K.

The first big advancement of Mavrovouniotis' work, coming nearly twenty years later from Jankowski et al. [68], was in fact largely down to a slight modification in the choice of groups, along with the addition of several more special correction terms for molecular features not previously accounted for. Jankowski also made use of software to estimate the acid disocciation constant of each compound, pK_a, which allowed the predominant form in aqueous solution to be determined before decomposing it into groups (biochemical reactants – mixtures of all species present – were still not accounted for in this work).

The most recent extension to the group contribution method came in 2012 from Noor et

[1] Although each group is assumed to contribute independently to the total $\Delta_f G$ of a compound, with no explicit corrections for group-group interactions being included, special corrections are defined that account for certain molecular features. For instance the *hydrocarbon correction* adds a specific value to the $\Delta_f G$ of all molecules composed only of carbon and hydrogen, while the *amide correction* contributes a certain energy for every occurrence in a molecule of a nitrogen atom attached to a carbonyl group. This process is not general, but requires specialized chemical knowledge.

al. [28]. This work developed the Pseudoisomeric Group Contribution (PGC) method, which essentially incorporates the work of Alberty on mixtures of biochemical species into the framework of Jankowski. The set of groups in [68] is expanded, mainly to include all protonated forms for these groups, so that the $\Delta_f G^0$ of each chemical *species* can be found by decomposing it into its functional groups. Each species has a different $\Delta_f G^0$ and since the distribution of species in a mixture will change with pH, through the Boltzmann ditribution of Eq. (3.6), this method allows the correct adjustment to be made when varying the pH, something not possible in the earlier works of [68–70].

In Section 4.5.1 we describe and employ a variant of these methods in our work. This will use the fact that molecules exist as a mixture of dissociated states in solution, and will also incorporate a second-order correction term to account for nearest-neighbour group interactions.

3.4 Summary

In this chapter we have covered the biochemistry and thermodynamics that will be needed in this thesis. In the next chapter, we will use this information to generate the full network of possible biochemical compounds and reactions that we will employ in Chapter 5 to investgate the possible alternatives to the highly conserved trunk pathway of glycolysis.

Chapter 4

Generating a network of all possible metabolites and reactions

4.1 Introduction

Our goal is to enumerate all biochemically feasible alternatives to the trunk pathway of glycolysis. This requires that we first computationally generate an appropriate network of metabolites and reactions. To do so, we will identify some simplifying properties that can be applied to the metabolites of central carbon metabolism, and specifically the trunk pathway. We will then use the list of all enzymatically catalyzed reactions known to occur on such metabolites to formulate a set of generic reaction mechanisms. In this way, we will be able to explore the full biochemical space available to the trunk pathway, and not be confined to only those metabolites and reactions that have been characterized in living organisms.

4.2 Generating the metabolites

Central carbon metabolism consists exclusively of reactions between "CHOP" molecules: those composed of carbon, hydrogen, oxygen and phosphorus atoms, with the latter being present only in phosphate groups. In the reactions of central carbon metabolism, nitrogen and sulfur are present only in cofactors such as ATP, NAD^+ and CoA [32, 72]. We will make the common distinction of separating the metabolites into two groups: *internal* and *external*. Here, internal metabolites correspond to model

-CH_3	**-CH_2(OH)**	**-COOH**	**-CHO**	**-CH_2p**	**-COp**
=CH_2	**=CH(OH)**	**=CO**	**=CHp**	-CH(OH)-	-CO-
-CHp-	-CH=	-C(OH)=	-Cp=	-CH_2-	

Table 4.1 The set of groups used to construct the internal metabolites in our network. Groups in bold indicate those which must be present at the ends of a molecule. Note that the phosphate group "p" and carboxyl group "COOH" both denote a mixture of protonated and deprotonated forms. This is taken into account in the calculation of free energies of formation (see Section 4.5).

variables, while external metabolites are parameters of the model and are present at fixed concentrations.

The internal metabolites consist of the CHOP molecules, those that may be intermediates during the conversion of G3P to pyruvate. The trunk pathway contains only unbranched aliphatic 3-carbon CHOP molecules. We therefore restrict our list of internal metabolites to unbranched, aliphatic 2, 3 and 4-carbon CHOP molecules. In our program, each internal metabolite is constructed systematically, in a modular fashion, as a linear assembly of the chemical groups shown in Table 4.1, with the restriction that we include no hydrocarbons (molecules consisting of only carbon and hydrogen) since these do not appear in central metabolism. Consider as an example the important metabolite pyruvate which is produced by combining, in order, the three groups CH_3-CO-COOH. Combining the groups of Table 4.1 in all possible permutations leads to a total of 1008 molecules.

An important physical property we consider is electrostatic charge. Small, uncharged molecules can have a very high membrane permeability and an electrostatic charge is an effective way to reduce this [12, 37]. The two ways in which this occurs is through the presence of either a dissociated carboxyl group (-COOH $\rightarrow$ -COO^-) or phosphate group and, in fact, all natural core metabolites contain at least one of these groups – all core molecules are acid derivatives [37]. Applying this restriction reduces the number of molecules to 828. Of these, 18 are molecules that contain 2 carbon atoms, 118 contain 3 carbons and 692 contain 4 carbons.

Note that we have also imposed some extra restrictions on our internal metabolites, based on some simplifying biochemical assumptions. In addition to hydrocarbons, we have excluded C-O-C ether links since they play no role in central carbon metabolism, along with C$\equiv$C triple bonds on the grounds of stability [72]. In our list of chemical groups, Table 4.1, we have not allowed for a single carbon atom to contain multiple hydroxyl or phosphate groups; these *geminal diols* were also excluded on the grounds of stability as they are never observed in metabolism.

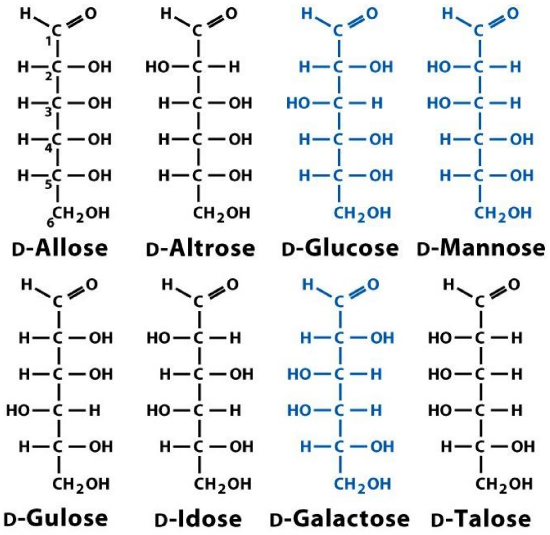


Figure 4.1 Stereochemistry and chirality are not included in or network. Shown here are the 8 (right-handed) D-aldohexoses, which all have the same chemical structure $CH_2(OH)$-CH(OH)-CH(OH)-CH(OH)-CH(OH)-CHO, but differ in the orientation of their bonds; they are stereoisomers. In addition, each has a chiral partner, corresponding to its mirror image, leading to the 8 (left-handed) L-aldohexoses. Our network includes all such molecules as a single metabolite.

The final point to note is that we do not include any notion of stereochemistry. Two molecules are stereoisomers if they have the same chemical formula and same set of chemical bonds, but differ in their three-dimensional orientation in space. Figure 4.1 shows the eight hexoses (6-carbon sugars) which all differ in their bond orientation. Though such compounds can have slightly different physical and chemical properties, and can have very different biological properties due to an enzyme's specificity for a particular form, this assumption is not likely to be a significant problem with respect to the thermodynamics of simple molecules. For instance; glucose, galactose and mannose (the three of the eight hexoses shown in Figure 4.1 that most commonly occur in nature), have $\Delta_f G^0$ values of -909, -901 and -901 kJ/mol respectively, a difference of less than 1% [68]. We thus treat all stereoisomers as a single molecule in our network – each internal metabolite is fully characterized by a "string" object, consisting of the groups in Table 4.1.

The external metabolites are those that participate in a great number of reactions, such as phosphates, CO_2, water and the metabolic cofactors ATP and NAD^+. The list of external metabolites included in our nework is given in Table 4.2. In the analysis of the following chapters, these external metabolites are assumed to exist at fixed concentrations, creating a pool into which the internal metabolites will be placed. These concentrations will be the parameters of our model, defining the physiological conditions of the cell. In solution, these compounds exist as a mixture of different

dissociated forms; this is taken into account when we come to calculate the Gibbs free energies of the reactions, as discussed in Section 3.3.3. Note that although water is listed here, it is not actually a parameter of the model. Since biochemical systems are assumed to be aqueous environments, the concentration of water is fixed to its molar concentration $[H_2O] = 55.5$ M.

ATP	ADP	AMP
NAD^+	NADH	$CO_2(aq)$
Pi	PPi	H_2O

Table 4.2 The set of external metabolites used in our analysis. Pi is orthophosphate and PPi is pyrophosphate.

4.3 Generating the reactions

To generate metabolic pathways from our chemical compounds we need to connect them through biochemical reactions. We do this in a systematic way by defining a set of allowed reaction mechanisms, which can be done by making use of the EC classification scheme (Section 3.2.1).

4.3.1 Reactions in our network

The reaction classes to be included in the network are described by just the first 3 EC numbers, since they describe mechanistically equivalent chemical transformations, carried out on different substrates. Twelve of these EC classes can be identified that together encompass all the reactions between CHOP molecules of length 2-4 carbons in core metabolism. These are shown in Table 4.3, alongside the specific molecular operations. For each reaction class we also include all known couplings to the external metabolites shown in Table 4.2. This means that a given pair of internal metabolites may be connected by more than one reaction (edge) in our network, e.g. the three different reactions coupling PEP and pyruvate in Figure 2.8. It is essential to include these different couplings, since they involve different free energy changes and will be affected by the cellular environment in different ways; something that is not considered in previous works [29, 73].

As an example of a generic reaction rule, consider the simple dehydration reaction in Table 4.3. The generalized reaction is encoded as follows: any molecule that posseses a hydroxyl group (OH) adjacent to a carbon with a hydrogen atom (H) can lose these

Reaction type	EC	Generalised reaction(s)
oxidation	1.1.1	R-CH_2(OH) + NAD^+ ⇌ R-CHO + NADH
	1.1.1	R-CH(OH)-R′ + NAD^+ ⇌ R-CO-R′ + NADH
	1.2.1	R-CHO + NAD^+ + H_2O ⇌ R-COOH + NADH
oxidation and phosphorylation	1.2.1	R-CHO + NAD^+ + Pi ⇌ R-COp + NADH
phosphate transfer	2.7.1	R-OH + ATP ⇌ R-p + ADP
	2.7.1	R=Cp-R′ + ADP ⇌ RH-CO-R′ + ATP
	2.7.2	R-COOH + ATP ⇌ R-COp + ADP
	2.7.9	R-OH + ATP + H_2O ⇌ R-p + AMP + Pi
	2.7.9	R-OH + Pi + ATP ⇌ R-p + AMP + PPi
	2.7.9	R-COOH + ATP + H_2O ⇌ R-COp + AMP + Pi
	2.7.9	R-COOH + Pi + ATP ⇌ R-COp + AMP + PPi
	2.7.9	R=CP-R′ + AMP + Pi ⇌ R-CO-R′ + ATP + H_2O
	2.7.9	R=Cp-R′ + AMP + PPi ⇌ R-CO-R′ + ATP + Pi
hydrolysis	3.1.3	R-p + H_2O ⇌ R-OH + Pi
decarboxylation	4.1.1	R-COOH + H_2O ⇌ R-H + CO_2(aq)
decarboxylation and phosphorylation	4.1.1	COOH-C(R′R″)-CO-R + Pi ⇌ C(R′R″)=Cp-R + CO_2(aq)
	4.1.1	COOH-C(R′R″)-CO-R + ATP ⇌ C(R′R″)=Cp-R + ADP + CO_2(aq)
dehydration	4.2.1	R(H)-(OH)R′ ⇌ R=R′ + H_2O
isomerization	5.3.1	R-CO-CH_2(OH) ⇌ R-CH(OH)-CHO
	5.4.2	R-CH(OH)-CH_2p ⇌ R-CHp-CH_2(OH)
	5.4.2	R=C(OH)-CH_2p ⇌ R=Cp-CH_2(OH)
	5.4.2	R-CHp-COOH ⇌ R-CH(OH)-COp
	5.4.2	R=Cp-COOH ⇌ R=C(OH)-COp
tautomerism	5.3.2	R=C(OH)-R′ ⇌ RH-CO-R′
	5.3.2	R=CH(OH) ⇌ RH-CHO
ATP-driven carboxylation	6.4.1	R-H + CO_2(aq) + ATP ⇌ R-COOH + ADP + Pi

Table 4.3 The set of reaction types included in our network as defined by the first 3 numbers of the EC classification along with all external metabolite couplings. Here, R and R′ denote arbitary chemical groups (not necessarily linear chains), Pi denotes orthophosphate, PPi denotes pyrophosphate, and p denotes the phosphate group of a phosphorylated substrate. All compounds are assumed to exist as an equilibrium mixture of protonated and deprotonated species in solution.

three atoms as a water molecule, resulting in the production of a C=C double bond. If a molecule has multiple hydroxyl groups, it may be possible for multiple dehydration reactions to occur, each producing a different metabolite depending on which hydroxyl group has been eliminated.

4.4 Network properties

By systematically generating all possible CHOP molecules of 4 carbons or less, and connecting them together with all possible reactions in our 12 classes, we generate a large reaction network. This network encompasses not only reactions that are found in nature but also others that, although biochemically feasible, have not been observed to occur in nature. Here we investigate some properties of this network.

The network generated in the process described above is shown in Figure 4.2C, and contains 828 electrostatically charged CHOP molecules of 4 carbons or less and 5326

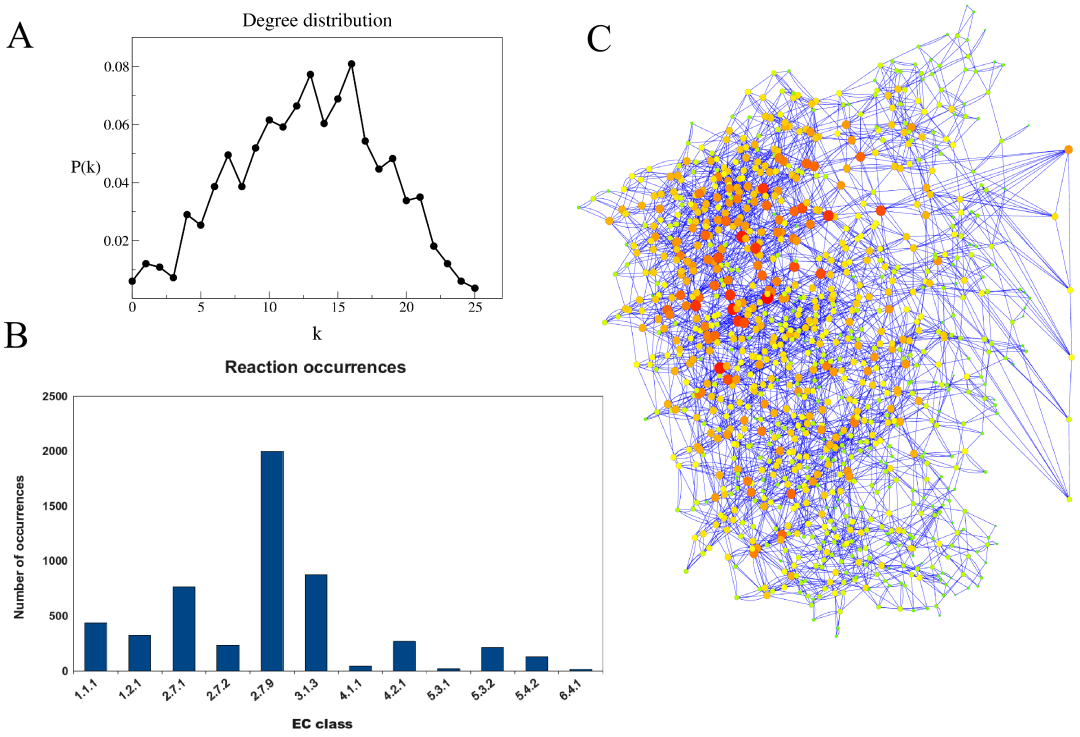


Figure 4.2 Some properties of our generated biochemical network. **(A)** The degree distribution: the fraction of nodes $P(k)$ that have degree k, i.e. participate in k reactions. **(B)** Histogram of reaction occurrences. Number of reactions from each EC class present in the network. **(C)** Visualization of generated network with trunk pathway pulled out on right hand side (dikinase reactions, EC 2.7.9, were excluded for visual clarity due to the large number of them, hence there being only one edge between the bottom two trunk nodes, PEP and pyruvate. Nodes of degree zero are also omitted). This network visualization was created using Cytoscape; node colour and size both correlate with node degree.

possible reactions. This visualization helps illustrate how densely connected the network is, and gives an impression of the number of alternative routes that might be available to the trunk pathway when the full space of chemicals and reactions can be utilized. Extending the network to molecules containing 5 and 6 carbon atoms results in more than 30,000 metabolites and 400,000 reactions, increasing the size of the network dramatically. Trying to capture the biochemistry of the full metabolic network then, i.e. including nitrogen and sulfur, and allowing cyclic compounds, will likely lead to millions of possible biochemical compounds and reactions.

In Figure 4.2A the degree distribution of the network is shown. The degree of a node (an internal metabolite) corresponds to the number of reactions that it can participate in. We see that there is a large range of possible values here, with a minimum degree of $k = 0$ and a maximum degree of $k = 25$. Note that the reason $k = 0$ metabolites exist is that we only include molecules possessing an electrostatic charge. For instance, if a metabolite can only dephosphorylate to an uncharged compound, then it will not

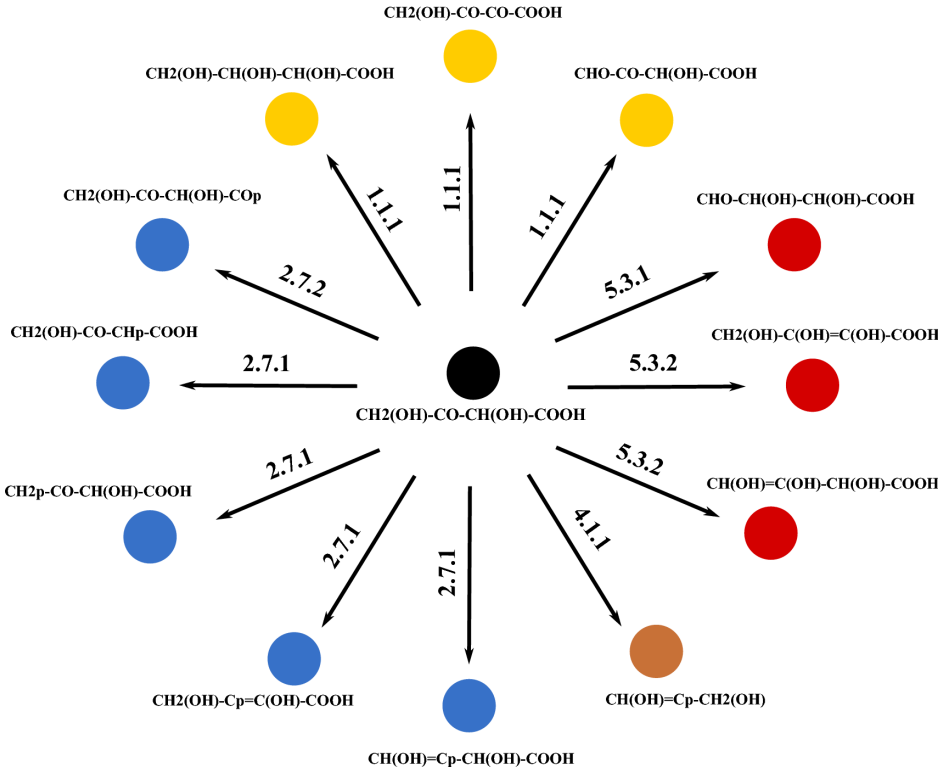


Figure 4.3 Multiple reactions can occur on a molecule. The central molecule can participate in 25 reactions, leading to 12 distinct compounds. Colours correspond to basic reaction reaction type: yellow, redox reactions; red, isomerizations; brown, decarboxylation; blue, phosphorylations. Note that some reactions can couple to different external metabolites (see step 10, Figure 2.8) – for example it is possible for the phosphorylations to be performed by EC 2.7.9, or in the case for the upper three, a (reverse) hydrolysis reaction via EC 3.1.3.

be connected to any other metabolites in the network.

One of the three compounds with the maximal degree of 25 is 2,4-dihydroxy-3-oxobutanoic acid, which has the chemical formula CH_2(OH)-CO-CH(OH)-COOH. Recall that there can be multiple reactions between the same two nodes, all coupling to different external metabolites. From the 25 reactions that this molecule can participate in, there are 12 distinct internal CHOP molecules that can be produced (Figure 4.3). The majority of reactions involve phosphorylations, which can occur in multiple ways. For example, the phosphorylation to the compound CH_2(OH)-CO-CH(OH)-COp can in principle be coupled to the transformation of ATP to ADP (EC 2.7.2), to the transformation of ATP to AMP in one of the dikinase reactions (EC 2.7.9) or through the straight acquisition of an orthophosphate (the reverse hydrolysis reaction of EC 3.1.3). Either of the two hydroxyl groups or the carboxyl group can in principle be phosphorylated.

The histogram of Figure 4.2B shows that the phosphorylation reactions make up a large

part of the reaction network. EC classes 2.7.1, 2.7.2, 2.7.9 and 3.1.3 together contain 3873 of the total 5326 reactions. It is interesting to note that although the dominant class in our network is EC 2.7.9, in reality only 5 enzymes have been observed in nature that carry out reactions in this class. It is not clear whether there is a biochemical or enzymatic reason for their rarity. It could be the case that these reactions require a very complex enzymatic mechanism and are thus rare, or it could be that there is a general lack of need for them in biology. Nonetheless, since there is no clear reason why these couldn't in principle replace any normal phosphorylation reaction, they are included in the network.

4.5 Reaction free energies: including thermodynamics

Having generated a network of compounds and reactions, we wish to be able to calculate the flux carried by the pathways contained within it. To do this, we need to know the thermodynamics of the network, i.e. the free energy change associated with each reaction. Since the majority of the compounds and reactions generated above are not known to exist in nature, thermodynamic data is not available for them. Thus we must somehow estimate these values from the available thermodynamic data, which we will do by applying a variant of the group contribution method [28, 68–70] as described in Section 3.3.7.

4.5.1 Group contribution method

Recall that the basic idea behind these group contribution methods is that any molecule can be split into a number of functional groups, with each group making a specific contribution to the total free energy of formation $\Delta_f G$ of the molecule. In all of these works it is assumed that the group contributions are independent of each other, but here we will extend this to allow for the possibility of group-group interactions as follows. Since the internal metabolite molecules are built as linear chains from the functional groups $\{g_i\}$ listed in Table 4.1, we write their free energy of formation as

$$\Delta_f G^{\mathrm{pH,I}} = E_0 + \sum_j E_1(g_j) + \sum_{<j,k>} E_2(g_j, g_k), \tag{4.1}$$

where E_0 is a constant (which is the same for all molecules), $E_1(g_j)$ is the contribution of group g_j and $E_2(g_j, g_k)$ is a small correction arising from neighbouring group-group interactions. Alberty [67] has listed the standard Gibbs free energies of formation

for every species that is significant between pH 5-9, for 199 compounds of biological interest, 56 of which are linear CHOP molecules as in our network. Values for E_0, the vector E_1 and matrix E_2 are obtained by performing a least-squares fit to this set of 56 data points. That is, Eq. (3.12) is used to calculate the free energies of formation for the 56 compounds in the "training set", under the biologically relevant conditions of pH = 7, $T = 298$ K, $I = 0.2$ M, with the assumption that all compounds are at 1 M concentration. The $\Delta_f G^{\mathrm{pH,I}}$ values of these compounds are then used to calibrate the values of E_0, E_1 and E_2 under these specified conditions. The values for E_0, E_1 and E_2 obtained by this fitting procedure can be found in Appendix A.2. These can then be used to calculate the $\Delta_f G^0$ of every compound in the network using Eqn. (4.1), and hence the energy change associated with every reaction can be calculated using $\Delta_r G^0 = \sum_{\mathrm{products}} \Delta_f G^0 - \sum_{\mathrm{substrates}} \Delta_f G^0$. I thank Bartek Waclaw, who wrote the computer program that performs this group contribution procedure.

Accuracy of group contribution method

For the 56-compound training set, we find that the error – the square root of the variance of absolute differences between the experimental $\Delta_f G^{\mathrm{pH,I}}$ values and those calculated using the group contribution method – to be 3.84 kJ/mol. The relevant error for us however is the resulting uncertainty in the free energies of the reactions, $\Delta_r G^0$. Although these errors in $\Delta_f G^{\mathrm{pH,I}}$ could, in principle, combine to produce larger errors in the reaction free energies, we find that in practice this is not the case. In Section 5.9 we discuss the effect of choosing different subsets of these training molecules and find that the overall outcome of the analysis presented in Chapter 5 is robust to such a choice.

4.6 Conclusion

In this chapter we have generated a network of biochemically feasible compounds and reactions that could potentially be implemented in real metabolism. We have employed a variant of the group contribution method to calculate the Gibbs free energy of each reaction in our network. This network has been restricted to the simple CHOP molecules observed in the metabolic trunk pathway of lower glycolysis, but already contains a large number of compounds and reactions, suggesting many biochemical alternatives may exist to the trunk pathway. In the next chapter we will investigate this by enumerating these alternative paths and will introduce a method to compare

them based on the metabolic flux that each pathway can sustain.

Chapter 5

The trunk pathway represents a maximal flux solution

5.1 Introduction

In the last chapter we generated a network of biochemical compounds and reactions, and used a variant of the group contribution method to calculate the free energy changes of each reaction. In this chapter, we will investigate the trunk pathway of lower glycolysis (G3P to pyruvate; Figure 2.8). We will use our generated network to enumerate all possible biochemical routes between G3P and pyruvate, finding that a great number of pathways exist. We will then introduce a method for comparing these pathways that will involve comparing the fluxes sustained by each of them under a variety of physiological conditions. We will consider in detail some of the alternatives found in our network, and also extend our investigation to pathways that are not restricted to end at pyruvate. Before doing this, I will introduce our method for calculating the biochemical flux of a linear pathway.

5.2 Calculating the flux of a linear pathway

To compare the pathways generated in our network we will calculate their steady state metabolic flux, by which we mean the rate of turn over of the molecules in the pathway. This is motivated by the fact that other work has suggested that maximization of ATP flux has been an important quantity under evolutionary selection [51, 52, 56].

To calculate the flux of a pathway we need to make assumptions about the enzyme kinetics. In this chapter we will describe a method that assumes "perfect enzymes". This approach offers a best case scenario for the flux and can be solved analytically for linear pathways. In the next chapter we will consider more complicated enzyme kinetics, those of the common modular rate law (Section 2.7.1). These kinetics can require intensive numerical calculations, but arguably provide a more realistic description and allow any network of reactions to be modelled, not only linear paths.

To compute the flux carried by a linear pathway we employ the method of Heinrich et al. [52]. Consider first an unbranched chain of n reactions with fixed concentrations of the initial substrate $[S_0]$ and final product $[S_n]$. For the simplest case, in which each reaction converts one substrate S_{i-1} into the next, S_i, and in which the flux v_i through reaction i obeys linear kinetics, we can write

$$v_i = k_i[S_{i-1}] - k_{-i}[S_i] = k_i\left([S_{i-1}] - \frac{[S_i]}{q_i}\right). \tag{5.1}$$

Here k_i and k_{i-1} are the first-order rate constants in the forward and backward directions and $q_i = e^{-\Delta_r G^0/RT} = k_i/k_{-i}$ is the thermodynamic equilibrium constant, where $\Delta_r G^0$ is the standard free energy change of the reaction as computed in our analysis using the group contribution method (Section 4.5.1). In our network we have many reactions that involve the conversion of external metabolites. A generic reaction that converts internal metabolite S_{i-1} to S_i, with a coupled conversion of external metabolite e_{i-1} to e_i, can be expressed as

$$S_{i-1} + e_{i-1} \underset{\kappa_{-i}}{\overset{\kappa_i}{\rightleftharpoons}} S_i + e_i. \tag{5.2}$$

By analogy with Eq. (5.1), we can write the flux v_i through this reaction as

$$v_i \;=\; \kappa_i[S_{i-1}][e_{i-1}] - \kappa_{-i}[S_i][e_i], \tag{5.3}$$

and since in this work we assume that all external metabolites are present at fixed concentrations $[e_i]$ (which determine the intracellular conditions), these concentrations can be absorbed into the rate constants to obtain the pseudo first-order rate equation

$$v_i \;=\; K_i[S_{i-1}] - K_{-i}[S_i] = K_i\left([S_{i-1}] - \frac{[S_i]}{q'_i}\right). \tag{5.4}$$

Here $K_i = \kappa_i[e_{i-1}]$, $K_{i-1} = \kappa_{-i}[e_i]$, and thus the *apparent equilibrium constant* $q'_i = K_i/K_{-i} = \kappa_i[e_{i-1}]/\kappa_{-i}[e_i] = q_i\frac{[e_{i-1}]}{[e_i]} = \frac{[e_{i-1}]}{[e_i]}e^{-\Delta_r G^0/RT}$. Note that q' incorporates the shift in the equilibrium constant that arises from taking into account the fixed external

metabolite concentrations defined in our model.

We now consider a pathway consisting of a linear chain of such reactions. In steady state, the flux carried by each reaction in the chain must be equal to the flux J through the pathway, i.e. $v_i = J$, for all i. Imposing this condition and rearranging Eq.(5.4) gives

$$[S_i] = q_i' \left([S_{i-1}] - \frac{J}{K_i} \right). \tag{5.5}$$

Iterating this equation, starting from $i = 1$, results in an expression for the steady state concentrations of the intermediate metabolites, in terms of the rate constants, $\Delta_r G^0$ values, and the concentration $[S_0]$ of the starting substrate:

$$[S_j] = [S_0] \prod_{i=1}^{j} q_i' - J \sum_{i=1}^{j} \frac{1}{K_i} \prod_{m=i}^{j} q_m'. \tag{5.6}$$

Setting $j = n$ in Eq. (5.6) and rearranging leads to an expression for the pathway flux J in terms of the rate constants, $\Delta_r G^0$ values, and the concentrations $[S_0]$ and $[S_n]$ of the initial and final substrates:

$$J = \frac{1}{D} \left([S_0] \prod_{i=1}^{n} q_i' - [S_n] \right), \tag{5.7}$$

where

$$D = \sum_{i=1}^{n} \frac{1}{K_i} \prod_{m=i}^{n} q_m'. \tag{5.8}$$

The flux calculated by Eqs. (5.7) and (5.8) is equivalent to that derived by Heinrich et al. [52], including the external metabolites through the alteration of the equilibrium constants of the reactions they participate in.

5.2.1 Perfect enzymes

Equations (5.7) and (5.8) require knowledge of the rate constants for all enzymes in the pathway. In reality, these are usually complicated functions involving multiple metabolite concentrations and kinetic parameters. To overcome this problem, we assume that all enzymes in our network behave as perfect catalysts [52, 74–76], where it is assumed that the reaction rate is limited only by the rate of diffusion of substrate to the enzyme. Because of this assumption, the flux obtained in our calculations can be thought of as representing the maximum possible flux sustainable by a given pathway.

Under this assumption the rate equation becomes [52, 74–76]

$$v_i = \frac{k_d[E_i]\left([S_{i-1}]q_i - [S_i]\right)}{1+q_i}, \tag{5.9}$$

where $[E_i]$ is the concentration of the enzyme catalyzing step i and k_d denotes the diffusion-limited rate constant. Deriving Eq. (5.9) requires several steps, which are given in Appendix A.3, following the derivation of Pettersson [76]. Complementary derivations of Eq.(5.9) have been presented by Heinrich et al. [74] and Albery et al. [75].

The form of Eq. (5.9) is analogous in form to Eq. (5.4) if we define the mapping

$$K_i = \frac{k_d[E_i]q_i}{1+q_i} \qquad , \qquad K_{-i} = \frac{k_d[E_i]}{1+q_i}, \tag{5.10}$$

where these rate constants now depend on the diffusion-limited rate k_d, the enzyme concentration and the equilibrium constant of the reaction. This mapping allows us to use the above derivation of Eqs. (5.7) and (5.8) for the flux of a pathway composed of these "perfect enzymes". After replacing q in Eq. (5.10) with q' for any reaction coupled to the conversion of external metabolites, substituting Eq. (5.10) into these two equations yields:

$$J = \frac{1}{D}\left([S_0]\prod_{i=1}^{n} q_i' - [S_n]\right), \tag{5.11}$$

where

$$D = \sum_{i=1}^{n} \frac{1}{K_i(E_i, q_i')} \prod_{m=i}^{n} q_m'. \tag{5.12}$$

Note that it is implicitly assumed in this calculation that the external metabolite concentrations do not affect the enzyme *kinetics*, but only the reaction thermodynamics. That is, the concentrations of the external metabolites enter only through the modification of the reaction's equilibirium constant and not through any kinetic terms. Although this may not be strictly true, it is interesting to note that ATP and NAD^+ are consistently saturating: in biology they are usually present at concentrations of more than an order of magnitude larger than the K_M values of their enzyme sites [77]. Hence these metabolites are not a limiting factor for the enzyme activity, somewhat justifying the above assumption. This assumption can be relaxed when we come to discuss other flux calculations, such as the common modular rate law in Chapter 6.

5.3 Enumerating all trunk pathways

Having generated a network of biochemically feasible compounds and reactions, the first question to ask is how many alternatives to the trunk pathway exist, i.e. how many biochemical routes are there between G3P and pyruvate? Using the depth-first search algorithm (described in Appendix A.4) to enumerate all pathways in the network, reveals a huge number of alternative routes which are consistent with the rules of biochemistry. The black line in Figure 5.1 shows an approximately exponential increase in this number as path length is increased. This represents the number of pathways based on the network topology alone – no thermodynamic constraints have yet been imposed. The primary reason for such a large number of pathways is the combinatorial nature of the biochemistry, whereby the same net transformation of one compound to another can be performed by the same set of individual reaction types but performed in a different order, making use of different intermediate compounds. There are also many alternative routes discovered that utilize a different set of the reaction types contained in the network. An example of each of these cases is shown in Figure 5.1. The pathway on the left makes use of the same 5 reaction types (defined by the first 3 EC numbers) as the real glycolytic trunk path, while the pathway on the right can be seen to use a different set, in which the final two reactions actually involve a carboxylation and decarboxylation through oxaloacetate (OXA) and thus involve a coupling to CO_2, which is not the case for the pathway on the left.

How many of these alternative trunk pathways are feasible under typical physiological conditions? We first demand that candidate pathways in the glycolytic direction should produce at least 2 ATP molecules. This is required to ensure that the full glycolytic pathway will produce a net ATP yield, since both the EMP and ED forms of glycolysis consume 1 ATP per G3P molecule in their upper half. If the trunk pathway yielded just 1 ATP molecule, the ATP production in the lower half would exactly match the consumption in the upper half and neither of the EMP or ED pathways would generate any ATP. The red line in Figure 5.1 shows the effect of the restriction to ATP$\geq$2. Note that although there exist 14 biochemical pathways of length 4 connecting G3P and pyruvate, none of these are able to generate 2 ATP molecules using the reaction mechanisms present in the network.

To further restrict our analysis we can demand that pathways must be thermodynamically feasible by using the flux calculation of Eq. (5.11) to check that the pathway can sustain a positive flux. This will depend on the cellular conditions, as defined by the concentrations of the external metabolites and the fixed concentrations of G3P and

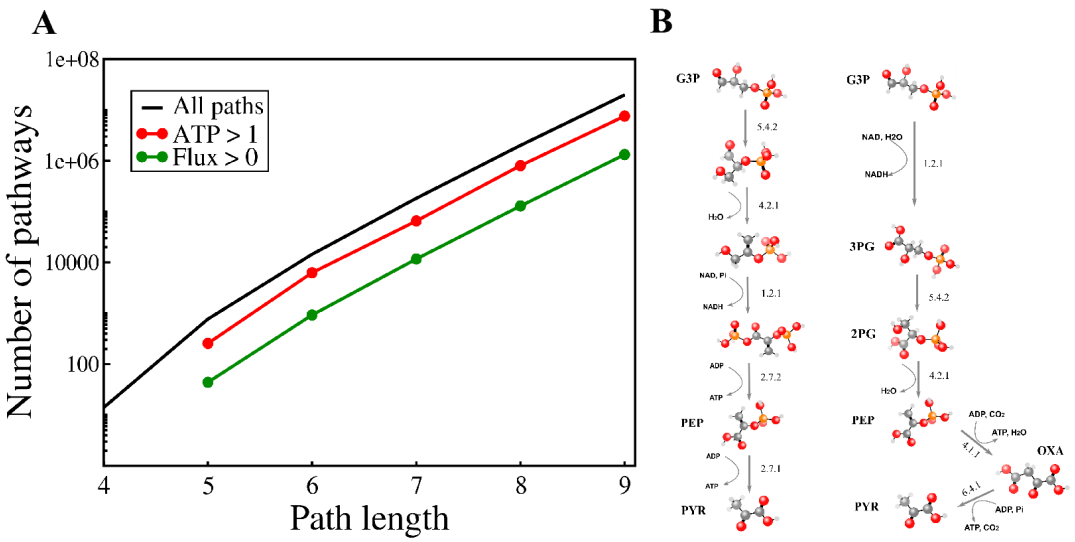


Figure 5.1 **A.** The number of pathways connecting G3P and pyruvate in our network increases exponentially as a function of the path length (number of reactions). Shown is the total number of biochemical routes (black), glycolytic paths that produce at least 2 ATP molecules (red) and those that sustain a positive flux under a certain set of external metabolite concentrations (green). These concentrations define the "typical physiological conditions" discussed in the text, corresponding to (all concentrations in mM) [G3P]=0.019, [PYR]=0.051 , [ATP]=9.6, [ADP]=0.56, [AMP]=0.28, [NAD^+]=2.6, [NADH]=0.083, [Pi]=1.0, [PPi]=0.001, and [CO_2]=1.0 with values taken from [77, 78] or estimated from a variety of sources. **B.** Examples of alternative glycolytic trunk pathways that use the same set of reaction types as the real trunk path (left), and a different set of reactions (right).

pyruvate, but we fix these to the typical cellular values given in the caption of Figure 5.1. To generate this plot we set all enzyme concentrations to $[E_i] = 0.1$ mM in the flux calculation of Eq. (5.11). The green line in Figure 5.1 shows that, given a particular cellular state, insisting on a positive flux can greatly reduce the number of candidate pathways but that there are still many.

5.4 Biophysical restrictions on the metabolite concentrations

As well as assessing the flux carried by our candidate metabolic pathways, we will also be interested in the steady state concentrations of the various intermediates, i.e. the $[S_i]$. Once the steady state flux of a pathway has been calculated using Eq. (5.11), the profile of steady state intermediate concentrations can be obtained with Eq. (5.6). When evaluating the performance of our various pathways, we will restrict these concentrations to lie within biophysically feasible limits. There are good reasons for imposing limits on these values. High metabolite concentrations can produce osmotic stress and increase the rate of disadvantageous spontaneous reactions, such as the non-

enzymatic conversion of dihydroxyacetone to the toxic intermediate methylglyoxal [79]. Because many enzymes have active sites that can bind to multiple substrates, and indeed many small molecules are sufficiently similar in shape to be recognized by the same enzyme, high concentrations can also reduce the rate of the primary reaction. On the other hand, metabolite concentrations that are too low are undesirable due to vanishing reaction fluxes. Experimentally it has been observed that metabolite concentrations can span several orders of magnitude, with most concentrations falling between 0.1 μM and 100 mM and the total intracellular metabolite pool, measured in *E. coli*, being around 300 mM [77].

In our initial study we define the *feasible range* of steady state concentrations for any internal metabolite to be between 1 nM and 500 mM. The lower bound here roughly corresponds to a concentration of 1 molecule per *E.coli* cell and the upper range is quite generous, corresponding to about five times the highest concentration measured in *E. coli* [77]. In the comparison that follows, if a candidate pathway has any steady state metabolite concentration that falls outside of this range, it is deemed infeasible and its flux is set to zero. In Sections 5.8.1 and 5.8.2 we will see the effect of altering these bounding values.

5.5 Maximizing flux: optimal enzyme distribution

To evaluate the performance of our candidate pathways, we choose to compare the maximal possible flux that they can sustain for a given set of values of the external metabolite concentrations and the concentrations of the starting and end metabolites G3P and pyruvate. We also assume that the total concentration of enzyme available to the cell is fixed, and that the cell is free to distribute the enzymes across the reactions in the pathway so as to maximize the total flux. Since enzyme production is costly this seems a reasonable assumption. Altering the distribution of enzymes along the pathway will alter the steady state flux as well as the metabolite concentrations, and so this optimization must be performed subject to the constraints on the intermediate concentrations. The task of calculating the flux of a pathway thus becomes an optimization problem in which the flux must be maximized subject to the constraint of a fixed total enzyme concentration, $[E_T]$, and the constraint that all steady state

metabolite concentrations must fall within some prescribed range. That is:

$$\begin{aligned}&\text{Maximize } J([E_i]), \text{ subject to:}\\&(1)\ \sum_i [E_i] = [E_T], \text{ and}\\&(2)\ S_{\min} \leq [S_i] \leq S_{\max}\ \forall i.\end{aligned}$$

The total concentration of enzymes does not affect the comparison and so is set to an arbitrary value $[E_T] = 1.0$ mM for all pathways. The limits on the metabolite concentrations $S_{\min}$ and $S_{\max}$ are initially set to 1 nM and 500 mM, and we will investigate the importance of these values in Sections 5.8.1 and 5.8.2. This multidimensional optimization was performed using a variant of Powell's method [80] which is described in Appendix A.5.

5.6 Optimal solution depends on cellular environment

Motivated by considerations of simplicity [29, 53] and protein cost [34, 60] we first restrict our analysis to those pathways with the minimal number of enzyme catalyzed steps. Consistent with the requirements that the glycolytic trunk path must produce at least 2 ATP molecules, and that the gluconeogenic pathways do not have a poorer yield of G3P per ATP than the real pathway, we find that the minimal length of pathways that can sustain positive flux under typical physiological conditions is 5 in both directions. Our network contains 202 candidate 5-step glycolytic paths and 300 candidate 5-step gluconeogenic paths that meet these requirements.

To evaluate the performance of these pathways generated by our network, we will compare their steady state metabolic fluxes. For glycolytic pathways this flux corresponds to the rate of ATP production, while for the gluconeogenic pathways, it corresponds to the rate of production of G3P (and any intermediate precursor metabolites). This flux depends not only on the total free energy across a given pathway, but also on the distribution of individual reaction free energies – the thermodynamic profile. This will in turn depend on the intracellular environment, via the concentration of the external metabolites. Since the concentration of these external metabolites can vary for different cells and under different environmental conditions, we wish to compare the performance of all pathways over a wide range of possible conditions.

Parameter	Range sampled
[source]	1μM to 1mM
[source]/[product]	0.001 to 1000
[ATP]/[ADP]	0.001 to 1000
[NAD^+]/[NADH]	0.001 to 1000
[AMP]	1μM to 1mM
[Pi]	1μM to 100mM
[PPi]	1nM to 10mM
[CO_2]	1μM to 1mM

Table 5.1 Parameters sampled independently in our analysis and the ranges over which they were sampled. The source and product concentrations correspond to G3P and PYR in the glycolytic direction and vice versa in the gluconeogenic direction.

5.6.1 Sampling the parameter space

To cover a range of environmental conditions we evaluated the flux of our candidate pathways for 10,000 different parameter sets, the values of which were sampled from a wide range. There are 10 parameters that define the state of the cell: the concentrations of the 8 external metabolites of Table 4.2 plus the concentration of G3P and pyruvate which must be fixed in the flux calculation. The parameter set is defined in Table 5.1 and each parameter is sampled over a range covering several orders of magnitude above and below its typical physiological concentration. Since these ranges cover many orders of magnitude, the sampling is performed uniformly from the logarithm of the ranges shown in Table 5.1. The biologically important ratios [ATP]/[ADP] and [NAD^+]/[NADH] relate to the energy state and redox state of the cell respectively, and it is these ratios that are treated as the parameters, with the values of [ATP] and [NADH] being set to an arbitrary value of 1.0 mM. With this choice, and the range of [AMP] shown in Table 5.1, the [ATP]/[AMP] ratio is thus being sampled from a range from 1 to 1000. A wider range was used for the concentrations of orthophosphate and pyrophosphate since available literature data for these values show much greater variability than the other parameters.

5.6.2 Comparative flux

To compare the performance of our candidate pathways, averaged over parameter space, we introduce a simple metric we denote the *comparative flux* (CF). The comparative flux of a pathway is defined as the flux of that pathway, divided by the maximum flux obtained amongst all pathways at that given point in parameter space. That is, for each point in parameter space sampled, the optimized flux J_i of each candidate

pathway is first found and then the comparative flux of path k is calculated as $\mathrm{CF}_k = J_k/\max\{J_i\}$. This CF_k is thus a value between 0 and 1: if a pathway has a value $\mathrm{CF}_k = 1$, it has the highest flux from all pathways under the given environmental conditions, while if a pathway scores $\mathrm{CF}_k = 0$, it is either thermodynamically infeasible or cannot satisfy the restrictions on the intermediate metabolite concentrations. Averaging this quantity across the whole parameter space gives a measure of the relative performance of each pathway. Such a *relative* measure is desirable for a fair comparison, since the optimized flux through a pathway can vary dramatically for different points in the parameter space. An important factor in this is that our flux calculation, Eq. (5.11), gives the flux through a linear pathway as being proportional to the concentration of the initial metabolite $[S_0]$ and this is being sampled from over 3 orders of magnitude. Any comparison using a simple measure of total flux, summed over all parameter points, would be highly biased to the regions of high flux (and so mainly high $[S_0]$) . To avoid such a bias, the relative measure of comparative flux is used: the performance of each path is only being compared with alternative pathways under equivalent conditions.

5.6.3 Comparing the alternative pathways

Taking our 202 candidate 5-step glycolytic pathways and 300 5-step gluconeogenic pathways, we randomly select 10,000 points from the 10-dimensional parameter space as described in Section 5.6.1, calculating the CF of each path at every point. Averaging this quantity across the whole parameter space, we find that the real glycolytic and gluconeogenic pathways perform remarkably well compared to the many alternatives (Fig. 5.2, left panels). For glycolysis (Fig. 5.2A), the natural trunk pathway (in green, indicated by the arrow) outperforms all the alternative pathways except one. For gluconeogenesis, the two natural variants, with different cofactor couplings for the pyruvate to PEP step, are ranked first (the pps route, shown in red in Fig. 5.2C) and 8th (the ppdk route, shown in black in Fig. 5.2C). These results strongly suggest that the natural trunk pathways carry a high flux compared to alternatives. Because we average the CF over a rather arbitrarily chosen region of parameter space, the results of such a comparison will depend on the range of parameter space that is sampled.

To avoid the potential bias associated with averaging over parameter space, we investigated in detail how each pathway performed *as a function of the parameters*, i.e. in different regions of parameter space. This allows us to understand how the performance of a given pathway depends on the intracellular environment. Our results show that different candidate pathways perform best in different regions of the

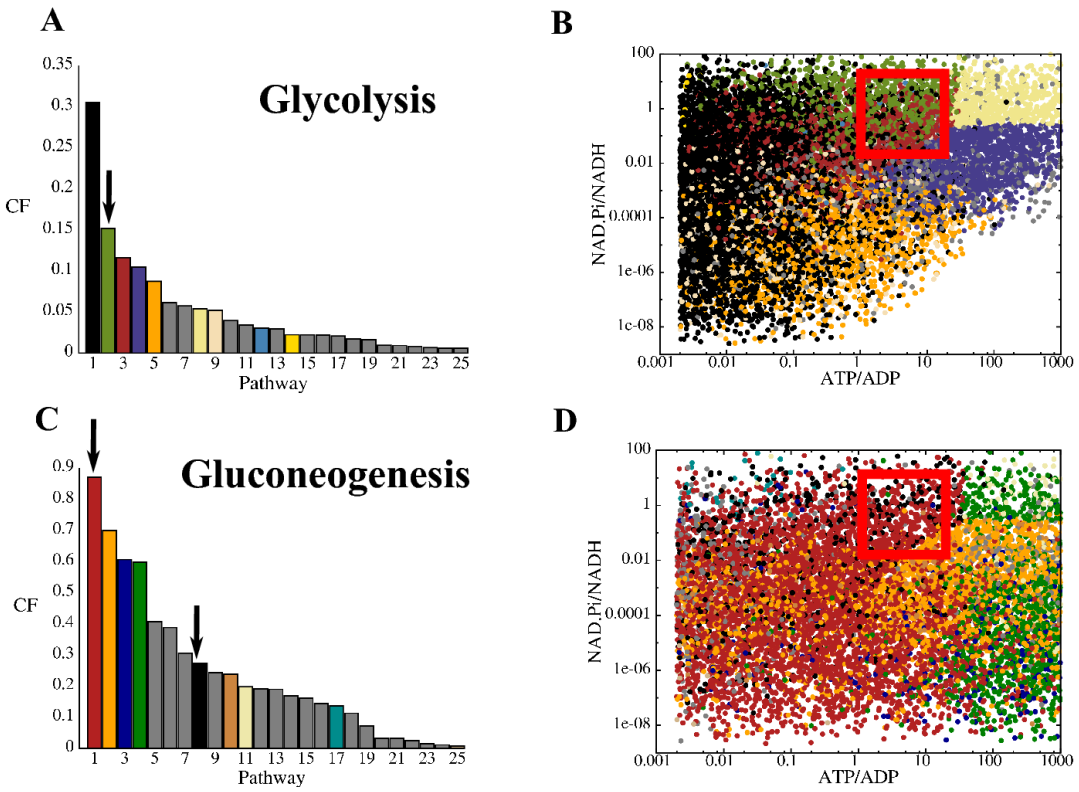


Figure 5.2 The real glycolytic and gluconeogenic trunk pathways represent maximal flux solutions. A, C: Alternative pathways generated in our analysis ordered by their comparative flux (see text), averaged across the whole 10-dimensional parameter space. The top 25 paths are shown. Panel A shows glycolytic pathways; the green bar (indicated by the arrow) is the real pathway. Panel C shows gluconeogenic paths; the red and black bars (indicated by arrows) are the real pps and ppdk routes. B, D: Pathway performance as a function of the intracellular environment. Each dot represents a randomly sampled point in parameter space; the colour of the dot indicates the candidate pathway which had the highest flux at that point in parameter space (colours as in panels A,C). The axes represent the redox state and energy state of the cell, $[NAD^+][Pi]/[NADH]$ and $[ATP]/[ADP]$. Panel B shows glycolytic pathways; real glycolysis (green) does best under typical physiological conditions (red boxes). Panel D shows gluconeogenic pathways; the two real routes (red and black) tend to perform best under typical cellular conditions.

parameter space. In particular, pathway performance is very sensitive to the cellular energy state, as measured by the ratio of the ATP and ADP concentrations, and redox state, as measured by the ratio $[NAD^+][Pi]/[NADH]$ (Fig. 5.2, right panels). Note that this latter choice just reflects the fact that it is this ratio of the external metabolites that appears in the real trunk pathway (in the glyceraldehyde 3-phosphate dehydrogenase reaction of EC 1.2.1.12, see Figure 2.8) and results in a clearer segregation of the pathways in Figures 5.2B and 5.2D. The red boxes in these panels indicate the typical physiological state of a cell [77, 81]: $[ATP]/[ADP]$ = 1-20 and $[NAD^+][Pi]/[NADH]$ = 0.01-10 M, based on $[NAD^+]/[NADH]$ = 10-100 and [Pi] = 1-100 mM. Focusing on the glycolytic pathways (Fig. 5.2B) we see that the natural trunk pathway (green dots) can outperform all the alternatives in the region of parameter space close to that found

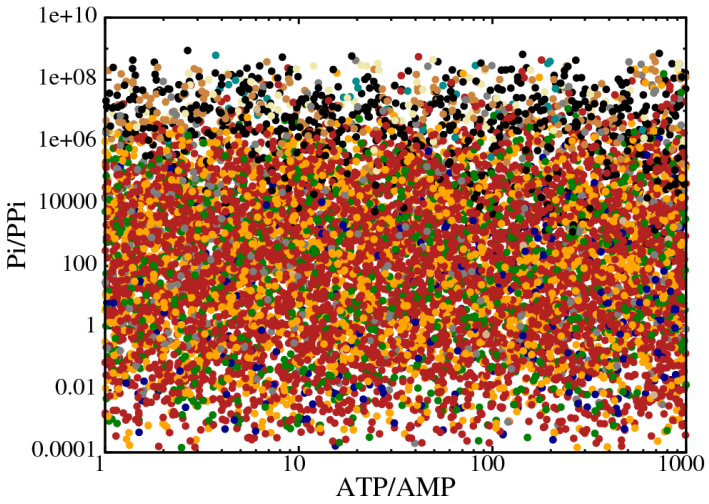


Figure 5.3 Scatter plot showing best performing gluconeogenic pathway in different regions of parameter space. Colours as in Figure 5.2, with this same data set used, but now plotting the [ATP]/[AMP] and [Pi]/[PPi] ratios. We find that the dominant factor distinguishing between the pps (red) and ppdk (black) gluconeogenic pathways is the ratio of orthophosphate to pyrophosphate. Note that although the ppdk path only appears for [Pi]/[PPi] ratios of greater than 10^5, this does not mean that the pathway is infeasible for values lower than this, just that for these lower values other pathways outcompete it with respect to the size of their flux.

in living cells (red box). This suggests that the glycolytic trunk pathway represents a maximal flux solution for the conversion of G3P to pyruvate, under typical intracellular conditions.

For the gluconeogenic pathways (Fig. 5.2D), a similar picture holds. Here, the two pathways found in nature, the pps-route (red) and the ppdk-route (black) introduced in Figure 2.8, both outperform the alternatives under typical physiological conditions (red box). In both these natural pathways, the transformation of pyruvate to PEP is coupled to the conversion of ATP to AMP. In the pps route, this releases orthophosphate (ATP+H_2O $\rightarrow$ AMP+Pi) via the phosphoenolpyruvate synthase enzyme, EC 2.7.9.2, $\Delta_r G^0 = -5.1$ kJ/mol. In the ppdk route, orthophosphate is consumed and pyrophosphate is released (ATP+Pi $\rightarrow$ AMP+PPi) by the pyruvate, phosphate dikinase enzyme, EC 2.7.9.1, $\Delta_r G^0 = +17.7$ kJ/mol. Figure 5.3 shows that the relative performance of these two natural variants in our analysis depends critically on the concentrations of orthophosphate and pyrophosphate. The reason for this is clear from consideration of the above external metabolite couplings. A high concentration of orthophosphate [Pi] will reduce the thermodynamic favourability of the pps reaction, while a high [Pi] (more specifically, a high [Pi]/[PPi] ratio) will in fact increase the favourability of the ppdk reaction.

5.7 Alternative pathways

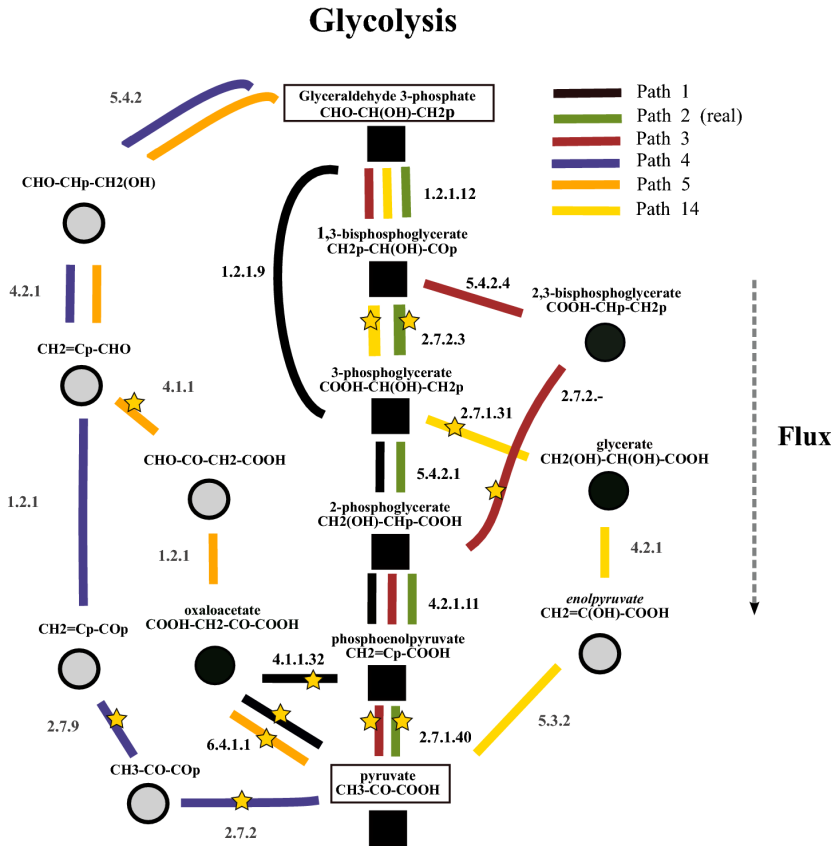


Figure 5.4 Schematic view of glycolytic pathways 1-5, 12 and 14 described in the text. Intermediate metabolites are shown by symbols; black squares indicate metabolites in the real trunk pathway, black circles indicate metabolites present in KEGG, and grey circles show metabolites generated by our program but not found in KEGG. Black EC numbers show reactions which exist in the KEGG database, grey numbers show those that do not, while yellow stars indicate ATP-producing reactions. Note that path 12, discussed in the text, follows the same route as the real path (path 2, green) with the exception that the final reaction is catalyzed by ppdk (EC 2.7.9.1), instead of pyruvate kinase (EC 2.7.1.40).

We will now look in more detail at some of the alternative routes generated in our analysis. Our analysis reveals several alternative pathways that can, under different intracellular conditions, outperform the true glycolytic and gluconeogenic pathways. For example, we see that in Figure 5.2B there are large patches where the real pathway (green) does not carry the highest flux; at low ATP/ADP ratios we find pathway 1 (black) dominating the scatter plot, while at high ATP/ADP pathways 4 (violet) and 8 (cream) perform best. Here we will investigate some of these potential alternatives, illustrated schematically in Figures 5.4 and 5.5 (for more details see Appendix A.7). These figures show the structure of a selection of our candidate pathways: glycolytic pathways 1-5, 12 and 14, together with gluconeogenic paths 1-5, all of which perform well in Figure 5.2 (or as we will see later, in Figure 5.7). Also shown are the EC numbers

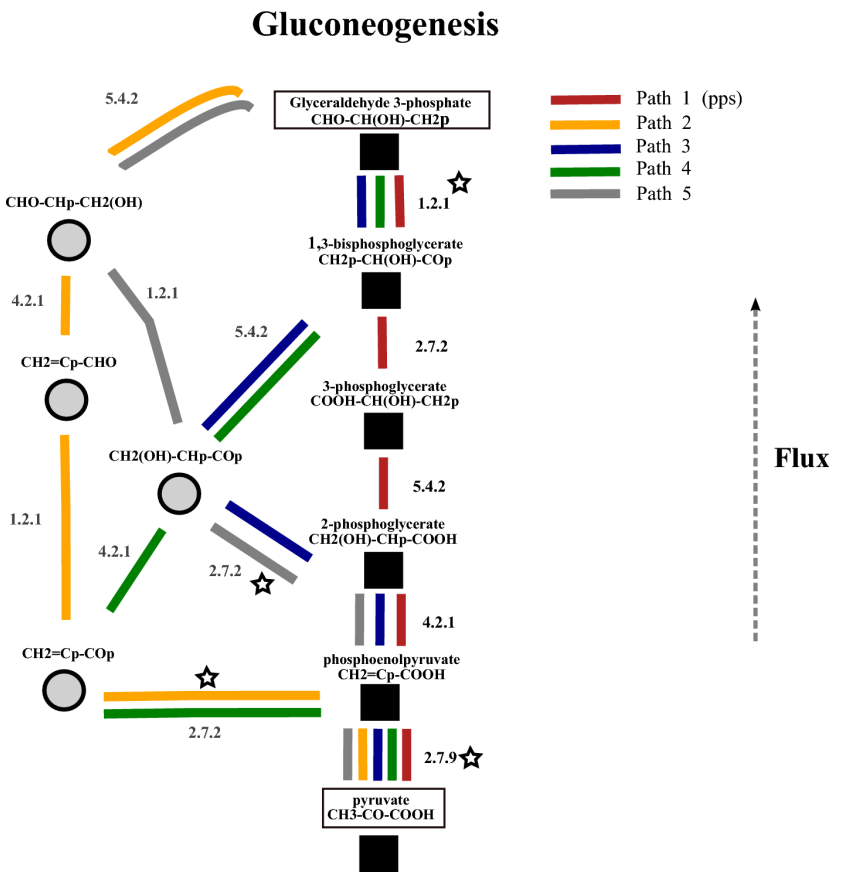


Figure 5.5 Schematic view of gluconeogenic paths 1-5 described in the text. Intermediate metabolites are shown by symbols; black squares indicate metabolites in the real trunk pathway, black circles indicate metabolites present in KEGG, and grey circles show metabolites generated by our program but not found in KEGG. Empty stars indicate ATP-consuming reactions.

of the corresponding reactions. Complete EC numbers (i.e. with all four component numbers) in black indicate that the enzyme appears in the KEGG database, having been observed in some organism. Incomplete grey numbers indicates their absence in KEGG – we assume that reactions not present in KEGG have not yet been shown to exist in nature.

5.7.1 Alternative glycolytic paths

In the glycolytic direction, path 1 (black) outperforms the real trunk pathway for low [ATP]/[ADP] ratios (Fig. 5.2B), resulting in its apparently better performance than the real pathway when averaged over the whole parameter space (Fig. 5.2A). This pathway differs from the real pathway in that it oxidizes G3P directly to 3-phosphoglycerate (3-PG) without the production of ATP (Fig. 5.4, compare black and green), resulting in it having a highly exergonic oxidation reaction ($\Delta_r G^0 = -42.6$ kJ/mol) as its first

step. Since in a linear pathway the initial reactions tend to exert the greatest control over the flux [52], this produces a large flux. However, pathway 1 is highly sensitive to the constraints on the intermediate metabolite concentrations. Because of its exergonic first reaction and subsequent less thermodynamically favourable reactions, it tends to accumulate high concentrations of downstream metabolites, causing it to be deemed infeasible in our analysis over a large part of the parameter space. Notice that all of the enzymes in pathway 1 are known to exist in biology, with the final two (PEP carboxylase, EC 4.1.1.32 and pyruvate carboxylase, EC 6.4.1.1) usually associated with a gluconeogenic role in eukaryotic cells, as in the standard textbook picture of gluconeogenesis [3], discussed in Section 5.10. It is possible that the problematic intermediate concentrations is the reason why this set of reactions has not been observed in nature.

Pathway 3 (red), is similar to the real glycolytic trunk path, except that 1,3 bisphosphoglycerate (1,3-BPG) is first isomerized (to 2,3-BPG) and then dephosphorylated (to 2-phosphoglycerate 2PG, with ATP generation), rather than being first dephosphorylated and then isomerized as in the real pathway. The large free energy change for the isomerization ($\Delta_r G = -28$ kJ/mol) means that pathway 3 can in principle carry a higher flux than the real pathway, but because of this large drop in free energy it tends to accumulate high concentrations of 2,3-BPG and therefore becomes less favourable at high concentrations of the starting metabolite G3P. Interestingly, a similar pathway exists in red blood cells, where 2,3-BPG is produced from 1,3-BPG via the Rapoport-Leubering shunt [82]. In these cells, however, the 2,3-BPG is hydrolysed to either 3-PG or 2-PG without ATP generation, thus sacrificing one ATP compared to the usual glycolytic pathway [82]. It is tempting to hypothesize that nature is forced to sacrifice an ATP molecule when using this shunt, to prevent the buildup of 2,3-BPG, which is already the most concentrated organophosphate in erythrocytes.

Pathway 4 (violet) outperforms the real pathway at high ATP concentrations (Fig. 5.2B); this is because its ATP producing steps are at the end (4th and 5th steps), in contrast to the real pathway where ATP is produced in the 2nd and 5th steps. Because later steps in a linear pathway tend to have less impact on the flux [52], this makes pathway 4 more tolerant of high ATP concentrations than the real pathway. Similarly, pathway 5 (orange) differs from the real pathway in that its oxidation step is moved to the end (the 4th step rather than the first as in the real pathway); this makes the performance of pathway 5 more thermodynamically tolerant of reducing conditions than the real pathway.

Two other alternative glycolytic trunk pathways, 12 and 14, are also of interest and are

shown in Fig. 5.4, even though they do not perform particularly well in our flux analysis of Figure 5.2. Pathway 12 is identical to the real glycolytic trunk pathway, except that in the final step, ATP is generated from AMP and pyrophosphate (via ppdk; pyruvate, phosphate dikinase) rather than from ADP, i.e. it is the exact reverse of the natural ppdk gluconeogenic pathway. This could be seen as resulting in a greater energetic yield than the natural glycolytic pathway, since a low energy AMP is being used directly to generate an ATP molecule, rather than being salvaged via the usual adenylate kinase reaction (EC 2.7.4.3) which involves the consumption of another ATP molecule [3]. This actually corresponds to a glycolytic variant observed in some anaerobic eukaryotes, in which glycolysis is the primary mode of ATP production [83]. Its flux and feasibility depend on the cellular [AMP] and [PPi] concentrations, which is not the case for the real pathway.

Pathway 14 converts 3-PG to glycerate, and then to enolpyruvate. Enolpyruvate is actually unstable in solution and converts spontaneously to pyruvate in a tautomerization reaction. (For a discussion of the stability of the compounds generated in our network, see Section 5.13.) It is likely therefore that this pathway would only require 4 enzymes instead of 5 as in the real pathway, whilst still producing 2 ATP molecules. This could also lead to a higher flux of ATP than suggested by our analysis, due to the rapid spontaneous depletion of the enolpyruvate "pulling" flux through the pathway. Moreover, all but one of the enzymes needed for this pathway are known to exist in nature. The only exception is the enzyme which catalyzes the dehydration reaction of glycerate, which is not present in KEGG. It is not unreasonable to suppose that such a simple dehydration could result from the uncharacterized, non-primary action of an existing enzyme.

5.7.2 Alternative gluconeogenic paths

In the gluconeogenic direction (Fig. 5.2, C and D; Figs. 5.5), pathways 1 (red) and 8 (black) correspond to the two prokaryotic trunk pathway variants found in nature, the pps and ppdk routes. It is apparent that the gluconeogenic alternatives are less diverse than those in the glycolytic direction. For instance, alternative pathways 2-7 all contain the same set of reaction types as the real pps pathway (Figure 5.5 and A.5), but carry out these reactions in different orders, making use of different intermediate metabolites. This affects their relative performance in different regions of the parameter space. For example, pathway 2 (orange) and pathway 4 (green) both differ from the real pps pathway in that ATP is consumed in the first two reactions (rather than in

the 1st and 4th reactions as in the real pps pathway, Fig. 5.5). This makes their flux more sensitive to the ATP concentration, explaining why they dominate at high ATP/ADP ratio (Fig. 5.2D). It would seem then, that there is a greater restriction on the chemistry facilitating thermodynamically feasible gluconeogenic pathways than on the glycolytic pathways. This might explain why nearly all of the variations found in the trunk pathway, discussed in Section 5.10, appear to be in the glycolytic direction, with gluconeogenesis being far more highly conserved [30, 31, 36].

An important point is that although certain sets of enzymes may be known to exist in nature, i.e. present in the KEGG database, what is actually important is whether or not they all exist in the same organism, and whether or not the genes controlling them are ever co-expressed. This requires organism-specific genome information to be available, in databases such as KEGG, along with gene expression data under specific experimental conditions. There are additional complexities too. The presence alone of a certain gene tells nothing of its product's function: genome analysis can be inconsistent with biochemical data since, from just a sequence comparison, there exists no definitive method for predicting the substrate specificity of enzymes belonging to large families [84].

5.8 Intermediate metabolite concentrations matter

We now investigate the role of the constraints on the metabolite concentrations imposed when evaluating pathway performance.

5.8.1 Effect of removing restrictions on metabolite concentrations

Imposing constraints on the intermediate internal metabolite concentrations is crucial to our analysis. Repeating our analysis without these constraints produces a very different outcome (Fig. 5.6). In the glycolytic direction, if the metabolite concentrations are unconstrained, glycolytic path 1 (black in Fig. 5.2A and B), which has a highly exergonic first reaction, produces the highest flux across the entirety of the parameter space, outperforming the real glycolytic pathway. Thus optimizing the flux alone, in the absence of metabolite concentration restrictions, cannot explain the selection of the real pathway by evolution. The fact that our results are strongly affected by these constraints highlights the importance of considering metabolite concentrations when using methods such as flux balance analysis [85] to study metabolic networks. Often these methods, though widely used, do not include any information on metabolite

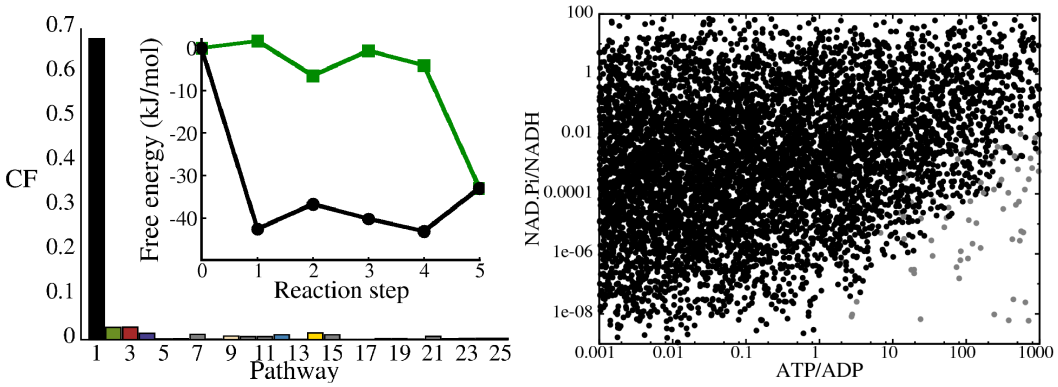


Figure 5.6 Constraints on metabolite concentrations are crucial. Repeating our analysis without these constraints, in the glycolytic direction, pathway 1 (black in Fig. 5.2A and B) dominates over the whole parameter space. The inset to the left panel shows the thermodynamic profile (cumulative standard free energy change) for pathway 1 (black) alongside that of the real path 2 (green).

concentrations or thermodynamics.

Interestingly, all of the enzymes required for glycolytic pathway 1 (Fig. 5.4) exist in nature, in various organisms, although this route is not known to be used as a glycolytic pathway. Our analysis hints that constructing this pathway could provide an interesting target for investigation in synthetic biology. This pathway produces all of the precursor metabolites defined in [29] and hence its high flux could in principle lead to an increase in the rate of production of these precursors as well as ATP. It possesses the most efficient thermodynamic design of all the pathways found and so could, in principle, provide a way to accelerate the growth of organisms useful for biotechnological applications, if suitable branching pathways were provided to prevent the buildup of downstream metabolites.

5.8.2 Effect of imposing tighter restrictions

The analysis was also repeated imposing a *narrower* concentration range of 10^{-7} to 10^{-2} M for the intermediate metabolite concentrations (rather than 10^{-9} to 5×10^{-1} M as in Section 5.6.3). This range is closer to the concentration ranges actually observed in cells and to those used in other works [34, 77, 86, 87]. Under these constraints, we find that the performance of the real glycolytic pathway is even better compared to the alternatives. Indeed, the real glycolytic path outperforms the alternatives over a wider range of parameter space. The restricted concentration range has only a small effect on the results for the gluconeogenic paths in the physiological region (Fig. 5.7). This further supports the picture emerging from our analysis; in the presence of reasonable limits on intermediate metabolite concentrations, the real glycolytic and gluconeogenic

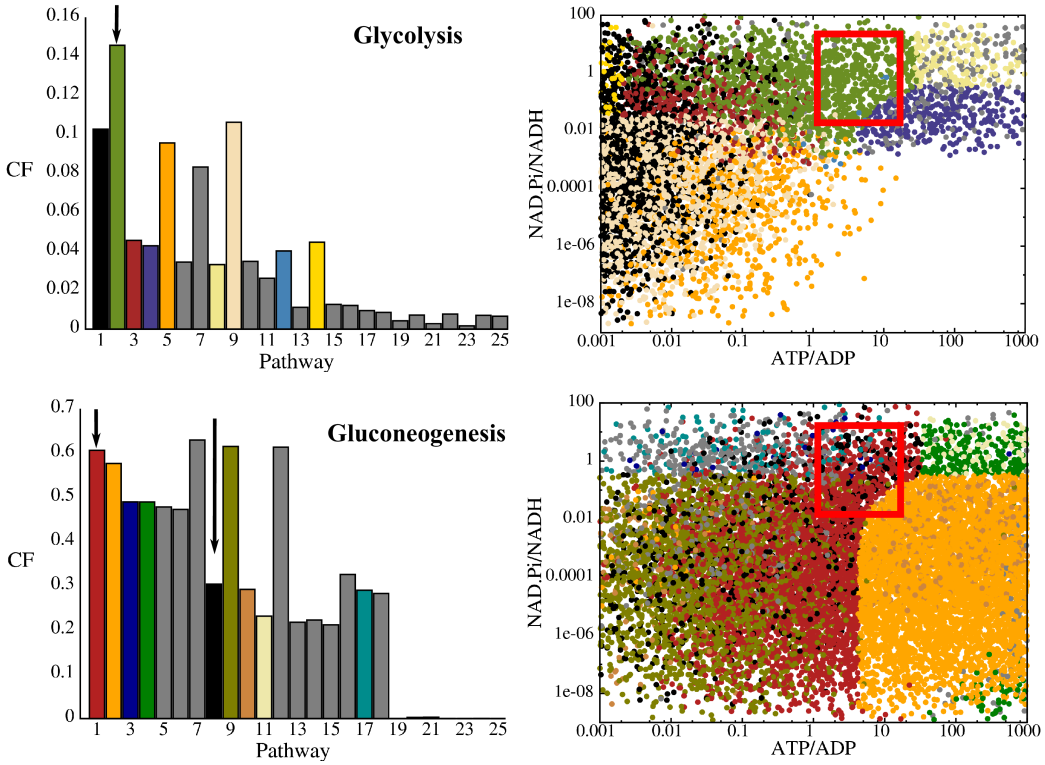


Figure 5.7 The results of the comparison, where now the lower and upper limits on the intermediate concentrations are 10^{-7} M and 10^{-2} M respectively. Imposing more realistic restrictions serves to improve the performance of the real paths relative to the alternatives. Colours as in Fig. 5.2. Arrows point at real glycolytic pathway (green) and gluconeogenic pps (red) and ppdk (black) pathways.

paths outperform the alternatives.

In summary, the overall picture so far is clear. With the goal of maximizing ATP production rate, the real glycolytic trunk pathway emerges as the best solution from a large number of alternative pathways, over a range of physiological conditions. In the region of parameter space corresponding closely to that typically found in cells, this pathway outperforms all others (Figure 5.2). Applying tighter biophysical constraints on the metabolite concentrations results in a further improvement in the performance of this natural pathway (Figure 5.7). Assuming maximal flux as the goal, in the gluconeogenic direction the common pps pathway performs best across the parameter space, while both the pps and ppdk routes appear as optimal solutions in the physiological region of parameter space (Figure 5.2), with the [PPi] concentration determining which of these performs best (Figure 5.3).

5.9 Robustness of results

5.9.1 Our results are robust to small changes in free energy

Using the group contribution method, the typical error in our calculation of the free energy of formation $\Delta_f G$ for a given molecule was 3.84 kJ/mol (Section 4.5.1). To check the robustness of our results to such errors, our entire analysis was repeated using $\Delta_f G$ values computed using different sub-sets of these training molecules, consisting of 90% and 80% of the molecules from the original set, chosen at random. Taking the 80% set, this resulted in the error rising to 5.82 kJ/mol, however the qualitative results using such networks were identical to those obtained from the full set of training compounds. For example, the top 25 glycolytic pathways obtained from the reduced set contained 23 out of the 25 pathways from the original analysis, containing all of the alternative pathways discusssed in Section 5.7. Although the rank order of the pathways in terms of comparative flux across the whole parameter space did differ between sets, the top 3 performing pathways were the same in both cases.

5.9.2 Flat enzyme distribution

In the flux calculation we used an optimization procedure to choose the concentrations of enzymes along the pathway that maximized the flux, with the restriction of a fixed total enzyme concentration and the requirement that all metabolite concentrations fall within some defined range. We have checked the outcome from relaxing this optimization, instead using a flat enzyme distribution, i.e. all enzymes are present at an equal concentration. Figure A.3 shows that our general results are unchanged by such a choice; the real pathways perform well when averaged over all parameter space and perform best in the region corresponding to typical cellular conditions.

5.9.3 Do the results hold for more realistic kinetics?

In the flux calculation we have assumed the rate of the reactions correspond to the rate of a "perfect enzyme". The use of this rate, along with the method of optimizing the distribution of enzyme concentrations, provides an upper limit; the maximum flux a pathway *could* sustain. However, this perfect enzyme rate lacks a feature that is of physical importance. It does not incorporate any form of substrate saturation, as in the Michaelis-Menten rate of Eq. (2.8), since the rate is linear with respect to

the substrate concentration. It also neglects any inhibition of the reaction through high product levels. This may not be a realistic assumption for many reactions under physiological conditions [77]. In reality, it has been measured that in glucose-fed *E. coli.*, approximately 59% of measured metabolites have a concentration 10-fold higher than their K_M value, demonstrating a trend towards saturation of most enzyme sites [77]. However it is also important to note that the trunk pathway of lower glycolysis was an exception in this study, where substrate concentrations were very similar to their K_M values. In addition, the cofactor molecules ATP and NAD^+ were consistently saturated, exceeding their K_M values more than 10-fold, implying that these molecules do not appear to limit enzymatic activity. This observation in fact holds over the majority of central carbon metabolism, partly justifying our method of incorporating these cofactors into the flux calculation solely through their modification of a reaction's equilibrium constant, rather than affecting any kinetic terms.

In the next chapter we will discuss possible ways to improve on the assumption of perfect enzymes. By using reversible Michaelis-Menten kinetics, we will see that using these more realistic enzyme kinetics can actually improve the performance of the real trunk pathway with respect to the alternatives.

5.10 Some natural variants do exists

Although the trunk pathway is highly ubiquitous in nature, some variants do exist. Here we analyze these and compare them to the alternative pathways that emerge from our analysis. We also assess what evolutionary goal functions (if any) these variant pathways might be optimizing.

5.10.1 Glycolytic variants

One variation can be found in plants and certain photosynthetic organisms [12, 88] and involves the bypass of 1,3-BPG through the essentially irreversible direct oxidation of G3P to 3PG as seen in the first step of glycolytic path 1 (Figure 5.4, EC 1.2.1.9). The use of this bypass results in no ATP production in glycolysis, but it is possible that organisms which do not rely on the degradation of sugars for energy, i.e. phototrophs, can sacrifice the glycolytic generation of ATP so as to retain a larger chemical motive force for higher fluxes, using glycolysis solely for biosynthetic purposes. This shortcut is also present in some bacteria and archaea [36, 38, 89].

Hyperthermophilic bacteria and archaea are known to use another distinct variation. At high temperatures metabolite stability, especially for phosphorylated compounds, can be very low (Section 5.13) with 1,3-BPG being one of the least stable metabolites [90]. The nonphosphorylating variant of the Entner-Doudoroff pathway present in hyperthermophiles almost entirely avoids phosphorylating the intermediates [89]. The unphosphorylated version of KDPG (KDG) is instead cleaved to form glyceraldehyde and pyruvate, with glyceraldehyde being dehydrated to glycerate before being phosphorylated to 2-PG and re-entering the trunk pathway. Again, there is no ATP production when using this path and so its existence must be explained by principles other than ATP production rate.

Finally, another atypical glycolytic pathway has been found in the bacterium *Clostridium thermocellum* [91]. This organism was found to lack a pyruvate kinase enzyme for the glycolytic conversion of PEP to pyruvate. The authors determined that the most likely route for this to occur was through the conversion of PEP to oxaloacetate, oxaloacetate to malate and finally malate to pyruvate and involves an additional two reactions over the usual pathway. This pathway is particularly interesting since it can actually be used to generate an extra NADPH than the more common trunk pathway, and generates GTP as well as ATP. It would be interesting to investigate if our analysis selects this pathway when considering a different goal function, one that incorporates the production of an additonal reducing agent. Perhaps a pathway containing 7 reactions is the minimal required to achieve this?

5.10.2 Gluconeogenic variants

In this work our analysis focused on prokaryotic metabolism. In prokaryotes, the gluconeogenic trunk pathway has 5 steps, with the initial step being the direct transformation of pyruvate to PEP coupled to the conversion of ATP to AMP (via the pps or ppdk enzymes). The initial steps of the textbook picture of eukaryotic gluconeogenesis is slightly different (Figure 5.8 inset). Here, pyruvate is converted to PEP in two steps: pyruvate carboxylase (EC 6.4.1.1) converts pyruvate to the 4-carbon molecule oxaloacetate, which is then simultaneously de-carboxylated and phosphorylated to form PEP by phosphoenolpyruvate carboxykinase (EC 4.1.1.32). The first step is driven by the dephosphorylation of ATP to ADP while the second step is driven by GTP. This pathway therefore has 6 steps and in total consumes 3 "energy units" (2 ATPs are converted to ADP while 1 GTP is converted to GDP).

Because this is a 6-step pathway, it was not considered in our main analysis. However,

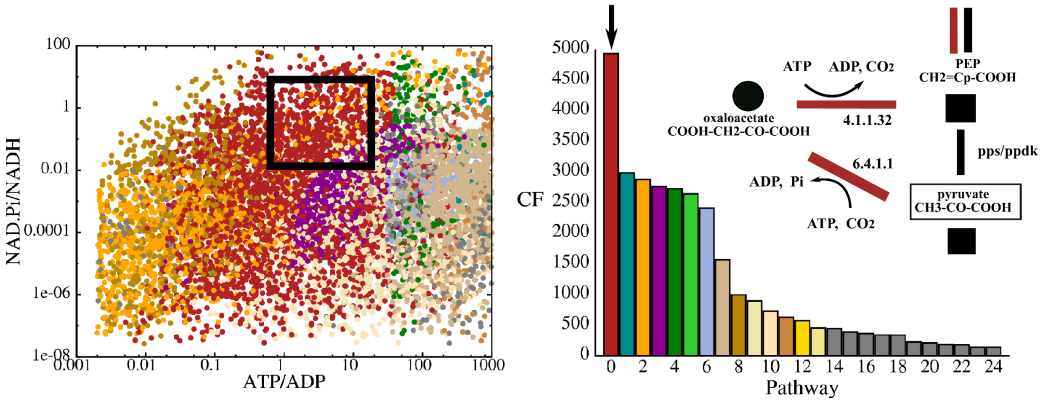


Figure 5.8 Comparison of gluconeogenic pathways without dikinase reactions (i.e. not allowing the conversion of ATP to AMP). The textbook 6-step gluconeogenic pathway (with ATP substituted for GTP) is now shown in red (indicated by the arrow) and the first two reactions, converting pyruvate to PEP, are shown in the inset. This pathway outperforms the alternatives when its comparative flux is averaged across the whole parameter space, and also performs best in the physiological region of parameter space.

when we repeat our analysis in the absence of dikinase reactions (i.e. not allowing the conversion of ATP to AMP), we find that 5-step gluconeogenic pathways are no longer feasible under typical physiological conditions since they can now only consume an equivalent 2 ATP molecules (both coupling to ATP $\rightarrow$ ADP). Figure 5.8 shows that in this case, the textbook pathway outperforms all alternatives under physiological conditions (note that since we do not include GTP in our analysis, the "natural" pathway here actually has 3 ATP $\rightarrow$ ADP conversion steps, however the energy released in the hydrolysis of ATP and GTP are very similar.

It is not clear why this pathway seems to be predominantly present in eukaryotic organisms, especially higher animals which are often lacking both the pps and ppdk genes[1]. If we perform a comparison between this length 6 gluconeogenic pathway and the length 5 pathways, the length 5 pathways are usually able to sustain a higher flux under typical physiological conditions, which poses the question of why use this oxaloacetate shunt? One possible benefit is that it serves to compartmentalize glycolysis and gluconeogenesis and facilitate the transport of NADH between the cytoplasm and mitochondria. In many eukaryotic cells, the conversion of pyruvate to oxaloacetate in the first step of glucuoneogenesis is carried out in the mitochondria (EC 6.4.1.1), while the other gluconeogenic enzymes reside in the cytosol [3]. Since the mitochondrial membrane does not contain a transporter for oxaloacetate, oxaloacetate is first reduced to malate (coupled to the conversion of NADH to NAD^+) and malate is transported into the cytosol and subsequently reoxidized to oxaloacetate (converting a NAD^+

[1]It is interesting to note that no mammals are in fact known to contain the gene for ppdk, which makes it an interesting target for drug development in certain disease-causing parasites whose growth depends on it [92].

to a NADH) before following the gluconeogenic pathway. This has the effect of maintaining the oxidizing conditions (higher NAD^+/NADH ratio) in the mitochondria favourable for glycolysis, while providing NADH for the reducing conditions that favour gluconeogenesis in the cytosol.

5.11 Can we do better with longer paths?

The glycolytic pathways compared in Figure 5.2 were restricted to length 5 and, as a result of the biochemistry of the network, could only produce 2 ATP molecules. It is interesting to speculate whether more ATPs could be made with longer pathways and if so, how well these could perform? If we no longer restrict the analysis to minimal length paths we find that, at least biochemically, there exist pathways between G3P and pyruvate that allow the production of more than 2 ATP molecules. Without considering the thermodynamics, extending to length 6 pathways reveals 41 paths that can create 3 ATPs, while enumerating paths of length 7 finds 19 pathways that can produce 4 ATP molecules. However, there is a finite amount of energy available for ATP production, which will depend on the cellular environment. Figure 5.9 shows a comparison of the length 6 pathways generated in our network that produce 3 ATP molecules, performed using the same sampling procedure and parameter ranges as in Figure 5.2. It can be seen that in general, for these pathways to be thermodynamically feasible the ATP/ADP ratio must be less than 1. The black pathway here, that performs best for the highest ATP/ADP values, actually corresponds to the exact reverse of the eukaryotic gluconeogenic pathway just discussed. However note that in the physiological region of parameter space it is not thermodynamically feasible.

5.12 Releasing the end-point from pyruvate

In the present study the analysis has been limited to pathways that start and end at G3P and pyruvate. However, there is no particular reason to think that evolution was restricted to these choices. Relaxing this requirement, and allowing further metabolism of pyruvate, would certainly lead to many more alternative pathways for the generation of energy, and for biosynthesis.

In biological cells, the pyruvate that is generated in glycolysis broadly has two fates. Under anaerobic conditions an organism will perform fermentation, in which pyruvate may be converted to a range of products that depends on the organism and the

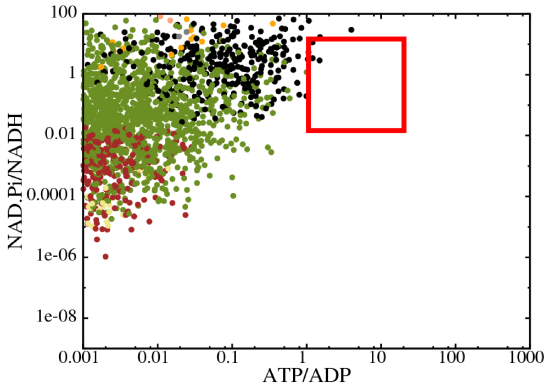


Figure 5.9 Comparison of length 6 glycolytic pathways that produce 3 ATP molecules (excluding dikinase reactions). For these pathways to be thermodynamically feasible a low ATP/ADP ratio is required. The black pathway corresponds to the reverse of the eukaryotic gluconeogenic pathway discussed in Section 5.10.2. In the physiological region of parameter space (red box) none of these pathways are feasible.

environment. Under these conditions the cell must maintain a state of redox balance so that there is no build up of the reducing agent, NADH. This requires that the NADH produced in the trunk pathway must be converted back to NAD^+ in the subsequent metabolism of pyruvate. In an organism operating in an aerobic environment this restriction does not apply, since it is possible that the NADH produced in glycolysis could be used to generate ATP through oxidative phosphorylation (Section 2.1.3). If we are to investigate pathways that extend beyond pyruvate, it might then be useful to distinguish between those that satisfy redox balance, and those that do not. For a pathway that does not obey redox balance and thus produces NADH, we assume that the organism is growing aerobically and can convert excess NADH into ATP through oxidative phosphorylation: any excess NADH molecules are counted as 2.5 ATP molecules, which is the ratio typically obtained in aerobic metabolism [3].[2]

We also note that there is no need for the end point of the glycolytic pathway to be charged, since this metabolite could be a waste product, such as ethanol CH_3-$CH_2(OH)$, which does not need to be retained by the cell. We repeated our enumeration of pathways from G3P but this time without restricting the end point to pyruvate. Many new pathways were generated, which we group in to categories as discussed below. The following sections enumerate potential pathways using the same set of conditions as defined in the caption of Figure 5.1.

[2]It is interesting to note that some cells do maintain and even prefer fermentation in aerobic conditions. In the presence of oxygen, and as long as a supply of sugars is available, yeast cells will preferentially use fermentation instead of aerobic respiration, known as the crabtree effect. The Warburg effect is a similar process that appears in many types of cancer cells, with the advantage believed to be that the higher rate, but less efficient, production of ATP in fermentation enables the rapid proliferation of these cells [93].

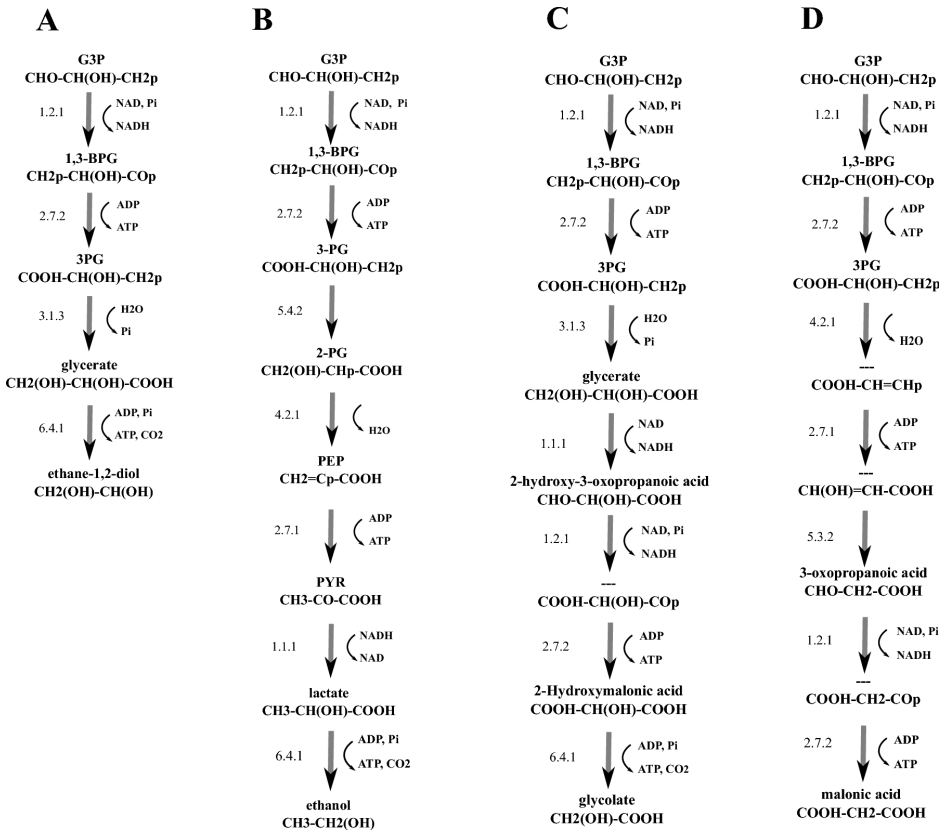


Figure 5.10 Pathways not constrained to end at pyruvate can produce a higher energy yield. Path A is an example of a length 4 pathway that can produce 2 ATP molecules from G3P. Paths B-D show that when the end-point is not restricted to pyruvate, pathways exist that can produce 3 ATP molecules as well as additional NADH: paths C and D produce 3 and 2 NADH molecules respectively.

Length 4 pathways producing 2 ATPs

Our first observation when performing this analysis is that shorter pathways than the real glycolytic trunk pathway exist that can produce 2 ATP molecules. One such pathway is shown in Figure 5.10A, terminating at the compound ethane-1,2-diol, $CH_2(OH)$-$CH_2(OH)$. Ethane-1,2-diol is weakly toxic and has a high affinity for water, making it a useful desiccant. It is however present in the KEGG database (compound c01380), appearing in the glyoxalate metabolism of some organisms. Pathway A in Figure 5.10 converts 3-PG to glycerate through a hydrolysis reaction and then uses an enzyme from class EC 6.4.1, coupling a decarboxylation and the formation of an ATP molecule in the final step. This route clearly misses out PEP and pyruvate (which may be needed by the cell as biosynthetic precursors [29]) but is interesting in that it produces 2 ATP molecules, using only 4 reactions, and can sustain a higher optimized-flux than the real trunk path under physiological conditions.

The fact that such paths are not apparently found in nature, even though our analysis suggests that they would perform well from a flux point of view, demonstrates that ATP production is likely not the sole function of lower glycolysis, but that other constraints are also imposed, such as integration with the rest of metabolism and the production of biosynthetic precursor metabolites.

Pathways that produce 3 ATPs at physiological conditions

When enumerating pathways with no fixed end-point, we also found a number of pathways of length 7 that are thermodynamically capable of extracting 3 ATP molecules from G3P under the typical physiological conditions defined in Figure 5.1. Figures 5.10B-D show some of these paths. Path B is a novel redox-balanced fermentation to ethanol through lactic acid. The two most common routes of fermentation involve either the reduction of pyruvate to lactic acid, or the decarboxylation of pyruvate to acetaldehyde followed by the reduction of acetaldehyde to ethanol. Pathway B is the common lactic acid fermentation with an additional step in which the lactic acid is decarboxylated to form ethanol with the coupling to ATP production from ADP and an orthophosphate. Such a reaction does not exist in the KEGG database.

Another route enumerated is pathway C of Figure 5.10. This is an example of a pathway that does not maintain redox balance but in fact produces 3 NADH molecules in addition to the 3 ATP molecules, giving the potential to produce a much higher glycolytic yield of ATP under conditions of respiration ($3 + 3 \times 2.5 \approx 10.5$). This

pathway makes use of several alternative compounds and has glycolate $CH_2(OH)$-COOH as its end product. The final route shown in Figure 5.10D produces 2 NADH molecules in addition to the 3 ATPs. It follows the first two steps of the trunk pathway before deviating to end at malonic acid. Note that neither of these pathways pass through pyruvate; despite their larger yield of ATP, they may not satisfy other basic requirements of the cell.

5.13 Are the chemical compounds generated by our network compatible with life?

Many of the compounds generated in our study do not have well-documented chemical properties and may never have occurred in biology. This raises the question of whether they would be stable in a biological cell, and whether they might in fact be toxic to a cell.

5.13.1 Stability

When considering the stability of our putative metabolites, it is important to bear in mind that some of the metabolites that actually are used in nature are relatively unstable in solution, spontaneously degrading into potentially unwanted side products, especially at high temperatures. To our knowledge, this issue has only been investigated experimentally for the thermophile *Sulfolobus solfataricus* [90]. Here it was shown that during gluconeogenesis, at a temperature of 70^oC, as much as 50% of the total carbon flow could be lost due to thermal degradation.

Some of the compounds used in our alternative pathways do raise concerns as to stability. In particular, glycolytic pathway 14 contained the metabolite enolpyruvate – a compound that is known to not exist freely in solution, undergoing a spontaneous tautomerization to pyruvate. It is generally the case that enols (compounds containing a hydroxyl group attached to a C=C double bond) are unstable in solution and undergo this tautomerization reaction (a keto-enol isomerization) to a carbonyl, since the keto form is generally more stable than the enol tautomer[3]. A similar argument can be made for ketenes (C=C double bond in which one is a carbonyl group) since these also

[3]In future works, any enol product of a reaction could be replaced with its keto tautomer so as not to assign any enzyme reaction kinetics to such a spontaneous reaction. Care would have to be taken with the reverse reaction however, and it is true that enzymes do exist that catalyze these tauomerizations (EC class 5.3.2), so it might not be an appropriate substitution in every case.

spontaneously convert, in a reaction with water, to a carboxylic acid [12]. However, none of our relevant alternative pathways contain ketenes.

5.13.2 Toxicity

With regard to toxicity, while it cannot be ruled out that some of our compounds are toxic, one can only make use of observed chemical principles and hope that they hold for similar molecules. For example, to a greater or lesser extent, all carbonyl groups present in aldehydes and ketones react with DNA, proteins and other macromolecules of metabolism, inducing mutations and inactivating enzymes [12]. Carbonyls can exhibit a range of activities however, and carbonyl-containing compounds such as glucose, G3P and DHAP clearly do not impose any serious problems for life, though they do have a measurable toxic effect [94]. Metabolites that contain two carbonyl groups, either on adjacent or next-adjacent carbons, tend to show a very high reactivity and can cause damage even at low levels. An example of this is methylglyoxal, the toxic product of the uncatalyzed dephosphorylation of DHAP. But despite its high toxicity, this metabolite does appear in real cells in the pathway known as the methyglyoxal shunt. It is for these reasons that we do not attempt to exclude any compounds from the network on the grounds of toxicity, but prefer to deal with them on a case-by-case basis. As it turns out, in all of the alternative pathways discussed in this chapter, only one compound is identified that contains two adjacent carbonyl groups: the 4-carbon molecule CHO-CO-CH2-COOH found in glycolytic pathway 5.

5.14 Conclusion

In this chapter we have studied the trunk pathway of glycolysis and gluconeogenesis. By utilizing our network of reactions, containing many that are not known to exist in biology, we have enumerated all possible biochemical alternatives to the trunk pathway. By comparing their fluxes under a range of physiological conditions we have shown that the real pathways perform well when averaged over a wide range of conditions and perform best in the region of parameter space corresponding closely to those typically found in cells. We also showed that this outcome is dependent on putting realistic constraints on the metabolite concentrations – without these, certain alternatives can dominate the comparison.

Our methodology is also able to suggest pathways that may be worth studying experimentally. Investigating the alternatives in more detail showed that some of the

alternatives contained enzymes already known to exist in nature, raising the question if any of these pathways could be operating in biology but have just yet to be observed.

In the next chapter I will present preliminary work on extending this methodology to include non-linear pathways (and thus the complete glycolytic pathway), and to improve on the "perfect enzyme" assumption through the use of generic reaction rates. I will also outline how this can be implemented in an evolutionary simulation that can be used to explore the biochemical space when the number of pathways is so great that enumeration may be computationally infeasible.

Chapter 6

Taking our analysis further: Simulating the evolution of metabolic pathways

6.1 Introduction

The flux calculation employed in the previous chapter arguably represents the maximum flux a linear pathway can sustain – it assumes that all enzyme catalyzed reactions are limited only by the diffusion of the substrates (in addition to allowing the distribution of enzymes along each pathway to be optimized so as to maximize the flux). This assumption may be unrealistic for most enzymes. In this chapter we extend our calculations, relaxing the assumption of perfect catalysis. Furthermore, the analytic expression for the flux, Eq. (5.7), also assumes that all pathways are linear. Not only could branched pathways exist in principle, but they actually become relevant if we take into account the entire glycolytic pathway. The presence of the cleavage reaction in glycolysis (step 4, Figure 2.8) means that the full glycolytic pathway is in fact branched. Since this reaction is bimolecular in the reverse direction, the analytical expression for the flux will be a quadratic equation [52]. Extending our analysis to allow pathways of arbitrary structures will prevent any analytical solution of steady state flux. Instead, the set of equations governing the dynamics of the network will need to be solved numerically (Section 6.2.1).

In addition, if one were to repeat the flux derivation for linear paths presented in the

previous chapter, but used the reversible form of the Michaelis-Menten equation in place of the perfect enzyme rate equation, the resulting expression for the flux would be a degree-n polynomial equation [95], where n is the number of reactions in the pathway. It is the linear form of the rate equation, Eq. (5.1), that enables an analytical expression for the flux to be obtained. However, this assumption of a linear rate equation corresponds to the condition that the metabolite concentrations are much lower than their corresponding Michaelis-Menten constants (taking $[S] \ll K_S$ and $[P] \ll K_P$ in Eq. (2.10)). This is certainly not always the case in metabolism, where metabolites are often present at concentrations an order of magnitude higher than their K_M values [77].

To address all these issues, we seek to develop a more general way to compute pathway fluxes. However in doing this many questions arise. For instance, what is the most appropriate form of enzyme mechanism to use and how should one go about choosing the parameter values? In this chapter I will describe preliminary work that is aimed to address this.

Such calculations will have two uses. Firstly, it is important to ask whether the results of the previous chapter hold when alternative assumptions are made about the enzyme kinetics. Secondly, we wish to incorporate these more realistic kinetics into an evolutionary simulation. Such a simulation could be used to explore the biochemical space available to pathways or networks of any structure, and to situations where the number of biochemical solutions may be so high as to prohibit enumeration. For example, beginning with glycolysis and the TCA cycle, it would be fascinating to test how many ways a cell can produce an equivalent or greater amount of energy. Even when applying restrictions, such as the requirement that the set of known essential metabolites must be produced, it is likely that many possibilities are open to a cell. Exploring these alternatives could shed light on evolutionary principles, as well as highlighting reactions and pathways that might already be operating unknown in biology.

6.2 Relaxing our assumptions on the flux calculation

In Section 2.7.1 we discussed the complications arising in reactions with multiple substrates; small changes in our assumptions about the enzyme mechanism can give rise to subtly different forms for the reaction rate. However, it is often the case that such precision is unnecessary in metabolic models, given that in certain parameter regimes

the different rate equations can behave similarily [39, 42, 96]. Since there are usually large uncertainties in the kinetic data, it has made sense to develop phenomenological "generalized" rate laws.

Recall the common modular rate law introduced in Section 2.7.1 and derived in Appendix A.1.3:

$$v(\mathbf{S},\mathbf{P}) = [E]\frac{k_{\text{cat}}^{+}\prod_i \widetilde{S}_i - k_{\text{cat}}^{-}\prod_j \widetilde{P_j}}{\prod_i(1+\widetilde{S}_i)+\prod_j(1+\widetilde{P_j})-1} \ . \tag{6.1}$$

This is a generic reaction rate for a reaction involving an arbitrary number of metabolites (i substrates and j products), and is based on the assumption that the substrates and products can bind and unbind to the enzyme in any order. It assumes that no simultaneous substrate-product binding can occur (Figure 2.10). Such a rate equation represents a more realistic rate function than that of the perfect enzyme, as it enables the inclusion of both substrate saturation and the rate inhibition arising from high levels of product. It does however involve a greater number of parameters – one for each metabolite that is involved in the reaction along with the rate constants k_{cat}^{+} and k_{cat}^{-}. These parameters are not independent but are connected through a constraint on the thermodynamics. Since at equilibrium the net rate of the reaction (6.1) is zero, the following relation must be satisfied:

$$k_{\text{cat}}^{+}\prod_i \left([S_i]/K_{S_i}\right) = k_{\text{cat}}^{-}\prod_j \left([P_j]/K_{P_j}\right) \ . \tag{6.2}$$

This can be rewritten using the equilibrium constant $q = \prod_j [P_j]_{eq} / \prod_i [S_i]_{eq}$ to give the Haldane relationship [47],

$$q = \frac{k_{\text{cat}}^{+}\prod_j K_{P_j}}{k_{\text{cat}}^{-}\prod_i K_{S_i}} \equiv e^{\frac{-\Delta_r G^0}{RT}} \ , \tag{6.3}$$

which can be used to eliminate one of the parameters since this is related to the standard free energy of the reaction. Eliminating k_{cat}^{-} allows the common modular rate law (6.1) to be re-expressed as

$$v = [E]k_{\text{cat}}^{+}\left(\frac{\prod_i \widetilde{S}_i}{\prod_i(1+\widetilde{S}_i)+\prod_j(1+\widetilde{P_j})-1}\right)\left(1-\frac{\Gamma}{q}\right) \ . \tag{6.4}$$

Eq. (6.4) still requires us to specify the parameters $[E]$, k_{cat}^{+} and the Michaelis constants K_{S_i} and K_{P_j} of all metabolites involved in the reaction. These parameters are not known for the hypothetical reactions which we generate in our analysis (and may not be available for many known reactions). One possible way to reduce the number

of parameters might be to follow the approach of the previous chapter and assume that the fixed-concentration external metabolites enter only through the shift they produce in the equilibrium constant of the reaction: i.e. $q' = \left(\prod_i [e_i] / \prod_j [e_j]\right) q$ as discussed in Section 5.2. This amounts to neglecting the kinetics associated with these metabolites and hence only the internal metabolites would be included in $\{S_i\}$ and $\{P_j\}$ in Eqn. (6.4). One further possible reduction is to combine $[E]$ and k_{cat}^+ into the single parameter governing the maximal forward rate of the reation, $V_f = [E]k_{\text{cat}}^+$.

There is a choice then, in how we approach the comparison of our alternative pathways using Eq. (6.4) for the reaction rates. The simplest would be to assign some typical values for each of the parameters and assume that all enzymes are identical. Alternatively one could include all parameters in an optimization procedure, choosing the set of values that maximizes the overall flux of the pathway.

6.2.1 Solving the system of ODEs

In either case, the system of ODEs governing the dynamics of the network will have no analytical steady state solution. Instead, we will need to numerically determine the steady state of the system and the corresponding fluxes. The concentration $[M_i]$ of metabolite i is governed by

$$\frac{d[M_i]}{dt} = \sum_k v_{k,\text{prod}} - \sum_k v_{k,\text{cons}} \ , \tag{6.5}$$

which is simply the sum over all reactions that produce or consume this metabolite. Each rate v_k depends on the choice of kinetics, where here we will be assuming the form of Eq. (6.4) for all reactions.

Chemical reaction systems often contain rate constants spanning several orders of magnitude and can exhibit a phenomenon known as stiffness. The notion of stiffness cannot be precisely mathematically defined, but arises in systems where some components of the solution decay much more rapidly than others and an extremely small integration step-size is required in comparison to the smoothness of the solution [97–99]. To see how this can arise in our equations, we rewrite the reaction rate of

Eq. (6.4) in the form:

$$v = \left(\frac{V_f}{\prod_i K_{S_i}}\right)\left(\frac{1}{\prod_i(1+\widetilde{S_i})+\prod_j(1+\widetilde{P_j})-1}\right)\left(\prod_i[S_i]-\frac{\prod_j[P_j]}{q}\right) \tag{6.6}$$

$$= R_f\prod_i[S_i] - R_b\prod_j[P_j]\ , \tag{6.7}$$

where $R_b \equiv R_f/q$ and $R_f = R_f(V_f, \{[S_i]\}, \{[P_j]\}, \{K_{S_i}\}, \{K_{P_j}\})$ is a function of the metabolite concentrations and the reaction parameters, being the product of the first two terms in Eq. (6.6). Depending on these values, R_f can span several orders of magnitude across the different reactions of a network. In nature, enzymes are characterized by k_{cat} values ranging from 10^{-2} s^{-1} to 10^5 s^{-1}, with K_M values between 10^{-8} M and 0.1 M [100]. For any single reaction, the equilibrium constant q, which depends on the fixed external metabolite concentrations, can in addition lead to the forward and backward "instantaneous rates", R_f and $R_b = R_f/q$ being orders of magnitude apart. Hence a range of timescales can exist in our networks of reactions, leading to the system of equations governing their dynamics becoming stiff.

Standard numerical techniques frequently fail when applied to such stiff systems, and often different methods of solution are necessary depending on the nature of the stiffness. Since we ultimately wish to be able to sample arbitrary networks of reactions in an automated way, we currently solve these systems of ODEs using Mathematica. Initial investigations have found that setting all $d[M_i]/dt = 0$ and using Mathematica's *FindRoot* function is the most robust and efficient method[1].

6.3 Application 1: repeating our analysis of the trunk pathway

The problem of calculating the fluxes is simplified in the case of the alternative linear trunk pathways enumerated in the previous chapter. Since each reaction is uni-uni molecular with respect to the internal metabolites, the common modular rate (6.1) reduces to that of the usual reversible Michaelis-Menten rate law (2.10), and each reaction has only 3 parameters associated with it: V_f, K_S and K_P. However, even

[1] Another possibility is to use the *NSolve* function, which evolves the system in time towards its steady state. This approach often leads to longer computational times however. A third option is to reformulate the problem as a minimization problem. By defining the functions $f_i(x) = \frac{d[M_i]}{dt}(x)$, where x represents all relevant metabolite concentrations, then the minimization of the function $F = \sum_i f_i^2(x)$ will correspond to the steady state. This approach can be performed using Mathematica's *FindMinimum* function, but inital tests show that this fails to find the solution more often than compared to the *FindRoot* and *NSolve* methods.

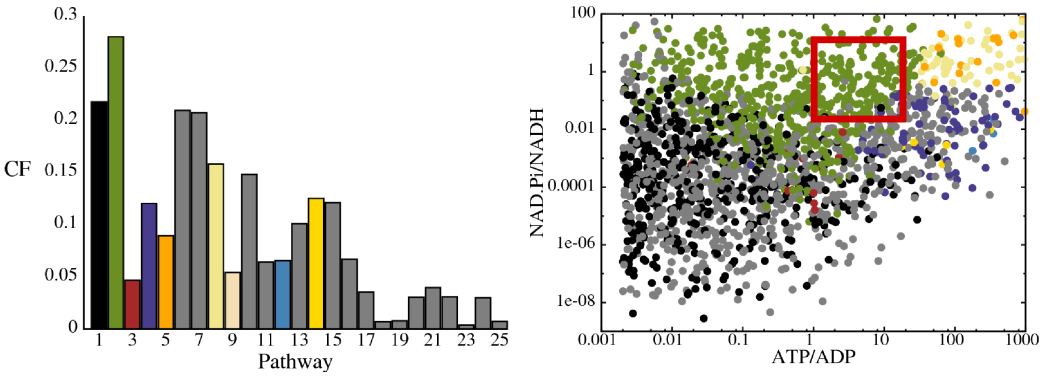


Figure 6.1 Comparison of glycolytic trunk pathways using reversible Michaelis-Menten kinetics. The colours and parameter sampling are the same as in Figure 5.2. All enzymes are assumed identical, with parameters $V_f = 10^{-3}$ Ms^{-1}, $K_S = K_P = 10^{-5}$ M. Although the ordering of the pathways does differ from that when assuming perfect enzymes, the real trunk pathway still performs best under typical physiological conditions.

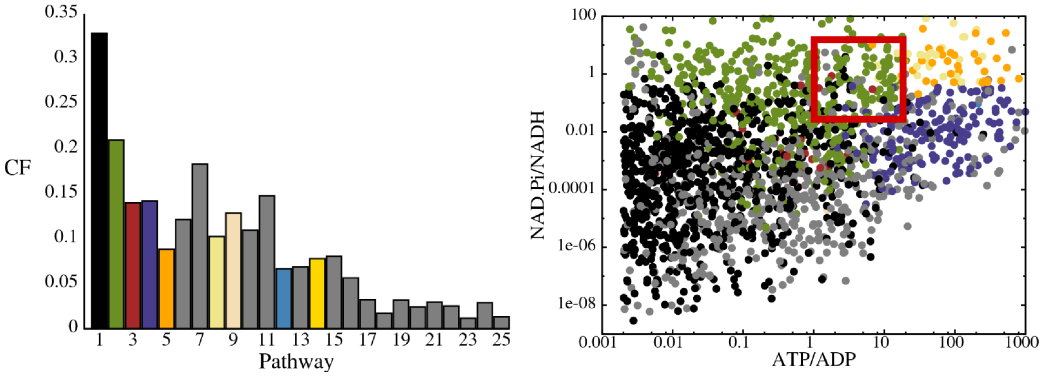


Figure 6.2 Further comparison of glycolytic trunk pathways using reversible Michaelis-Menten kinetics. Again, all enzymes are assumed identical, with parameters $V_f = 10^{-3}$ Ms^{-1}, $K_S = 10^{-6}$ M, $K_P = 10^{-3}$ M.

for these short linear pathways, the actual values of these parameters along with the concentrations of the external metabolites can lead to the system of equations presenting a numerically non-trivial problem. Due to the moderate computation time of solving the equations, and the large number of times these equations must be solved in the Powell optimization procedure, the optimization of the reaction parameters has not yet been performed. This is ongoing work – ideally the 3 parameters for every reaction in a pathway would be chosen so as to maximize the flux, allowing us to compare the best case scenario for each pathway.

To explore the effect this choice of kinetics has on the pathway comparison of Section 5.6.3, we instead assume that all enzymes are equivalent. Figure 6.1 shows the result of our comparison for the same glycolytic pathways shown in Figure 5.2, when all enzymes are described by identical parameter values; $V_f = 10^{-3}$ Ms^{-1}, $K_S = K_P = 10^{-5}$ M. We find that the outcome is altered with respect to the ordering of pathways in the plot of

the cumulative flux CF. However, the performance of the real trunk pathway (path 2, green) is enhanced, performing best over a much wider region of parameter space.

Setting $K_S = K_P$ assumes that the enzyme has the same affinity for the substrate and product metabolites. Figure 6.1 repeats the above analysis for the case $K_S = 10^{-6}$ M, $K_P = 10^{-3}$ M, whereby the enzyme now has a higher affinity for the substrate metabolite. This will have the effect that the rate of the reaction will not be as strongly decreased when the product concentration is high, and can be interpreted as each enzyme now favouring the *forward* reaction. Again we see that although the rank order of the pathways is changed, the real pathway still performs best under typical physiological conditions.

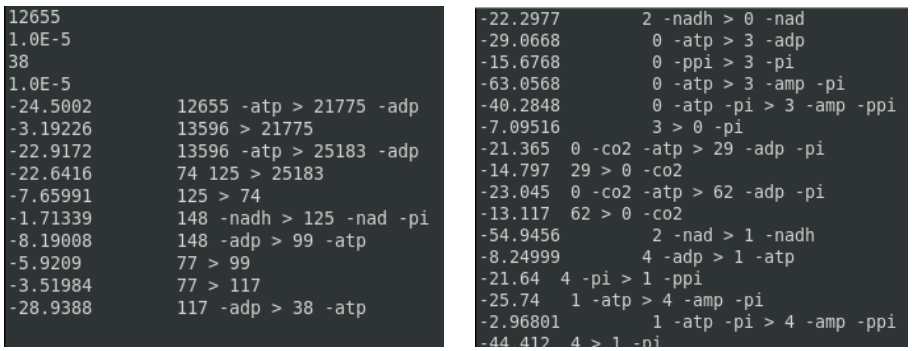
Though it will be interesting in future work to see how this is affected when the reaction parameters in each pathway are optimized, the general outcome seems to be largely unchanged: when realistic biophysical constraints are imposed on the intermediate metabolite concentrations, the real trunk pathway performs best under typical physiological conditions.

6.4 Application 2: Development of a framework for performing evolutionary simulations

I will now describe a program we are developing that we intend to use to perform evolutionary simulations. These simulations will allow "pathways" (or subnetworks of arbitrary structure) to be sampled from the full biochemical space, and allow a range of enzyme kinetics to be implemented. Beginning from some intial network configuration, and given a list of fixed-concentration substrate and product molecules along with a user-defined goal function, the network evolves through a series of mutations. An evolutionary simulation will thus require new features in addition to the flux calculation. For instance, it requires one to define the procedure through which mutations can occur. In this section I will describe the basic structure of the program, followed by a discussion of some of the technical details and future extensions.

6.4.1 Basic structure of program

The program begins by reading in two reaction lists: (i) the "super network" which contains all possible reactions and their standard free energies $\Delta_r G^0$, and (ii) the "initial network", which is preceded by a list of *source* and *product* metabolites and their fixed



(a) **(b)**

Figure 6.3 Examples of the input format for the initial network (a) and super network (b). The initial network has four additional lines at the start, the first and third will list all source and product metabolites respectively, while the second and fourth lists their (fixed) concentrations (M). In each reaction entry the first field is the free energy of the reaction $\Delta_r G^0$ (kJmol^{-1}) while the second column contains a reaction string, with substrates and products seperated by a ">". Note that a dash preceding any metabolite name indicates that it is an external metabolite, while all internal metabolites are assigned an integer label.

concentrations (required in the flux calculation). The format for these inputs is shown in Figure 6.3, and is described in the caption. Again, we have divided the set of metabolites into two groups: *internal* and *external*. The list of external metabolites and their fixed concentrations is stored in a separate file.

Once the user has decided on the form of the enzyme kinetics, the intial network undergoes a series of mutations involving the addition of reactions from the super network, or reaction deletions. Defining some objective function, for example ATP flux J_{ATP}, the program then repeatedly mutates the network, each time calculating this objective. Networks that exceed some defined performance threshold are stored as a list of reactions and parameters, and an image of the network can be generated using the open source program Graphviz [101]. The program then continues the evolutionary procedure.

6.4.2 Calculating the objective

The program requires a goal function to be defined that is related to the function of the network. This might correspond to the yield or rate at which ATP is produced, or the production of any other important metabolite. One could also include a cost that is associated with the number of reactions present in the network. To calculate the objective function requires first that we find the steady state solution of the ODEs representing the network dynamics, for an arbitraty network generated by the program. Due to the potential stiffness of the ODEs, and since we wish to be able

to sample arbitrary networks of reactions in an automated way, our program currently utilizes the *FindRoot* function of Mathematica to determine the steady state metabolite concentrations. From these, the distribution of fluxes and the desired objective function can then be calculated.

Simplifying the network

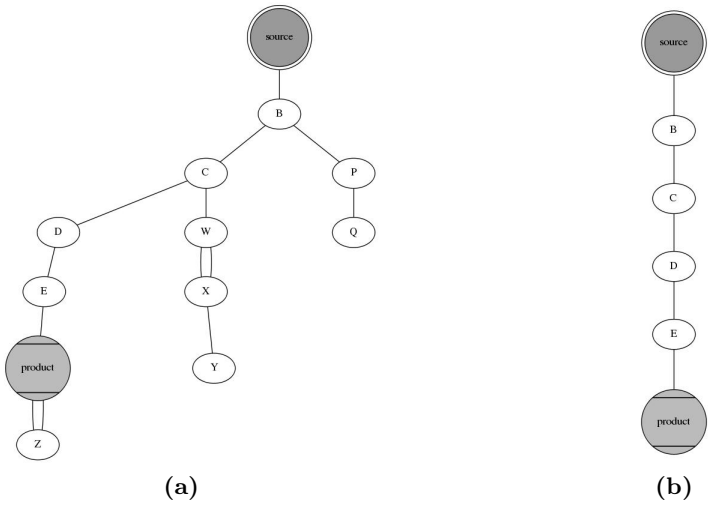


Figure 6.4 Network simplification procedure. If the evolutionary simulation leads to the network shown in (a), before calculating the steady state fluxes the network is minimized by removing all dead-end branches and two-node loops (b). Dead-end branches must be in equilibrium and have zero metabolic flux, while the two-node loops may be held out of equilibrium in a futile cycle.

To aid in the determination of the steady state, we also perform a network simplification before each call to Mathematica. This procedure prunes the network by removing all unnecessary reactions – any dead-end branches and dead-end two-node loops as shown in Figure 6.4. In steady state, dead-end branches must be in equilibrium and thus carry no flux. Any loops that arise in the mutation procedure will be held out of equilibrium through the consumption of some external metabolite. For example, if in a two-node loop one reaction is a phosphorylation coupled to the consumption of ATP, while the other is a dephosphorylation through a hydrolysis reaction, this two-node loop creates a futile cycle, resulting only in the consumption of ATP. In reality such cycles do occur, as can be seen in glycolysis, Figure 2.8. Although these cycles can play a role in metabolic regulation and thermal homeostasis [102], in general, cells will reciprocally regulate these reactions, preventing them from being simultaneously active [3]. Dead-end reactions and two-node loops are thus recursively removed until the minimal network is reached (Figure 6.4).

Reaction directionality

There is one more consideration with respect to the flux calculation in the evolutionary simulations. Each reaction has two rate constants associated with it, k_{cat}^{+} and k_{cat}^{-}, Eq. (6.1). However, these are not independent but are related through the thermodynamic constraint of Eq. (6.2). The form of the rate then only explicitly includes one of these values as a parameter, which in Eq. (6.4) we have chosen to be that associated with the *forward* direction, k_{cat}^{+}. Thus each reaction in a given network has a directionality associated with it. Considering Eq. (6.7), this means that although the ratio $R_f/R_b = q$ at equilibrium, and so is not affected by the choice of *forward* or *backward* directionality, the actual values of these two quantities will be. This is an issue, since when coupled with other reactions as part of a pathway, this choice of directionality will affect the kinetics of each reaction and thus affect the steady state flux. In a network with a non-trivial structure, it is not obvious what direction the flux will flow through any given reaction knowing only the $\Delta_r G^0$ values and topology. To be consistent then, can require two separate calculations in the determination of the objective function. First, the steady state is found in which all reactions have a directionality as they appeared in the previous network, or if they have just been added to the network, as they appear in the input file. From this, one can obtain the direction of flux through each reaction in steady state. A second calculation is then performed, if needed, in which all reactions have been re-ordered so that their *forward* direction corresponds to a positive flux.

6.4.3 Mutating the network

The program can currently implement two different evolutionary algorithms. The choice of how to mutate the network will depend on what feature one is trying to capture: is the simulation attempting to resemble the process of evolution, or to fully explore the biochemical space?

Random mutations

The first evolutionary procedure involves the random addition and deletion of reactions from the network. At each mutation, with probability p a reaction that shares at least one metabolite with the current network is added from the super network, or with probability $(1 - p)$ a reaction is deleted from the network. Any reaction deletion is subject to some restrictions: the network must be able to produce all defined product

metabolites, but is not required to utilize all defined source metabolites. That is, any deletion attempt that disconnects any of the products is disallowed, but as long as all products are still being produced, a deletion may result in one of the substrates no longer being consumed. One can also include additional mutation acceptance criteria, for instance any mutation that results in a large decrease in the objective function could be disallowed. However doing so can result in a decrease in the extent and speed at which the biochemical space is explored. The choice of strict acceptance criteria could potentially result in a situation in which the evolutionary procedure cannot generate all possible biochemical pathways from a given initial configuration.

Network growth and contraction

In order to ensure that the full biochemical space is accessible the program can implement an alternative evolutionary procedure. This involves the repeated growth and contraction of the network. From an initial configuration, a large number of reactions are added from the super network in a process similar to that of [55]; each reaction added must connect to a metabolite already present in the network. This expanded network is then contracted through the random deletion of reactions. After each deletion, the program checks that all products are still produced, using the depth-first search algorithm to check that all products can be reached from at least one source metabolite. It then simplifies the network (Figure 6.4) and checks for the existence of a steady state, calculating the objective function if one exists. Once no more reactions can be deleted, the network is expanded and the process repeated.

Cleavage reactions

The network can be expanded to 6-carbon CHOP molecules so as to capture the full glycolytic pathway. This also requires that the aldehyde-lyase enzymes (EC 4.1.2), that catalyze the cleavage reactions such as that of fructose-1,6-bisphosphate into G3P and DHAP (step 4, Figure 2.8), be included in the network. These reactions are not uni-molecular with respect to the internal metabolites and care must be taken over how a mutation adds these to the network. If a mutation were to result in the addition of such a reaction, and connect only one of the cleavage products to the existing network, the second would exist as a dead-end metabolite. This could lead to a very large steady state concentration, one which could be far outside the allowed biophysical limits and one that could result in the cleavage reaction becoming thermodynamically unfavourable. One solution to this is to ensure that upon addition of such a reaction,

both products are forced to be connected to the network through the inclusion of a second reaction from the super network. As occurs in glycolysis, one could check first for an isomerization reaction between the two metabolites.

6.4.4 Example of procedure

Figure 6.5 illustrates one iteration of the the network growth and contraction evolutionary procedure, for the simple case of glycolysis. Here, only the lower trunk pathway is allowed to evolve (that is, the super network used is restricted to molecules of 4 carbons or less). The initial network (the EMP glycolytic pathway) is shown in 6.5A. This initial network then undergoes the growth phase; Figure 6.5B shows the resulting network after 30 reaction additions followed by the network pruning procedure. Such expanded networks serve as the starting point to generating new networks. Reactions are deleted until no more can be removed, with the network pruning being performed at each stage and the objective function being calculated. Figure 6.5C shows a network generated by this process. Note that if either of the reactions in the diamond motif are deleted, the network will be in its minimal state and the procedure will be repeated.

6.4.5 Discussion

The hope is that these evolutionary simulations will allow arbitrary networks to be mutated and their performance assessed against some goal function. Whether this is computationally feasible for large networks remains to be seen, but initial investigations using glycolysis suggest that it will at least be useful for individual pathways. Given the vast scale of the biochemical space potentially available to life it seems an important question to ask why only such a highly restricted subset has ever been utilized. Could improvements to certain pathways be realised by exploring this space? Such a question is relevant from both an evolutionary point of view and in practical applications. Indeed, bioengineers are already investigating the possibility of entirely *de-novo* pathways towards compounds of industrial importance [73, 103–105]. It is also exciting to speculate whether certain reactions may be occurring in nature unknown to researchers, due to them being the secondary or tertiary reaction of an enzyme known for an entirely different function. It is widely recognized that many metabolic pathways may still await detection [106]. This is true even for the most highly studied core metabolic pathways, with a new reaction being discovered in the Rapoport-Leubering shunt of glycolysis as recently as 2008 [82]. The ideas and process of the previous chapters could aid in the discovery of such as-yet uncharacterized

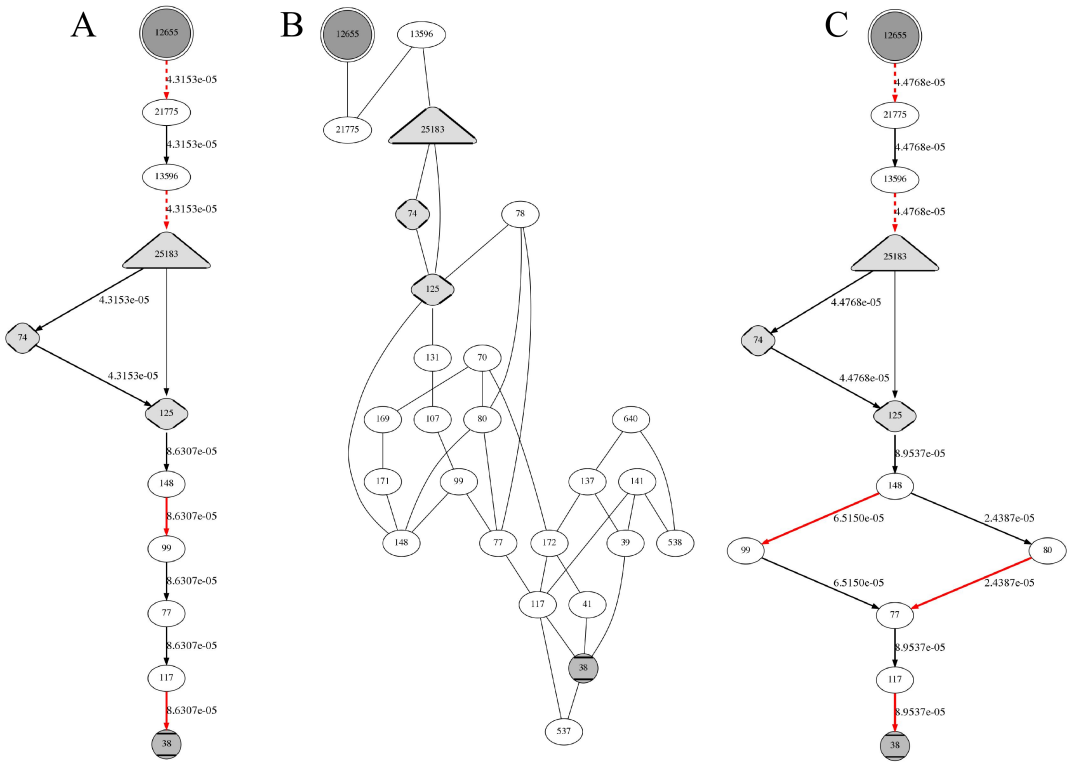


Figure 6.5 A simple illustration of the network growth and contraction procedure. **A:** The initial configuration is the EMP glycolytic pathway. **B:** The network generated after the growth phase; some arbitrary number of reactions is added from the supernetwork and the network is pruned. **C:** The contraction phase proceeds by the deletion of reactions until some minimal network is reached. Here, only one more reaction may be removed before a minimal network is reached. In this example, if a reaction from the left branch is removed, we end at a different minimal glycolytic pathway. If a reaction from the right branch is removed we are returned to the initial network. The program generates network images using Graphviz [101], an open-source network visualization tool. Reactions that produce and consume ATP are shown as solid and dashed red lines respectively. Cleavage reactions are indicated by altered node shapes, while source and product metabolites (here 12655 = glucose and 38 = pyruvate) are shown as double circles and boxed circles respectively. Edge labels indicate steady state fluxes.

pathways.

6.5 Future extensions

6.5.1 Open source program

Our program currently uses Mathematica to find the steady state of the network. Similar functions exist in other libraries, such as the *mpmath* Python library [107] which is freely available. Future versions of the code will be modified so as to use only open-source software.

6.5.2 Varying the environmental conditions

Since organisms exist in a vast range of environments, it would be desirable to be able to vary the environmental conditions of the generated reaction network. The intracellular pH of bacteria and yeast can range between 4.8-9.0 and 5.5-7.5 respectively [28], though it tends to be between 6 and 7.5 for most organisms[2]. Altering the pH or ionic strength is easily achieved through Eq. (3.12). The thermodynamic data in [67] allows this to be done for any pH value between 5 and 9, and ionic strengths of up to 0.35 (the limit of applicability of the extended Debye-Hückel theory). Since it is known that the pH value can have a significant affect on the $\Delta_r G$ of some reactions in solution, the ppdk reaction being one of these [28], future work will repeat the analysis of Chapter 5 using networks generated at a range of pH values. It will be interesting to see if any of our candidate pathways are greatly affected when varying the pH.

The ability to vary the temperature of our network would be a desirable property for future work too, since it is an important environmental factor, and one for which cellular life has little control. Since $\Delta G = \Delta H - T\Delta S$, the temperature dependence of the free energy of formation is less straight forward, requiring the enthalpies of formation of reactants to be known. Although such data is much less widely available, the enthalpies of formation have been measured for 94 of the reactants in [67]. This will enable future work to explore the combined effects that temperature and pH have on our pathways.

6.5.3 Expanding the molecular representation

The approach used in this thesis has used a simple "String" representation of molecules as combinations of well defined chemical groups. This methodology was sufficient for our study of the trunk pathway and is extendable to branched compounds and to networks containing extra elements such as nitrogen and sulfur. However, in metabolism many important compounds contain cyclic structures – aromatic rings for example. Such cyclic compounds cannot be easily incorporated in our work and so future extensions that allowed for these would have to adopt a different formalism.

An alternative way to generically represent molecule structures is the bond electron matrix (BEM) [109], where a molecule is represented by a matrix in which every row and column corresponds to an individual atom. Each entry ij denotes the number of

[2]Many acidophiles, some obtaining optimal growth in environments with a pH value of less than 2, have been found to maintain an intracellular pH close to neutrality [108].

covalent bonds between atom i and atom j and diagonal elements ii denote the number of non-bonded valence electrons of atom i. Each generalized reaction mechanism is also defined as an "operator" matrix, so that taking the BEM of a substrate molecule and adding the operator matrix results in the BEM of the product. This is the formalism employed by Hatzimanikatis in the BNICE software package [73], which unfortunately has not been made publically available.

Another option would be to make use of the SMILES (Simplified Molecular-Input Line-Entry System) notation developed by David Weininger and Daylight Chemical Information Systems Inc. [110]. This offers a line notation for specifying the structure of chemical molecules, also represented by ASCII strings, and has a defined vocabulary to represent cyclic compounds. It also has the benefit that algorithms exist to canonicalize SMILES strings (since there can exist many ways to represent the same molecule using a String notation). The existence of the reaction transform language SMIRKS combined with the language SMARTS which allows molecular substructures to be searched, would allow general reaction mechanisms to be described. This approach has been used in previous works, in research aiming to generate novel pathways for the synthesis of important chemicals [104].

Either of these approaches would allow the network to be expanded to all types of chemical compounds and reactions. This will be necessary so that future work can investigate all pathways of metabolism, many of which include large molecules with aromatic groups. We are in fact currently collaborating with Vincent Danos' group in the informatics department at the University of Edinburgh, who are developing their own graph-based approach to the problem of generating all possible molecules from the level of single atoms.

6.5.4 Including additional enzymatic knowledge

One additional feature that could be incorporated in future studies is the observation that certain types of reactions can be slower than others, in terms of their k_{cat} value [100]. For example, it is known that ligase reactions (EC 6.x.x.x) are, on average, an order of magnitude slower than isomerization reactions (EC 5.x.x.x). It would be interesting to see how including such information could alter the results of our pathway comparison – pathways with a better thermodynamic design may be penalized if they utilize the less efficient reaction types. Our glycolytic pathway 1 (Figure 5.4), for example, makes use of a ligase reaction whereas the real glycolytic trunk pathway does not.

Part II

Chapter 7

The contact process and spatial models in biology

7.1 Introduction

In this chapter I will discuss the importance of spatial structure and stochasticity in the modelling of biological systems. I will discuss one particularly favoured method for incorporating these features, namely lattice models. As well as their importance in a biological and ecological setting, these models have proved greatly interesting to mathematicians and physicists for a number of reasons. Firstly, it has not been possible to derive analytic solutions to even the simplest systems, requiring new mathematical techniques to be developed to approximate and understand their behaviour. Secondly, interest from the physics community stems from the fact that many of these systems exist far from equilibrium and contain various types of nonequilibrium phase transitions. I will thus review the theory behind equilibrum and non-equilibrium phase transitions, and introduce as an example the contact process (CP), the canonical lattice model for the spread of an infectious disease. I will then describe how such systems are modelled using computer simulation and finally I will discuss some extensions to the standard CP, providing the biological motivation behind our own novel extension that will be the subject of the next chapter.

7.2 Spatial models in biology

The vast proportion of biological and ecological models have been founded on the *mean-field assumption*. This invokes mass-action kinetics, i.e. the population is homogeneous and well mixed, and thus ignores all spatial information. This assumption is clearly not the case for many biological systems, in which spatial structure is apparent and its effects have been shown to play an important role in the dynamics [111, 112]. In any population where dispersal is finite, there must arise some spatial aggregation. The benefit of neglecting this aggregation is that mean field models are usually analytically tractable, and their dynamics and general behaviour is easier to understand than in their spatial counterparts.

However, it is often the case that a spatial stochastic model and a mean-field approach can produce conflicting conclusions while purportedly modelling the same system [113, 114]. This can manifest itself quantitatively, say through the prediction of different species abundances [115], but it can also drastically change the qualitative outcome, with models making different predictions for which species will survive and which will become extinct [116]. In one study of host-pathogen ecologies [114], the authors find that their spatially explicit lattice model predicts that an upper limit exists on the transmissibility of the pathogen; higher levels of transmission become detrimental to the pathogen itself. Their mean field analysis however entirely misses this feature.

Clearly then, if a model is to capture the behaviour of a real system, care must be taken to identify the most appropriate type of model to apply. On the other hand, it can be equally important to study generic models, since these allow the general principles that underpin certain systems to be established. A scan of the ecological and epidemiological literature might leave one with a perceived dichotomy: models in which interactions are strongly local, and models where the system is well mixed so that the interactions are effectively infinite. It is clearly the case that most real-world systems will fall somewhere between these two extremes, however the study of such extremes allows the basic implications of spatial structure to be examined [117].

7.2.1 Lattice models

One of the ways in which the general effects of spatial correlations have been studied has been through the use of lattice systems, or interacting particle systems [118, 119]. These systems discretize space, and individuals (referred to as particles or agents) inhabit a

grid of sites with a well defined topology, often a two-dimensional square lattice. Agents can interact through a set of lattice update rules, each with an associated rate, with the system evolving stochastically (or deterministically in the case of the traditional cellular automata [120]). These update schemes usually depend on the states of neighbouring particles, and are most often taken to be Markov processes, characterized by their lack of memory; the future evolution of the system depends only on its current state and not on its history.

Such models have been utilized in the study of a host of natural systems, including bacterial [121], animal [122] and plant [123] populations, the dynamics of infectious diseases [124] and forest fires [125]. Their applicability has not just been limited to biological systems, but has also found use in the social sciences and economics [126]. Hence a general understanding of their behaviour and complexities is of some importance.

7.2.2 Other ways of including space

I will conclude this section by briefly mentioning other ways in which space can be incorporated into models. There are three common ways in which space is taken into account [114]; the lattice models described above, "patch" models and reaction-diffusion equations, with a range of phenomena being captured by each approach. Reaction-diffusion equations are partial differential equations which incorporate space in a continuous manner, representing the population as a density field in time and space. Their importance has been demonstrated in a variety of contexts which exhibit spatial pattern formation [127] or wave-like phenomena in range expansions [128]. The patch models consist of a network of separated "patches" or populations, each undergoing some internal dynamics, which are connected by a network of edges governing the patch-patch interactions. As such they have found wide use in the field of epidemiology, where it is crucial to study the effect that factors such as transportation or social networks have on disease spread [129]. In fact, networks of connections offer a generalization to the lattice approach. Recently a large amount of research has focused on more realistic networks of interaction than that of a regular lattice, enabling the importance of the topology of these networks to be studied [130].

From within the physics community, interest in the lattice models has stemmed from the fact that these systems are often nonequilibrium in nature, and represent the simplest examples of models exhibiting nonequilibrium phase transitions. I will now review the importance of phase transitions and the distinction between the equilibrium and

nonequilibrium cases.

7.3 Equilibrium systems

A system in thermodynamic equilibrium contains no net flow of matter or energy and would undergo no change if it was isolated from its surroundings. Much is known about the physics of systems in thermodynamic equilibrium and many powerful tools have been established to facilitate their analysis, in large part through the works of Boltzmann and Gibbs. This has led to a generic statistical description which describes equilibrium systems in terms of the probabilities of the microstates – for instance the probability of being in a state of energy E_i is well known to be given by the Boltzmann distribution $P_i = \exp(-E_i/kT)/Z$ where $Z = \sum_j \exp(E_j/kT)$ is the partition function. In addition, thermodynamic potentials have been derived along with corresponding extremum principles. For example, in a system at constant T and V the Helmholtz free energy $F = U - TS$ is minimized at equilibrium; at constant T and P we know that it is the Gibbs free energy $G = H - TS$ that is minimal (where T, P, V are temperature, pressure and volume and U, H and S are the internal energy, enthalpy and entropy); and for any isolated system the equilibrium state is characterized by maximal entropy.

An important concept in the definition of equilibrium is that of *detailed balance*, which states that at equilibrium, every process should be equilibrated by its reverse process. In terms of a probabilistic process then, one can write formally

$$\pi_i T_{ij} = \pi_j T_{ji} \ , \tag{7.1}$$

where π_i represents the stationary state (equilibrium) probability for the system to be in state i, and T_{ij} is the transition matrix element corresponding to the probability of moving from state i to j. If this holds for every pair of states then detailed balance is said to be satisfied and the system is in equilibrium.

In contrast, if a system does not obey detailed balance, it is by definition a nonequilibrium system and between some pairs of microstates there can be a non-vanishing flow of probability. Nonequilibrium systems are thus characterized by the presence of currents or fluxes (even when in steady state) and are driven by the environment, requiring an influx of matter or energy. Despite a large amount of interest in recent decades, no general formalism like that developed under equilibrium has been established for the nonequilibrium case. Since nearly all real world processes occur far

from equilibrium, and indeed this must be one of the defining features of life, there is a pressing desire to obtain a deeper and more general understanding of nonequilibrium processes.

7.4 Phase transitions

One aspect to this is the investigation of transitions between different phases in a system. Phase transitions have been studied extensively in equilibrium systems, and are accompanied by singularities in the free energy and its derivatives, which in turn give rise to discontinuities in certain physical observables. They fall into two categories. *First-order* phase transitions involve a discontinuity in one or more state variables, such as density, as a function of some thermodynamic variable, such as temperature. They are associated with a latent heat – the absorbtion or release of heat at a fixed temperature – and a regime in which the system is in a mixed phase during this transition. The boiling of water to vapour is an example of a first-order phase transition, in which there is a discontinuity in the density. *Second-order*, or *continuous*, phase transitions on the other hand, are described by an order parameter that usually has a non-zero value in the ordered phase while vanishing in the disordered phase, varying continuously between the two regimes. Perhaps the best known example of this is the spontaneous magnetization in the Ising model, a collection of interacting spins on a lattice. In the Ising model, the order parameter is the magnetization and the thermodynamic variable is the temperature. At temperatures above some critical value T_c, the spins are randomly orientated and the system is in a disordered state with no net magnetization. As the system is cooled below this value, it becomes energetically favourable for the particle spins to align and the magnetization takes a non-zero value.

An important discovery is that in the case of these continuous phase transitions, the behaviour of certain physical properties are characterized by power laws. For example, the magnetization in the Ising model behaves, close to the critical point, as $m \sim |T - T_c|^{\beta}$, where β is the associated *critical exponent* for the order parameter m. Crucially, as criticality is approached, the correlation length becomes infinite and the long range correlations lead to the nature of the transition depending only on the global properties of the system, such as the spatial dimensionality and the symmetry of the order parameter. The nature of the transition becomes independent of the microscopic details of the model. As a result, models of very different phenomena are found to be described by the same set of critical exponents, and can be correspondingly grouped into universality classes. The set of six critical exponents associated with equilibrium phase

Relation	Description
$m(T,0) \sim \tau^{\beta}$	$m(T,h)$ is the order parameter at zero external field, $h = 0$
$m(T,h) \sim \lvert h\rvert^{1/\delta}$	For a non-zero external field $h \to 0$
$\chi(T,0) \sim \lvert\tau\rvert^{-\gamma}$	χ is the order parameter susceptibility
$C(T,0) \sim \lvert\tau\rvert^{-\alpha}$	C is the specific heat
$\xi(T,0) \sim \lvert\tau\rvert^{-\nu}$	ξ is the spatial correlation length
$G(r;T_c,0) \sim r^{2-d-\eta}$	$G(r;T,h)$ is the correlation function and d the spatial dimensionality.

Table 7.1 The six critical exponents of equilibrium phase transitions: β, δ, γ, α, ν, η. The reduced temperature $\tau = (T - T_c)/T_c$ measures the distance from criticality.

transitions are shown in Table 7.1. Remarkably, the six critical exponents associated with equilibirum transitions can be shown to be related by four scaling laws [131]:

$$\alpha + 2\beta + \gamma = 2 \qquad \alpha + \beta(\delta + 1) = 2 \tag{7.2}$$

$$(2 - \eta)\nu = \gamma \qquad \nu d = 2 - \alpha \tag{7.3}$$

meaning that the full critical behaviour of models belonging to the same universality class can be determined from just two critical exponents.

7.4.1 Nonequilibrium phase transitions and directed percolation class

This extremely powerful notion of universality has also been found to be applicable in the case of nonequilibrium phase transitions. Of particular interest has been the class of continuous phase transitions from an active to an absorbing state [131–133], an absorbing state being one from which the system cannot escape. The existence of a completely frozen absorbing state clearly breaks detailed balance and so such transitions are necessarily nonequilibrium, arising in models of populations, epidemics and enzyme biology. Several universality classes have been identified for such transitions, with the most prominent by far being that of directed percolation (DP). In fact the repeated observance of this class and its apparent robustness led to what is known as the DP-conjecture [131, 134]. This states that a model should belong to the DP universality class if the following conditions are met:

- The model displays a continuous phase transition from a fluctuating active phase into a unique absorbing state.
- The transition is characterized by a non-negative one-component order parameter.
- The dynamical rules are short ranged.
- The system has no unconventional symmetries, conservation laws, or quenched

Exponent Relation	$d=1$	$d=2$	$d=3$	$d \geq 4$
β	0.276	0.583	0.813	1
$\nu_\perp$	1.097	0.733	0.584	1/2
$\nu_\parallel$	1.734	1.295	1.111	1

Table 7.2 The critical exponents of the DP universality class in all dimensions.

randomness.

It turns out in fact that more complicated, multicomponent systems can still belong to the DP class [135] as well as some models that actually break the above conditions, for instance systems containing an infinite number of absorbing states [136].

For the sake of simplicity, let us consider some generic process belonging to the DP class, in which lattice sites can be in two states, activated or deactivated, and change between states with some defined transition rates (perhaps a simple birth-death process, though the details are not important at this stage). The defining critical exponents governing the DP class are given in all dimensions in Table 7.2. The exponents correspond to the following behaviour near criticality:

$$\rho \sim (\lambda - \lambda_c)^\beta \quad , \quad \xi_\perp \sim (\lambda - \lambda_c)^{-\nu_\perp} \quad , \quad \xi_\parallel \sim (\lambda - \lambda_c)^{-\nu_\parallel} \, , \tag{7.4}$$

where the order parameter ρ is the density of active sites, $\xi_\perp$ and $\xi_\parallel$ are the correlation length and time respectively and λ is the parameter controlling the transition which occurs at λ_c. Note that in the equilibrium case there was one (spatial) correlation length that diverged at criticality. In nonequilibrium systems, since time is an independent degree of freedom we have the two correlations as above. The exponents $\nu_\perp$ and $\nu_\parallel$ are in general different, and one often sees the dynamical exponent defined $z = \nu_\parallel / \nu_\perp$ governing the dynamical scaling relation

$$\xi_\parallel \sim \xi_\perp^z \, . \tag{7.5}$$

Note that for dimensions above or equal to the *upper critical dimension* $d_c = 4$, diffusive mixing is strong enough to overcome the correlations in the system and the exponents take on the values predicted by mean field theory, irrespective of the dimensionality. Below this d_c, fluctuation effects and correlations strongly affect the critical behaviour, and the values of the critical exponents depend on the dimensionality of the system.

It is worth noting that despite a large amount of theoretical work demonstrating the generality of this class, there has been no unambiguous measurement of DP-like critical

behaviour in an experimental system[1]. On top of the difficulties of obtaining sufficiently accurate exponents, a possible reason for this lack of experimental observation is the sensitivity these systems show to disorder. Real systems are likely to contain at least some impurities or defects, and a body of theoretical work has demonstrated the effect that even a small amount of disorder can have on the critical behaviour of a system (discussed further in see Section 7.6).

An open question is whether an equally small number of universality classes will be discovered in nonequilibrium situations, which necessarily involve richer behaviour. One reason for this richer behaviour emerges because unlike the equilibrium scenario, in nonequilibrium systems time is an independent degree of freedom. Some of the earliest nonequilibrium studies were to investigate relaxation to equilibrium states and phase ordering kinetics [138, 139]. Once an initial state is specified, this leads to multiple sets of dynamical critical exponents being defined. For example, starting from a single active site at some origin allows the following spreading exponents to be defined at criticality [131]:

$$P(t) \sim t^{-\delta}, \quad \mathrm{n(t)} \sim \mathrm{t}^{\eta}, \quad \mathrm{R}^2(\mathrm{t}) \sim \mathrm{t}^{2/\mathrm{z}}\,, \tag{7.6}$$

with $P(t)$ being the survival probability at time t, $n(t)$ the mean number of active sites at time t and $R^2(t)$ the mean-square displacement of active sites from the origin. Other dynamic exponents can be found by considering alternative initial conditions. For example, from an initially fully active lattice, at criticality the active sites decay algebraically as $\rho(t) \sim t^{-\alpha}$ and the correlation length grows as $\xi_{\perp}(t) \sim t^{1/z}$. A complete summary of all critical exponents, static and dynamic, conventionally associated with transitions to an absorbing state can be found in Table 4.1 of [131].

A complicating factor is the observation that the universality classes of static models can actually be split by the dynamical exponents [132]. That is, despite two models being characterized by the same static critical exponents, consideration of their dynamical exponents shows them to belong to different nonequilibrium universality classes. In the next section I will review the contact process, perhaps the simplest example of a system containing a nonequilibrium phase transition to an absorbing state, and one that belongs to the universality class of DP. We will also see how this canonical model has been used to study the effects of disorder on critical behaviour.

[1]Perhaps with the exception of [137] which showed evidence of a DP transition between two different turbuent states in nematic liquid crystals.

7.5 The contact process

The contact process (CP) [131, 140] is a stochastic interacting particle system defined on a regular lattice and was introduced by Harris [141] as a simple model for the spread of an infectious disease. In the CP, each lattice site can take two states, 0 or 1, with 0 corresponding to a healthy individual and 1 corresponding to an infected individual. (Alternatively 0 and 1 could be interpreted as empty or occupied sites.) Infected sites recover at constant rate d, while healthy (or susceptible) sites become infected at a rate proportional to the number of infected nearest neighbours. Thus the dynamics can be expressed:

$$1 \to 0 \quad \text{with rate} \quad d \tag{7.7}$$

$$0 \to 1 \quad \text{with rate} \quad b(n_1/z)\,, \tag{7.8}$$

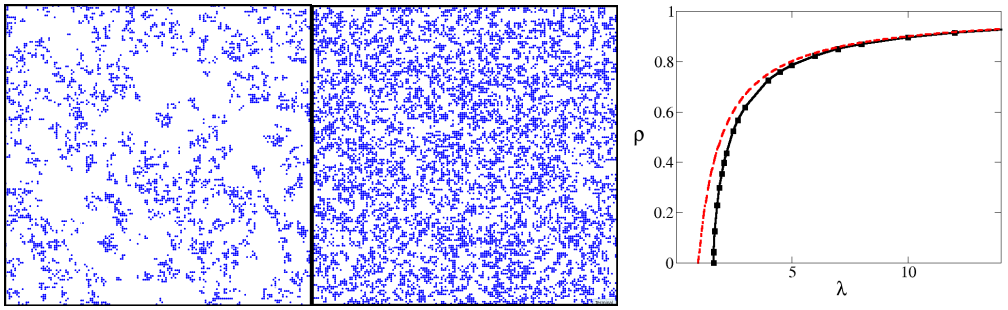
where n_1 is the number of infected neighbouring sites and z is the coordination number of the lattice; the total number of neighbours each site has. If an uninfected site's neighbours are all infected, then the infection rate takes the maximum value b. Note that the CP corresponds to the susceptible-infected-susceptible (SIS) model in epidemiology, in which recovered individuals do not gain any resistance to the infection but become immediately susceptible again.

Interest in the CP from within the physics community stems from the fact that it is one of the simplest models exhibiting a nonequilibrium phase transition to an absorbing state. Defining the ratio $\lambda = b/d$, there is some critical value λ_c which separates two states: the absorbing state for $\lambda < \lambda_c$ in which no infection (or population) can be maintained, and a fluctuating active state for $\lambda > \lambda_c$ in which there exists a steady state level of infection (or non-zero population density). This transition belongs to the directed percolation universality class (Section 7.4.1) where the order parameter, the density of infected sites, behaves as the power law

$$\rho \sim |\lambda - \lambda_c|^{\beta} \tag{7.9}$$

near criticality. The values of λ_c and β in two dimensions have been determined numerically as 1.649 and 0.583 respectively, though note that only the critical exponent β is a universal quantity; the position of the critical point can vary depending on the details of the model.

We will see how to represent such systems in a formal mathematical way in Section



(a) (b) (c)

Figure 7.1 Typical steady-state lattice configurations for the contact process at (a) $\lambda = 1.7$ and (b) $\lambda = 2.0$. As λ is decreased in (c) we observe a phase transition from an active fluctuating state, $\rho > 0$, to the absorbing state $\rho = 0$. Black symbols show results from numerical simulations ($\lambda_c \sim 1.65$, $\beta \sim 0.58$) and red dashes shown mean field theory prediction ($\lambda_c^{\mathrm{MFT}} = 1$, $\beta^{\mathrm{MFT}} = 1$).

8.2.1, but it is straightforward at this stage to write down the mean field equation for the contact process. To do this we simply neglect the local structure by replacing the local density of infected sites n_1/z in (7.8) with the global density, allowing us to write the time evolution of the density of active sites $\rho(t)$ as

$$\frac{d\rho(t)}{dt} = b\rho(t)(1-\rho(t)) - d\rho(t)\ . \tag{7.10}$$

Solving for steady state gives $\rho = 1 - 1/\lambda$ and hence we see that for the mean field system, at the critical value $\lambda_c^{\mathrm{MFT}} = 1$ the absorbing state $\rho = 0$ is reached. We can expand the expression for the steady state density close to criticality by writing $\rho = 1-1/\lambda = 1-(1+[\lambda-\lambda_c^{\mathrm{MFT}}])^{-1} \sim (\lambda-\lambda_c^{\mathrm{MFT}})$. Hence by comparing this to Eq. (7.9) we see that the mean field exponent is $\beta^{\mathrm{MFT}} = 1$. The differences between the simulated system and the mean field approximation can be seen in Figure 7.1c. Notice that at high values of λ the agreement in the steady state density is good between the two approaches. However, as λ is decreased this discrepancy grows, due to the diverging correlation length as criticality is approached. Also shown in Figure 7.1 are two typical steady-state lattice configurations of the CP at different values of λ.

7.6 Models with disorder

Disorder is an important phenomenon which is proposed to be responsible for the lack of experimental verification of the DP transition. Nearly all real systems will contain some sort of impurity. The CP provides a useful way to study the effects of disorder because it is one of the simplest and most highly studied models belonging to the DP

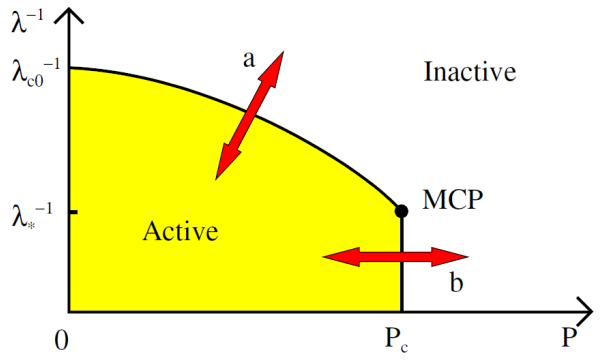


Figure 7.2 The schematic phase diagram of the diluted contact process, taken from [142]. The multicritical point (MCP) separates the two distinct classes of phase transition contained in the model, as discussed in the text.

class.

The Harris criterion [132, 143] states that disorder will be relevant, that is will change the nature of the transition and thus its critical exponents, if $d\nu_\perp \leq 2$, where d is the spatial dimensionality and $\nu_\perp$ is the exponent of the correlation length in the standard process without disorder (the *pure* process). Since it has been determined numerically that $\nu_\perp \simeq 0.73$ in the 2D CP, the Harris criterion predicts that quenched disorder will affect the exponents of the CP transition. Various approaches have been utilized for introducing disorder into the CP, and it has been demonstrated that even weak disorder can disrupt the DP nature of the transition.

The most commonly studied method for including quenched disorder (disorder frozen in space) is to entirely remove a fraction of the lattice sites from participating in the dynamics [144–146]. This leads to what is known as the diluted contact process and is achieved by introducing a new parameter p, the probability of removing any given site. Note that as long as the fraction of sites removed does not exceed the lattice percolation threshold, there is the possibility of an active state. Above this value there is no percolating cluster of lattice sites remaining, and the disconnected, finite clusters will always hit the absorbing state through some fluctation. The most obvious effect this will have is to raise the critical value of $\lambda = b/d$ compared to the pure CP, since as more sites are removed from the dynamics it will become harder to maintain the same level of infection. It turns out however that the diluted CP has very rich behaviour. Figure 7.2 shows the schematic phase diagram of the diluted contact process, taken from [142], and here I briefly outline the most important features. We will find that a knowledge of the effects of disorder can be useful in understanding the critical behaviour contained in more complicated models, such as our model of stacked contact processes

introduced in the next chapter.

First consider the case where we are below the percolation threshold and decrease λ to enter the absorbing phase. This corresponds to the transition marked a in Figure 7.2. It turns out that although the static exponents change from the nondiluted case, their values do not actually depend on the amount of dilution over a large range of dilution fractions, revealing a novel universality class for the disordered contact process [144, 147]. Notably, there is a marked difference in the dynamical critical behaviour between the pure and the disordered systems. In the diluted contact process, the power law dynamical scaling relation of Eq. (7.5) gets replaced by activated scaling, of the form $\ln(\xi_{\parallel}) \sim \xi_{\perp}^{\psi}$, characterized by a new exponent ψ, leading to an ultraslow decay in the density, $\rho(t) \sim [\ln(t)]^{-\bar{\delta}}$. Without a full mathematical derivation, intuitive reasons for why this change occurs can be gained by considering the Griffiths phase, where it becomes clear that rare favourable configurations dominate the statistics. The Griffiths phase [147, 148] is an example of the new physics that can arise from the presence of disorder. In the region $\lambda < \lambda_c^0$, where λ_c^0 is the critical value of the pure CP, the density of active sites will decay exponentially towards the absorbing state. However, when λ is between the critical values of the pure and disordered systems, $\lambda_c^0 < \lambda < \lambda_c$, although the process is globally subcritical and will decay to the absorbing state, there can exist rare large regions in which the fraction of diluted sites is low, leading to the process being locally supercritical. The decay of these regions is extremely slow since it requires large density fluctuations to push them to extinction. These rare regions actually dominate the long-time averages of the critical behaviour and the long relaxation times make the study of the phase transitions in the disordered CP a computationally intensive task [145, 147].

The second interesting feature of the diluted CP is the occurrence of a multicritical point (MCP) at $\lambda = \lambda_*$ separating the above case from a second type of phase transition, labelled b in Figure 7.2. Here, dilution has left the lattice at its percolation threshold, and any value $\lambda > \lambda_*$ can maintain a non-zero density of active sites. The transition occurs by varying p, and is thus driven entirely by the geometry of the lattice [142]. The authors show that although this transition shares some features with transition a, such as activated scaling and logarithmically slow decay in density, transition b belongs to a distinct universality class with unique exponents.

Disorder can also be introduced into the CP through varying either the creation or annihilation rates for some sites. In studies that allowed sites to take two different death rates it was shown that the random distribution of these throughout the lattice again led to critical behaviour that deviated from that of the DP class [149, 150]. In

this case, the behaviour did not appear to belong to a distinct universality class, but instead the critical exponents were found to vary continuously with the strength of the disorder. We will return to the effects of disorder in the next chapter when we come to analyse our model of stacked contact processes (Section 8.5), which we will find incorporates disorder in a natural yet highly complex manner.

Despite the large interest in lattice models, even in the case of the simple CP the presence of local interactions and the strong spatial correlations that ensue has prevented an analytical solution. Some approximations have been developed that allow lattice systems to be approximated through sets of coupled differential equations, and in Section 8.4 we will describe one of these in detail. These can be cumbersome however, and as we will see, are not guaranteed to be even qualitatively correct. Instead, much work in this area has made use of numerical simulations, which will be the main approach taken in the next chapter. In the following section I will describe how this can be done.

7.7 Simulating interacting particle systems

There are a number of techniques available for simulating stochastic interacting particle systems, the appropriate choice depending on the details of the system being modelled. In this thesis I have made use of an event-driven Monte Carlo simulation [151], that is based on the fact that the process being modelled is Markovian, with events behaving as Poisson processes. A Markov process is characterized by its lack of memory, in that the future evolution of the system is determined only by its present state and not on its history. In this section I will briefly review Poisson processes and how to calculate random waiting times for them. I will then describe the event-driven Monte Carlo method for simulating stochastic processes.

7.7.1 Poisson processes

The Poisson probability distribution $P(n)$ for n events is:

$$P(n) = \frac{\bar{n}^n e^{-\bar{n}}}{n\,!}, \tag{7.11}$$

where $\bar{n}$ is the average number of events. If the Poisson process occurs at a rate r over a time interval T, then the average number of events expected is $\bar{n} = rT$ and we have

$$P(n) = \frac{(rT)^n e^{-rT}}{n\,!}. \tag{7.12}$$

If the considered time interval $T = \Delta t$ is sufficiently small, so that $r\Delta t \ll 1$, expanding the above expression to first order in $r\Delta t$ gives the probability that exactly one event occurs in this interval $P(1) = r\Delta t$, and the probability that no event occurs $P(0) = 1 - r\Delta t$. Neglecting higher orders of $r\Delta t$ effectively assumes that no more than one event can occur in such an infinitesimal time interval.

The time distribution between successive events for a Poisson process is a simple exponential decay. Given that an event occurred at time t, the probability $P(\tau)$ that the next event occurs within the interval $t+\tau$ to $t+\tau+d\tau$ is simply the probability that no event takes place between t and $t+\tau$ (obtained from Eq. (7.12) to be $e^{-r\tau}$), multiplied by the probability that exactly one event occurs between $t+\tau$ and $t+\tau+d\tau$ (which is just $rd\tau$ for an infinitesimal interval): thus $P(\tau) = e^{-r\tau} r d\tau$. The probability density function $p(\tau)$, that is, the distribution of waiting times between successive events, is thus governed by the exponentially decaying function

$$p(\tau) = r\, e^{-r\tau}\ , \tag{7.13}$$

from which it can be calculated that the mean waiting time between events is $\langle\tau\rangle = 1/r$.

It is possible to generate sample numbers at random from any probability distribution if we know its cumulative distribution function (CDF). Consider the CDF for the exponential distribution of Eq. (7.13), i.e. the probability that an event occurs before time τ: $P(t < \tau) = \int_0^\tau r e^{-rt'} dt' = 1 - e^{-r\tau}$. Solving for τ gives $\tau = (-1/r)\ln(1-P)$. Substituting x in place of $(1-P)$, if one samples x uniformly from between 0 and 1, one will be sampling τ from the exponential distribution of Eq. (7.13). This will be required in the Monte Carlo simulations – to randomly select a waiting time for any Poisson process with rate r, one just need select a random number x uniformly from between 0 and 1, and calculate the associated waiting time using

$$\tau = \frac{-1}{r} \ln x\,. \tag{7.14}$$

7.7.2 Event-driven Monte Carlo simulation

The simulation method employed in this thesis is technically an asynchronous, continuous-time, event-driven Monte Carlo simulation. Many names exist for slight variants of what are essentially the same technique, but the important features to note here are as follows. Firstly, the lattice site updates are performed asynchronously, one at a time, rather than synchronously as is typical in the traditional cellular automata [120]. This is clearly a more realistic choice when modelling any real process. Secondly, the time variable is continuous, being advanced by random amounts as calculated by Eq. (7.14) rather than by discrete units (though when done correctly these two methods ought to give identical results [151]). Finally, *event-driven* simply refers to the fact that time is advanced from one event to the next, with each site having its own time variable determining when its next event will occur. This will become clear now, as I describe the general procedure in detail.

Step 1. Initialize the $t = 0$ lattice with some configuration.

Step 2. For each site i, calculate the total rate $R_i = \sum_p r_p$ of all events by summing the rates of all possible processes that can occur on that site. Select a random number $x \in (0, 1)$ and calculate a waiting time from the exponential distribution, $\tau_i = (-1/R_i) \ln x$. For each site i, store the *absolute* time $t_i = t + \tau_i$ at which the next event at this site will occur (t is the current time).

Step 3. Find the site k with the smallest time t_k. Pick an event q to occur at this site with probability r_q/R and update the state of site k accordingly.

Step 4. Update the current time $t \to t_k$. For the updated site k and all of its neighbours, perform step 2: calculate and store the new time at which the next event will occur at each of these sites.

Step 5. Repeat steps 3 and 4 until time some defined end-time T_f is exceeded.

7.8 Extensions to the contact process

The contact process is the canonical model for investigating the role of spatial structure in population and infection dynamics. As such it has formed the basis of numerous models that extend the CP to include additional complexities that more closely resemble real systems. For example, the CP has been generalized to contain multiple species and used to investigate species abundances [152] and species-area relationships [153]

in ecological models. A two species CP has been introduced motivated by symbiosis; both species behave as standard CPs but when members from each species occupy the same lattice site, there is a symbiotic interaction that leads to a reduced death rate for both individuals [154]. Variations of the CP have also included a maturation period in which newly occupied juvenile sites must wait some time interval before they are able to reproduce [155, 156], as well as a more general notion of aging in which sites pass through a series of temporal stages with varying reproduction rates [157]. I will now motivate our own extension to the contact process.

7.8.1 Vertical and horizontal transmission

The CP in its original form represents a fixed population in which some infectious disease is spreading by the direct contact of individuals. Many infectious processes can however undergo two modes of transmission: *horizontal* transmission, when infection occurs between two members of the population as in the CP, and *vertical* transmission, where offspring inherit the disease or genetic trait from their parents. There are numerous examples of human diseases which occur via both mechanisms, such as the HIV and several hepatitis viruses, as well as certain parasitic infections, including helminths [158] and microsporidians [159]. Here the dynamics of the infection and the vital dynamics (birth and death) of the host population may be coupled in a non-trivial way, and an interesting interplay between vertical and horizontal transmission can emerge.

Both modes of transmission have played important roles in evolution. Traditionally, evolution was percieved to be the result of descent with modification: natural selection acting on traits passed vertically from parent to offspring. It is now known however that this vertical transmission is not the only method of acquiring genetic information, and in many systems in the microbial world, may not even be the dominant one. Horizontal gene transfer (HGT) allows genes to be passed between individuals in a community [160]. One way in which this occurs is through *conjugation*, where genetic material in the form of a loop of DNA, known as a plasmid, is transferred to neighbouring bacteria [161]. The genetic information being shared can enable the spread of traits such as antibiotic resistance, and can enable recipients to occupy new ecological niches.

Despite the abundance of examples in which both vertical and horizontal transmission take place, a surprisingly small number of models in the ecological literature take both into account. One reason for this is that in many of the most infectious diseases, the timescale on which the infection is gained and lost is much shorter than the

lifespan of an individual. Apart from [162] (in which it is proposed that the model discussed is the first spatial model to analyze vertical transmission), none of the other models investigating the general interplay between vertical and horizontal transmission included spatial structure, instead employing a deterministic mean field approach [163–167]. The small number of exceptions encountered were for system-specific models, such as the first spatially explicit, lattice based models used to study plasmid transfer [121, 168], and a model of virus-infected fungus on chestnut trees [169].

Motivated by this, the aim of the following chapter is to introduce and explore a novel, yet natural extension to the original contact process. Briefly, the model will resemble *stacked contact processes*, wherein each dynamic process obeys the update rules of the standard CP but exists on the lattice of active sites generated by the lower-level CPs. For example, the lowest level CP may govern the dynamics of a population of organisms, while the next CP level describes the dynamics of a parasite that can only exist on sites in which there is a host organism, and that is vertically transferred when its host organism reproduces. The model thus captures the crucial features of the above processes, namely the presence of both vertical and horizontal transmission within a spatial and stochastic framework. This model will be the subject of the next chapter.

Chapter 8

Stacked contact processes: Extending the contact process to include both horizontal and vertical transfer

8.1 Introduction

In this chapter I will first define our model of *stacked contact processes*, introduced in the last chapter, appealing to the biological context of a parasitic infection. Through this novel extension to the standard contact process, i.e. stacking contact processes one on top of the other, we develop a generic statistical physics model that captures the interplay between vertical and horizontal transmisson in a spatially explicit, stochastic model. We will first analyze the steady state behaviour of the two-level model and show that even in the simplest case, where the infection does not incur any cost, we uncover a new effect: a nonmonotonic dependence of parasite prevalence on host turnover. By calculating both the basic mean-field theory, and an improvement on this known as the ordinary pair approximation, we show that this behaviour arises from the strong spatial correlations in the lattice model. This exposes a highly nontrivial coupling between the spatial and temporal scales, since host population turnover affects the spatial clustering of infected organisms, which in turn affects the prevalence of the infection within the population. I will then discuss the potentially rich behaviour of the phase transitions in this model, and relate this to the literature on disordered contact processes. Finally,

inspired by natural examples of hyperparasitism (parasites of parasites), we will extend our model to include multiple levels of contact processes, and identify a transition between the maintenance of a finite and infinite number of levels, which we conjecture is connected to a roughening transition in models of surface-growth.

8.2 Modelling a two-level stacked contact process

We begin by considering the dynamics of a host and a single parasite which can be transmitted vertically or horizontally: a two-level stacked CP on a two-dimensional square lattice. In this model, we assume that lattice sites can be either empty or occupied by a host organism. Occupied lattice sites become empty at rate d_0, due to death of hosts, and host organisms attempt to replicate into neighbouring empty sites at rate b_0. Note that the attempt rate in any direction is thus b_0/z, where z is the coordination number of the lattice. In this model $z = 4$ since we use the standard von Neumann neighbourhood, in which the neighbourhood of a site consists of the four sites located directly north, south, east and west.

Host organisms may also carry some transmissible trait; for instance a parasite, or a plasmid. For simplicity we will assume here that this will incur no fitness cost or benefit – the trait is assumed to be neutral in the sense that organisms with and without will have identical birth and death rates[1]. Infected individuals pass the parasite to uninfected neighbours at rate b_1, and the parasite is lost at a constant rate d_1, through some natural process or immune response. The offspring of an infected organism is always infected via vertical transmission, and when an infected host dies the parasite dies with it. Thus the infection dynamics consists of a secondary CP, with parameters b_1 and d_1, which sits on top of the host's birth-death CP. Note that if the host population does not occupy the whole lattice, then the secondary CP does not take place on a regular lattice but rather on the irregular network of occupied sites which changes stochastically due to the dynamics of the host population.

The processes that occur in the model are illustrated in Figure 8.1. The model is simulated using an event-driven kinetic Monte Carlo algorithm (Section 7.7.2), in which all events occur as Poisson processes, on a 100×100 lattice with periodic boundary conditions. In this thesis we will be concerned with the steady state behaviour of the system, i.e. with the question of whether or not the parasite (or genetic element) is

[1] Although this is often not the case in reality – host-parasite interactions can range from antagonistic to mutalistic – for many parasites it is reasonable that they might benefit by minimizing any damage to their host.

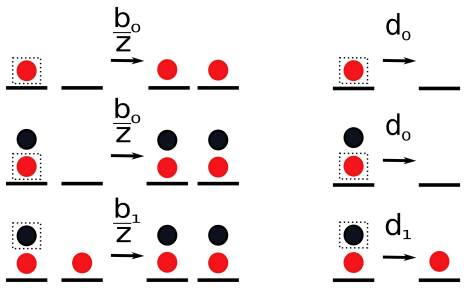


Figure 8.1 The events that constitute the 2-level stacked CP. Red and black circles represent host organisms and parasites respectively. A site can be empty, occupied by a susceptible host or occupied by an infected host. The left panels represent reproduction of a susceptible or infected host (top and middle, rate b_0), and transmission of the parasite (bottom, rate b_1). The right panels represent death of a susceptible or infected host (top and middle, rate d_0), and loss of the parasite (bottom, rate d_1). The coordination number of the lattice z is the number of nearest neighbour sites. The dotted squares indicate the level of the CP at which the event happens.

maintained in the population at steady state. We will therefore compute averages of quantities like the host and parasite densities, where all averages are calculated after an initial transient to allow the steady state to be reached.

8.2.1 Mathematical representation of the system

In this chapter we will derive some mathematical relations to describe the behaviour of our model. To represent this system mathematically, it is useful to define indicator variables to represent the states of the lattice sites. Defining $\tau_i = 0$ or 1 if site i is empty or occupied respectively, and $\sigma_i = 1$ if site i contains an infected host and $\sigma_i = 0$ otherwise (i.e. if site i is either unoccupied or contains a susceptible host) ensures that the configuration $\mathcal{C}$ of the lattice is fully specified by the set of microscopic variables τ_i and σ_i, $\mathcal{C} = \mathcal{C}(\{\tau, \sigma\})$. For a general observable A that depends on the state of the system, its ensemble average is

$$\langle A \rangle = \sum_{\mathcal{C}} P(\mathcal{C}, t) A(\mathcal{C}) \equiv \sum_{\{\tau,\sigma\}} P(\{\tau, \sigma\}, t) A(\{\tau, \sigma\}) \,, \tag{8.1}$$

where $P(\mathcal{C}, t)$ is the probability of the system being in configuration $\mathcal{C}$ at time t. The time derivative of A can be expressed as

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle A \rangle = \sum_{\{\tau,\sigma\}} \sum_{\{\tau',\sigma'\}} P(\{\tau, \sigma\}, t) \omega(\{\tau, \sigma\} \to \{\tau', \sigma'\}) \left[A(\{\tau', \sigma'\}) - A(\{\tau, \sigma\}) \right] \tag{8.2}$$

where $\omega(\{\tau,\sigma\} \to \{\tau',\sigma'\})$ is the transition rate between two microscopic configurations. This can be re-expressed as a sum over the set of elementary transitions $\mathcal{T} = \{\tau,\sigma\} \to \{\tau',\sigma'\}$ that can take place (the transitions shown in Figure 8.1). Thus

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle A\rangle = \sum_{\mathcal{T}} \big\langle I(\mathcal{T})\omega(\mathcal{T})\delta A(\mathcal{T})\big\rangle \tag{8.3}$$

where $I(\mathcal{T})$ is, for a given microscopic configuration $\mathcal{C}$ in the average, an indicator variable that equals unity if the transition $\mathcal{T}$ is permitted and zero otherwise, and $\delta A(\mathcal{T})$ is the change in the observable A due to the transition $\mathcal{T}$.

In order to list all possible transitions in the system in a clear manner, we introduce an additional notation. This will simplify things later (Section 8.4.1), and is simply a set of site labels $0, 1, 2$ where 0 represents an empty site, 1 represents a site occupied by a parasite-free host and 2 is a site occupied by a parasite-infected host. With this choice, the elementary update rules for the two-level process (Figure 8.1) can be expressed:

$$10 \overset{b_0/z}{\to} 11 \qquad 20 \overset{b_0/z}{\to} 22 \qquad 21 \overset{b_1/z}{\to} 22 \qquad 1 \overset{d_0}{\to} 0 \qquad 2 \overset{d_0}{\to} 0 \qquad 2 \overset{d_1}{\to} 1. \tag{8.4}$$

To clarify, consider how these two notations relate to each other. Since τ_i is the indicator variable for the presence of any host organism (infected or uninfected) on site i, and σ_i is the indicator for the presence of a parasite (i.e. infected host), we have

$$\tau_i = \begin{cases} 1 & \text{site } i \text{ is labelled 1 or 2} \\ 0 & \text{otherwise} \end{cases} \qquad \sigma_i = \begin{cases} 1 & \text{site } i \text{ is labelled 2} \\ 0 & \text{otherwise.} \end{cases} \tag{8.5}$$

The corresponding indicator variable for label 1, the *uninfected* host, is thus the difference $(\tau_i - \sigma_i)$. For label 0, an empty site, it is $(1 - \tau_i)$ for which we introduce the notation $\bar{\tau}_i = (1 - \tau_i)$.

Using this 0, 1, 2 site label notation, Table 8.1 lists all possible single site transitions along with the corresponding values of the quantities appearing in Eq. (8.3). Note that in the transition rates, the sum is over the z neighbouring sites, j, of site i.

8.2.2 First moments

We are interested in calculating the steady state densities of the host and parasite populations, captured by the averages of τ_i and σ_i across all lattice sites; the first moments of the system. Performing the sum of Eq. (8.3), using the quantities in Table

$\mathcal{T}$	$\omega(\mathcal{T})$	$I(\mathcal{T})$	$\delta\tau_i(\mathcal{T})$	$\delta\sigma_i(\mathcal{T})$
$0 \to 1$	$\frac{b_0}{z}\sum_j(\tau_j - \sigma_j)$	$\bar{\tau}_i$	1	0
$0 \to 2$	$\frac{b_0}{z}\sum_j \sigma_j$	$\bar{\tau}_i$	1	1
$1 \to 2$	$\frac{b_1}{z}\sum_j \sigma_j$	$\tau_i - \sigma_i$	0	1
$1 \to 0$	d_0	$\tau_i - \sigma_i$	-1	0
$2 \to 0$	d_0	σ_i	-1	-1
$2 \to 1$	d_1	σ_i	0	-1

Table 8.1 All single site transitions in the 2-level stacked CP along with their corresponding rate, indicator variable and resulting change in observables as defined in Eq. (8.3). Note that $\bar{\tau}_i \equiv (1 - \tau_i)$.

8.1 first for the case where the observable A is τ_i, and secondly for σ_i, yields:

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\tau_i\rangle = \frac{b_0}{z}\sum_j\langle\bar{\tau}_i\tau_j\rangle - d_0\langle\tau_i\rangle \tag{8.6}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\sigma_i\rangle = \frac{b_0}{z}\sum_j\langle\bar{\tau}_i\sigma_j\rangle + \frac{b_1}{z}\sum_j\langle(\tau_i - \sigma_i)\sigma_j\rangle - (d_0 + d_1)\langle\sigma_i\rangle\ . \tag{8.7}$$

The angle brackets denote averages over multiple realizations of the dynamics. These equations make intuitive sense. The first term on the r.h.s. of Eq. (8.6) corresponds to the birth of a host: an empty site is filled by replication of a host, while the second term represents the death of a host organism. The first term on the r.h.s. of Eq. (8.7) corresponds to vertical transmission: an empty site is filled by replication of an infected host. The second term represents horizontal transmission: a susceptible host, denoted by $(\tau_i - \sigma_i)$, is infected by a parasite-carrying neighbour. The final term corresponds to the combined loss of parasite and death of an infected host. These equations can in principle be simply written down by inspecting the elementary rules of the model, however the systematic procedure presented above is useful to ensure that all transitions are accounted for when working with higher order moments, as we will see in Section 8.4.1.

8.3 Phase diagram: Host dynamics influences parasite persistence

The first question we can ask of the model is under what conditions is a non-zero steady state density of parasite maintained? In this section we explore the parameter space

and construct the phase diagram for the two-level model. As a result we discover that the parasite persistence is actually affected by the underlying host dynamics, in what turns out to be a highly nontrivial manner.

In this model, the host population undergoes a simple CP, whose stationary properties are fully determined by the single parameter $\lambda_0 = b_0/d_0$ (Section 7.5). If λ_0 is less than a critical value, $\lambda_{\rm crit}$, the only steady state is an empty lattice and the population rapidly becomes extinct. If $\lambda_0 > \lambda_{\rm crit}$, a non-zero steady state host population can be maintained (though technically in any finite system this will eventually become extinct due to a rare large fluctuation). The transition between these two regimes is second order with the values of $\lambda_{\rm crit}$ and the associated critical exponents having been characterized in detail [131] (Section 7.5). Expanding Eq. (8.6), we see that the dynamics of the host population obeys

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\tau_i\rangle = \frac{1}{z}\sum_j b_0\langle\tau_j\left(1-\tau_i\right)\rangle - d_0\langle\tau_i\rangle \ . \tag{8.8}$$

Eq. (8.8) is thus the equation associated with the standard contact process. The steady state solution of Eq. (8.8) satisfies $\langle\tau_i\rangle = (\lambda_0/(\lambda_0-1))\langle\tau_i\tau_j\rangle$. Since we are interested in the steady state in a spatially homogeneous system, $\langle\tau_i\rangle$ should be independent of both time and space, and one can simply write $\langle\tau\rangle$ wherever any average of a single τ_i variable appears. Likewise, in Eq. (8.8) the second moment $\langle\tau_i\tau_j\rangle$ appears only for sites i and j that are nearest neighbours, and takes a common value for any i and j which we denote $\langle\tau\tau\rangle$. If we choose to ignore correlations between neighbouring sites by assuming that $\langle\tau_i\tau_j\rangle \approx \langle\tau_i\rangle\langle\tau_j\rangle$, we arrive at the mean-field result for the steady state host density:

$$\langle\tau\rangle^{\rm MF} = 1 - (1/\lambda_0). \tag{8.9}$$

Consider now the dynamics of the parasite. Expanding Eq. (8.7) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\sigma_i\rangle = \frac{1}{z}\sum_j\left[b_0\langle\sigma_j\rangle + (b_1-b_0)\langle\tau_i\sigma_j\rangle - b_1\langle\sigma_i\sigma_j\rangle\right] - (d_0+d_1)\langle\sigma_i\rangle \tag{8.10}$$

and comparing this with Eq. (8.8), we see that the parasite dynamics is similar to that of a standard CP but with an extra term: the cross-correlation $\langle\tau_i\sigma_j\rangle$ perturbs the parasite dynamics from that of a standard CP. Interestingly, we see that if $b_1 = b_0$ (i.e. if the rates of horizontal and vertical transmission are equal), this term vanishes and the form of Eq. (8.10) becomes that of a standard CP with parameter $\lambda_{\rm eff} = b/(d_0+d_1)$. Later in the chapter we will see that this has consequences for the maintenance of multiple parasite levels.

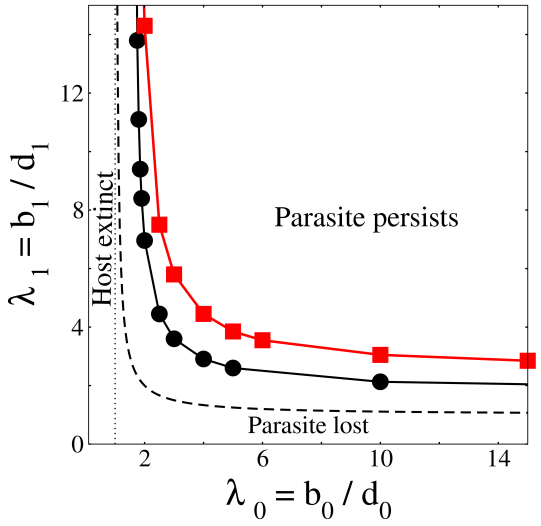


Figure 8.2 Phase diagram of the two-level stacked CP. Symbols: simulation data showing boundary of parasite persistence, for $d_1 = 1$ and $d_0 = 1$ (circles) or $d_0 = 20$ (squares). Dashed line: mean-field theory predicts parasite persistence for $\lambda_1 > \lambda_0/(\lambda_0 - 1)$.

In the mean-field limit, where spatial correlations are neglected, we have that $\langle \tau_i \sigma_j \rangle \approx \langle \tau_i \rangle \langle \sigma_j \rangle$ and $\langle \sigma_i \sigma_j \rangle \approx \langle \sigma_i \rangle \langle \sigma_j \rangle$. Applying this to Eq. (8.10), and defining $\lambda_1 = b_1/d_1$, yields a steady state solution for the density of the parasite:

$$\langle \sigma \rangle^{MF} = 1 - (1/\lambda_0) - (1/\lambda_1) \equiv 1 - (1/\lambda_{\text{eff}}), \tag{8.11}$$

where $\lambda_{\text{eff}} = \lambda_0 \lambda_1/(\lambda_0 + \lambda_1)$. In the mean field theory then, the parasite density is controlled solely by the two parameters λ_0 and λ_1 which uniquely relate to the lower and upper CPs respectively.

Eqs. (8.10) and (8.11) thus allow us to plot a phase diagram of the model, shown in Figure 8.2 as a function of λ_0 and λ_1. Three steady state scenarios are possible: (i) the host population is extinct, (ii) the host population is finite but the parasite is extinct, and (iii) the parasite persists within a finite host population. Since the dynamics of the host population is a standard CP, the host population is extinct, scenario (i), if $\lambda_0 < \lambda_{\text{crit}}$. For $\lambda_0 > \lambda_{\text{crit}}$, the host population persists, with or without the parasite. The condition for parasite persistence predicted by the mean field theory, $\lambda_1 > \lambda_0/(\lambda_0 - 1)$, is obtained by setting $\langle \sigma \rangle = 0$ in Eq. (8.11), and is shown by the dashed line in Figure 8.2. As one might expect, as λ_0 decreases, the density of the host population decreases, and a higher rate of horizontal transmission λ_1 is needed to maintain the parasite.

The symbols in Figure 8.2 show the boundaries between parasite persistence and loss

obtained from our Monte Carlo simulations for two values of the host death rate d_0. The fact that the simulation data are shifted upwards and to the right of the mean-field prediction shows that spatial correlations make it harder to maintain the parasite. Note also that different results are obtained when different values of d_0 are used, meaning that the parameters λ_0 and λ_1 do not fully determine the phase behaviour of the real system, as predicted by the mean field theory. For fixed $\lambda_0 = b_0/d_0$, a higher turnover rate of the host population (i.e. higher d_0 and b_0) appears to make it harder to maintain the parasite.

In these simulations, the critical value of λ_1 (with $d_1 = 1$) for a given λ_0 value was first estimated by eye. Then, starting well below this estimated value in the absorbing phase, from an initial condition of a fully parasite-occupied lattice, three runs were performed for a series of increasing λ_1 values. Defining a large cut-off time t_f (corresponding to about 10^9 lattice updates) the system was deemed to be in the active state if all three runs allowed the parasite to persist past t_f. Checking the parasite density time-series for these runs confirmed that a steady state exists. These low densities will be prone to extinction due to fluctuations however, so the above procedure necessarily overestimates the true critical value. However the quantitative accuracy is more than enough to produce a reliable phase diagram. For the majority of λ_0 values, it is easily possible to obtain the critical point to an accuracy of one decimal place using this method (the two persistence curves differ by a much greater amount than this). This error grows on the left hand side of the plot, where approaching the critical λ_0 value, the small host population experiences greater fluctuations.

8.3.1 Parasite density depends on host population turnover rate as a result of spatial clustering

When we investigate the behaviour of the model in more detail using simulations, it turns out that the situation is more complex. Figure 8.3a shows a more comprehensive investigation of the steady state parasite density $\langle\sigma\rangle$ as a function of the host turnover rate. In these simulations, b_0 and d_0 are varied while keeping $\lambda_0 = b_0/d_0$ fixed, meaning that the steady state density of the host population is constant since it depends only on λ_0. Remarkably, we see that for some parameter combinations, the density of the parasite actually depends nonmonotonically on the host turnover rate. This nonmonotonic trend implies that if the host dynamics is slow, increasing the turnover rate increases the parasite density, while if the host dynamics is fast, the parasite density decreases with the turnover rate. We checked this behaviour for a range of lattice sizes.

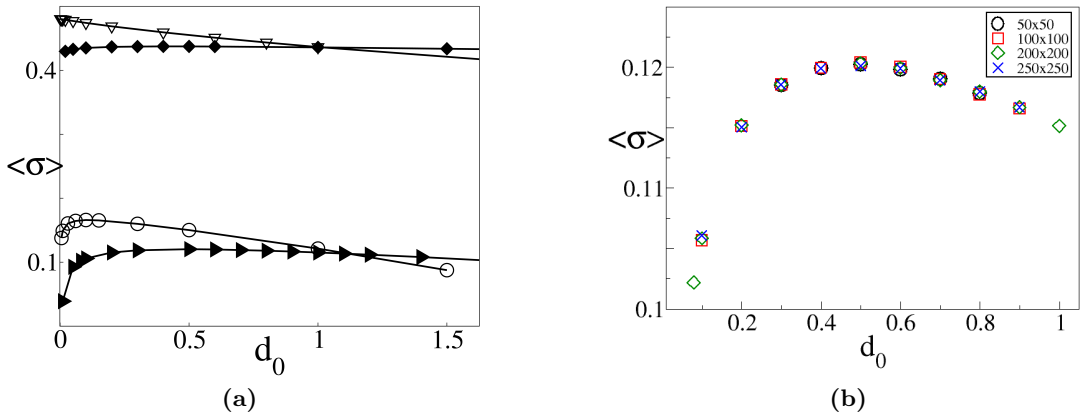


Figure 8.3 (a) Parasite density $\langle\sigma\rangle$ as a function of host turnover rate d_0 for fixed λ_0 (and hence fixed host density). Curves top to bottom correspond to $(\lambda_0, b_1, d_1) = (5, 4, 1)$,$(2.5, 15, 1)$,$(2, 8, 1)$ and $(1.8, 15, 1)$. (b) The $(\lambda_0, b_1, d_1) = (1.8, 15, 1)$ curve for parasite density plotted for various lattice sizes shows that this nonmonotonicity is not due to any finite size effect but is a feature of the model.

Figure 8.3b shows that the nonmonotonic trend is preserved, demonstrating that this is not a finite-size effect arising from our choice of a 100×100 grid, but is a feature of our model.

It is important to note that this nonmonotonicity is not predicted by the mean field theory, which predicts instead a constant steady state value for $\langle\sigma\rangle$, given by Eq. (8.11). The nonmonotonicity must therefore be a consequence of the spatial correlations present in the lattice system. To investigate this in our simulations, we make two measurements that capture information about the spatial distribution of the parasite. The first measurement is the radial distribution function, given by

$$g(r) = \frac{\langle\sigma_r\sigma_0\rangle}{\langle\sigma\rangle} - \langle\sigma\rangle. \tag{8.12}$$

Here, the first term is the conditional probability of finding a parasite at a distance r from some origin 0, given that there is a parasite at 0. The second term ensures that $g(r) \to 0$ as $r \to \infty$, where r is the Euclidean distance between two lattice sites. Plotting $g(r)$ from our simulations, we find that the correlations decay exponentially with distance. We therefore define a correlation length ξ, which indicates the degree of parasite clustering, through

$$g(r) = Ae^{-r/\xi}\ , \tag{8.13}$$

where A is some constant. From the plot of $g(r)$ for each value of d_0, the correlation length was extracted by fitting the above exponential form to the data using the *Regression* method in the Xmgrace plotting software. Figure 8.4a plots the correlation

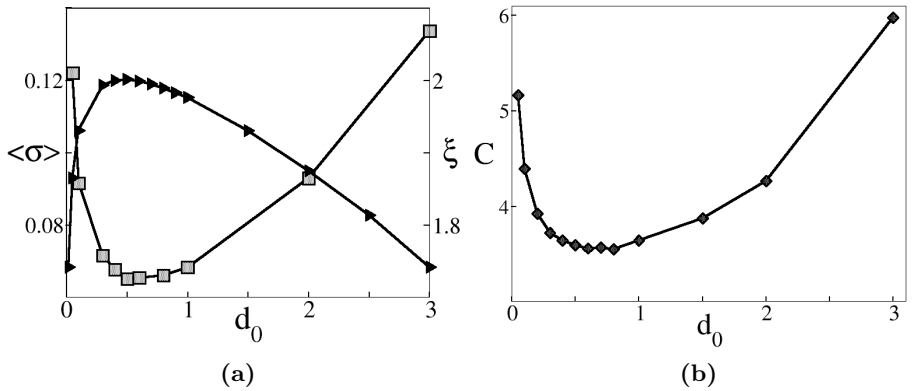


Figure 8.4 Parasite clustering leads to lower parasite levels. (a) Parasite density $\langle\sigma\rangle$ (triangles, left axis) and spatial correlation length ξ (squares, right axis) for $(\lambda_0, b_1, d_1) = (1.8, 15, 1)$. (b) Measure of local clustering, $C = \langle\sigma_i\sigma_j\rangle/\langle\sigma\rangle^2$, where i and j are neighbouring sites.

length ξ against d_0, for a fixed $\lambda_0 = b_0/d_0$. Strikingly, the correlation length is minimal at the turnover rate where the parasite density is maximal. This makes intuitive sense: horizontal transmission requires contact between infected and uninfected hosts, so a greater spatial clustering of the parasite population should result in a lower parasite density, since this clustering will reduce the number of contacts between infected and uninfected individuals.

To gain further insight, we also investigate the local parasite clustering, C, that we define by the function

$$C = \frac{\langle\sigma_i\sigma_j\rangle}{\langle\sigma\rangle^2} \,, \tag{8.14}$$

where i and j are neighbouring sites. Comparing this to Eq. (8.12) we can see that see that C is related to $g(r)$ through $C = (g(1) + \langle\sigma\rangle)/\langle\sigma\rangle$. In the absence of spatial structure, $\langle\sigma_i\sigma_j\rangle = \langle\sigma\rangle^2$ and thus $C = 1$; this is the lowest value C can take. If there are any spatial correlations, then $\langle\sigma_i\sigma_j\rangle > \langle\sigma\rangle^2$ and C will increase from 1 as the degree of these correlations increases. Figure 8.4b shows that this alternative measure of parasite clustering also supports the hypothesis that high clustering results in lower parasite density.

Visualizing the simulations allows us to understand why the turnover rate of the host affects the spatial clustering of the parasite. At very high host turnover rates, the simulations show that local patches of host organisms rapidly grow from a single seed and vanish by stochastic extinction, on a timescale much faster than that of horizontal transmission. Because offspring are always of the same type as their parents, organisms in one of these patches are either all infected or all uninfected. Thus as the host turnover rate increases, the infected population becomes more highly clustered, leading to lower

rates of horizontal transmission and lower parasite density.

In contrast, when the host population turnover is slow, the parasite dynamics constitutes a CP on the effectively frozen, disordered network of lattice sites that are occupied by host organisms. This network contains clusters of sites that may be poorly connected to, or entirely disconnected from, the rest of the host population which creates the possibility of local, stochastic parasite extinctions. In this regime, an increase in the host turnover rate serves as a mixing mechanism, by more homogeneously distributing the host population across the lattice. This provides a greater opportunity for the parasite to spread by connecting clusters before any local extinction occurs. This is most obviously the case in the low λ_0 regime where the host density is low, but it is still true for higher host densities. Also, the nonmonotonicity will be most pronounced when the value of λ_0 is low enough to give a steady state host density $\langle\tau\rangle$ that is below the lattice percolation threshold. When this is the case, in the limit $d_0 \to 0$ there is no percolating host cluster for the parasite to exist upon, and so it must be that $\langle\sigma\rangle = 0$.

In summary, the nonmonotonic dependence of the parasite density on the host turnover rate arises from a coupling of the spatial fluctuations of the underlying host CP to the parasite dynamics, and can be viewed as a competition between the mixing effect of the host birth-death process at low turnover rates and the population segregation arising from parent-offspring clustering at high turnover rates. It is a direct consequence of the spatial structure of the model and the inclusion of vertical transmission. To our knowledge this interplay between the spatial structure and timescales of the processes is new, and has not been discussed before in such CP inspired models. In an attempt to better understand this behaviour we now present an extension to the mean field theory; one which captures some spatial structure.

8.4 Ordinary pair approximation: improvement to mean field theory

Lattice simulations are computationally expensive, even for modest lattice sizes, often requiring multiple realizations to obtain the statistics of interest. Hence finding analytic approximations is desirable, as they allow the behaviour of the model to be explored many orders of magnitude faster, and can facilitate a better understanding in general. It is often the case, and as we have seen here, that the presence of spatial correlations can give rise to phenomena not captured by the mean field approximation, which entirely

neglects these correlations. We therefore try to explain analytically the nonmonotonic behaviour of Figure 8.4 by extending the mean field theory to include some degree of spatial structure.

An important improvement to the mean field assumption for a lattice system is the ordinary pair approximation (OPA) [170]. This approximation extends the mean field equations so as to capture first-order correlations in space. In many models, this is sufficient to qualitatively capture the most important behaviour, and often yields quantitatively useful results. Eqs. (8.8) and (8.10) show that the single site densities depend on the densities of pairs of sites. In the next section, when we write down the equations governing these pairs, we will find that they in turn depend on triplets of sites. This is known as the closure problem: to fully describe the state of the system requires an infinite hierarchy of equations, involving ever higher orders of correlation. To obtain an analytic approximation then, one needs to close these equations at some level. The OPA is one closure method, and involves expressing the triplet densities in terms of single-site and pair densities. It thus accounts for local correlations but neglects any higher order clustering.

Presented below is the OPA for our two-level stacked contact process, along with its predictions for parasite density, clustering and the phase diagram of the system. We now look at how to derive the equations governing the two-site densities.

8.4.1 Two-point functions

The two-point functions are necessarily more complicated than the one-point functions previously discussed. Consider two neighbouring sites labelled i and j. A configuration XY indicates that lattice site i is in state X and site j is in state Y and recall that each site is labelled 0, 1 or 2 if it is empty, occupied by an uninfected host or occupied by an infected host respectively. It will be useful to write down the conditions under which the various products, that is the indicator variables for the two-point functions $\tau_i\tau_j$, $\tau_i\sigma_j$ and $\sigma_i\sigma_j$ are equal to 0 or 1:

$$\tau_i\tau_j = \begin{cases} 0 & \text{if the configuration contains a 0} \\ 1 & \text{otherwise} \end{cases} \tag{8.15}$$

$$\tau_i\sigma_j = \begin{cases} 1 & \text{for configurations 12 and 22} \\ 0 & \text{otherwise} \end{cases} \tag{8.16}$$

$$\sigma_i\sigma_j = \begin{cases} 1 & \text{for configuration 22} \\ 0 & \text{otherwise.} \end{cases} \tag{8.17}$$

An equivalent table to Table 8.1, but showing all possible transitions that can occur when considering pairs of sites, is given in Appendix B.1. Using this table allows the enumeration of all possible transitions between two-site configurations, and the corresponding quantities that appear in Eq. (8.3). Performing this sum for the nearest neighbour two-point functions leads to the following equations:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\langle\tau\tau\rangle &= \frac{2b_0}{z}\big[(z-1)\langle\tau\bar{\tau}\tau\rangle + \langle\bar{\tau}\tau\rangle\big] - 2d_0\langle\tau\tau\rangle = 0 && (8.18)\\
\frac{\mathrm{d}}{\mathrm{d}t}\langle\tau\sigma\rangle &= \frac{2b_0}{z}\big[(z-1)\langle\tau\bar{\tau}\sigma\rangle + \langle\bar{\tau}\sigma\rangle\big] + \frac{b_1}{z}\big[(z-1)\langle\tau(\tau-\sigma)\sigma\rangle + \langle(\tau-\sigma)\sigma\rangle\big] \\
&\quad - (2d_0+d_1)\langle\tau\sigma\rangle = 0 && (8.19)\\
\frac{\mathrm{d}}{\mathrm{d}t}\langle\sigma\sigma\rangle &= \frac{2b_0}{z}\big[(z-1)\langle\sigma\bar{\tau}\sigma\rangle + \langle\bar{\tau}\sigma\rangle\big] + \frac{2b_1}{z}\big[(z-1)\langle\sigma(\tau-\sigma)\sigma\rangle + \langle(\tau-\sigma)\sigma\rangle\big] \\
&\quad - 2(d_0+d_1)\langle\sigma\sigma\rangle = 0\,. && (8.20)
\end{aligned}$$

Together with the one-point functions that we derived earlier

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\langle\tau\rangle &= (b_0-d_0)\langle\tau\rangle - b_0\langle\tau\tau\rangle = 0 && (8.21)\\
\frac{\mathrm{d}}{\mathrm{d}t}\langle\sigma\rangle &= b_0\langle\bar{\tau}\sigma\rangle + b_1\langle(\tau-\sigma)\sigma\rangle - (d_0+d_1)\langle\sigma\rangle = 0\,, && (8.22)
\end{aligned}$$

Eqs. (8.18)–(8.22) provide an exact description of the two-level stacked CP. Here the three-point averages $\langle xyz\rangle$ relate to quantities along a chain of three contiguous sites *in any configuration*. A possible improvement on this would be to use a different average for the distinct cases of a straight three-site chain and an L-shaped chain. However, in the following we will use a moment closure approximation which is not sensitive to such differences in chain morphology.

8.4.2 Moment closure strategy

Eqns. (8.18)–(8.22) demonstrate the closure problem mentioned before: the equations for the two-point functions depend on three-point functions (and the equations for these three-point functions will in turn depend on the four-point functions and so on). To close the equations, we need a suitable strategy for replacing the three-point functions with combinations of one- and two-point functions. The approach taken here is of the ordinary pair approximation, whereby in any chain of three sites there is no explicit dependence of the state of the third site on that of the first; it only enters implicitly through the intermediate site.

To make this clear, let x, y and z be random variables associated with three consecutive sites. Then, we approximate the joint probability distribution $P(x,y,z) =$

$P(x)P(y|x)P(z|x,y)$ with

$$P(x,y,z) \approx P(x)P(y|x)P(z|y) \approx P(y)P(x|y)P(z|y) \approx P(z)P(y|z)P(x|y) \; . \tag{8.23}$$

For indicator variables, x, y, z taking values 0 or 1, it is clear that for a single site $P(x) = \langle x \rangle$. For a pair of sites, $\langle xy \rangle$ is the probability that one site is x and the other is y, which means that since $P(x,y) = P(x)P(y|x)$, we have that $P(y|x) \equiv \langle xy \rangle / \langle x \rangle$. Thus in terms of our three-point correlations we may write from, Eq. (8.23),

$$\langle xyz \rangle \approx \langle x \rangle \frac{\langle xy \rangle}{\langle x \rangle} \frac{\langle yz \rangle}{\langle y \rangle} \approx \langle y \rangle \frac{\langle xy \rangle}{\langle y \rangle} \frac{\langle yz \rangle}{\langle y \rangle} \approx \langle z \rangle \frac{\langle yz \rangle}{\langle z \rangle} \frac{\langle xy \rangle}{\langle y \rangle} \; . \tag{8.24}$$

Examining Eq. (8.24) we see that all three possible ways of breaking down the joint probability distribution into conditional distributions give the same answer under this approximation:

$$\langle xyz \rangle \approx \frac{\langle xy \rangle \langle yz \rangle}{\langle y \rangle} \; . \tag{8.25}$$

Eq. (8.25) forms the basis of the OPA, enabling the three-point functions to be expressed as one- and two-point functions. The OPA reduces to the traditional mean field case if the two-point functions also factorize:

$$\langle xyz \rangle \approx \frac{\langle x \rangle \langle y \rangle \langle y \rangle \langle z \rangle}{\langle y \rangle} = \langle x \rangle \langle y \rangle \langle z \rangle \; . \tag{8.26}$$

In this case, the three-point functions also factorize showing that such an approximation is at least logically consistent. Where the barred quantity $\bar{\tau}$ appears as the central term in a three-point average, this moment closure procedure gives

$$\langle x \bar{\tau} y \rangle \approx \frac{\langle x \bar{\tau} \rangle \langle \bar{\tau} y \rangle}{\langle \bar{\tau} \rangle} = \frac{\big[\langle x \rangle - \langle x \tau \rangle\big]\big[\langle y \rangle - \langle \tau y \rangle\big]}{1 - \langle \tau \rangle} \; . \tag{8.27}$$

8.4.3 Exact solution to the OPA equations

Using this moment closure procedure it is actually possible to find an exact analytical solution to the steady state of Eqns. (8.18)–(8.22). It is rare to find exact solutions to these approximations, but it is desirable as it avoids any complications that may arise when solving the equations of motion numerically and is obviously much more computationally efficient. Considering just the host to begin with, we have from Eq. (8.21)

$$\langle \tau \tau \rangle = \left(1 - \frac{d_0}{b_0}\right) \langle \tau \rangle \; . \tag{8.28}$$

Thus

$$\langle \tau\bar{\tau}\tau\rangle \approx \frac{\left[\langle\tau\rangle - \langle\tau\tau\rangle\right]^2}{1-\langle\tau\rangle} = \frac{1}{1-\langle\tau\rangle}\left(\frac{d_0\langle\tau\rangle}{b_0}\right)^2 . \tag{8.29}$$

Substituting this expression into Eq. (8.18) and eliminating the two-point functions using Eq. (8.28) gives

$$\frac{b_0}{z}\left[\frac{z-1}{1-\langle\tau\rangle}\left(\frac{d_0\langle\tau\rangle}{b_0}\right)^2 + \frac{d_0}{b_0}\langle\tau\rangle\right] - d_0\left(1-\frac{d_0}{b_0}\right)\langle\tau\rangle = 0 \tag{8.30}$$

$$\Longrightarrow \left(1-\frac{1}{z}\right)\frac{1}{\lambda_0}\langle\tau\rangle + \frac{1}{z}\left(1-\langle\tau\rangle\right) - \left(1-\frac{1}{\lambda_0}\right)\left(1-\langle\tau\rangle\right) = 0 \tag{8.31}$$

where we have eliminated the trivial solution $\langle\tau\rangle = 0$ and introduced $\lambda_0 = \frac{b_0}{d_0}$. The solution of the resulting linear equation is

$$\langle\tau\rangle = \frac{1-\frac{1}{\lambda_0}-\frac{1}{z}}{1-\frac{1}{z}\left(1+\frac{1}{\lambda_0}\right)} = \frac{\lambda_0 z - \lambda_0 - z}{\lambda_0 z - \lambda_0 - 1} . \tag{8.32}$$

Three important outcomes can be observed at this stage. Firstly, the critical value of λ is shifted *upwards* from the traditional mean-field value of 1 to $\lambda_0^{\rm crit} = \frac{z}{z-1}$. This shift is qualitatively consistent with the lattice simulations (Figure 8.2). Secondly, as $z \to \infty$, $\langle\tau\rangle \to 1 - \frac{1}{\lambda_0}$, the traditional mean-field value, and the correlation function, Eq. (8.28), factorizes. Thirdly, $0 < \langle\tau\rangle \leq 1$ for $\lambda_0 > \lambda_0^{\rm crit}$ and so the host density is still bounded between zero and one.

Turning now to the equations involving the parasite, we note that Eq. (8.22) involves the three remaining unknowns ($\langle\sigma\rangle$, $\langle\tau\sigma\rangle$ and $\langle\sigma\sigma\rangle$) but neither of the knowns ($\langle\tau\rangle$ or $\langle\tau\tau\rangle$). However, dividing Eq. (8.22) through by $\langle\sigma\rangle$ yields

$$\frac{\langle\sigma\sigma\rangle}{\langle\sigma\rangle} = \frac{b_0 - d_0 - d_1}{b_1} + \left(1-\frac{b_0}{b_1}\right)\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle} . \tag{8.33}$$

Now, by using the moment-closure approximation in Eq. (8.19) and dividing through by $\langle\sigma\rangle$, we see that

$$\begin{aligned}&\frac{2b_0}{z}\left((z-1)\frac{\langle\tau\rangle - \langle\tau\tau\rangle}{1-\langle\tau\rangle}+1\right)\left(1-\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle}\right)\\ &\quad + \frac{b_1}{z}\left[(z-1)\left(\frac{\langle\tau\tau\rangle}{\langle\tau\rangle}\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle} - \frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle}\frac{\langle\sigma\sigma\rangle}{\langle\sigma\rangle}\right) + \frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle} - \frac{\langle\sigma\sigma\rangle}{\langle\sigma\rangle}\right] - (2d_0+d_1)\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle} = 0 .\end{aligned} \tag{8.34}$$

Using Eqs. (8.28) and (8.33) to eliminate $\langle\tau\tau\rangle/\langle\tau\rangle$ and $\langle\sigma\sigma\rangle/\langle\sigma\rangle$, we can rewrite this

as the quadratic equation

$$A\left(\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle}\right)^2 + B\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle} + C = 0\,, \tag{8.35}$$

with the coefficients

$$\begin{aligned} A &= (b_0 - b_1)(z-1)/z & (8.36)\\ B &= \frac{-2(z-1)d_0}{z}\frac{\langle\tau\rangle}{1-\langle\tau\rangle} + \frac{b_1(z-1)}{z}\left(1-\frac{d_0}{b_0}\right) - \frac{z-1}{z}(b_0 - d_0 - d_1) \\ &\quad -(2d_0 + d_1) - \frac{b_0}{z} & (8.37)\\ C &= \frac{1}{z}\left[2d_0(z-1)\frac{\langle\tau\rangle}{1-\langle\tau\rangle} + b_0 + d_0 + d_1\right]\,, & (8.38) \end{aligned}$$

the solution of which is

$$\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}\,. \tag{8.39}$$

Once the appropriate branch has been chosen, one can find $\frac{\langle\sigma\sigma\rangle}{\langle\sigma\rangle}$ from Eq. (8.33). Then, by applying the moment-closure approximation, Eq. (8.25), to Eq. (8.20), we find

$$\begin{aligned} &\frac{2b_0}{z}\left[(z-1)\frac{(1-\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle})^2}{1-\langle\tau\rangle}\langle\sigma\rangle + 1 - \frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle}\right] + \\ &\qquad \frac{2b_1}{z}\left[(z-1)\left(\frac{\langle\tau\sigma\rangle}{\langle\tau\rangle}\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle} - \frac{\langle\sigma\sigma\rangle^2}{\langle\sigma\rangle^2}\right) + \frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle} - \frac{\langle\sigma\sigma\rangle}{\langle\sigma\rangle}\right] - \\ &\qquad\qquad 2(d_0+d_1)\frac{\langle\sigma\sigma\rangle}{\langle\sigma\rangle} = 0 \end{aligned} \tag{8.40}$$

where we have eliminated the trivial solution $\langle\sigma\rangle = 0$. Noting that the $\langle\tau\sigma\rangle/\langle\tau\rangle$ in the second term above can be expressed $(\langle\tau\sigma\rangle/\langle\sigma\rangle)\cdot(\langle\sigma\rangle/\langle\tau\rangle)$ gives

$$\langle\sigma\rangle = \frac{D}{E} \tag{8.41}$$

where D and E have the forms

$$\begin{aligned} D = &\frac{2b_0}{z}\left(\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle} - 1\right) + \frac{2b_1}{z}\left[(z-1)\left(\frac{\langle\sigma\sigma\rangle}{\langle\sigma\rangle}\right)^2 + \frac{\langle\sigma\sigma\rangle}{\langle\sigma\rangle} - \frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle}\right]\,, \\ &+ 2(d_0+d_1)\frac{\langle\sigma\sigma\rangle}{\langle\sigma\rangle} \end{aligned} \tag{8.42}$$

$$E = \frac{2b_0(z-1)}{z}\frac{\left(1-\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle}\right)^2}{1-\langle\tau\rangle} + \frac{2b_1(z-1)}{z}\left(\frac{\langle\tau\sigma\rangle}{\langle\sigma\rangle}\right)^2\frac{1}{\langle\tau\rangle}\,. \tag{8.43}$$

Despite the unwieldy nature of the latters constituent terms, it is satisfying that

Eqs. (8.32) and (8.41) provide an analytic form for the ordinary pair approximation, and one that does not depend only on λ_0 and λ_1 in contrast to the mean field result. As a consistency check, when we calculate the parasite density from Eq. (8.41) we find that as $z \to \infty$, the solution approaches the mean field prediction $\langle\sigma\rangle = 1 - 1/\lambda_0 - 1/\lambda_1$.

One final point to note is that when solving the OPA, we must choose which branch of $\langle\tau\sigma\rangle/\langle\sigma\rangle$ to take from the quadratic Eq. (8.39). It can be the case that both solutions to this quadratic are real and positive. Choosing either of these does not qualitatively change the outcome of the next section, but it is interesting to note that by always choosing the largest of the two roots, the analytical solution exactly matched the result when we solved the OPA numerically (i.e. Eqs. (8.18)–(8.22), where the moment closure approximation has been applied, solved using Euler's method. This also provides a check that the above analytical solution is indeed correct). For this reason, all of the plots in the next section were produced using this choice of the largest root.

8.4.4 Comparison to simulation

To test the usefulness of the OPA, we now compare its predictions to those of our stochastic simulations.

OPA phase diagram

Figure 8.5 shows the phase diagram obtained from the OPA. To compute this we choose a fixed value for d_0 and set $d_1 = 1$. For each value of λ_0, λ_1 is then incremented from zero until a positive parasite density is obtained from Eq. (8.41). We plot the transition lines corresponding to the non-spatial mean field (dashed line), OPA (solid lines) and our simulations (symbols) and find that the OPA does offer an improvement over the mean field theory. The OPA captures the correct dependence on host turnover, although not to as great an extent as in the simulations, where the difference between these two curves is much more pronounced. Simulation results shown are for $d_0 = 1$ and $d_0 = 20$ as before.

Parasite density and clustering

Figure 8.6 shows the parasite density and local clustering C as calculated in the OPA, along with the mean field and simulation results. The data shown is for the curve discussed in Figure 8.4; $\lambda_0 = 1.8$, $b_1 = 15$, $d_1 = 1$. The quantitative accuracy

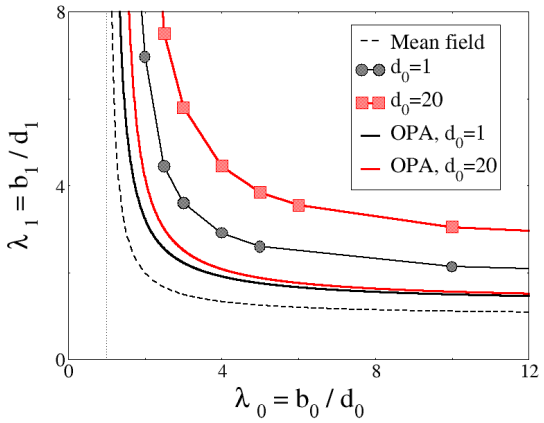


Figure 8.5 The ordinary pair approximation offers an improvement to the phase diagram of the mean field theory. As in Figure 8.2, symbols show simulation data for the boundary separating parasite persistence and extinction, for $d_1 = 1$ and $d_0 = 1$ (circles) or $d_0 = 20$ (squares). Dashed line shows mean field theory prediction $\lambda_1^{\text{crit}} = \lambda_0/(\lambda_0 - 1)$ and solid black and red lines show the OPA approximation's prediction for $d_0 = 1$ and $d_0 = 20$ respectively.

of the OPA for these parameters is poor, and worse still it is not able to capture the nonmonotonic behaviour we are most interested in. It does however offer an improvement over the mean field approximation (which predicts a constant parasite density as d_0 is varied, $\langle\sigma\rangle = 1 - 1/\lambda_0 - 1/\lambda_1$) as it captures the decrease in parasite density as the rate of host turnover is increased.

Figure 8.6b shows the local clustering measure $C = \langle\sigma_i\sigma_j\rangle/\langle\sigma\rangle^2$, Eq. (8.14), as determined by the OPA and our simulations. Recall that $C = 1$ in the mean field theory. Again, the quantitative match is rather poor, but strikingly the OPA does capture the nonmonotonicity: there is a minimum in C at $d_0 \approx 1$ in the OPA compared with the minima at $d_0 \approx 0.6$ found in the lattice simulation.

One reason for the poor fit is the choice of parameters. At $\lambda_0 = 1.8$ the host coverage of the lattice is sparse and there is higher parasite clustering as a result. The OPA method takes into account nearest neighbour pairs but neglects any higher order of clustering, so it is perhaps no surprise that it performs poorly in this regime. However investigating other parameter sets in the OPA have found none that reproduce the nonmonotonicity in the parasite density, though better quantitative agreement between simulation and the OPA can be achieved when the parameters used do not lead to such low densities (Appendix B.2).

In principle one could go on to take into account higher order correlations, going to three-point correlation functions for instance, but note that in the case of the 2-point functions our model contained 36 possible transitions (Table B.1). Extending to triplets

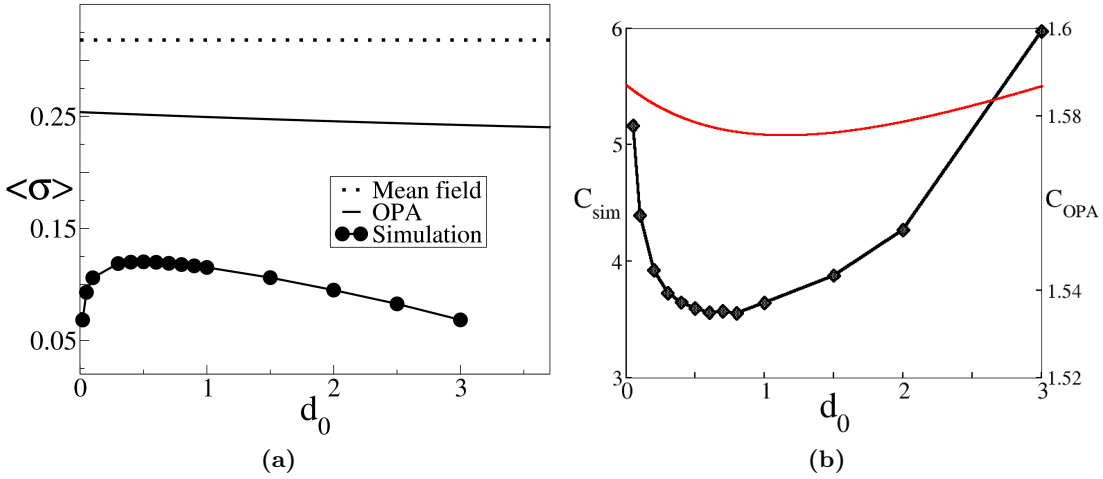


Figure 8.6 (a) Parasite density found by non-spatial mean field, OPA method and simulation for $\lambda_0 = 1.8$, $d_1 = 1$, $b_1 = 15$. The OPA does not capture the nonmonotonicity in the parasite density, but does offer an improvement over the non-spatial theory, capturing the decrease in parasite density for faster rates of host turnover, though the quantitative agreement is poor for this choice of parameters. (b) Local clustering measure $C = \langle\sigma_i\sigma_j\rangle/\langle\sigma\rangle^2$ obtained in simulation and from the OPA. Interestingly the OPA does manage to capture a minimum in the local clustering.

would result in 216 distinct transitions, which already becomes analytically intractable. Other factors introduce significant extra complexity. For example, closure of the 4-point functions would suffer to a greater degree from the approximation in the OPA that sites are in a contiguous chain. One may need to take into account linear chains, L-shapes and square configurations separately.

8.5 Phase transitions and disorder in the two-level model

So far we have considered the steady state behaviour of the two-level system. I believe it could be of interest in future work to study the nature of the phase transitions contained within the model. The original CP has a nonequilibrium phase transition to an absorbing state. By reference to the work on disordered processes, in this section I will explain why the stacked CP model likely contains a rich variety of phase behaviour, and why characterizing this would be a numerically challenging task. Figure 8.7 shows again the phase diagram of the two-level stacked CP as obtained by numerical simulation, for two values of d_0. The arrows marked A and B and the dot marked C represent three distinct manners of transition to the aborbing state *for the upper process* (parasite extinction, $\langle\sigma\rangle \to 0$, but host may live on, $\langle\tau\rangle \neq 0$). We will consider these in turn.

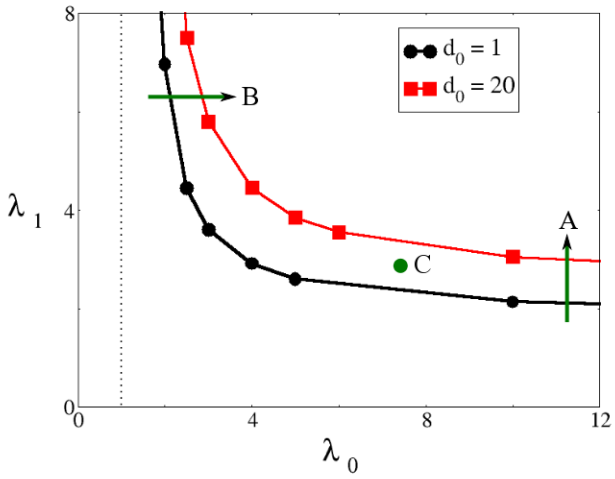


Figure 8.7 Possible phase transitions in the two-level stacked CP. A, B and C represent three distinct types of phase transition out of the parasite-free absorbing state and are described in detail in the text.

A. The transition at A is perhaps intuitively the easiest to understand so we begin with it. In this transition we are at a fixed value of λ_0 (governing the host dynamics) and vary the value of λ_1 (the infectious process). If we imagine the extreme case, where A is shifted right, $\lambda_0 \to \infty$, and also $d_0 \to 0$, then the parasite dynamics approaches that of the standard CP – the parasite sits on top of a fully occupied host lattice, which does not undergo any birth-death dynamics. Thus the transition should belong to the class of directed percolation[2].

On the other hand, if we remain in the $d_0 \to 0$ regime, but reduce λ_0 to some finite value, we are left with the parasite dynamics undergoing an exact CP (since there is no vertical transmission) but on an incomplete host lattice. Since the empty sites are essentially removed from the dynamics, this resembles the diluted contact process discussed in Section 7.6, which is known to belong to a different universality class from that of directed percolation.

Now, if we increase d_0 from zero we move from a system resembling the diluted CP to one which resembles the other main class of disordered systems; one in which the reproduction rates vary from site to site. A site containing a parasite reproduces into an empty site with rate b_0/z and into a site occupied by an uninfected host with rate b_1/z. Note the difference though, that this disorder is not quenched (fixed) as in the disorder literature, Section 7.6, but changes with the lattice configuration. Note also that it is not randomly distributed, but is correlated through the spatial distribution

[2]However it should be noted that it is not obvious how close to $d_0 = 0$ one needs to be for this to be true. It could be the case that any deviation from zero will be enough to damage the DP-like behaviour.

of the parasites. As λ_1 is increased along the transition line of A, for fixed λ_0 then, we could interpret this as the magnitude of the disorder itself being continuously varied. This transition could thus behave in an entirely different way than those previously studied.

B. The transition marked B presents a different scenario, and occurs when we vary the host dynamics, λ_0, while keeping λ_1 fixed. Again, consider first the case where $d_0 \to 0$ so that there is essentially no host dynamics. Just left of the transition then, we have a non-zero host population but no parasite can be maintained. In this regime, increasing λ_0 is increasing the density and connectivity of the host population, until it is sufficient to support the parasite given λ_1. Interestingly, this resembles a diluted CP in which the disorder is being gradually reduced (i.e. allowing more and more of the removed sites to partake in the dynamics by filling them with host organisms). Thus we see some sort of geometrically-driven phase transition, as discussed in Section 7.6 and [142]. Again however, as we increase the fixed value of d_0 from zero, we move away from this analogy with the diluted CP and possibly cross over to another class of transitions.

C. The transition marked C is slightly more unusual. It is not represented well pictorially but is simple to describe in words. This transition occurs when both λ_0 and λ_1 are fixed but we vary the turnover rate of the host lattice; that is, increasing from $d_0 = 1$ to $d_0 = 20$, while at fixed λ_0. This will move the system from a regime in which there is a non-zero steady state parasite density, to one in which the parasite cannot be maintained. To see this, imagine sliding the black line in Figure 8.7 up onto the red line, crossing the point marked C in the process. Thus it seems that this transition is somehow driven by the timescales of the processes, which are connected to the spatial correlations. This type of transition does not obviously resemble any that I have encountered in the literature.

I have until now neglected to mention one other intriguing point. In generating the phase diagram, we held d_0 and d_1 fixed, while varying b_0 and b_1 to alter the ratios λ_0 and λ_1 respectively. If one were to instead fix $b_0 = b_1$ and vary the values of d_0 and d_1, then recall that by setting $b_1 = b_0$ in Eq. (8.10), the parasite dynamics reduces to that of an exact contact process with an effective $d_{\text{eff}} = d_0 + d_1$. Might this then avoid some of the complicating issues and return the critical behaviour of the system (at least that of transition A) to that of directed percolation?

With such a complicated phase diagram then, and the computational difficulties of studying disordered transitions [144, 147], the task of characterizing the critical

behaviour of the stacked CP seems like a mammoth task. What is appealing is that the disorder in this system is not added arbitrarily, but emerges as a consequence from the underlying dynamics of the system. However, what might intially seem appealing turns out to lead to a surprisingly rich set of complexities. One final point to make is that it seems a sensible choice to parameterize the phase diagram of the system by the ratios λ_0 and λ_1 (indeed the mean field theory predicts that this is sufficient), since these are conceptually linked to the host and infection dynamics respectively. However, the presence of the vertical transfer, wherein varying the host dynamics also has an effect on the parasite transmission, raises the question of whether a better parameterization for this system exists.

Up to now we have limited the number of stacked CPs to two levels. Even in the simplest case, in which the parasite or infection incurs no cost or benefit, we have found this model to contain non-trivial behaviour. Due to an intricate coupling between spatial and temporal scales, we have observed a nonmonotonic dependence of the parasite level on the rate of host turnover for some parameter values. Deriving the ordinary pair approximation for the two-level system, we have shown that this effect is not captured by either this, or the naive mean field theory, suggesting that a deep understanding of the fluctuations in the host CP is required to predict the steady state properties of the parasite population. In the next section, we will extend our stacked CP model to multiple levels of contact processes, inspired by natural examples of hyperparasitism.

8.6 Multiple stacked contact processes and hyperparasitism

Parasitic infections are not always limited to two levels. A number of hyperparasitic relations are known to occur in nature, in which an organism carrying a primary parasite is susceptible to a secondary parasite [171]. A recently discovered example of this was is the found for the behaviour-manipulating fungus that leads to "zombie ants", this fungus itself being parasitized by a second fungus [172]. Hyperparasitism can be also harnessed as a biocontrol mechanism – examples include viral infections of the fungus Cryphonectria parasitica that causes chestnut blight [173], and cytoplasmic RNA elements that infect the fungus causing Dutch elm disease [174]. Other examples of such multi-tiered systems can be found in ecology: that of plant-inhabiting aphids, harbouring symbiotic bacteria which in turn possess a plasmid [175]; or a mammal population carrying ticks, which in turn may be infected by the bacteria *Rickettsia Rickettsi* [176]. Inspired by these scenarios, we extend our model to a multilevel stacked contact process, in which individuals carrying a primary infection may be susceptible

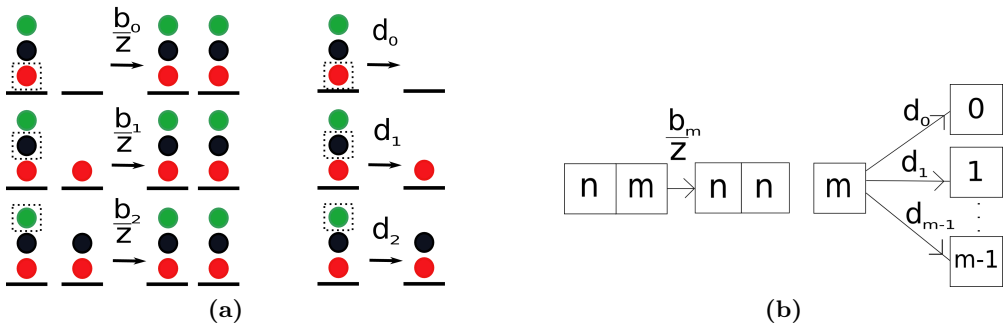


Figure 8.8 Possible transitions for the multi-level CP. (a) Transitions for a site with label $m = 3$. The symbols are as in Fig. 8.1, with green circles representing the secondary parasite. Birth processes are shown on the left: host can reproduce into a neighbouring empty site (top left); transmit its primary parasite to a susceptible neighbour (middle left - note the secondary parasite is also transmitted); or transmit its secondary parasite to a susceptible neighbour (bottom left). Death processes are shown on the right: death of the host (top right); loss of the primary parasite (middle right); or loss of the secondary parasite (bottom right). (b) Shows schematically all possible transitions and their rates for the general case of a site with label m, where $n > m$.

to secondary infections, and those carrying the secondary infection may be susceptible to tertiary infections, etc. This *parasites on parasites* scenario brings to mind the words of Johnathon Swift: "So, naturalists observe, a flea, has smaller fleas that on him prey, and these have smaller still to bit 'em, and so on proceed *ad infinitum*" [177]. This picture raises a number of questions: How does the dynamics of any parasite level couple to the others, and how many levels of parasites are sustainable in a population?

Motivated by these examples, we now extend our model to investigate a *multi-level* stacked CP. Extending the terminology introduced earlier in the chapter, in the M-level stacked CP a site is labelled 0 if it is empty, 1 if it contains an uninfected host, and $m = 2, 3, \ldots, M$ if it contains a host infected with $m-1$ levels of parasites. For example, a site labelled 2 contains a host infected with only a primary infection (parasite) while a site labelled 3 contains a host which is infected with primary and secondary infections (a parasite and a hyperparasite). Note that the presence of an infection at level m implies the presence of all lower-level infections, e.g. one cannot have a secondary infection without a primary infection. Figure 8.8 illustrates the dynamical processes that can occur in this model. An empty site may be occupied at rate b_0/z by reproduction of a neighbouring site of any label (top left), or a site labelled m may be infected at rate b_m/z by higher-level parasites from a neighbouring site with label $n > m$; its label is then promoted to n. The host organism of a level m site can die, at rate d_0 , or experience loss of one of its parasites, at rate d_ℓ (where $\ell \leq m$ relates to the level of parasite that is lost). When the latter happens, all higher parasites are also lost and the site is demoted to level ℓ. The parameters of the multi-level stacked CP

model are the number of levels M, and the set of level-dependent birth and death rates $\{b_0, \ldots, b_{M-1}; d_0, \ldots, d_{M-1}\}$.

It turns out that the dynamics of this system can be understood as a standard CP in the case where the birth rates at all levels are equal; $b_n = b \; \forall \; n$. To see this we divide lattice sites into two sets: sites that have labels greater than or equal to m (here denoted $\mathcal{M}_+$), and sites that have labels less than m (here denoted $\mathcal{M}_-$). To clarify, the density of $\mathcal{M}_+$ sites corresponds to the density of the organism represented by the site label m, since all higher level sites $(m+1), \ldots, M$ require the presence of this level (as well as all lower levels). Consider the dynamics between these two sets. An $\mathcal{M}_-$ site, with label ℓ, can become $\mathcal{M}_+$ through infection by a neighbouring $\mathcal{M}_+$ site: this occurs at rate $b_\ell/z = b/z$ regardless of the level of the $\mathcal{M}_+$ site. An $\mathcal{M}_+$ site can become $\mathcal{M}_-$ by the death of the host, or through the loss of any of the parasites at levels $n = 2 \ldots m$. This occurs at total rate $d = \sum_{\ell=0}^{m-1} d_\ell$ where the sum is over levels up to $m-1$, as illustrated in Figure 8.8b. Thus the dynamics of the $\mathcal{M}_+$ density, and thus the organism labelled m, is exactly equivalent to that of a standard CP with parameter $\lambda_{\text{eff}} = b/\sum_{\ell=0}^{m-1} d_\ell$.

As a concrete example, let us take an $m = 3$ labelled site and introduce a new indicator variable $\eta_i = 1$ if site i contains a hyperparasite and $\eta_i = 0$ otherwise. By inspection of Figure 8.8a one can simply write down the equation of motion for the density of the hyperparasite analogous to Eqs. (8.6) and (8.7):

$$\begin{aligned}\frac{d}{dt}\langle\eta_i\rangle &= \frac{b_0}{z}\sum_j\langle(1-\tau_i)\eta_j\rangle + \frac{b_1}{z}\sum_j\langle(\tau_i-\sigma_i)\eta_j\rangle + \frac{b_2}{z}\sum_j\langle(\sigma_i-\eta_i)\eta_j\rangle - (d_0+d_1+d_2)\langle\eta_i\rangle \\ &= \left(b_0 - \sum_{\ell=0}^{2} d_\ell\right)\langle\eta\rangle + (b_1-b_0)\langle\tau\eta\rangle + (b_2-b_1)\langle\sigma\eta\rangle - b_2\langle\eta\eta\rangle \; . \end{aligned} \tag{8.44}$$

Note again that what $\langle\eta_i\rangle$ represents is the density of the hyperparasite, i.e. the density of all sites with label $m \geq 3$, since all higher level layers can only exist on a site where the hyperparasite is present. It is then easy to see that the cross-level correlations $\langle\tau_i\eta_j\rangle$ and $\langle\sigma_i\eta_j\rangle$ vanish when the birth rates of all levels are equal, and the equation takes the form of Eq. (8.21); the form of a CP with parameter $\lambda_{\text{eff}} = b/\sum_{\ell=0}^{2} d_\ell$.

This observation provides us with important insight into the number of levels that are sustainable in a stacked CP. For the standard (one-level) CP, a finite density of occupied lattice sites can only be sustained for $\lambda > \lambda_{\text{crit}}$, where $\lambda_{\text{crit}} \approx 1.649$ for a 2D square lattice [131]. In the stacked CP with equal birth rates we therefore expect the density of $\mathcal{M}_+$ sites, those with label $n \geq m$, to be non-zero only if $\lambda_{\text{eff}} = b/\sum_{\ell=0}^{m-1} d_\ell > \lambda_{\text{crit}}$. Figure 8.9 shows simulation results for the average density of lattice sites with label

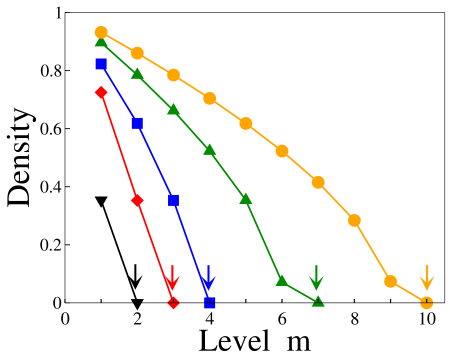


Figure 8.9 Density of sites with label m or higher, as a function of level m, for a stacked CP with equal birth rates $b_n = b$ and death rates $d_n = 1$. Left to right correspond to simulations with b = 2,4,6,10 and 15 respectively. Arrows indicate $m^* = \lceil b/(\lambda_{\rm crit} d) \rceil$, the boundary for sustainability as predicted by mapping to the standard CP.

$n \geq m$, as a function of m, for a stacked CP with equal birth rates $b_n = b$ and death rates $d_n = d = 1$, for several values of b. By mapping onto a standard CP with $\lambda_{\rm eff} = b/(md)$, we predict that all levels $n \geq m^* = \lceil b/(\lambda_{\rm crit} d) \rceil$ have zero density and are thus unsustainable. The notation $\lceil x \rceil$ deotes the ceiling function, the smallest integer that is not less than x. Our simulation results, shown in Figure 8.9, bear this out: the system indeed only sustains a finite number $(m^* - 1)$ of levels.

Are there any circumstances where a stacked CP can be sustained for an infinite number of levels? This is indeed possible if either the birth rate b_n increases, or the death rate d_n decreases, sufficiently strongly with n. We first suppose that the birth rate $b_n = b$ is constant but the death rate decreases by a factor f at successive levels: $d_n = f d_{n-1} = f^n d_0$ where $0 < f < 1$. In this case, the average density of sites with label $n \geq m$ is given by that of a standard CP with

$$\lambda_{\rm eff} = \frac{b}{\sum_{\ell=0}^{m-1} d_\ell} = \frac{b}{d_0(1 + f + f^2 + ... + f^{m-1})} = \frac{b(1-f)}{d_0(1-f^m)} \ .$$

To sustain an infinite number of levels would require that $\lambda_{\rm eff}^{m\to\infty} = b(1-f)/d_0 \geq \lambda_{\rm crit}$, which implies that $f < 1 - (d_0 \lambda_{\rm crit}/b)$. Our simulation results, Fig. 8.10a, show that, with $d_0 = 1$ and $b = 5$, the system indeed sustains an infinite number of levels for values of f below the critical value of $f = 0.67$.

We also explore the case where, instead of varying the death rate between levels, we instead increase the birth rate by a factor $1/f$, i.e. setting $b_n = b_{n-1}/f$. This case is interesting since the dynamics does not map onto that of a standard CP; the cross-level correlations terms, for example $(b_1 - b_0)\langle \tau \eta \rangle$ in Eq. (8.44), no longer vanish. Nevertheless, our simulations show qualitatively similar results (dashed lines in Figure 8.10a).

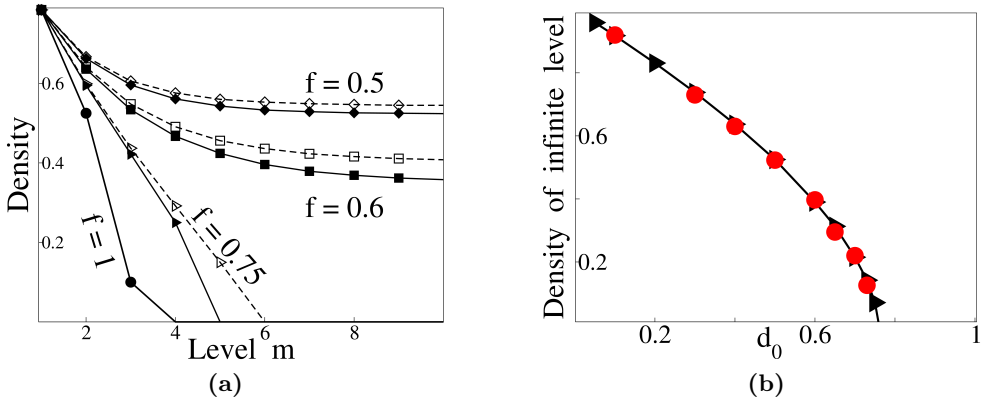


Figure 8.10 (a) Density of sites with $n \geq m$, as a function of level m, in the stacked CP for decreasing death rates $d_n = f^n d_0$, with $d_0 = 1, b_n = 5$ (solid lines) and for increasing birth rates $b_n = f^{-n} b_0$, with $b_0 = 5, d_n = 1$ (dashed lines). In the former case, the exact CP predicts a nonzero density as $m \to \infty$ only for $f < 0.67$. (b) Density of the mth level in the limit $m \to \infty$ with $b_n = 5, d_n = f^n d_0, f = 0.75$ as a function of host death rate d_0 from simulations (circles) and as predicted by a standard CP (triangles) with $\lambda_{\text{eff}}^{m\to\infty} = b(1-f)/d_0$. The simulation data is obtained by measuring the plateau values in plots like panel (a), for m values up to 20. The mapping to a standard CP predicts that only a finite number of levels are sustainable for $d_0 > d_0^* \approx 0.758$.

Interestingly, in the case where the values of b and f are given, there exists a critical value of the host death rate d_0 which separates regimes where the stacked CP can and cannot sustain an infinite number of levels. This is given by $d_0^* = b(1-f)/\lambda_{\text{crit}}$. Figure 8.10b shows, as a function of d_0, the predicted density of sites with label $n \geq m$, as $m \to \infty$: this is given by the density of a standard CP with $\lambda_{\text{eff}} = b(1-f)/d_0$. The density as $m \to \infty$ indeed falls to zero at $d_0 = d_0^*$, and in a manner reminiscent of a second order phase transition.

8.6.1 Connection to a roughening transition

The transition to an infinite number of levels of parasites that we observe in this model appears to be related to a transition that occurs in models for the growth of interfaces by deposition of particles on surfaces [178–180]. In these models, the interface is modelled as a lattice, with a given height at each lattice site. According to the ballistic deposition rule for surface growth [181], particles fall onto the lattice from above and only stick if the neighbouring lattice site already contains a particle. In some models, particles can also desorb from the surface [180]. In the stacked CP model, we can think of the label m of a given lattice site as corresponding to the local height of the interface. The transmission of higher-level parasites to a site with label m, from a neighbour with $n > m$ then corresponds to ballistic deposition, while death of a host and loss of parasites loosely correspond to desorption of particles. This apparent mapping

is intriguing because these models for surface growth show a roughening transition [178–180]: when the deposition rate is low, the interface remains *smooth*, that is, the width of the surface layer remains finite in the thermodynamic limit, whereas when the deposition rate is high, the surface layer grows and *roughens* over time, that is, arbitrarily large differences in surface height can arise in the thermodynamic limit. These two cases correspond to finite and infinite hierarchies of parasites in our stacked CP model. Although the mapping to the surface growth models is not exact, one might expect to see the same phenomenology, e.g. in terms of critical exponents, for the transitions. For example a generic model for coupled directed percolation processes with unidirectional coupling between adjacent levels [182] can show a different β exponent at different levels when the critical points coincide. It would be interesting to see if this is true for models such as ours which show bidirectional coupling between all levels.

8.7 Discussion

The contact process is delibrately simple, incorporating space and stochasticity into a model of population or infection dynamics in a general way. Despite its simplicity, the CP still exhibits non-trivial behavour arising from the strong spatial correlations, and undergoes a nonequilibrium phase transition to an absorbing state. As such it has served as a framework to aid in the development of analytical tools in the approximation of lattice systems, and formed the baseline from which more detailed models have been proposed, finding application in many branches of science.

In this chapter we introduced an extension to the CP which was motivated by the observation that in reality, many diseases or genetic traits undergo both vertical and horizontal transmission. This extension took the form of *stacked contact processes*, and motivated by examples of hyperparasitism we extended the model to include an arbitrary number of CP levels. We demonstrated that although this extension is conceptually simple, even in the two-level model the spatial correlations and competing timescales produced a complex behaviour; behaviour which neither the mean field approach or pair approximation was able to fully describe. Through the discussion of disorder, we also highlighted the likelihood that a rich variety of phase transitions are contained in this model, and characterizing these could provide interesting future work.

8.7.1 Other spatial approximations

We presented the ordinary pair approximation and derived an exact analytical solution for the steady state of the two-level stacked contact process. Through this approach we were able to obtain a much improved quantitative description than that of the mean field theory, and one that qualitatively captured the decrease in parasite density with increasing host turnover. However the nonmonotonic behaviour could not be reproduced using this method. Various other approaches exist to approximate the spatial correlations in lattice models. For example the equations governing triplets of sites could be derived, but it is often the case that this extra spatial information does not yield a great enough improvement to justify the added complexity. There is a desire to obain an analytical approximation to lattice models from both a fundamental mathematical point of view, as well as the increased efficiency this would bring to studying such models in biology and ecology. For a summary of some of this work I refer you to [170, 183–185]. Some of the approaches involve model-specific heuristics and require parameters that are obtained from the actual lattice simulations, which seems to defeat the ultimate purpose. We did not pursue any improvement on the OPA in this thesis, but it would be satisfying to at least qualitatively capture the full behaviour of the stacked CPs. Extending to triplets could be a worthwhile improvement.

8.7.2 A well-studied host-parasite model

It was surprising to find such a small amount of literature examining the combined effect of vertical and horizontal transmission in spatially explicit models; perhaps the stacked contact processes could offer some insight into general ecological principles that can arise from the interplay between the two modes. Since the stacked contact process model presented here has been mainly framed in terms of parasite dynamics, it is worth briefly discussing some of the existing models. It turns out the majority of the models for host-parasite or host-pathogen dynamics also neglect the role of vertical transmission.

An early spatial host-parasite model, and one that many subsequent studies were based upon, is that originally developed in [184]. In this lattice model, once a host is infected by a parasite it cannot recover, and in addition it is immediately sterilized (hence no vertical transfer). The reason for this was perhaps a mathematical simplification for the improved pair approximation the authors were developing. Sterilization was also assumed in many follow up works: in [186] where the evolution of virulence is

studied; in [187] studying the effect of short and long range interactions; and in [188] where the authors analyze the phase transitions of this model and find that both the host and parasite-extinction transitions belong to the universality class of directed percolation – it seems then that the presence of vertical transmission leads to much richer dynamics. The extension in [189] actually relaxes the sterilization condition, but assumes that infected individuals only give birth to uninfected offspring and hence includes no vertical transmission either. It would be interesting to see how any of the ecological principles derived from the previous models change with the inclusion of vertical transfer. Changing the basic structure of the models in such a fundamental way, it is likely that at least some of the key results will differ.

The model encountered that most closely resembles the one developed in this chapter was that of Schinazi [162], in which they claim their model is the first spatial model to analyze vertical transmission. Here, an imperfect vertical transmission was included, so that infected individuals could produce either infected or uninfected offspring with some probability. However this model did not allow for any recovery, meaning that the only way in which the parasite could be lost from the system was through the death of its host.

Chapter 9

Conclusion

The study of biological evolution encompasses many fields of science, requiring knowledge of processes at the level of genetics to ecosystems. In the first part of this thesis we considered evolution at the fundamental level of metabolism and asked why biology utilizes such an apparently small subset of the biochemistry available to it, and whether the universal nature of metabolism is the result of an evolutionary optimization or historical contingency. To investigate such questions, we generated a network of chemicals and reactions based on basic biochemical rules and considered the trunk pathway of glycolysis, the most conserved section of metabolism. We found that even for such a simple pathway, the number of possible alternatives is huge when one allows access to this full biochemical network. By comparing these alternative pathways in terms of their metabolic flux capabilities, we demonstrated that the naturally occurring pathways can sustain the highest metabolic fluxes under typical physiological conditions, when realistic thermodynamic and biophysical constraints are applied. This adds to the body of work suggesting that existing metabolic pathways may be understood as an evolutionary optimization problem.

This outcome is robust to whether the enzyme concentrations are optimized or assumed equal in the flux calculation and still hold when more realistic kinetics, namely reversible Michaelis-Menten kinetics, are used in place of the perfect enzyme assumption. Future work will extend this test of robustness to investigate different intracellular pH values and temperatures, as it is known that these can have a significant effect on the free energies of biochemical reactions.

Our analysis also highlighted alternative pathways which utilize reactions already known to exist in nature, but that are not usually associated with the trunk pathway.

Some of these had the potential to perform relatively well with respect to their fluxes, and raise the question of whether they may in fact be operating, unknown in nature. Such questions could be addressed by studying organism- and environment-specific transcriptomics. In addition, many enzymes have the ability to catalyze multiple reactions inside the cell, but the characterization of such functions is far from complete. It is thus possible that the missing reactions in some of the alternative pathways could be the result of secondary enzyme function. Enumerating pathways with the methodology developed in the previous chapters will highlight such possibilities, and these could be further investigated experimentally. Besides from the fundamental desire to fully understand extant metabolism and its evolution, such techniques can also be applied in a bioengineering setting, in constructing *de novo* pathways towards medically or industrially important compounds.

By expanding our network of reactions, future work should study other pathways or subnetworks of central carbon metabolism, with the exploration of the enormous biochemical space being aided by evolutionary simulations as described in Chapter 6. In [29], Noor et al. showed that central carbon metabolism exists as a minimal walk between biosynthetic precursor molecules. However, they also found that it was often the case that multiple alternative minimal walks existed. Requiring that this same set of essential metabolites be produced, and assuming a goal function such as energy production or maximal precursor production, how many viable alternatives to central carbon metabolism exist, and how do they compare? The approach discussed in this thesis thus offers a chance to test questions for which the widely used flux balance analysis models would not be as applicable. For instance, in our pathway comparison of Section 5.6.3, each alternative pathway is a linear path of 5 reactions. Since the standard FBA models do not take thermodynamics into account but depend only on the network stoichiometry, all pathways utilizing the same reaction types but in a different order would be equivalent in such models.

I believe that future metabolic studies will benefit by considering the full range of biochemistry and not restricting to only those reactions and compounds known to exist. Studying the metabolic pathways that were never realised by life could give new insights into existing paths and the evolutionary pressures that produced them.

In the second part of this thesis we developed a generic statistical physics model of an infectious process undergoing both horizontal and vertical transmission. To do this, we extended the canonical contact process to a model of stacked contact processes and found that this intuitively simple model contains a rich behaviour. In extending this study of the stacked contact processes, an attractive challenge would be to persue the

characterization of the phase transitions and an understanding of the nature of disorder in the system. From our discussion of disorder, it seems that this simple extension to the CP will likely contain a variety of critical behaviour. It would also be satisfying to improve our ordinary pair approximation to one that fully captures the behaviour of the system.

The evolutionary implications arising from the interplay between horizontal and vertical transfer have not been widely studied. In terms of the host-parasite systems, the majority of models leave out vertical transfer altogether and many also neglect spatial structure. Our generic model could thus be used to test the ecological and evolutionary predictions of such models, since the combined effects of space and vertical transfer will certainly alter the dynamics.

It has been shown experimentally that in plasmid-bearing bacterial populations there is a trade-off between the vertical and horizontal modes of transmission [161]. The cellular activities of a plasmid that lead to an increased rate of horizontal transmission are usually damaging to the host bacteria, leading to a lower rate of reproduction and thus vertical transfer. It would be interesting to modify our model to investigate the implications of this trade-off. For instance, by allowing the horizontal transmission rate b_1 to evolve while imposing a corresponding trade-off in the vertical transmision rate b_0, are there conditions under which the population evolves to favour vertical transmission over horizontal, or vice versa? It might be expected that at high host densities, where there are many opportunities for infection, an increase in the horizontal transmission will outweigh the small reduction in host reproduction. At low host densities, where the plasmid does not encounter as many opportunities for horizontal transfer, it might be more beneficial for the plasmid to reduce its infectivity and thus increase its vertical transmission. Future work could explore these alternative scenarios.

Appendix A

Appendix to Chapters 2–6

A.1 Derivation of Michaelis-Menten equation

The Michaelis-Menten equation is the canonical form for the reaction rate of an enzyme-catalyzed reaction and has been used widely in biochemistry. In its original form, it applies only to irreversible reactions. Here, I give the two separate derivations to this canonical form and extend to the case of a reversible reaction. I also outline how these derivations can be extended to other reaction stoichiometries and show that this results in the common modular rate law of Eq. (2.12).

A.1.1 Irreversible Michaelis-Menten

Consider the irreversible reaction

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_{\mathrm{cat}}} E + P \; . \tag{A.1}$$

The canonical Michaelis-Menten equation is derived as an approximation from this process, and gives an expression for the rate v of an irreversible reaction:

$$v = \frac{k_{\mathrm{cat}}[E]_t[S]}{K_M + [S]}, \tag{A.2}$$

where k_{cat} corresponds to the enzyme turn-over number, $[E]_t$ is the *total* enzyme concentration (free plus substrate-bound), $[S]$ is the substrate concentration and K_M is the substrate concentration at which the reaction proceeds with half its maximal rate. Eq. (A.2) has been derived under two different assumptions, both of which are

presented here.

Equilibrium approximation

In the original derivation, Michaelis and Menten [45] make the assumption that the substrate is in instantaneous equilibrium with the substrate-enzyme complex, i.e. $k_1[E][S] = k_{-1}[ES]$. Note that on the timescale of the reaction the total enzyme concentration $[E]_t$ is constant and is the sum of the free and substrate-bound enzyme forms, $[E]_t = [E] + [ES]$. Combining these two equations allows us to write

$$[ES] = \frac{[E]_t[S]}{\frac{k_{-1}}{k_1} + [S]}. \tag{A.3}$$

This gives the final expression for the total rate of the reaction:

$$v = \frac{d[P]}{dt} = k_{\text{cat}}[ES] = \frac{k_{\text{cat}}[E]_t[S]}{K_S + [S]}, \tag{A.4}$$

where $K_S = k_{-1}/k_1$ is the dissociation constant for the enzyme-substrate complex.

Quasi-steady-state approximation

An alternative derivation was proposed by Haldane and Briggs [46] which assumes that on the time-scale of product formation, the concentration of the enzyme-substrate complex is constant. Applying the law of mass action to the scheme depicted in (A.1), this assumption corresponds to:

$$\frac{d[ES]}{dt} = 0 = k_1[E][S] - k_{-1}[ES] - k_{\text{cat}}[ES]. \tag{A.5}$$

Combining this with the relation for fixed total enzyme yields

$$[ES] = \frac{k_1[E]_t[S]}{k_{-1} + k_{\text{cat}} + k_1[S]} = \frac{[E]_t[S]}{\frac{k_{-1}+k_{\text{cat}}}{k_1} + [S]}, \tag{A.6}$$

allowing the total rate to be written in the same form as before:

$$v = \frac{k_{\text{cat}}[E]_t[S]}{K_M + [S]}. \tag{A.7}$$

Note that $K_M = \frac{k_{-1}+k_{\text{cat}}}{k_1}$ is called the Michaelis-Menten constant and has a different form from the dissociation constant derived under the equilibrium assumption. Since both assumptions lead to the same form however, the underlying assumptions of each

derivation are often ignored when modelling chemical reactions.

A.1.2 Reversible Michaelis-Menten equation

The canonical Michaelis-Menten equation is only valid in the limit that the reaction is irreversible. This could be applicable if the substrate concentration is much higher than the product concentration $[S] \gg [P]$, or the reaction has a very large and negative free energy $\Delta_r G \ll 0$. In most biological reactions this will not be the case, with many reactions operating reversibly, having some non-zero steady-state product concentration. We can derive an extended, reversible form for the Michaelis-Menten equation by considering the following reaction scheme:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k^-_{\text{cat}}}{\overset{k^+_{\text{cat}}}{\rightleftharpoons}} EP \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P. \tag{A.8}$$

The *equilibrium assumption* now assumes that the association and dissociation of both substrate and product molecules is much faster than the interconversion between ES and EP. For the substrate and product respectively, this rapid equilibrium corresponds to $k_1[E][S] = k_{-1}[ES]$ and $k_3[EP] = k_{-3}[E][P]$. Thus we can write the dissociation constants for the substrate and product, now denoted K_S and K_P, as

$$K_S = \frac{k_{-1}}{k_1} = \frac{[E][S]}{[ES]} \rightarrow [ES] = \frac{[E][S]}{K_S}, \tag{A.9}$$

$$K_P = \frac{k_3}{k_{-3}} = \frac{[E][P]}{[EP]} \rightarrow [EP] = \frac{[E][P]}{K_P}. \tag{A.10}$$

This allows the conserved total enzyme $[E]_t$ to be written as

$$[E]_t = [E] + [ES] + [EP] = [E]\left(1 + \frac{[S]}{K_S} + \frac{[P]}{K_P}\right), \tag{A.11}$$

which can be substituted into the right hand sides of Eqns. (A.9) and (A.10) to give the concentration of enzyme complexes in terms of total enzyme concentration:

$$[ES] = \frac{[E]_t[S]}{K_S(1 + \frac{[S]}{K_S} + \frac{[P]}{K_P})} \qquad \text{and} \qquad [EP] = \frac{[E]_t[P]}{K_P(1 + \frac{[S]}{K_S} + \frac{[P]}{K_P})}. \tag{A.12}$$

The rate of the forward reaction, $v = k^+_{\text{cat}}[ES] - k^-_{\text{cat}}[EP]$, can now be written as

$$v = \frac{k^+_{\text{cat}}[E]_t\frac{[S]}{K_S} - k^-_{\text{cat}}[E]_t\frac{[P]}{K_P}}{1 + \frac{[S]}{K_S} + \frac{[P]}{K_P}}. \tag{A.13}$$

Noting that the upper limit for the forward reaction rate is $k^+_{\text{cat}}[E]_t = V_f$ and for the backwards reaction is $k^-_{\text{cat}}[E]_t = V_r$ gives

$$v = \frac{V_f \frac{[S]}{K_S} - V_r \frac{[P]}{K_P}}{1 + \frac{[S]}{K_S} + \frac{[P]}{K_P}} = \frac{V_f \frac{[S]}{K_S}\left(1 - \frac{V_r K_S [P]}{V_f K_P [S]}\right)}{1 + \frac{[S]}{K_S} + \frac{[P]}{K_P}}. \tag{A.14}$$

At chemical equilibirum the denominator of Eqn. (A.14) must be equal to zero and so the equilibirum constant of the reaction is

$$q = \frac{[P]_{eq}}{[S]_{eq}} = \frac{V_f K_P}{V_r K_S}. \tag{A.15}$$

Inserting this into Eqn. (A.14) and denoting $\Gamma = [P]/[S]$, gives finally the reversible Michaelis-Menten equation:

$$v = V_f \frac{\frac{[S]}{K_S}}{1 + \frac{[S]}{K_S} + \frac{[P]}{K_P}} \left(1 - \frac{\Gamma}{q}\right). \tag{A.16}$$

This expression can be decomposed into three terms, following Noor et al. [47]:

$$v = V_f \times \kappa \times \gamma. \tag{A.17}$$

The first term, $V_f = k^+_{\text{cat}}[E]_t$ is the *capacity* term, or the maximal rate; the second term, $\kappa = \frac{[S]}{K_S}/(1 + \frac{[S]}{K_S} + \frac{[P]}{K_P})$ is the *fractional saturation* term and relates to the level of enzyme saturation; and the third term $\gamma = (1 - \Gamma/q)$ is the *thermodynamic* term, relating to the thermodynamic driving force of the reaction.

A.1.3 Derivation of the common modular rate law

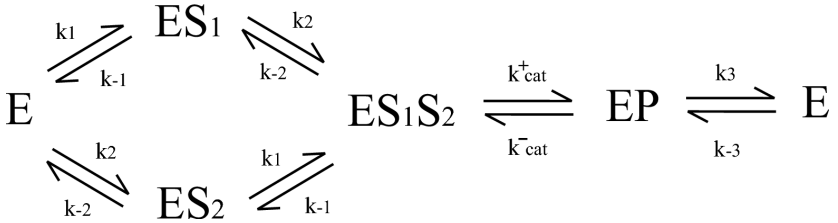


Figure A.1 Reversible random-order binding for a reaction with two substrates and one product.

The derivation of the common modular rate law is essentially the same as that of the reversible Michaelis-Menten rate. Consider the reversible, random-order binding mechanism depicted in Figure A.1. It is straightforward to derive the rate of this

reaction using the rapid-equilibrium derivation, as used for the reversible Michaelis-Menten presented in the last section. To follow the derivation of the reversible Michaelis-Menten equation, note that the total enzyme concentration is now $[E]_t = [E] + [ES_1] + [ES_2] + [ES_1S_2] + [EP]$ and there are now three dissociation constants: $K_{S_1} = k_{-1}/k_1$, $K_{S_2} = k_{-2}/k_2$ and $K_P = k_3/k_{-3}$. The net rate of the reaction is given by the interconversion between the fully-bound substrate-enzyme complex and enzyme-product complex; $v = k^+_{\text{cat}}[ES_1S_2] - k^-_{\text{cat}}[EP]$, resulting in:

$$v = V_f \frac{\frac{[S_1][S_2]}{K_{S_1}K_{S_2}}}{1 + \frac{[S_1]}{K_{S_1}} + \frac{[S_2]}{K_{S_2}} + \frac{[S_1][S_2]}{K_{S_1}K_{S_2}} + \frac{[P]}{K_P}} \left(1 - \frac{\Gamma}{q}\right). \tag{A.18}$$

Some simplifying assumptions are made in this enzyme process, the first being that substrates (products) can bind (unbind) from the enzyme in any order (random-order binding mechanism). Note also that the binding of the substrate is assumed to be unaffected by the presence of another substrate molecule – e.g. S_1 binds to the free enzyme E with the same affinity as it does to the enzyme-substrate complex ES_2. In addition, only one molecule of each metabolite can bind (i.e. no enzyme complex with two S_1 substrates bound), corresponding to the assumption that there is a specific enzyme-binding site for each substrate. In addition, there are no substrate-product-enzyme complexes, such as ES_1P_1.

Performing the derivation for other reaction stoichiometries (two substrates to two products, etc.) reveals a clear pattern in the denominator: each enzyme form present in the reaction scheme leads to a corresponding term in the denominator [42, 43]. Introducing the notation $\widetilde{S}_i = [S_i]/K_{S_i}$ yields the common modular rate of Eq. (2.12).

A.1.4 Flux balance analysis

For completeness, we note that many studies of metabolic networks do not use kinetic approaches but instead make use of flux balance analysis (FBA). Briefly, this method neglects the reaction kinetics entirely and instead depends only upon the topological structure of the network. Let N be the stoichiometric matrix in which each row corresponds to a species of metabolite and every column to a reaction, so that the entry N_{ij} corresponds to the stoichiometric coefficient (the number of moles consumed) of metabolite i in reaction j. Then, if $\mathbf{x}$ is the vector of metabolite concentrations, and

$\mathbf{v}$ the vector of reaction fluxes, the evolution of the system is described by

$$\frac{d}{dt}\mathbf{x} = N.\mathbf{v} \ . \tag{A.19}$$

FBA involves assuming a steady state distribution of fluxes by setting the left-hand side of this equation to zero. The resulting linear algebra problem is solved to obtain all possible sets of fluxes $\mathbf{v}$ that satisfy this equation. This results in multiple solutions, which can be reduced by requiring the simultaneous optimization of some objective function, such as the maximization of ATP production or the maximization of flux of an additional reaction that represents growth. There are many good introductions to FBA [40, 190].

Finding the solutions to these systems is much less computationally and experimentally intensive than is the case with detailed kinetic models, not require the determination of the enzyme kinetic parameters as input to the model. FBA models have been successful in predicting organism growth rates on varying media, the effects of single and multiple gene knockouts and in assessing the ability of a cell to produce compounds of industrial importance [40]. Extensions have also been developed which allow thermodynamic and gene regulation data to be incorporated into these models [190].

A.2 Group contribution results: E_0, E_1 and E_2 values obtained from least-squares fitting

The values of E_0, E_1 and E_2 obtained using the 56-compound training set of linear CHOP molecules under conditions of $T = 298$ K, pH = 7 and $I = 0.2$ M are as follows.

$E_0 = -227.22$ kJmol^{-1}.

E_1: vector of group contributions (kJmol^{-1}). The first column shows each molecular group, with the second column showing its contribution to the free energy of formation of a molecule.

$$E_1 = \begin{pmatrix} \text{-CH3} & 218.75 \\ \text{-CH2(OH)} & 67.712 \\ \text{-COOH} & -232.22 \\ \text{-CHO} & 31.93 \\ \text{-CH2p} & -825.42 \\ \text{-COp} & -1090.9 \\ \text{=CH2} & 236.12 \\ \text{-CH2-} & 82.90 \\ \text{=CH(OH)} & 51.21 \\ \text{-CH(OH)-} & -72.63 \\ \text{=CO} & 178.95 \\ \text{-CO-} & -110.57 \\ \text{=CHp} & -816.33 \\ \text{-CHp-} & -950.69 \\ \text{-CH=} & 83.713 \\ \text{-C(OH)=} & -99.314 \\ \text{-Cp=} & -966.85 \end{pmatrix}$$

E_2: matrix of corrections arising from group-group interactions (kJmol^{-1}). First row and first column indicate the molecular groups.

$$E_2 = \begin{pmatrix} & \text{-CH3} & \text{-CH2(OH)} & \text{-COOH} & \text{-CHO} & \text{-CH2p} & \text{-COp} & \text{=CH2} & \text{-CH2-} & \text{=CH(OH)} & \text{-CH(OH)-} & \text{=CO} & \text{-CO-} & \text{=CHp} & \text{-CHp-} & \text{-CH=} & \text{-C(OH)=} & \text{-Cp=} \\ \text{-CH3} & 16.99 & 3.41 & -7.25 & 0.39 & 0 & -7.8 & 0 & 8.76 & 0 & -0.08 & 0 & -7.26 & 0 & 0 & 0 & 0 & 0 \\ \text{-CH2(OH)} & 3.41 & 0 & -17.84 & 0 & 0 & 0 & 0 & -1.4 & 0 & -1.34 & 0 & -0.13 & 0 & 0.37 & 0 & 0 & 0 \\ \text{-COOH} & -7.25 & -17.84 & 14.71 & -1.14 & 0 & 0 & 0 & -1.58 & 0 & 5.46 & 0 & 9.75 & 0 & 0.37 & 0.25 & 0.37 & 0.37 \\ \text{-CHO} & 0.39 & 0 & -1.14 & 0 & 0 & 0 & 0 & 2.08 & 0 & 5.49 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{-CH2p} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3.26 & 0 & -1.49 & 0 & 0 & 0 & 0 & 0 \\ \text{-COp} & -7.8 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 12.55 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{=CH2} & 0 & 0 & 0 & 0 & 0 & 0 & 0.75 & 0 & 0.75 & 0 & 0.75 & 0 & 0.75 & 0 & 0 & 0.37 & 0.37 \\ \text{-CH2-} & 8.76 & -1.4 & -1.58 & 2.08 & 0 & 0 & 0 & -0.48 & 0 & -5.18 & 0 & -5.1 & 0 & 0 & 0 & 0 & 0 \\ \text{=CH(OH)} & 0 & 0 & 0 & 0 & 0 & 0 & 0.75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{-CH(OH)-} & -0.08 & -1.34 & 5.46 & 5.49 & -3.26 & 12.55 & 0 & -5.18 & 0 & -0.92 & 0 & 7.95 & 0 & 0 & 0 & 0 & 0 \\ \text{=CO} & 0 & 0 & 0 & 0 & 0 & 0 & 0.75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{-CO-} & -7.26 & -0.13 & 9.75 & 0 & -1.49 & 0 & 0 & -5.1 & 0 & 7.95 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{=CHp} & 0 & 0 & 0 & 0 & 0 & 0 & 0.75 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{-CHp-} & 0 & 0.37 & 0.37 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{-CH=} & 0 & 0 & 0.25 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0 & 0 \\ \text{-C(OH)=} & 0 & 0 & 0.37 & 0 & 0 & 0 & 0.37 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \text{-Cp=} & 0 & 0 & 0.37 & 0 & 0 & 0 & 0.37 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

A.3 Perfect enzyme calculation

The reaction rate we have used in this work, corresponding to a perfect enzyme [52, 74–76], is:

$$v = \frac{k_d[E]_t \left([S_{i-1}]q_i - [S_i]\right)}{1 + q_i}. \tag{A.20}$$

Here we show how this can be obtained by considering a special case of the simple enzyme mechanism depicting the reversible binding of the substrate S and product P to the enzyme E:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} EP \underset{k_{-3}}{\overset{k_3}{\rightleftharpoons}} E + P \tag{A.21}$$

The steady state of this process is given by [74]

$$v = \frac{[E]_t(k_1k_2k_3S - k_{-1}k_{-2}k_{-3}P)}{k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3 + (k_1k_{-2} + k_1k_3 + k_1k_2)S + (k_{-1}k_{-3} + k_2k_{-3} + k_{-2}k_{-3})P} \tag{A.22}$$

Now, we follow the approach of Pettersson [76] and consider an enzyme operating via scheme (A.21) which, in the absence of product ($P \to 0$ in Eq. A.22), will conform to an irreversible Michaelis-Menten rate equation:

$$v = \frac{[E]_t(k_1k_2k_3[S])}{k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3 + (k_1k_{-2} + k_1k_3 + k_1k_2)[S]} \equiv \frac{[E]_t k_{\text{cat}}[S]}{K_M + [S]} \tag{A.23}$$

yielding the following relationships between the physiological parameters and the microscopic rate constants

$$k_{\text{cat}} = \frac{k_2k_3}{k_{-2} + k_3 + k_2} \quad \text{and} \quad K_M = \frac{k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3}{k_1(k_{-2} + k_3 + k_2)}. \tag{A.24}$$

In the earliest stages of enzyme evolution it is likely that enzymes will have shown a very low affinity for their substrates and had low rates of catalytic action, implying that k_2,$k_{-2} \ll k_{-1}$,k_3 and thus $k_{\text{cat}} \approx k_2$ and $K_M \approx k_{-1}/k_1$. In Pettersson it is assumed that for enzymes operating in scheme (A.21), the substrate and product association rate constants k_1 and k_{-3} are unaffected by evolution and take the value $k_d = 10^9$ $\text{M}^{-1}\text{s}^{-1}$ often attributed to the rate constant for a diffusion-controlled binding, k_d. Since k_1 is assumed fixed, then the decrease in K_M during early times must have be due to a decrease of k_{-1}. The accompanying evolutionary increase of k_2 will then lead to a situation in which $k_2 \gg k_{-1}$. Writing

$$\frac{k_{\text{cat}}}{K_M} = \frac{k_1k_2k_3}{k_{-1}k_{-2} + k_{-1}k_3 + k_2k_3} \tag{A.25}$$

and multiplying top and bottom by $1/k_1k_2k_3$, noting that the equilibrium constant of the reaction is $q = k_1k_2k_3/k_{-1}k_{-2}k_{-3}$, gives

$$\frac{k_{\text{cat}}}{K_M} = \frac{1}{\frac{1}{qk_{-3}} + \frac{k_{-1}}{k_1k_2} + \frac{1}{k_1}}. \tag{A.26}$$

Applying the diffusion limit $k_1 = k_{-3} = k_d$ we find that the regime in which $k_2 \gg k_{-1}$

leads finally to the expression

$$\frac{k_{\text{cat}}}{K_M} \approx \frac{k_d}{1+q^{-1}} = \frac{k_d q}{1+q} \tag{A.27}$$

Recall that for the irreversible reaction $S \rightarrow P$ the Michaelis-Menten rate is $v = k_{\text{cat}}[E]_t[S]/(K_M + [S])$. At low substrate concentrations, the Michaelis-Menten rate equation takes the linear form $v \approx \left(\frac{k_{\text{cat}}[E]_t}{K_M}\right)[S]$. If we assume that the substrate and product metabolites have the same affinity, and assume the reaction is operating in this linear regime, we can use this approximation to write

$$v = k_{\text{forw}}[S] - k_{\text{back}}[P] = \left(\frac{k_d[E]_t}{1+q^{-1}}\right)[S] - \left(\frac{k_d[E]_t}{1+q}\right)[P] \tag{A.28}$$

$$v = k_d[E]_t\left(\frac{[S]q - [P]}{1+q}\right) \tag{A.29}$$

which is Eqn. A.20 for a perfect catalyst.

A.4 Depth-first search algorithm

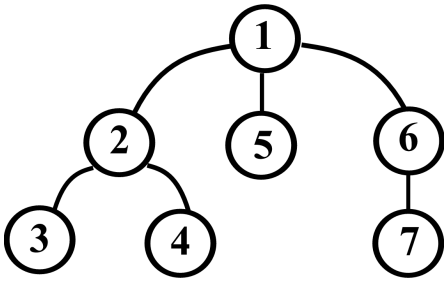


Figure A.2 Illustration of the ordering of node visits in the depth-first search (DFS) algorithm.

There are many algorithms available for searching or spanning graphs, one of the most common being the depth-first search (DFS) algorithm. The DFS algorithm was used a number of times in this thesis; in the creation of all molecules from the list of chemical groups, in the enumeration of all possible metabolic pathways, and as a tool to check network connectivity in the evolutionary simulations. It is an algorithm that traverses a graph exhaustively, i.e. in all possible ways, with the traversal performed in a depth-first, as opposed to a breadth-first, manner. That is, routes are explored as *deeply* as possible, until either no more steps can be taken or some defined maximal cut-off length is reached, before back-tracking and trying alternative branches.

The process is simple to explain pictorially for a tree (a graph with no closed loops), as

in Figure A.2. The number of the nodes indicates the order in which they are visited. Assuming that the left-hand node is always visited first, then starting at node 1 the algorithm visits node 2 and then 3, at which point no more nodes can be reached. The algorithm then back-tracks to node 2 and finds one possible route, moving to node 4. From here, no more nodes can be visited, and the algorithm must back-track to node 1 before finding any possibilities. Choosing the left node, the algorithm moves to node 5, before back-tracking again to node 1 and visiting finally nodes 6 and then 7. The network of reactions generated in the present work is not of a tree structure but contains many loops. To avoid these, one simply maintains a list of all visited nodes in the current route and rejects any moves that would lead to these. Below is the pseudo-code for the recursive implementation of the DFS algorithm (recursive implies that the method repeatedly calls itself).

```
takeStep( ... ){

   if( pathLenth == desiredLength ):
        return path;
   else:
      for( each neighbour of current node):
           add neighbour to visited list if unvisited;
           make neighbour current node;
           pathLength++;

           takeStep( ... );

           remove last node visited;
           pathLength--;
}
```

A.5 Powell's method

In calculating the flux of each pathway we fixed the total enzyme concentration $[E_T]$ but distributed the enzymes along the pathway so as to maximize the flux, with the constraint that all steady-state intermediate metabolite concentrations fell within some given interval $[S_{\min}, S_{\max}]$. The flux calculation thus became the following optimization problem,

$$\begin{aligned} &\text{Maximize } J([E_i]), \text{ subject to:} \\ &(1)\ \sum_i [E_i] = [E_T], \text{ and} \\ &(2)\ S_{\min} \leq [S_i] \leq S_{\max}\ \forall i\ , \end{aligned}$$

which was solved using Powell's method. Powell's method is an algorithm for finding the local maximum of a function in multiple dimensions, in which derivatives of the

function need not be taken. A detailed description can be found in [80], from where we obtained the computer code to perform the optimization. Powell's method is a *direction set method*, utilizing multiple one-dimensional line optimizations. The basic procedure behind the algorithm is that it takes in some initial point $\mathbf{P_0}$ (corresponding to some initial set of enzyme concentrations) and set of direction vectors $\mathbf{u_i}$. These direction vectors are initially the basis vectors corresponding to the parameters of the model (the individual enzyme concentrations).

The algorithm performs a line maximization along each of the direction vectors in turn, ending at point $\mathbf{P_N}$. This line maximization is performed using the *golden section search method*, also described in [80]. It then constructs a new vector, $\mathbf{P_N} - \mathbf{P_0}$, being the net direction travelled after trying all possible directions, and performs a line maximization along this direction. The procedure decides whether to keep the same set of direction vectors in the next iteration (the case if for example the resulting increase in our flux function was not primarily due to any one direction in particular) or to replace the vector along which our function made the largest increase during the last iteration with this newly calculated direction vector (this choice is to minimize the buildup of linear dependence between the direction vectors, since this discarded vector is likely to be a large component of the new direction vector). The algorithm proceeds until the change in the flux function in any iteration falls below some tolerance value, and iterates from multiple starting points to ensure the global maximum is found. In the case of our trunk pathways, initializing from 100 starting points was enough to obtain maximal fluxes consistent to within 1%, much less than the typical differences between pathways.

The algorithm is easily modified to respect the enzyme and metabolite constraints. For the fixed enzyme concentration, the number of parameters in the optimization is simply reduced by one and the remaining enzyme concentration calculated as $[E_n] = [E_T] - \sum_{i=1}^{i=n-1}[E_i]$. To respect the restrictions on the intermediate metabolite concentrations, every time this constraint is not satisfied the flux of the pathway is set to zero.

A.6 Flat enzyme distribution

In Section 5.9 we discussed the effect of not optimizing the enzyme distribution, but instead assuming that all enzymes are at an equal concentration when calculating the flux through our candidate pathways. Figure A.3 shows the result of the comparison when this is the case. Comparing this to Figure 5.2 we see that the general outcome

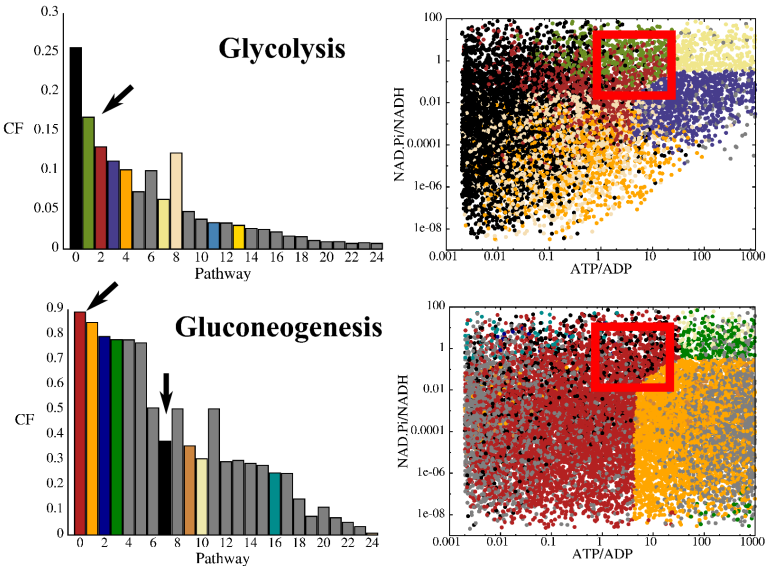


Figure A.3 Comparative flux, CF, for each pathway averaged over all parameter space (left) and best performing pathway in different regions of parameter space (right) where now the enzyme distribution has not been optimized to maximize flux, but instead all enzymes of the pathway are present at the same concentration. The real pathways (shown by arrows in the left panels) outperform the alternatives in the physiological region of parameter space (red boxes in right panels). The colour coding is as in Figure 5.2.

of our analysis is unchanged.

A.7 Alternative paths

Shown are the alternative glycolytic (Fig. A.4) and gluconeogenic (Fig. A.5) pathways discussed in Section 5.7 in more mechanistic detail.

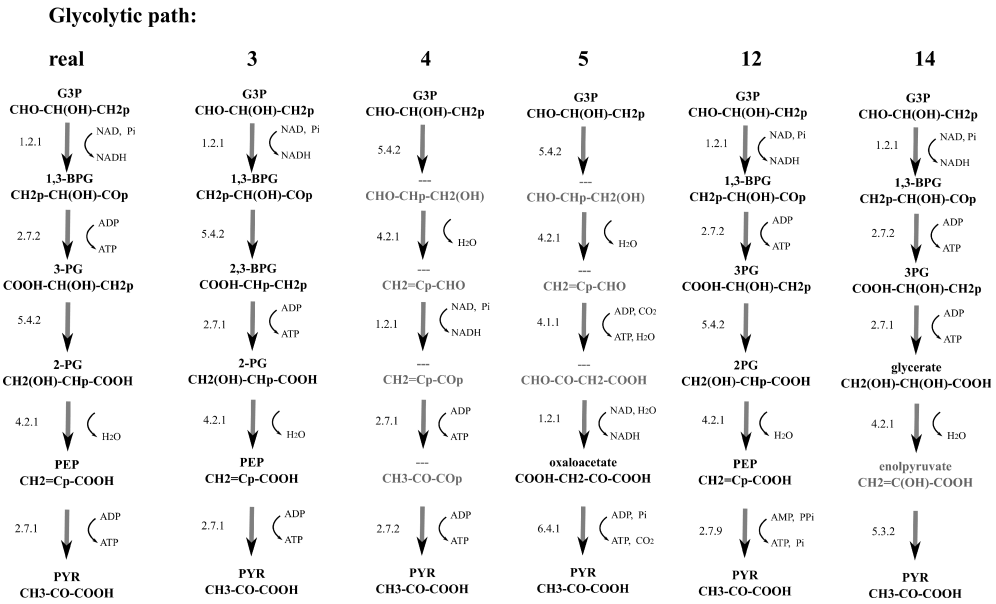


Figure A.4 Alternative glycolytic pathways found in our analysis. Glycolytic pathways 3, 4, 5, 12 and 14 are shown, labelling all external metabolite couplings and first 3 EC numbers. Compound names and chemical formulae shown in bold are present in the KEGG database.

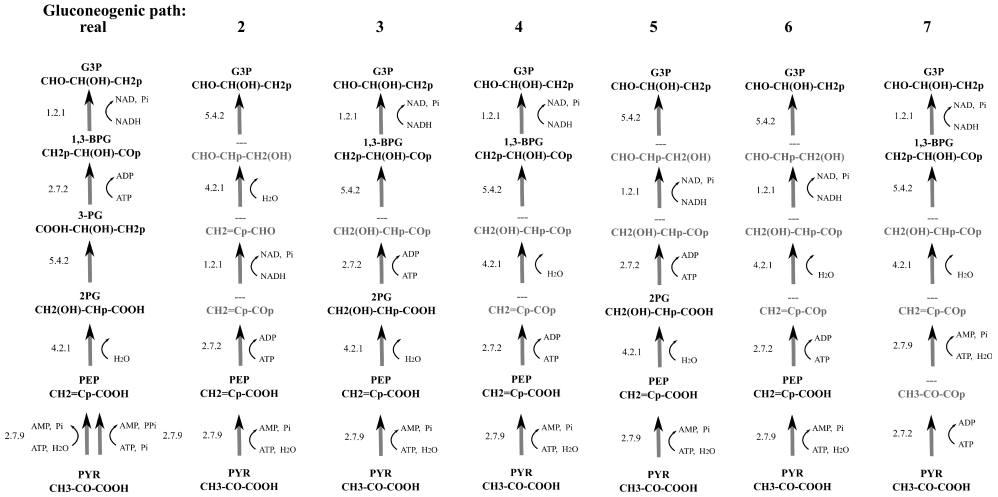


Figure A.5 Gluconeogenic pathways 2-7, showing all external metabolite couplings and first 3 EC numbers . Compound names and chemical formulae in black are present in KEGG database, grey are not. Note that these top-performing gluconeogenic pathways all utilize the same set of reaction classes.

Appendix B

Appendix to chapters 7 and 8

B.1 Two-point functions for the OPA

Table B.1 lists all possible transitions of neighbouring lattice sites ($XY \to X'Y'$ where $X, Y \in \{0, 1, 2\}$), along with the transition rate, indicator variable and resulting change in the observables of the system. It is important to note that these configurations are ordered pairs. This has two crucial consequences: (i) In these expressions, the sum over k is over the $z-1$ nearest-neighbours of site i that are *not* site j while the sum over ℓ is over the $z-1$ nearest neighbours of site j that are *not* site i. The additional $+1$ contributions to some of the transition rates (e.g., that of $10 \to 11$) come from the interaction between site i and site j. (ii) The change in the observable depends on the ordering. For instance, $\delta\tau_i\sigma_j(\mathcal{T})$ has a contribution of -1 from the transition $12 \to 11$ but a contribution of 0 from the transition $21 \to 11$.

$\mathcal{T}$	$\omega(\mathcal{T})$	$I(\mathcal{T})$	$\delta\tau_i\tau_j(\mathcal{T})$	$\delta\tau_i\sigma_j(\mathcal{T})$	$\delta\sigma_i\sigma_j(\mathcal{T})$
$00 \to 10$	$\frac{b_0}{z}\sum_k(\tau_k - \sigma_k)$	$\bar{\tau}_i\bar{\tau}_j$	0	0	0
$00 \to 20$	$\frac{b_0}{z}\sum_k \sigma_k$	$\bar{\tau}_i\bar{\tau}_j$	0	0	0
$00 \to 01$	$\frac{b_0}{z}\sum_\ell(\tau_\ell - \sigma_\ell)$	$\bar{\tau}_i\bar{\tau}_j$	0	0	0
$00 \to 02$	$\frac{b_0}{z}\sum_\ell \sigma_\ell$	$\bar{\tau}_i\bar{\tau}_j$	0	0	0
$10 \to 00$	d_0	$(\tau_i - \sigma_i)\bar{\tau}_j$	0	0	0
$10 \to 20$	$\frac{b_1}{z}\sum_k \sigma_k$	$(\tau_i - \sigma_i)\bar{\tau}_j$	0	0	0

$\mathcal{T}$	$\omega(\mathcal{T})$	$I(\mathcal{T})$	$\delta\tau_i\tau_j(\mathcal{T})$	$\delta\tau_i\sigma_j(\mathcal{T})$	$\delta\sigma_i\sigma_j(\mathcal{T})$
$10 \to 11$	$\frac{b_0}{z}\left[\sum_\ell(\tau_\ell - \sigma_\ell) + 1\right]$	$(\tau_i - \sigma_i)\bar{\tau}_j$	1	0	0
$10 \to 12$	$\frac{b_0}{z}\sum_\ell \sigma_\ell$	$(\tau_i - \sigma_i)\bar{\tau}_j$	1	1	0
$20 \to 00$	d_0	$\sigma_i\bar{\tau}_j$	0	0	0
$20 \to 10$	d_1	$\sigma_i\bar{\tau}_j$	0	0	0
$20 \to 21$	$\frac{b_0}{z}\sum_\ell(\tau_\ell - \sigma_\ell)$	$\sigma_i\bar{\tau}_j$	1	0	0
$20 \to 22$	$\frac{b_0}{z}\left[\sum_\ell \sigma_\ell + 1\right]$	$\sigma_i\bar{\tau}_j$	1	1	1
$01 \to 00$	d_0	$\bar{\tau}_i(\tau_j - \sigma_j)$	0	0	0
$01 \to 02$	$\frac{b_1}{z}\sum_\ell \sigma_\ell$	$\bar{\tau}_i(\tau_j - \sigma_j)$	0	0	0
$01 \to 11$	$\frac{b_0}{z}\left[\sum_k(\tau_k - \sigma_k) + 1\right]$	$\bar{\tau}_i(\tau_j - \sigma_j)$	1	0	0
$01 \to 21$	$\frac{b_0}{z}\sum_k \sigma_k$	$\bar{\tau}_i(\tau_j - \sigma_j)$	1	0	0
$11 \to 01$	d_0	$(\tau_i - \sigma_i)(\tau_j - \sigma_j)$	-1	0	0
$11 \to 21$	$\frac{b_1}{z}\sum_k \sigma_k$	$(\tau_i - \sigma_i)(\tau_j - \sigma_j)$	0	0	0
$11 \to 10$	d_0	$(\tau_i - \sigma_i)(\tau_j - \sigma_j)$	-1	0	0
$11 \to 12$	$\frac{b_1}{z}\sum_\ell \sigma_\ell$	$(\tau_i - \sigma_i)(\tau_j - \sigma_j)$	0	1	0
$21 \to 01$	d_0	$\sigma_i(\tau_j - \sigma_j)$	-1	0	0
$21 \to 11$	d_1	$\sigma_i(\tau_j - \sigma_j)$	0	0	0
$21 \to 20$	d_0	$\sigma_i(\tau_j - \sigma_j)$	-1	0	0
$21 \to 22$	$\frac{b_1}{z}\left[\sum_\ell \sigma_\ell + 1\right]$	$\sigma_i(\tau_j - \sigma_j)$	0	1	1
$02 \to 00$	d_0	$\bar{\tau}_i\sigma_j$	0	0	0
$02 \to 01$	d_1	$\bar{\tau}_i\sigma_j$	0	0	0
$02 \to 12$	$\frac{b_0}{z}\sum_k(\tau_k - \sigma_k)$	$\bar{\tau}_i\sigma_j$	1	1	0
$02 \to 22$	$\frac{b_0}{z}\left[\sum_k \sigma_k + 1\right]$	$\bar{\tau}_i\sigma_j$	1	1	1
$12 \to 10$	d_0	$(\tau_i - \sigma_i)\sigma_j$	-1	-1	0
$12 \to 11$	d_1	$(\tau_i - \sigma_i)\sigma_j$	0	-1	0
$12 \to 02$	d_0	$(\tau_i - \sigma_i)\sigma_j$	-1	-1	0
$12 \to 22$	$\frac{b_1}{z}\left[\sum_k \sigma_k + 1\right]$	$(\tau_i - \sigma_i)\sigma_j$	0	0	1
$22 \to 02$	d_0	$\sigma_i\sigma_j$	-1	-1	-1

$\mathcal{T}$	$\omega(\mathcal{T})$	$I(\mathcal{T})$	$\delta\tau_i\tau_j(\mathcal{T})$	$\delta\tau_i\sigma_j(\mathcal{T})$	$\delta\sigma_i\sigma_j(\mathcal{T})$
$22 \to 12$	d_1	$\sigma_i\sigma_j$	0	0	-1
$22 \to 20$	d_0	$\sigma_i\sigma_j$	-1	-1	-1
$22 \to 21$	d_1	$\sigma_i\sigma_j$	0	-1	-1

Performing the sum (8.3) we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\tau_i\tau_j\rangle = \frac{b_0}{z}\left[\left\langle\tau_i\bar{\tau}_j\left(\sum_\ell\tau_\ell+1\right)\right\rangle+\left\langle\bar{\tau}_i\tau_j\left(\sum_k\tau_k+1\right)\right\rangle\right]-2d_0\langle\tau_i\tau_j\rangle \tag{B.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\tau_i\sigma_j\rangle = \frac{b_0}{z}\left[\left\langle\tau_i\bar{\tau}_j\sum_\ell\sigma_\ell\right\rangle+\langle\sigma_i\bar{\tau}_j\rangle+\left\langle\bar{\tau}_i\sigma_j\left(\sum_k\tau_k+1\right)\right\rangle\right]+ \frac{b_1}{z}\left[\left\langle\tau_i(\tau_j-\sigma_j)\sum_\ell\sigma_\ell\right\rangle+\langle\sigma_i(\tau_j-\sigma_j)\rangle\right]-(2d_0+d_1)\langle\tau_i\sigma_j\rangle \tag{B.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\sigma_i\sigma_j\rangle = \frac{b_0}{z}\left[\left\langle\sigma_i\bar{\tau}_j\left(\sum_\ell\sigma_\ell+1\right)\right\rangle+\left\langle\bar{\tau}_i\sigma_j\left(\sum_l\sigma_k+1\right)\right\rangle\right]+ \frac{b_1}{z}\left[\left\langle\sigma_i(\tau_i-\sigma_j)\left(\sum_\ell\sigma_\ell+1\right)\right\rangle+\left\langle(\tau_i-\sigma_i)\sigma_j\left(\sum_k\sigma_k+1\right)\right\rangle\right]- 2(d_0+d_1)\langle\sigma_i\sigma_j\rangle\ . \tag{B.3}$$

Expanding these out and performing the sums leads finally to the equations of motion for the two-point functions, Eqns. (8.18)–(8.20). In doing so, one must take care to recall the site ordering: in any triplet average above, the indicator variables must appear in the correct order with respect to their indices, i.e. kij or ijl. Also, note again that the sums over k and ℓ are over the $(z-1)$ nearest neighbours of sites i and j respectively.

Notice that we had 6 one-point function transitions and 36 two-point transitions. Applying the same procedure above to higher order correlations already becomes analytically intractable, with a total of 216 distinct transitions for the three-point functions.

B.2 OPA results

The ordinary pair approximation has been shown to capture some of the behaviour of the 2-level stacked contact processes but provided a poor quantitative description for

the parameter set discussed in the main text. Here I show the correspondence between the simulations and the OPA for parameter sets in which the OPA provides a good match to the simulation data.

In the two plots of Figure B.1, the λ_0 values lead to high host densities. At small d_0 this host lattice is essentially frozen and thus the parasite dynamics will be close to that of a standard CP. The OPA actually gives quite accurate predictions for the standard CP (especially at higher densities) which explains the good agreement at low d_0. However as we increase d_0 the increased vertical transfer (birth of infected individuals) leads to higher parasite clustering, which the OPA does not capture since it only takes into account pairs of neighbouring sites. Thus as this effect becomes stronger for larger d_0 values, the discrepancy between the simulations and the OPA continues to grow.

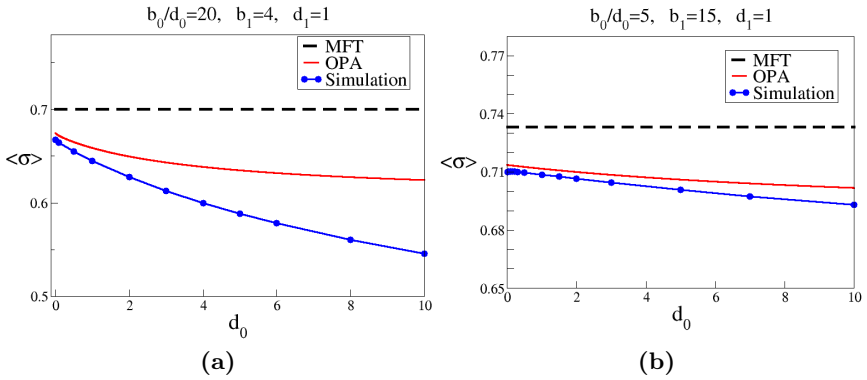


Figure B.1 Mean field theory (MFT), ordinary pair approximation (OPA) and simulation results for different parameter values, varying host turn over rate with fixed $\lambda_0 = b_0/d_0$, b_1 and d_1.

Bibliography

[1] Court, S. J, Blythe, R. A, & Allen, R. J. (2013) Parasites on parasites: Coupled fluctuations in stacked contact processes. *EPL (Europhysics Letters)* **101**, 50001.

[2] Court, S. J, Waclaw, B, & Allen, R. J. (2014) Lower glycolysis carries a higher flux than any biochemically possible alternative (`arXiv:1404.2900`).

[3] Berg, J, Tymoczko, J, & Stryer, L. (2002) *Biochemistry.* (W H Freeman and Company), 5th edition.

[4] (Oct 1st, 2013) Kyoto encyclopedia of genes and genomes, http://www.genome.jp/kegg/.

[5] Chang, R. (2005) *Physical Chemistry for the Biosciences.* (University Science Books), 1st edition.

[6] Trefil, J, Morowitz, H. J, & Smith, E. (2009) The origin of life. *American Scientist* **97**, 206–213.

[7] Koppenol, W. H, Bounds, P. L, & Dang, C. V. (2011) Otto Warburg's contributions to current concepts of cancer metabolism. *Nat Rev Cancer* **11**, 325–337.

[8] Kroemer, G & Pouyssegur, J. (2008) Tumor cell metabolism: Cancer's achilles' heel. *Cancer Cell* **13**, 472 – 482.

[9] Ro, D.-K, Paradise, E. M, Ouellet, M, Fisher, K. J, Newman, K. L, Ndungu, J. M, Ho, K. A, Eachus, R. A, Ham, T. S, Kirby, J, Chang, M. C. Y, Withers, S. T, Yoichiro, S, Sarpong, R, & Keasling, J. D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943.

[10] Radakovits, R, Jinkerson, R. E, Darzins, A, & Posewitz, M. C. (2010) Genetic engineering of algae for enhanced biofuel production. *Eukaryotic Cell* **9**, 486–501.

[11] Lemaux, P. G. (2008) Genetically engineered plants and foods: A scientist's analysis of the issues (part I). *Annual Review of Plant Biology* **59**, 771–812.

[12] Bar-Even, A, Flamholz, A, Noor, E, & Milo, R. (2012) Rethinking glycolysis: on the biochemical logic of metabolic pathways. *Nat Chem Biol* **8**, 509–517.

[13] Klamt, S, Haus, U.-U, & Theis, F. (2009) Hypergraphs and cellular networks. *PLoS Comput Biol* **5**, e1000385.

[14] Wagner, I & Musso, H. (1983) New naturally occurring amino acids. *Angewandte Chemie International Edition in English* **22**, 816–828.

[15] Peregrin-Alvarez, J. M, Sanford, C, & Parkinson, J. (2009) The conservation and evolutionary modularity of metabolism. *Genome biology* **10**, R63.

[16] Barve, A & Wagner, A. (2013) A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature* **500**, 203–206.

[17] Orth, J. D, Conrad, T. M, Na, J, Lerman, J. A, Nam, H, Feist, A. M, & Palsson, B. O. (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism – 2011. *Molecular Systems Biology* **7**.

[18] Fink, T, Bruggesser, H, & Reymond, J.-L. (2005) Virtual exploration of the small-molecule chemical universe below 160 daltons. *Angewandte Chemie* **117**, 1528–1532.

[19] Jeong, H, Tombor, B, Albert, R, Oltvai, Z. N, & Barabasi, A.-L. (2000) The large-scale organization of metabolic networks. *Nature* **407**, 651–654.

[20] Albert, R & Barabási, A.-L. (2002) Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97.

[21] Wagner, A & Fell, D. A. (2001) The small world inside large metabolic networks. *Proc. R. Soc. Lond. B* **268**, 1803–1810.

[22] Albert, R, Jeong, H, & Barabasi, A.-L. (2007) Error and attack tolerance of complex networks. *Nature* **406**, 378–382.

[23] Barabási, A.-L & Albert, R. (1999) Emergence of scaling in random networks. *Science* **286**, 509–512.

[24] Ravasz, E, Somera, A. L, Mongru, D. A, Oltvai, Z. N, & Barabási, A.-L. (2002) Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555.

[25] Arita, M. (2004) The metabolic world of Escherichia coli is not small. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 1543–1547.

[26] Montañez, R, Medina, M. A, Solé, R. V, & Rodríguez-Caso, C. (2010) When metabolism meets topology: Reconciling metabolite and reaction networks. *BioEssays* **32**, 246–256.

[27] Tanaka, R. (2005) Scale-rich metabolic networks. *Phys. Rev. Lett.* **94**, 168101.

[28] Noor, E, Bar-Even, A, Flamholz, A, Lubling, Y, & Davidi, D. (2012) An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics* **28**, 2037–2044.

[29] Noor, E, Eden, E, Milo, R, & Alon, U. (2010) Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Molecular Cell* **39**, 809 – 820.

[30] Bilgen, T. (2005) *Metabolic Evolution and the Origin of Life.* (John Wiley & Sons, Inc.), pp. 557–582.

[31] Romano, A & Conway, T. (1996) Evolution of carbohydrate metabolic pathways. *Research in Microbiology* **147**, 448 – 455.

[32] Morowitz, H. J. (1999) A theory of biochemical organization, metabolic pathways, and evolution. *Complexity* **4**, 39–53.

[33] Morowitz, H. J. (1992) *Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis*. (Yale University Press), 2nd edition.

[34] Flamholz, A, Noor, E, Bar-Even, A, Liebermeister, W, & Milo, R. (2013) Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proceedings of the National Academy of Sciences*.

[35] Peekhaus, N & Conway, T. (1998) What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. *Journal of Bacteriology* **180**, 3495–3502.

[36] Verhees, C. H, Kengen, S. W. M, Tuininga, J. E, Schut, G. J, Adams, M. W. W, de Vos, W. M, & van der Oost, J. (2003) The unique features of glycolytic pathways in Archaea. *Biochem. J.* **375**, 231–246.

[37] Srinivasan, V & Morowitz, H. J. (2009) Analysis of the intermediary metabolism of a reductive chemoautotroph. *Biological Bulletin* **217**, 222–232.

[38] Ronimus, R. S & Morgan, H. W. (2002) Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea* **1**, 199–221.

[39] Fell, D. (1996) *Understanding the Control of Metabolism (Frontiers in Metabolism)*. (Portland Pr), 1st edition.

[40] Orth, J. D, Thiele, I, & Palsson, B. O. (2010) What is flux balance analysis? *Nature Biotechnology* **28**, 245–248.

[41] Kauffman, K. J, Prakash, P, & Edwards, J. S. (2003) Advances in flux balance analysis. *Current Opinion in Biotechnology* **14**, 491 – 496.

[42] Liebermeister, W & Klipp, E. (2006) Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theoretical Biology and Medical Modelling* **3**.

[43] Liebermeister, W, Uhlendorf, J, & Klipp, E. (2010) Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics* **26**, 1528–1534.

[44] Stanford, N. J, Lubitz, T, Smallbone, K, Klipp, E, Mendes, P, & Liebermeister, W. (2013) Systematic construction of kinetic models from genome-scale metabolic networks. *PLoS ONE* **8**, e79195.

[45] Johnson, K. A & Goody, R. S. (2011) The original Michaelis constant: Translation of the 1913 Michaelis-Menten paper. *Biochemistry* **50**, 8264–8269.

[46] Briggs GE, H. J. (1925) A note on the kinetics of enzyme action. *Biochem J.* **19**, 338–339.

[47] Noor, E, Flamholz, A, Liebermeister, W, Bar-Even, A, & Milo, R. (2013) A note on the kinetics of enzyme action: A decomposition that highlights thermodynamic effects. *FEBS Letters* **587**, 2772 – 2777. A century of Michaelis-Menten kinetics.

[48] Teusink, B, Passarge, J, Reijenga, C. A, Esgalhado, E, van der Weijden, C. C, Schepper, M, Walsh, M. C, Bakker, B. M, van Dam, K, Westerhoff, H. V, & Snoep, J. L. (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry* **267**, 5313–5329.

[49] van Eunen, K, Kiewiet, J, Westerhoff, H. V, & Bakker, B. M. (2010) The use of in vivo-like enzyme kinetics in a computational model of yeast glycolysis. Proceedings of the 4th international ESCEC symposium on experimental standard conditions of enzyme characterizations.

[50] Tummler, K, Lubitz, T, Schelker, M, & Klipp, E. (2014) New types of experimental data shape the use of enzyme kinetics for dynamic network modeling. *FEBS Journal* **281**, 549–571.

[51] Ebenhöh, O & Heinrich, R. (2001) Evolutionary optimization of metabolic pathways. Theoretical reconstruction of the stoichiometry of ATP and NADH producing systems. *Bulletin of Mathematical Biology* **63**, 21–55.

[52] Heinrich, R, Montero, F, Klipp, E, Waddell, T. G, & Meléndez-Hevia, E. (1997) Theoretical approaches to the evolutionary optimization of glycolysis: Thermodynamic and kinetic constraints. *European Journal of Biochemistry* **243**, 191–201.

[53] Meléndez-Hevia, E & Isidoro, A. (1985) The game of the pentose phosphate cycle. *Journal of Theoretical Biology* **117**, 251–263.

[54] Bar-Even, A, Noor, E, Lewis, N. E, & Milo, R. (2010) Design and analysis of synthetic carbon fixation pathways. *Proceedings of the National Academy of Sciences* **107**, 8889–8894.

[55] Handorf, T, Ebenhöh, O, & Heinrich, R. (2005) Expanding metabolic networks: Scopes of compounds, robustness, and evolution. *Journal of Molecular Evolution* **61**, 498–512.

[56] Schuetz, R, Kuepfer, L, & Sauer, U. (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology* **3**, 119.

[57] Soyer, O. S & Pfeiffer, T. (2010) Evolution under fluctuating environments explains observed robustness in metabolic networks. *PLoS Comput Biol* **6**, e1000907.

[58] Khandelwal, R. A, Olivier, B. G, Röling, W. F. M, Teusink, B, & Bruggeman, F. J. (2013) Community flux balance analysis for microbial consortia at balanced growth. *PLoS ONE* **8**, e64567.

[59] Schuetz, R, Zamboni, N, Zampieri, M, Heinemann, M, & Sauer, U. (2012) Multidimensional optimality of microbial metabolism. *Science* **336**, 601–604.

[60] Dekel, E & Alon, U. (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–592.

[61] Webb, E. (1992) *Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes.* (Academic Press, San Diego), 6th edition.

[62] (Oct 1st, 2013) Enzyme nomenclature. http://www.chem.qmul.ac.uk/iubmb/enzyme/.

[63] Alberty, R. A. (1998) Calculation of standard transformed entropies of formation of biochemical reactants and group contributions at specified pH. *The Journal of Physical Chemistry A* **102**, 8460–8466.

[64] Alberty, R. (1983) Chemical thermodynamic properties of isomer groups. *Industrial & Engineering Chemistry Fundamentals* **22**, 318–321.

[65] Alberty, R & Goldberg, R. (1992) Standard thermodynamic formation properties for the adenosine 5'-triphosphate series. *Biochemistry* **31**, 10610–10615.

[66] Alberty, R. A. (2005) Mathematical functions for thermodynamic properties of biochemical reactants (`http://library.wolfram.com/infocenter/Demos/5704`).

[67] Alberty, R. A. (2006) *Biochemical Thermodynamics: Applications of Mathematica*, Methods of Biochemical Analysis.

[68] Jankowski, M. D, Henry, C. S, Broadbelt, L. J, & Hatzimanikatis, V. (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophysical Journal* **95**, 1487 – 1499.

[69] Mavrovouniotis, M. L. (1990) Group contributions for estimating standard gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnology and Bioengineering* **36**, 1070–1082.

[70] Mavrovouniotis, M. L. (1991) Estimation of standard gibbs energy changes of biotransformations. *Journal of Biological Chemistry* **266**, 14440–14445.

[71] van Krevelen, D & Chermin, H. (1951) Estimation of the free enthalpy (gibbs free energy) of formation of organic compounds from group contributions. *Chemical Engineering Science* **1**, 66 – 80.

[72] Morowitz, H. J, Kostelnik, J. D, Yang, J, & Cody, G. D. (2000) The origin of intermediary metabolism. *Proceedings of the National Academy of Sciences* **97**, 7704–7708.

[73] Hatzimanikatis, V, Li, C, Ionita, J. A, Henry, C. S, Jankowski, M. D, & Broadbelt, L. J. (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics* **21**, 1603–1609.

[74] Heinrich, R & Hoffmann, E. (1991) Kinetic parameters of enzymatic reactions in states of maximal activity; an evolutionary approach. *J. theor. Biol* **151**, 249–283.

[75] Albery, J. W & Knowles, J. R. (1976) Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* **15**, 5631–5640.

[76] Pettersson, G. (1989) Effect of evolution on the kinetic properties of enzymes. *European Journal of Biochemistry* **184**, 561–566.

[77] Bennett, B. D, Kimball, E. H, Gao, M, Osterhout, R, Van Dien, S. J, & Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol* **5**, 593–599.

[78] Garrett, R & Grisham, C. (2007) *Biochemistry*. (Belmont, CA: Thomson Brooks/Cole), 3rd edition, p. 584.

[79] Phillips, S. A & Thornalley, P. J. (1993) The formation of methylglyoxal from triose phosphates. *European Journal of Biochemistry* **212**, 101–105.

[80] Press, W. H, Teukolsky, S. A, Vetterling, W. T, & Flannery, B. P. (2007) *Numerical Recipes: The Art of Scientific Computing*. (Cambridge University Press, New York, NY, USA), 3rd edition.

[81] Berg, J, Hung, Y. P, & Yellen, G. (2009) A genetically encoded fluorescent reporter of ATP:ADP ratio. *Nature Methods* **6**, 161–166.

[82] Cho, J, King, J. S, Qian, X, Harwood, A. J, & Shears, S. B. (2008) Dephosphorylation of 2,3-bisphosphoglycerate by MIPP expands the regulatory capacity of the Rapoport-Luebering glycolytic shunt. *Proceedings of the National Academy of Sciences* **105**, 5998–6003.

[83] Feng, X.-M, Cao, L.-J, Adam, R. D, Zhang, X.-C, & Lu, S.-Q. (2008) The catalyzing role of ppdk in giardia lamblia. *Biochemical and Biophysical Research Communications* **367**, 394 – 398.

[84] Berg, I. A, Kockelkorn, D, Ramos-Vera, W. H, Say, R. F, Zarzycki, J, Hugler, Michael Alber, B. E, & Fuchs, G. (2010) Autotrophic carbon fixation in archaea. *Nature Reviews Microbiology* **8**, 447–460.

[85] Varma, A & Palsson, B. O. (1994) Metabolic flux balancing - basic concepts, scientific and practical use. *Biotechnology* **12**, 994–998.

[86] Haraldsdóttir, H, Thiele, I, & Fleming, R. (2012) Quantitative assignment of reaction directionality in a multicompartmental human metabolic reconstruction. *Biophysical Journal* **102**, 1703–1711.

[87] Noor, E, Bar-Even, A, Flamholz, A, Reznik, E, Liebermeister, W, & Milo, R. (2014) Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Computational Biology* **10**.

[88] Bustos, D. M, Bustamante, C. A, & Iglesias, A. A. (2008) Involvement of non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase in response to oxidative stress. *Journal of Plant Physiology* **165**, 456–461.

[89] Ahmed, H, Ettema, T. J. G, Tjaden, B, Geerling, A. C. M, van der Oost, J, & Siebers, B. (2005) The semi-phosphorylative Entner-Doudoroff pathway in hyperthermophilic archaea: a re-evaluation. *Biochem. J.* **390**, 529–540.

[90] Kouril, T, Esser, D, Kort, J, Westerhoff, H. V, Siebers, B, & Snoep, J. L. (2013) Intermediate instability at high temperature leads to low pathway efficiency for an in vitro reconstituted system of gluconeogenesis in *Sulfolobus solfataricus*. *FEBS Journal* **280**, 4666–4680.

[91] Zhoua, J, Olsona, D. G, Argyrosc, D. A, Denga, Y, van Gulikd, W. M, van Dijkend, J. P, & Lynda, L. R. (2013) Atypical glycolysis in *Clostridium thermocellum*. *Applied and Environmental Microbiology* **79**, 3000–3008.

[92] Raverdy, S, Foster, J. M, Roopenian, E, & Carlow, C. K. (2008) The *Wolbachia* endosymbiont of *Brugia malayi* has an active pyruvate phosphate dikinase. *Molecular and Biochemical Parasitology* **160**, 163–166.

[93] Diaz-Ruiz, R, Rigoulet, M, & Devin, A. (2011) The Warburg and Crabtree effects: On the origin of cancer cell energy metabolism and of yeast glucose repression. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1807**, 568–576. Bioenergetics of Cancer.

[94] Thornalley, P. J, Langborg, A, & Minhas, H. S. (1999) Formation of glyoxal, methylglyoxal and 3-deoxyglucosone in the glycation of proteins by glucose. *Biochem. J.* **344**, 109–116.

[95] Heinrich, R, Schuster, S, & Holzhütter, H.-G. (1991) Mathematical analysis of enzymic reaction systems using optimization principles. *European Journal of Biochemistry* **201**, 1–21.

[96] Sauro, H. (2011) *Enzyme Kinetics for Systems Biology*. (Ambrosius Publishing).

[97] Young, T. R & Boris, J. P. (1977) A numerical technique for solving stiff ordinary differential equations associated with the chemical kinetics of reactive-flow problems. *The Journal of Physical Chemistry* **81**, 2424–2427.

[98] Shampine, L. F & Gear, C. W. (1979) A user's view of solving stiff ordinary differential equations. *SIAM Review* **21**, 1–17.

[99] Éric Walter. (2014) *Numerical methods and optimization: A consumer guide*. (Springer International Publishing).

[100] Bar-Even, A, Noor, E, Savir, Y, Liebermeister, W, Davidi, D, Tawfik, D. S, & Milo, R. (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410.

[101] Ellson, J, Gansner, E. R, Koutsofios, E, North, S. C, & Woodhull, G. (2003) *Graphviz and dynagraph – static and dynamic graph drawing tools*. (Springer-Verlag), pp. 127–148.

[102] Qian, H & Beard, D. (2006) Metabolic futile cycles and their functions: a systems analysis of energy and control. *Systems Biology, IEE Proceedings* **153**, 192–200.

[103] Medema, M. H, van Raaphorst, R, Takano, E, & Breitling, R. (2012) Computational tools for the synthetic design of biochemical pathways. *Nat Rev Micro* **10**, 191–202.

[104] Cho, A, Yun, H, Park, J. H, Lee, S. Y, & Park, S. (2010) Prediction of novel synthetic pathways for the production of desired chemicals. *BMC Systems Biology* **4**.

[105] Henry, C. S, Broadbelt, L. J, & Hatzimanikatis, V. (2010) Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. *Biotechnology and Bioengineering* **106**, 462–473.

[106] Kotera, M, Tabei, Y, Yamanishi, Y, Tokimatsu, T, & Goto, S. (2013) Supervised de novo reconstruction of metabolic pathways from metabolome-scale compound sets. *Bioinformatics* **29**, i135–i144.

[107] Johansson, F et al. (2013) *mpmath: a Python library for arbitrary-precision floating-point arithmetic (version 0.18)*. `http://mpmath.org/`.

[108] Baker-Austin, C & Dopson, M. (2007) Life in acid: pH homeostasis in acidophiles. *Trends in Microbiology* **15**, 165 – 171.

[109] Ugi, I, Bauer, J, Brandt, J, Friedrich, J, Gasteiger, J, Jochum, C, & Schubert, W. (1979) New applications of computers in chemistry. *Angewandte Chemie International Edition in English* **18**, 111–123.

[110] Weininger, D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences* **28**, 31–36.

[111] Tilman, D & Kareiva, P. (1997) *Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions*, Monographs in population biology. (Princeton University Press).

[112] Durrett, R & Levin, S. (1994) The importance of being discrete (and spatial). *Theoretical Population Biology* **46**, 363 – 394.

[113] Krone, S. (2004) Spatial models: stochastic and deterministic. *Mathematical and Computer Modelling* **40**, 393–409.

[114] Rand, D. A, Keeling, M, & Wilson, H. B. (1995) Invasion, stability and evolution to criticality in spatially extended, artificial host-pathogen ecologies. *Proceedings: Biological Sciences* **259**, pp. 55–63.

[115] Levin, S. A. (1974) Dispersion and population interactions. *The American Naturalist* **108**, 207–228.

[116] Markham, D. C, Simpson, M. J, Maini, P. K, Gaffney, E. A, & Baker, R. E. (2013) Incorporating spatial correlations into multispecies mean-field models. *Phys. Rev. E* **88**, 052713.

[117] Webb, S. D, Keeling, M. J, & Boots, M. (2007) Host-parasite interactions between the local and the mean-field: How and when does spatial population structure matter? *Journal of Theoretical Biology* **249**, 140–152.

[118] Durrett, R & Levin, S. A. (1994) Stochastic spatial models: A user's guide to ecological applications. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **343**, 329–350.

[119] Liggett, T. (1985) *Interacting Particle Systems.* (Springer-Verlag, New York).

[120] Wolfram, S. (1983) Statistical Mechanics of Cellular Automata. *Reviews of Modern Physics* **55**, 601–644.

[121] Krone, S, Lu, R, Fox, R, Suzuki, H, & Top, E. (2007) Modelling the spatial dynamics of plasmid transfer and persistence. *Microbiology* **153**, 2803 – 2816.

[122] Tainaka, K, Hayashi, T, & Yoshimura, J. (2006) Sustainable sex ratio in lattice populations. *EPL (Europhysics Letters)* **74**, 554.

[123] Harada, Y & Iwasa, Y. (1994) Lattice population dynamics for plants with dispersing seeds and vegetative propagation. *Researches on Population Ecology* **36**, 237–249.

[124] Rhodes, C & Anderson, R. (1996) Persistence and dynamics in lattice models of epidemic spread. *Journal of Theoretical Biology* **180**, 125–133.

[125] Drossel, B & Schwabl, F. (1992) Self-organized critical forest-fire model. *Phys. Rev. Lett.* **69**, 1629–1632.

[126] Louzoun, Y, Solomon, S, Goldenberg, J, & Mazursky, D. (2003) World-size global markets lead to economic instability. *Artificial Life* **9**, 357–370.

[127] Cross, M. C & Hohenberg, P. C. (1993) Pattern formation outside of equilibrium. *Rev. Mod. Phys.* **65**, 851–1112.

[128] Venegas-Ortiz, J, Allen, R. J, & Evans, M. R. (2014) Speed of invasion of an expanding population by a horizontally transmitted trait. *Genetics* **196**, 497–507.

[129] Keeling, M. J & Eames, K. T. (2005) Networks and epidemic models. *Journal of The Royal Society Interface* **2**, 295–307.

[130] Newman, M. E. J. (2002) Spread of epidemic disease on networks. *Phys. Rev. E* **66**, 016128.

[131] Henkel, M, Hinrichsen, H, & Lubeck, S. (2008) *Non-equilibrium phase transitions, Volume 1: Absorbing phase transitions.* (Springer).

[132] Ódor, G. (2008) *Universality in Nonequilibrium Lattice Systems: Theoretical Foundations.* (World Scientific).

[133] Marro, J & Dickman, R. (1999) *Nonequilibrium Phase Transitions in Lattice Models.* (Cambridge University Press).

[134] Janssen, H. (1981) On the nonequilibrium phase transition in reaction-diffusion systems with an absorbing stationary state. *Zeitschrift f'ur Physik B Condensed Matter* **42**, 151–154.

[135] Grinstein, G, Lai, Z.-W, & Browne, D. A. (1989) Critical phenomena in a nonequilibrium model of heterogeneous catalysis. *Phys. Rev. A* **40**, 4820–4823.

[136] Muñoz, M. A, Grinstein, G, Dickman, R, & Livi, R. (1996) Critical behavior of systems with many absorbing states. *Phys. Rev. Lett.* **76**, 451–454.

[137] Takeuchi, K. A, Kuroda, M, Chaté, H, & Sano, M. (2007) Directed percolation criticality in turbulent liquid crystals. *Phys. Rev. Lett.* **99**, 234503.

[138] Hohenberg, P. C & Halperin, B. I. (1977) Theory of dynamic critical phenomena. *Rev. Mod. Phys.* **49**, 435–479.

[139] Binder, K & Stauffer, D. (1974) Theory for the slowing down of the relaxation and spinodal decomposition of binary mixtures. *Phys. Rev. Lett.* **33**, 1006–1009.

[140] Konno, N. (1994) *Phase transitions of interacting particle systems.* (World Scientific).

[141] Harris, T. (1974) Contact interactions on a lattice. *Annals of Probability* **2**, 969–988.

[142] Vojta, T & Lee, M. Y. (2006) Nonequilibrium phase transition on a randomly diluted lattice. *Phys. Rev. Lett.* **96**, 035701.

[143] Harris, A. B. (1974) Effect of random defects on the critical behaviour of Ising models. *J. Phys. C: Solid State Phys* **7**, 1729–1741.

[144] de Oliveira, M. M & Ferreira, S. C. (2008) Universality of the contact process with random dilution. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P11001.

[145] Moreira, A. G & Dickman, R. (1996) Critical dynamics of the contact process with quenched disorder. *Phys. Rev. E* **54**, R3090–R3093.

[146] Dickman, R & Moreira, A. G. (1998) Violation of scaling in the contact process with quenched disorder. *Phys. Rev. E* **57**, 1263–1268.

[147] Vojta, T & Dickison, M. (2005) Critical behavior and Griffiths effects in the disordered contact process. *Phys. Rev. E* **72**, 036126.

[148] Griffiths, R. B. (1969) Nonanalytic behavior above the critical point in a random Ising ferromagnet. *Phys. Rev. Lett.* **23**, 17–19.

[149] Fallert, S. V, Kim, Y. M, Neugebauer, C. J, & Taraskin, S. N. (2008) Contact process in disordered and periodic binary two-dimensional lattices. *Phys. Rev. E* **78**, 041117.

[150] C.J. Neugebauer, S. F. (2006) Contact process in heterogeneous and weakly disordered systems. *Physical Review E* **74**, 1729–1741.

[151] Rajamani, K, Pate, W. T, & Kinneberg, D. J. (1986) Time-driven and event-driven Monte Carlo simulations of liquid-liquid dispersions: a comparison. *Industrial & Engineering Chemistry Fundamentals* **25**, 746–752.

[152] Weitz, J. S & Rothman, D. H. (2003) Scale-dependence of resource-biodiversity relationships. *Journal of Theoretical Biology* **225**, 205–214.

[153] Cencini, M, Pigolotti, S, & Muñoz, M. A. (2012) What ecological factors shape species-area curves in neutral models? *PLoS ONE* **7**, e38232.

[154] de Oliveira, M. M, Dos Santos, R. V, & Dickman, R. (2012) Symbiotic two-species contact process. *Phys. Rev. E* **86**, 011121.

[155] Krone, S. M. (1999) The two-stage contact process. *The Annals of Applied Probability* **9**, 331–351.

[156] Foxall, E. (2014) New results for the two-stage contact process (`arXiv:1401.2570`).

[157] Deshayes, A. (2014) The contact process with aging (`arXiv:1405.6153`).

[158] Shoop, W. (1991) Vertical transmission of helminths: Hypobiosis and amphiparatenesis. *Parasitology Today* **7**, 51–54.

[159] Smith, J & Dunn, A. (1991) Transovarial transmission. *Parasitology Today* **7**, 146–148.

[160] Ochman, H, Lawrence, J. G, & Groisman, E. A. (2005) Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.

[161] Turner, P. E, Cooper, V. S, & Lenski, R. E. (1998) Tradeoff between horizontal and vertical modes of transmission in bacterial plasmids. *Evolution* **52**, 315–329.

[162] Schinazi, R. (2000) Horizontal versus vertical transmission of parasites in a stochastic spatial model. *Mathematical Biosciences* **168**, 1–8.

[163] Busenberg, S, Cooke, K. L, & Pozio, M. A. (1983) Analysis of a model of a vertically transmitted disease. *Journal of Mathematical Biology* **17**, 305–329.

[164] Lipsitch, M, Nowak, M, Ebert, D, & May, R. (1995) The population dynamics of vertically and horizontally transmitted parasites. *Proc. R. Soc. B* **260**, 321–327.

[165] Lipsitch, M, Siller, S, & Nowak, M. (1996) Evolution of virulence in pathogens with vertical and horizontal transmission. *Evolution* **50**, 1729–1741.

[166] Jones, E, White, A, & Boots, M. (2011) The evolution of host protection by vertically transmitted parasites. *Proc. R. Soc. B* **278**, 863–870.

[167] S. Faeth, K. Hadeler, H. T. (2007) An apparent paradox of horizontal and vertical disease transmission. *Journal of Biological Dyanmics* **1**, 45–62.

[168] R.Fox, X. Zhong, S. E. (2008) Spatial structure and nutrients promote invasion of IncP-1 plasmids in bacterial populations. *The ISME Journal* **2**, 1024–1039.

[169] Liu, Y.-C, Durrett, R, & Milgroom, M. G. (2000) A spatially-structured stochastic model to simulate heterogenous transmission of viruses in fungal populations. *Ecological Modelling* **127**, 291 – 308.

[170] Filipe, J & Gibson, G. (2001) Comparing approximations to spatio-temporal models for epidemics with local spread. *Bulletin of Mathematical Biology* **63**, 603–624.

[171] Sullivan, D. J & Völkl, W. (1999) Hyperparasitism: Multitrophic ecology and behavior. *Annual Review of Entomology* **44**, 291–315.

[172] Andersen, S. B, Ferrari, M, Evans, H. C, Elliot, S. L, Boomsma, J. J, & Hughes, D. P. (2012) Disease dynamics in a specialized parasite of ant societies. *PLoS ONE* **7**, e36352.

[173] Milgroom, M. G & Cortesi, P. (2004) Biological control of chestnut blight with hypovirulence: A critical analysis. *Annual Review of Phytopathology* **42**, 311–338.

[174] Swinton, J & Gilligan, C. A. (1999) Selecting hyperparasites for biocontrol of Dutch elm disease. *Proceedings: Biological Sciences* **266**, 437–445.

[175] Douglas, A. E. (1998) Nutritional interactions in insect-microbial symbioses: Aphids and their symbiotic bacteria buchnera. *Annual Review of Entomology* **43**, 17–37.

[176] Wang, H.-H, Grant, W. E, & Teel, P. D. (2012) Simulation of climate–host–parasite–landscape interactions: A spatially explicit model for ticks (Acari: *Ixodidae*). *Ecological Modelling* **243**, 42–62.

[177] Swift, J. (1733) On poetry: A rhapsody.

[178] Alon, U, Evans, M. R, Hinrichsen, H, & Mukamel, D. (1996) Roughening transition in a one-dimensional growth process. *Phys. Rev. Lett.* **76**, 2746.

[179] Alon, U, Evans, M. R, Hinrichsen, H, & Mukamel, D. (1998) Smooth phases, roughening transitions, and novel exponents in one-dimensional growth models. *Phys. Rev. E* **57**, 4997–5012.

[180] Blythe, R. A & Evans, M. R. (2001) Slow crossover to Kardar-Parisi-Zhang scaling. *Phys. Rev. E* **64**, 051101.

[181] Meakin, P, Ramanlal, P, Sander, L. M, & Ball, R. C. (1986) Ballistic deposition on surfaces. *Phys. Rev. A* **34**, 5091–5103.

[182] Täuber, U. C, Howard, M. J, & Hinrichsen, H. (1998) Multicritical behavior in coupled directed percolation processes. *Phys. Rev. Lett.* **80**, 2165–2168.

[183] Filipe, J. (1999) Hybrid closure-approximation to epidemic models. *Physica A* **266**, 238–241.

[184] Sato, K, Matsuda, H, & Sasaki, A. (1994) Pathogen invasion and host extinction in lattice structured populations. *Journal of Mathematical Biology* **32**, 251–268.

[185] Hiebeler, D. E & Millett, N. E. (2011) Pair and triplet approximation of a spatial lattice population model with multiscale dispersal using markov chains for estimating spatial autocorrelation. *Journal of Theoretical Biology* **279**, 74 – 82.

[186] Haraguchi, Y & Sasaki, A. (2000) The evolution of parasite virulence and transmission rate in a spatially structured population. *Journal of Theoretical Biology* **203**, 85–96.

[187] Webb, S. D, Keeling, M. J, & Boots, M. (2007) Host-parasite interactions between the local and the mean-field: How and when does spatial population structure matter? *J. Theor. Biol.* **249**, 140–152.

[188] Hasegawa, T, Konno, N, & Masuda, N. (2011) Numerical study of a three-state host-parasite system on the square lattice. *Phys. Rev. E* **83**, 046102.

[189] Schinazi, R. (1996) On an interacting particle system modeling an epidemic. *Journal of Mathematical Biology* **34**, 915–925.

[190] Raman, K & Chandra, N. (2009) Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics* **10**, 435–449.