Of systems and cancer

Joost Boele

Members of the doctoral examination committee:

prof.dr. Jaap Heringa
VU University Amsterdam
dr. Michiel de Hoon
RIKEN Yokohama Institute
prof.dr. Antoine van Kampen
Academic Medical Center, University of Amsterdam
dr. Richard Notebaart
Nijmegen Centre for Molecular Life Sciences
prof.dr. Martine J. Smit
VU University Amsterdam
dr. Manfred Wuhrer
Leiden University Medical Center

Cover design: Elise Marcus – http://www.elisemarcus.com
Printed by: Gildeprint – Enschede
ISBN: 9789461087911

The work in this thesis was carried out in the Systems Bioinformatics group of the Amsterdam Institute for Molecules, Medicines and Systems at the Vrije Universiteit Amsterdam, and was supported by the department of Computer Science.

Copyright © 2014 Joost Boele. All rights reserved. No part of this publication may be reproduced in any form without written permission from the author.

VRIJE UNIVERSITEIT

Of systems and cancer

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Exacte Wetenschappen
op woensdag 5 november 2014 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Joost Boele

geboren te Amsterdam

promotor:

prof.dr. B. Teusink

Table of contents

1 Introduction **1**

2 Isoform-specific regulation of oncomiR miR-21 **31**

3 Differential CXCR4/CXCR7/CXCL12-axis signaling in breast cancer cells **45**

4 FAME, the Flux Analysis and Modeling Environment **63**

5 An interactive graphical map of metabolism **73**

6 The biomass composition of human cells **87**

7 General discussion **111**

References **127**

Nederlandse samenvatting **147**

English summary **151**

Acknowledgments (dankwoord) **155**

List of publications **159**

Chapter 1

Introduction

In collaboration with Azra Mujic-Delic, Filipe Santos, Frank J. Bruggeman, Martine J. Smit, and Bas Teusink.

Adapted from: *Meth Enz* (2011), **500**:509–532.

The once underappreciated role of metabolism in cancer

The past years have seen research efforts into the metabolic aspects of cancer reinvigorated. This renewed attention is not very surprising, as we do not yet understand why the metabolic make-up of highly proliferative cells is so different from that of their healthy counterparts. Among the most prominent differences is the Warburg effect, the tendency of many cancers to forgo the oxidation of glucose, even when they have enough oxygen available [1]. In addition, recent discoveries of regulatory functions of metabolites and metabolic enzymes [2,3] suggest a previously underappreciated role of metabolism in oncogenesis.

In the study of the Warburg effect, significant progress came from the discovery of mutations and alternative splice isoforms of metabolic enzymes [4], as well as from the elucidation of regulatory pathways that are dysfunctional in cancerous tissue [5]. Recent data even indicates that metabolism can also influence signaling dynamics, which is in contrast to more classical views that suggest that signaling networks are in full control of metabolism. Evidence for these regulatory interactions include the discoveries that certain small organic acids have additional roles as allosteric effectors of signaling molecules [6,7], and that some metabolic intermediates act as phosphate donors for protein kinases. The latter phenomenon can influence the phosphorylation of target proteins [3], or directly alter their function [2].

These new discoveries illustrate that the interface between metabolism and signaling is not strictly one-way. Nevertheless, the signaling machinery traditionally receives more attention in the context of cancer. Disturbed signaling networks are pivotal in all of the hallmarks of cancer that were originally proposed by Hanahan & Weinberg in 2000 [8], and under normal circumstances, signaling networks are tightly regulated. However, mutations in key signaling proteins can lead to overexpression of oncogenes and silencing of tumor suppressor genes, which can result in deregulation of the entire signaling network. Examples of over- or underexpression of signaling proteins leading to dysfunction are as numerous as they are detrimental [9]. For instance, tumor suppressor p53 is known to be mutated or deleted in more than 50% of human cancers [10,11], and amplification of the epidermal growth factor family member gene HER2 (human epidermal growth factor receptor 2, also known as ErbB2) has been found in 20-25% of all breast cancers [12], as well as in other cancer types like gastric cancer [13]. Additionally, several members of the PI3K/Akt signaling pathway and the MAPK/ERK signaling pathway are

often found to have mutations in cancer [9]. Both pathways are initiated by the activation of receptor tyrosine kinases (RTKs) and G protein-coupled receptors (GPCRs) by their respective ligands, and can trigger downstream responses that lead to proliferation and cell survival.

Although on paper it may seem that alterations to the signaling machinery can by themselves explain the cancer phenotype, decades of experience with hit-or-miss therapeutic interventions that target signaling components speak to the contrary. Moreover, whereas every other paper about cancer starts by remarking that the disease is "multifaceted" or "complex", very few of these papers actually deal with more than one facet. Especially aspects that are as broadly defined as "metabolism" and "signaling" are seldom viewed from an integrative perspective, and as a result, the interface between these domains is poorly understood. However, in recognition of the necessity of gaining such understanding, "reprogramming of energy metabolism" was added to the list of cancer hallmarks when it was recently revised [14].

Here, we will discuss the reciprocal relationship between signaling and metabolism in the context of cancer, as well as the computational techniques that can be employed to elucidate the intricacies of these large domains. First, let us look at both domains separately.

Oncogenic adaptations of core metabolism

Very generally speaking, metabolism is reponsible for the conversion of what the cell finds in its environment to what the cell needs to function or reproduce. "What the cell needs" is a very broad definition, and it could range from the production of ATP, amino acids, and lipids for maintenance (or for the production of biomass), to the storage of carbohydrates in the form of glycogen, or to the performance of some function that is expected of that particular cell (type). The latter category includes e.g. hormone production by glandular cells, detoxification processes in the liver, or ATP production by muscle cells that require this energy to contract when they are asked to. Proliferation places a great demand on the production of biomass components (and energy to put them together), and consequently, cancer cells display increased metabolic activity. However, even though this distinctive phenotype of cancer cells was first described almost a century ago [1], it has mostly managed to stay below the scientific radar – until recently. Recent research efforts have picked up this topic and identified many alterations of core metabolic enzymes in an attempt to explain the Warburg effect. After decades of research, there is still no single

satisfying explanation, but we will give a brief overview of the most prominent current views in this section.

Is metabolism a subordinate or a major player?

Otto Warburg envisioned metabolism at the center of it all. His hypothesis for the observation that cancer cells forgo the full oxidation of glucose in the TCA cycle in favor of a quick 2 ATP per glucose by producing lactate [15] was that cancer cells do not have functional mitochondria anymore [16]. With the advent of modern molecular biology and the subsequent shift in focus towards genetics, this view of metabolism as the driving force behind cells becoming cancerous has been surpassed by one in which metabolism is subordinate to signal transduction cascades and transcriptional regulation.

Viewing cancer cells as cells where a single runaway system can steer the entire cellular machinery towards unchecked replication may seem elegant (albeit not very robust), but can we be sure that this is the whole story? Some fairly recent insights suggest metabolism plays a more central role. For one, famously, enhanced anaerobic metabolism is a common feature of fast-growing cancers [17], but also of unicellular organisms growing at high substrate availability (e.g. the Crabtree effect in *S. Cerevisiae* [18] and the "bacterial Crabtree effect" in *E. Coli* [19]). Thus, cancer's low-energy yield aerobic fermentation of glucose into lactate may reflect a mammalian metabolic program that is also used by healthy proliferating cells. The recently renewed interest in cancer metabolism has led to knowledge of various individual adaptations that contribute to the Warburg effect, and this knowledge can be used to shed light on such fundamental issues. We will review these insights below.

A metabolic perspective on the Warburg effect

Perhaps the most prominent example of glycolytic enzyme dysregulation in cancer is pyruvate kinase (PK), the final enzyme in glycolysis (see Figure 1.1A). In many cancers, the dominant form of PK is its embryonic splice variant, known as pyruvate kinase M2 (PKM2) [20]. PKM2 can form either homodimers or -tetramers [20]. The tetrameric form, which is the predominant form in differentiated cells, has a high affinity for its substrate phosphoenolpyruvate (PEP), and exemplifies that PKM2 is not restricted to cancer cells [21]. In contrast, dimeric PKM2 is much less efficient at converting PEP to pyruvate than its tetrameric and non-embryonic counterparts [20]. Interestingly,

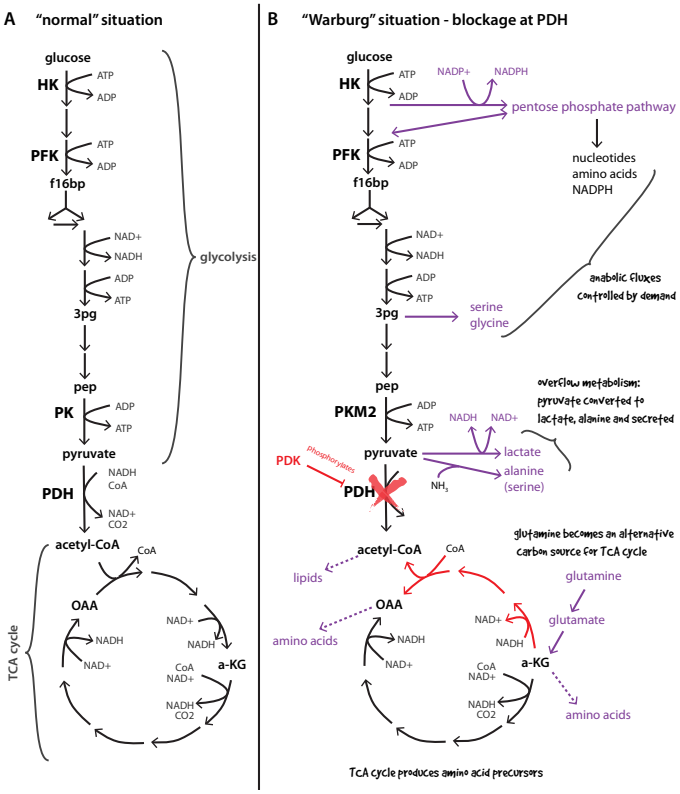


Figure 1.1: (A) In normal situations, glucose (primarily) flows towards pyruvate which is then converted by PDH to acetyl-CoA, which subsequently enters the TCA-cycle to produce energy andor metabolic intermediates. (B) In cancerous and other highly proliferative tissues, a different situation occurs. PDK phosphorylates PDH, inactivating it and cutting off flow towards the TCA cycle. Pyruvate and the metabolites upstream of it accumulate and can be converted to amino acids, or to other biomass precursors in the pentose phosphate pathway. Flux through the majority of these pathways is controlled by the demand side (i.e. the rate of biosynthesis) and cannot be considered "overflow metabolism"; instead, superfluous carbon or ammonium are secreted in the form of lactate and alanine, respectively. Meanwhile, the TCA cycle is fueled by glutamine and yields αKG or oxaloacetate (OAA), which can in turn be used in amino acid synthesis.

the formation of PKM2 tetramers is interfered with by phosphorylation of the Y105 residue of the protein [22]. This phosphorylation is common in human cancers [22], and consequently, the "inefficient" dimeric PKM2 is often found in cancer cells.

The decreased substrate affinity of dimeric PKM2, compared to tetrameric PKM2 or regular PK, presents a paradox: how can cancer cells have increased

glycolytic flux and enhanced lactate production if the final step of glycolysis is being slowed down? One explanation is the simultaneously enhanced expression of multiple glycolytic enzymes in cancer cells. Indeed, transcription factors with such broad impact on metabolism are known, and we will discuss them in detail in a later section. On a more theoretical note, the impact that individual enzyme activities have on pathway flux has been extensively studied within the theoretical framework of metabolic control analysis (MCA) [23]. In this light, the fact that PKM2 can have lower activity while the pathway flux increases, suggests that pyruvate kinase does not control flux through the pathway, and that mutations in this enzyme serve another purpose.

Metabolic enzymes and intermediates have alternative roles in cancer

In spite of these insights from MCA, one functional consequence of dimeric PKM2 that has been suggested is that it causes congestion in the glycolytic pathway, with the build-up of glycolytic intermediates eventually overflowing into other pathways, such as anabolic routes that lead to the formation of building blocks for amino acids, nucleotides, and more (Figure 1.1B) [20]. This hypothesis, known as "overflow metabolism", goes against metabolic control logic: As the flux through glycolysis is much higher than through anabolic pathways, and glucose is not limiting in most cases, one would expect "control by demand" (as shown by detailed supply-demand analyses carried out by Hofmeyr and colleagues [24], and explained in Box 1). In such a system the product of a biosynthetic pathway inhibits the first step of that pathway to make sure supply matches the demand for the product — a design commonly observed for e.g. amino acid biosynthesis. If anabolism would drain glycolytic intermediates, the high flux of glycolysis should simply be able to replenish these pools very effectively. Consequently, the flux through a biosynthetic pathway originating from glycolysis is expected to be regulated through "pull" by the demand side, not by "push" from the supply side (Box 1).

Box 1: Supply and demand in metabolic control analysis

With respect to any biomass precursor, one can view the reactions that produce that metabolite as "supply", and reactions that use it to form biomass as "demand". Under homeostatic conditions, the fluxes (J) through supply and demand are

equal, and the concentration of the biomass precursor does not change. Typically, the reactions that produce, say, an amino acid from a glycolytic intermediate, are highly sensitive to their own product, causing production of that amino acid to be inhibited unless it is used up by other (demand-side) reactions (Figure 1.2A). As a consequence, the flux through the supply reactions (v1) should be unaffected by the glycolytic intermediate concentration, and the only way to increase flux through the pathway is by increasing the demand (v2): this causes the amino acid's concentration to decrease and reduces the inhibition of v1. This is illustrated in Figure 1.2B: If demand (v2) increases, supply (v1) will be able to match it, while the amino acid's concentration may be affected slightly to mitigate its inhibition of v1. However, trying to increase flux through v1 by increasing the glycolytic intermediate concentration is ineffective, as the rate of v1 drops rapidly with the resulting increased amino acid concentration (Figure 1.2C). Supply-demand theory is reviewed in [24].

In bacteria, glycolytic intermediates have been suggested to act as "flux sensors" that help determine the metabolic status of the cell. For instance, high fructose 1,6-bisphosphate levels appear to indicate a high glucose flux state [25]. A build-up of glycolytic intermediates, such as the one we would expect to be caused by an inefficient enzyme at the end of the pathway (e.g. dimeric PKM2), could entail a slew of allosteric interactions that tell the cell what (not) to do. For instance, in tumor cells, the glycolytic intermediate 3-phosphoglycerate (3PG) has been shown to inhibit 6-phosphogluconate dehydrogenase, an enzyme in the pentose phosphate pathway [26]. 2-phosphoglycerate (2PG), which is one enzymatic step "downstream" of 3PG, in turn provides feedback control of 3PG levels by activating 3-phosphoglycerate dehydrogenase [26]. Similarly, 1,3-bisphosphoglycerate (1,3BPG) can spontaneously react with select lysine residues in proteins to form 3-phosphoglyceryl-lysine, a process that can result in the inhibition of glycolytic enzymes [27]. These phenomena are not limited to glycolytic intermediates: The TCA cycle intermediates succinate and fumarate have been shown to affect glycogen synthesis by allosteric interactions with glycogen synthase [27]. These and hundreds of other known allosteric interactions are collected in the "Allosteric Database" [28], a useful resource and an apt illustration of the extent to which these mechanisms extend their influence.

Although the above could make metabolism seem self-contained and self-regulating, it is still strongly dependent on external checks and balances. For instance, if pyruvate is not converted to lactate, it can enter the TCA cyle

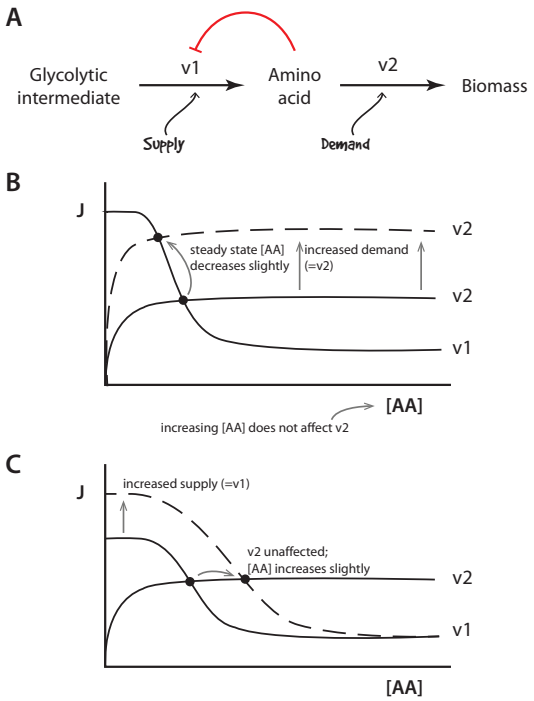


Figure 1.2: Supply and demand in metabolic control analysis. The legend for this figure is included in Box 1.

after being converted to acetyl-CoA by pyruvate dehydrogenase (PDH, Figure 1.1A). As the link between glycolysis and the TCA cycle, PDH is not only subject to transcriptional regulation, which is relatively slow (minutes to hours), but it can also be phosphorylated by PDH kinase (PDK), which reduces PDH's activity and is a much faster mechanism (seconds). Having the activity of PDH reduced significantly restricts flux from pyruvate to acetyl-CoA, and thus from glycolysis to the TCA cycle [29]. PDK has four known isoforms, of which two important ones (PDK2 and PDK4) are regulated by peroxisome proliferator-activated receptor γ (PPARγ), which is technically a nuclear receptor, but can also be considered a nutrient signaling transcription activator in this context [30]. PDH kinase 1 itself can also be directly phosphorylated, which enhances its activity [31]. Thus, unsurprisingly, alterations to PDH and the machinery that phosphorylates PDH are also commonly linked to cancerous phenotypes [31].

The TCA cycle sustains the production of biomass, but not of energy in cancer cells

Downstream of glycolysis, contrary to Warburg's initial hypothesis, most cancer cells do possess functional mitochondria, although the relative fluxes of the substrates entering the TCA cycle are often different between cancerous and healthy cells (see e.g. [32]). Notably, cancer cells are known to rely heavily on glutamine for sustaining their metabolism, to the point that many cancers depend on the availability of glutamine to survive [33]. After glutaminolysis, in which glutamine is converted into glutamate and glutamate is in turn converted into alpha-ketoglutarate (αKG; also known as 2-oxoglutarate), the glutamine-derived αKG can be used to produce amino acids (histidine, proline, and arginine) and the antioxidant peptide glutathione, or it can enter the TCA-cycle [20]. Recent work suggests that *de novo* production of lipids in cancer cells almost exclusively uses glutamine-derived carbon [34], and that the TCA cycle runs in the direction of reductive carboxylation (the opposite of oxidative phosphorylation) to make this carbon available as acetyl-CoA [35,36]. As enhanced proliferation is contingent on the availability of lipids and amino acids, it is likely that cancer cells have to maintain at least a basal level of TCA cycle activity to survive, with oxygen-dependent energy production a convenient side-effect of this rather than a primary goal. Accordingly, lactate formation reduces the cells' reliance on the availability of oxygen: (abundant) carbon can be "wastefully" metabolized to produce 2 ATP per glucose, while preserving oxygen for use by other cellular processes.

Functional and biochemical perspectives combined: overflow metabolism

So what happens if glucose is not fully oxidized? There are several views on this, but what they have in common is that the blockage of TCA entry (by the inhibitory phosphorylation of PDH) forces the cell to look elsewhere for the energy it requires. With the main carbon supply of the TCA cycle cut off, the cell must resort to increasing the glycolytic flux to produce sufficient ATP to support maintenance (and growth). This has a number of consequences. Famously, the large quantities of carbon that the cell imports must go somewhere, and the most commonly observed solution for this is the secretion of lactate [1]. The production of lactate from pyruvate does not use or produce ATP, but it does defray the NADH investment that is made upstream in the glycolytic pathway (Figure 1.1B). The production and secretion of lactate to

rid the cell of carbon and alanine to rid the cell ammonium are examples of what is known as "overflow metabolism", which is commonly considered to be a feature of fast-growing cancers [17]. As described above and in Box 1, it is unlikely that increased concentrations of glycolytic intermediates can "push" flux into biosynthetic pathways, as control of these processes is usually demand-driven [24]. Therefore, fluxes towards biomass components cannot be considered overflow mechanisms, although they are often elevated in proliferating cells.

It is still slightly paradoxical that inhibition of PDH and pyruvate kinase can lead to an enhanced glycolytic flux, but in part, control by demand and negative feedback can explain this. Upon PDH inactivation, lactate formation at a mere 2 ATP per glucose is the major source of ATP. Given this much lower yield, at the same glycolytic flux and unaltered ATP demand, ATP would quickly drop to unhealthy levels. As glycolysis responds to ATP demand (e.g. the enzyme phosphofructokinase is particularly sensitive to the energy charge [37]), glycolytic flux will increase via allosteric feedback mechanisms, strictly at the metabolic level. However, as enzymes get saturated and thus have an inherently limited capacity, the expression levels of these enzymes must increase if the cell is to sustain its enhanced glycolytic flux – and glycolytic flux must be quite high to make up for the diminished ATP yield of this pathway when lactate is the end point. Kacser et al. showed that to increase a flux through a pathway several-fold while maintaining metabolite levels at more or less constant levels, multisite modulation (the regulation of the levels of multiple enzymes in a pathway, as opposed to regulation of a single "key" enzyme) is required [38,39]. In the extreme case, in a linear pathway, doubling all the enzyme levels at once would double the pathway flux without changing any metabolite level. In cancer cells, hypoxia-inducible factor 1α (HIF1α) can achieve this in its role as a master regulator of glycolytic enzyme expression [40] (see below).

Pyruvate is a central molecule in metabolism, and it can be used for more than fueling the TCA cycle or producing lactate: Among other things, it can also act as a precursor for the amino acids alanine and serine. Alanine is often secreted by cancer cells [41,42], which may be a mechanism to get rid of excess nitrogen via a harmless amino acid with minimal carbon content, instead of via the toxic ammonia. Serine, in contrast, is not secreted, but its synthesis is still often increased in cancer compared to healthy tissue; this is particularly notable in breast cancer cells [32]. Of course the amount of serine that can be put into protein production must still be proportional to the other amino

acids, so any excess serine must go somewhere else. Indeed, a link between the biosynthesis routes that produce serine and αKG was observed in the same study [32]. Alternatively, serine can be converted to the amino acid glycine in a one step enzymatic process that also yields a one carbon unit to folate, which is required for proliferation [43]. Interestingly, serine is also an allosteric activator of PKM2 [44], and a relationship between PKM2 activity and serine and glycine consumption was recently reported [44]. In this study, researchers showed that deprivation of both amino acids from colon cancer cells leads to inhibition of pyruvate kinase activity and reduced lactate production, as well as increased mitochondrial citrate production [44].

Similarly to serine, the role of glycine in cancer is not fully understood, although new technologies offer tantalizing insights. In a recent study, metabolite profiling of a panel of 60 human cancer cell lines [45] indicated a surprising correlation between glycine consumption and cellular proliferation rates, apparently signifying that glycine requirements exceed the endogenous production capabilities of rapidly proliferating cells. Accordingly, the enzymes involved in the mitochondrial glycine biosynthetic pathway were found to have high expression levels, but more importantly, breast cancer patients with high expression levels of these enzymes (SHMT2, MTHFD2 and MTHFD1L) had higher mortality [45].

Expanding our view to include the rest of core metabolism, studies indicate substantial flux into the pentose phosphate pathway [46, 47] to produce nucleotide and amino acid precursors as well as NADPH, an energy source driving many anabolic/biosynthetic processes, and – potentially more important – a substrate for the reduction of reactive oxygen species. A final carbon sink of note is αKG. Intriguingly, in addition to its familiar roles as a TCA cycle intermediate and amino acid precursor, it also interacts with prolyl hydroxylases EGLN1-3 to destabilize the transcription factor HIF1α [48]. This transcription factor steers cells towards a Warburg-like metabolism if left unchecked [49], and we will discuss it in detail in the next section.

The influence of signaling pathways on metabolic regulation in cancer

The relation between signaling and tumorigenesis has been extensively studied, but even so, we are far from having a comprehensive understanding of the signaling network like we do for metabolism [50]. In this sense, it is illustrative that the means by which the signaling network and the metabolic network

are connected have started receiving attention only very recently [14, 51–54]. Though the full list of signaling proteins that influence metabolism is extensive, in this section we will highlight the ones that appear the most important in the context of cancer.

HIF1α is a key signaling molecule in the regulation of metabolism

As its name implies, hypoxia-inducible factor 1α is normally active in cells that undergo hypoxia. It regulates many genes involved in tumorigenesis, such as those governing angiogenesis, metastasis and invasion, genetic instability and, not surprisingly, glucose and energy metabolism [40]. In energy metabolism, HIF1α is known to regulate the expression of most glycolytic genes, including hexokinase 1 and 2 (HK1 and HK2), enolase 1 (ENO1), pyruvate kinase M2 (PKM2), and lactate dehydrogenase A (LDHA) [55, 56]; it also regulates the transcription of pyruvate dehydrogenase kinase 1 (PDK1) and glucose transporters GLUT1 and GLUT3 [57] (Figure 1.3).

But HIF1α is not always active: In the presence of oxygen, hydroxylation of proline residues targets the protein for proteasomal degradation [52]. This process cannot take place under hypoxic conditions, as the hydroxylation reaction utilizes oxygen as a cosubstrate [52], so HIF1α is rescued from degradation when there is no oxygen. Interestingly, some prolyl hydroxylases are αKG-dependent [7], effectively enabling feedback from the TCA cycle to a regulator of core metabolism. Conversely, prolyl hydroxylases can be inhibited by 2-hydroxy-glutarate, which is produced from αKG by mutated forms of isocitrate dehydrogenase (IDH1/2), a TCA-cycle enzyme [58, 59] (Figure 4). Furthermore, even in the presence of oxygen, lactate and pyruvate have been shown to cause HIF1α accumulation [5], which implies the existence of an intriguing (and dangerous!) feed-forward loop from (aerobic) glycolysis to the machinery enhancing it. This situation where HIF1α regulates metabolism, while metabolism regulates HIF1α, leads to the question of which of the two is in charge – but while we currently have quite some knowledge about HIF1α's antics, a clear-cut answer to this question still lies beyond our reach.

Since HIF1α plays such a pivotal role in the regulation of metabolism and influences practically all enzymes in glycolysis, it should not be surprising that it also has oncogenic potential. Indeed, mutants of the protein have been reported that have enhanced stability in aerobic conditions, compared to wild type HIF1α [60]. Another means of oxygen-independent preservation of HIF1α is through mTOR (mammalian Target Of Rapamycin)-induced protein synthesis, which occurs in multiple tumor types (Figure 1.3) [61]. Deregulation

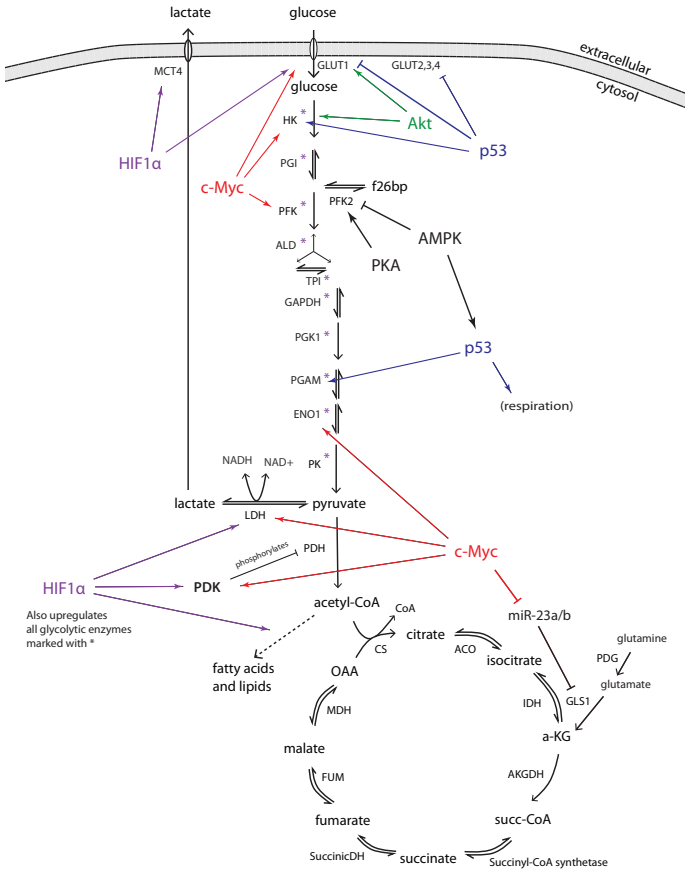


Figure 1.3: Core metabolism is constitutively active, but it is subject to strict hierarchical regulation by various transcription factors and signaling molecules. Ones that are particularly important in the context of cancer are HIF1α, which regulates all glycolytic enzymes (marked by *) and some non-glycolytic ones, c-Myc, Akt, and p53.

of mTOR, and the subsequent synthesis of HIF1α, can be caused by loss of function of PTEN (phosphatase and tensin homolog) or by HER2-induced overexpression, among other causes [55].

PTEN is an important tumor suppressor, whose surrounding signaling network has received much attention [62]. It is also a confirmed target of the microRNA miR-21 [63], one of the archetypical *onco*miRs [64]. We discovered that the isoform landscape of this miRNA, which was previously thought to be of little consequence, is a product of a novel regulatory mechanism which can induce the degradation of miR-21 (Chapter 2). This additional layer of regulation offers the cell more fine-grained control over its regulatory apparatus, but complicates matters further for those intending to understand it.

The rerouting of metabolic flux to a less oxygen-dependent mode fits well with HIF1α becoming active under hypoxic conditions, but metabolism is not the only cellular function HIF1α tunes towards low-oxygen conditions. Recently, the protein was reported to promote angiogenesis by enhancing the transcription of the microRNA miR-210 [65], and of the chemokine receptors CXCR4 and CXCR7 (C-X-C chemokine receptor types 4 and 7, respectively) [65–67]. CXCR7 is activated by CXCL12 (C-X-C motif chemokine 12, also known as SDF1α), the same ligand that activates CXCR4, but in contrast to CXCR4, the signal transduction pathways downstream of CXCR7 are poorly understood. However, their coordinated regulation by HIF1α suggests the interesting possibility of the two receptors working in harmony, e.g. by promoting angiogenesis in response to hypoxia. Indeed, in a high-throughput reverse phase protein array (RPPA) time course study, we found many signaling intermediates that seem to be affected by both receptors. This is described in Chapter 3.

Alterations in the transcription factor c-Myc contribute to mitochondrial function and glutaminolysis

Similarly to HIF1α, the transcription factor c-Myc is involved in the regulation of several glycolytic genes, including hexokinase 2 (HK2), enolase I (ENO1), lactate dehydrogenase A (LDHA) and phosphofructokinase (PFK) [68] (Figure 1.3). It is often altered, amplified, or chromosomally translocated in cancer: By estimation, increased c-Myc expression contributes to the cause of at least 40% of all human cancers [68]. Furthermore, c-Myc transcriptionally activates genes that increase mitochondrial biogenesis and mitochondrial function [53]. In addition to upregulating the glucose transporter GLUT1 [69], c-Myc is

known to directly induce the expression of glutamine transporters SLC5A1 and SLC7A1, and thus to also play a role in the regulation of glutaminolysis [70]. Interestingly, the same study showed that c-Myc is able to repress the expression of microRNAs miR-23a and miR-23b, and to thereby increase the levels of glutaminase 1 (GLS1) and contribute to another layer of control on glutaminolysis (Figure 1.3).

Since c-Myc and HIF1α seem to activate the same set of genes, one may wonder whether they do this concurrently and whether they are able to regulate and interact with one another. Interestingly, both are true. When c-Myc is overexpressed it cooperates with HIF1α to alter cellular metabolism and upregulate glycolytic enzymes [71]. Conversely, HIF1α is also known to antagonize c-Myc function by interacting with Myc-associated proteins, whereas its family member HIF2 seems to collaborate with c-Myc rather than to oppose it [71].

Signaling enzymes and transcription factors are intimately intertwined to allow for fine-grained control, but this can make it difficult to assign a single role to a single actor. For instance, much like p53, liver kinase B1 (LKB1) is known for its activity as a tumor suppressor, but also for connecting cell growth and energy metabolism by directly activating AMP-activated protein kinase (AMPK) [72]. AMPK acts as a sensor of intracellular ATP, ADP and AMP levels, and can serve as a central metabolic switch by promoting the generation of ATP by catabolic pathways and inhibiting anabolic pathways [73]. It can induce acute metabolic changes by phosphorylating specific metabolic enzymes, but it also exerts control on metabolism by modulating enzyme transcription [74]. Interestingly, AMPK is able to control cell growth and tumorigenesis by suppressing mTOR dependent transcriptional regulators [74], thus providing feedback from energy metabolism to the cell growth machinery.

On the interface between signal transduction and metabolism

As we noted above, metabolism is traditionally described in literature as a simple subordinate to signaling. But there are various ways in which metabolism influences signaling, ranging from allosteric interactions by metabolites to unexpected signaling functionality of metabolic enzymes. We will list the highlights of this emerging field here.

Metabolic intermediates can allosterically affect signaling

The most straightforward way for a metabolite to influence a process it does not itself participate in is through an allosteric interaction with the enzyme that catalyzes the reaction. While such interactions are abundant and well-studied in the case of metabolites influencing metabolic enzymes [28], common metabolites have also been shown to allosterically interact with signaling enzymes. For example, glutamine-derived αKG from the TCA cycle has been shown to activate mTORC1 through interactions with prolyl hydroxylases [7, 75] (Figure 1.4). Similarly, p53 activity has been reported to be suppressed by increased glucose metabolism [76]. Although this suppression appears to coincide with the expression of the glucose transporter GLUT1 and hexokinase [76], the precise cause is yet unknown.

Furthermore, lactate and pyruvate, important metabolites in core metabolism, have been shown to induce the accumulation of HIF1α [5]. The specific mechanism underlying this is currently unclear, but we do know that lactate can inhibit prolylhydroxylase 2 activity, and thereby stabilize HIF1α [77]. This was only observed in normoxic oxidative tumor cells and not in tumor cells with a Warburg-phenotype. Another recently described mechanism by which metabolites can influence signaling or even act as signaling molecules themselves is through metabolite-sensing GPCRs [78]. As such a sensor exists for lactate, this offers the interesting possibility for cancer cells to send signals to themselves by secreting lactate or other "waste products", such as alanine.

A major challenge at this point is that current knowledge about these interactions is of a qualitative nature. Quantitative (kinetic) data, such as those available for allosteric interactions in metabolism, will greatly aid our understanding of the interplay between metabolism and signaling, as they can be used to quantify the responses to various stimuli at a high resolution. However, unfortunately, straightforward assays of transcription factor or signaling enzyme activity are not yet available, which makes the study of this "upstream" feedback a technological challenge as well as a conceptual one.

Signaling molecules and metabolic enzymes can modulate each other through phosphorylation

Feedback from metabolism to other modules is not limited to allosteric interactions, as even metabolic enzymes may have functional significance in the context of signaling. Recently, phosphoenolpyruvate (PEP) was shown to not only provide its phosphate group to ADP to yield ATP in the "normal" course

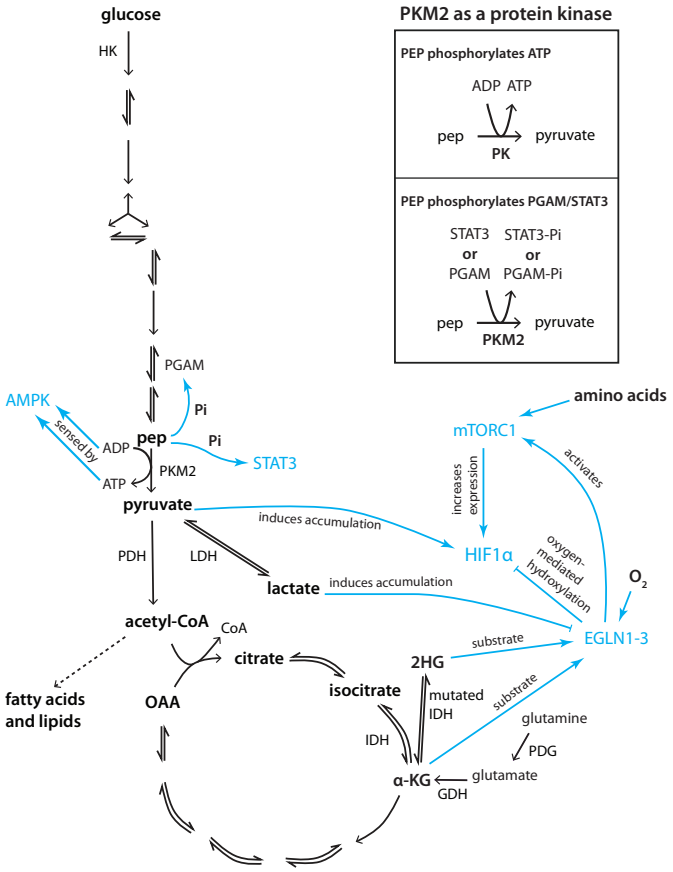


Figure 1.4: Various metabolic intermediates are known to influence the regulatory machinery. Notable among them are PEP, which can donate phosphate not only to ADP (yielding ATP) but also to STAT3 (a signaling enzyme) and PGAM (a glycolytic enzyme). These reactions that are both catalyzed by PKM2 (inset). Various metabolites influence the levels of HIF1α. Pyruvate and lactate induce its accumulation, but downstream of them, αKG and 2HG (a mimic of αKG that is produced by mutated IDH), are substrates for prolyl hydroxylases like EGLN1-3, which destabilize HIF1α. The prolyl hydroxylase family also activates mTORC1, which increases the expression of HIF1α.

of glycolysis, but to also function as a phosphate donor in enzyme-mediated reactions that phosphorylate other enzymes [2]. In this context, PKM2 has been shown to act as a (signaling) kinase that phosphorylates STAT3 [3], and PEP has also been shown to be involved in the phosphorylation of phosphoglycerate mutase 1 (PGAM1) in PKM2 expressing cells [2]. Both observations support our suspicion that the significance of PKM2 lies in the signaling rather than the metabolic realm (Figure 1.4).

PKM2 is not alone in this sense: Other metabolic enzymes have been shown to have so-called "moonlighting activity" as well. Notably, GAPDH has long been known to act as a phosphatidylserine-binding protein with a role in membrane fusion [79], and enolase has been detected on the cell surface where it may have a role in sensing plasminogen [80]. In a similar vein, various other metabolic enzymes, including PKM2 itself [81], LDHA [82], and the PDH complex [81] can be phosphorylated, but the functional significance of these modifications varies from enzyme to enzyme.

For example, hexokinase (HK) can be phosphorylated by a variety of proteins in a variety of species, inducing a wide range of effects. Phosphorylation of serine 14 in yeast affects HK2 shuttling between the nucleus and the cytoplasm [83], hexokinase phosphorylation in frog increases the enzyme's substrate affinity and activity [84], and HK2 phosphorylation by Akt protects against the consequences of ischemic injury and oxidative stress in murine cardiomyocytes [83].

Another illustration of the wide range of effects of enzyme phosphorylation is found in phosphofructokinase 2 (PFK2). Normally, this enzyme phosphorylates fructose-6-phosphate to yield fructose-2,6-bisphosphate, which in turn allosterically activates PFK1, an important main-branch glycolytic enzyme (Figure 1.1A). PFK2 can be phosphorylated by Protein Kinase A, inactivating it [85], but phosphorylation of a different amino acid residue by AMPK instead activates the enzyme [86].

If the above shows anything, it is that simply listing the knowledge that exists on various network components will not get us very far towards understanding said network. Modeling these components as an integrated system is an obvious remedy, and several methods to accomplish this are available, each with its own benefits and drawbacks. While the work described in later chapters of this thesis makes use of a broad spectrum of modeling options, genome-scale stoichiometric models constitute the bulk of the modeling effort we present. Stoichiometric modeling is generally considered to be the only feasible option for modeling metabolism on the genome scale, and we shall

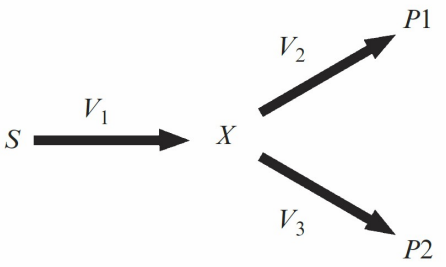


Figure 1.5: Simple metabolic network consisting of three reactions, three external metabolites (S, P1 and P2) that act as source or sink, and one internal metabolite (X).

describe it and its place in the spectrum of modeling options in detail in this next and final section.

Genome-scale metabolic models: their place in the spectrum of modeling options

Two opposing strategies have been generally employed to model biological systems. The so-called "bottom-up" approach focuses mostly on the very detailed description of the different components of the system, while the "top-down" approach looks for (uneducated) correlations between different variables of the system. Genome-scale metabolic models can be seen as a "middle-out" strategy because they first delimit the genomic pool of the system, and from there model the potential interactions of these components.

It should be clear that all modeling strategies are suitable for specific tasks, but have limited capabilities for others, and so it is important to understand the possibilities, limitations and underlying assumptions before embarking on any of them. The direct comparison of genome-scale metabolic models with other modeling approaches places them in the spectrum of modeling options, and helps understand their value.

Bottom-up (kinetic) models versus genome-scale metabolic models

The structure of a kinetic model is actually quite simple. Suppose we have a simple system with two branches (Figure 1.5).

We often assume an environment in which S is some infinite source with a constant concentration (often referred to as an external metabolite [87, 88]),

e.g. glucose in the bloodstream, or a nutrient in a chemostat. Similarly, P1 and P2, the products of this pathway, have constant concentrations and are also external metabolites. X is an internal metabolite, the concentration of which depends on the rates of production and consumption. In mathematical terms, this can be described as:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = v_1 - v_2 - v_3 \tag{1.1}$$

The rates of the enzymes, v_i, are a function (f) of the kinetic parameters $\mathbf{p}$ (e.g. Michaelis-Menten constants V_{max} and K_m values) and concentrations of the metabolites, in this case:

$$V_i = f(\mathbf{p}, X, S, P1, P2) \tag{1.2}$$

Thus, Eq 1.1 is a differential equation, the solution of which is a function that describes the time-course of the concentration of X, i.e.:

$$X(t) = f(X_0, \mathbf{p}, S, P1, P2) \tag{1.3}$$

The X_0 indicates that the actual behaviour is dependent on the initial concentration of X. Since the rates depend on $\mathbf{p}$, S, P1 and P2, so does X. If X is computed at each time point, Eq 1.2 can be solved to find the rates of all the enzymes in time, i.e. $v_i(t)$. Thus, such kinetic models can be constructed and validated through time-series of metabolites and flux measurements.

Detailed kinetic models can be used to explore the dynamics of a system, as well as the control structure of a pathway [89]. Once a model is available, parameters and conditions (such as the concentration of S in our example) can be altered to study biologically relevant properties, such as bistabilities [90], oscillations [91], and robustness or homeostasis [92, 93], just to mention a few. In a practical sense, this understanding makes it possible to predict which steps to enhance, or which feedbacks to remove, for a particular modification of the pathway's behavior (e.g. enhanced production of a valuable product, or improving a diseased state). But even though there are examples of successful applications of these types of kinetic models [94, 95], they face some serious limitations.

First, the kinetic parameters $\mathbf{p}$ are very rarely available for all enzymes in a pathway, although databases such as BRENDA [96] and Sabio-RK [97] are a great help. However, these kinetic parameters are often measured at different or non-physiological conditions, such as at the optimal pH of the enzymes, instead of the physiological pH. As a result, it remains quite controversial to

what extent the *in vitro* kinetics reflect the *in vivo* kinetics [98]. An alternative to the *in vitro* kinetic parameter determinations is to estimate "*in vivo* kinetic parameters" from time courses of metabolites after short time-scale perturbations of the metabolic network [99–101]. Parameter estimation is a field onto itself, and not an easy task. We refer to a recent review on the topic [102].

Second, kinetic models necessarily represent relatively small, isolated metabolic pathways. Since these pathways are embedded in a large metabolic network, the boundary conditions, i.e. the exchanges of information with the rest of the system, become critically important to the success of the model predictions. It was shown that including the boundaries explicitly, even with uncertain parameter values, can improve the predictive power of the model outcome [103]. Several approaches have been described that take the uncertainty in kinetic parameters explicitly into account, generating an ensemble of model outcomes that can be inspected for robust and more uncertain model predictions [104–107]. Eventually such new approaches, together with further developments in metabolomics research and computing power, may substantially increase the size of systems for which kinetic models can be constructed by reverse engineering [108, 109]. However, at the moment these approaches scale poorly, and genome-sized ones are still a long way ahead.

Genome-scale metabolic models, on the other hand, do cover the total metabolic potential that is encoded in the genome of an organism, but this comes at a cost. They share with kinetic models the structure, i.e. the stoichiometry of all reactions, but they leave out (almost) all of the kinetic details. This is possible because many pathways, such as the one depicted in Figure 1.5, when analyzed over a sufficiently long time will reach a steady state, i.e. a state in which all internal metabolites will be balanced by the producing and consuming reactions. In our example, this means that, in steady state:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = v_1 - v_2 - v_3 = 0 \tag{1.4}$$

Eq 1.4 basically states that the concentration of X is constant in time, as it is balanced by its production and consumption rate. When the kinetics of the reactions are known, we can compute the steady state concentration of X, and the steady state rates through the enzymes, which we then, in steady state, call fluxes:

$$X(t \to \infty) = X_{ss} = f(\mathbf{p}, S, P1, P2) \tag{1.5}$$

$$V_{i,ss} = f(X_{ss}, \mathbf{p}, S, P1, P2) \tag{1.6}$$

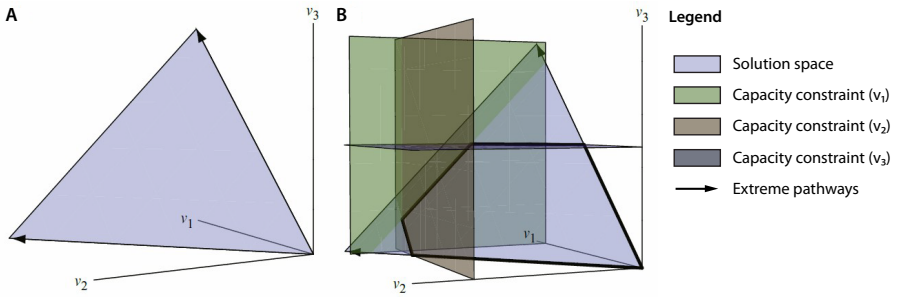


Figure 1.6: Illustration of the concept of solution space, and how mass-balance and capacity constraints result in a bounded space of possible flux states. (A) The steady state solution space is two-dimensional (a plane), because there are three unknowns (3 fluxes) and 1 equation: the mass balance of X, i.e. Eq. 1.4. This plane stretches out to all directions, but is depicted only in the positive quadrant for comparison with panel B. The arrows represent "extreme pathways", basis vectors that span the solution space, see [111] for a review on the topic. (B) Specific capacity constraints on the fluxes v_1, v_2, and v_3 turn the plane into a bounded space within which all feasible flux states should lie. For each flux we have set capacity constraints, $0 \leq v_i \leq v_{i,max}$.

Eq 1.4 leads to a set of linear equations that link the rates v_i with each other, which does not require the parameters **p**, S, P1 and P2. However, to truly *predict* the rates v_i or the steady-state metabolite concentration from the conditions defined by S, P1 and P2, one does need this kinetic information, as specified in Eqs 1.5 and 1.6. Via Eq 1.4 we only look at steady-state stoichiometric interrelationships between (steady-state) fluxes, nothing more. This difference with kinetic modelling is crucial for understanding the limitations of the analyses of genome-scale metabolic models discussed later on. Imposing steady-state relationships between rates (referred to as mass-balance constraints [110], for obvious reasons), constrains the possible states the system can be in only with respect to fluxes through the network, also referred to as the solution space (illustrated in Figure 1.6).

The other type of constraint used in genome-scale metabolic modeling, is called the capacity constraint. Such constraints also have a link with kinetics, as they represent any limitation that can be imposed on individual rates. Such limitations can be measured V_{max} values that form an upper limit through the reaction catalyzed by the enzyme. More often, it contains a constraint on the directionality of a reaction, i.e. the assessment whether a reaction can work in both directions under the physiological range of metabolite concentrations that is assumed (or sometimes measured) in the organism under study [112–114]. Mass balance and capacity constraints together limit the solution space of all

possible flux distributions (Figure 1.6).

The analysis of genome-scale metabolic models, applying mass-balance and capacity constraints, is collectively named constraint-based modeling. Note that a detailed kinetic model will predict a specific steady state with fluxes that lie inside the solution space of the corresponding stoichiometric network [115]. Leaving out the kinetic details in constraint-based modeling comes at the cost of being able to predict the space of feasible states, but not the specific state. Also, information about the sensitivity of the steady state to parameters changes, as well as the trajectory by which this state is reached, is not available through stoichiometry alone.

Top-down (biostatistical) models versus genome-scale metabolic models

Despite the -omics data explosion, it is clear that for true systems-level understanding, the abundance of relevant data, unfortunately, is matched by an abundance of less relevant data. This has to do with the inductive and open-ended nature of the typical -omics experiment. There is nothing wrong with this (well, a little perhaps [116]), and especially when done in a systematic way with good experimental design, there are good methods to turn data into knowledge [117, 118]. Statistics-based data analysis often gives qualitative interaction networks, or candidate components (genes, transcription factors, metabolites) that are calculated to be related to the phenomenon under study. For many applications this may be sufficient, because these leads can be followed up by validation experiments. Also, the -omics data can simply be used for diagnostic purposes in biotechnology, for example to indicate a particular limitation or stress during a fermentation process [119, 120], or to predict likely outcomes of cancer treatment [121], just to mention a few of the many applications of recent years.

Attempts to enhance the purely statistical analyses by integrating different data sets and a priori knowledge lie within the realm of integrative bioinformatics. We will not make an attempt to give a full account on that field, but focus on where metabolic information was, or can be, used for integration. The construction of genome-scale models starts from the genome: using bioinformatic approaches, the sequenced genome is scanned for regions that encode proteins with enzymatic activity [122–124]. Based on the presence of such genes, and supplemented with known biochemical and physiological information, putative metabolic pathways can be inferred. Thus, genome-scale metabolic models do not only contain reaction information (as was used

above in comparison with kinetic models), but also the often many-to-many relationships between genes, proteins and reactions [123]. The mapping of genes to proteins and to reactions allows for integration of these levels. Transcriptome data (gene level) can be mapped on metabolic maps in this way, providing a visual, metabolic context for the interpretation of the transcriptome data [125, 126]. By visual inspection, this may give leads to parts of the metabolic network that are regulated. For instance, the need for CO_2 in fermentations of *L. plantarum* was identified in this way [127].

A computational counterpart of visual inspection was developed as reporter metabolites for proteome or transcriptome data [128]. In these analyses, the differential expression of reactions that produce or consume a particular metabolite is scored for each metabolite, and compared to an expected score (based on chance alone). Those metabolites, whose surrounding reactions change significantly, are then reporters for metabolic effects. CO_2 would be such a reporter metabolite in the *L. plantarum* example above. Reporter reactions are computed in a similar vein from metabolome data [129].

Also, the metabolic association of genes, via their mapping to the reaction network, has been used for functional association tests. Thus, genes close to each other on the metabolic map were found to have a stronger correlation in gene expression. Using constraint-based modeling, via the so-called flux coupling analysis [130], this relationship was resolved with a much higher functional resolution [131, 132]. The key to this higher resolution is the appreciation that two reactions in series in a linear pathway are stoichiometrically fully coupled, whereas two reactions diverging from the same metabolite (such as reactions 2 and 3 in Figure 1.5) are uncoupled, i.e. they can carry flux in steady state independently from each other [132].

Thus, genome-scale metabolic models provide a "context for content" [133], and they are very useful in -omics data integration, especially when results can be visually interpreted (Chapter 5). Reconstructions of other networks, such as transcription regulation networks [134], signal transduction networks [135] or the translation machinery [136], allow for further extension of top-down analyses in which biological data is integrated in a systematic way.

Flux balance analysis: the work horse of constraint based modeling

In many of the applications of genome-scale models, flux balance analysis (FBA) is used, and this is by far the most popular constraint-based modeling technique [137] (for a more extensive primer on FBA, we refer to [138]). FBA uses optimization of a certain objective function to find a subset of optimal

states in the large solution space of possible states that is shaped by the mass balance and capacity constraints (see above, [110, 138]). We purposely state subset of optimal states, because the flux value of the objective function is unique, but the pathway distributions that give this optimal flux value may not be. One who is interested in the set of flux distributions that give rise to an optimal value of the objective function can perform a Flux Variability Analysis (FVA) to check the uniqueness of the flux distribution [139]. FVA maximizes and minimizes each reaction rate in the network at the optimal objective flux value: The resultant range in fluxes gives an indication of how flexible the network is in reaching the optimum.

Flux balance analysis basically tries to find a state (or sets of states) that maximizes (or minimizes) some desired flux or linear combination of fluxes. The approach is an optimization technique, and hence it implies that the modeled system (e.g. organism) behaves optimally (by evolution). This may limit the understanding of how microorganisms behave in experiments, as they may not have had time to adapt to the environment. There are three ways to deal with the optimality issue.

First, the optimality issue can just be ignored, simply because one does not care about a specific, quantitative prediction. For example, one may want to know whether the model can produce biomass, or some compound, not necessarily how much. This goes for (synthetic) lethality predictions, and predictions of evolution of metabolic networks.

Second, we can allow cells to adapt to their (constant) environment by laboratory evolution experiments. Aiding this approach, FBA can be used to predict evolutionary engineering outcomes, which may be guided by *in silico* analysis of specific deletions to enhance productivity [140, 141]. It also provides optimal yields with which experimentally obtained yields can be benchmarked.

Third, extensions to FBA have been proposed to circumvent the need for evolution, notably MOMA (minimization of metabolic adjustments [142]). MOMA requires a well-defined reference state, and then tries to predict the response of the network to a perturbation (change in external conditions, deletion of a reaction) by assuming that the network is robust. Technically, it tries to minimize the (Euclidian) distance between the reference flux distribution and the new solution space (which has changed by the perturbation). In our experience, the major problem with MOMA is that it weighs all fluxes equally with respect to biological relevance, which is not very likely to be correct. Moreover, the response is dominated by high fluxes. Thus, MOMA predic-

tions in our hands do not produce proper specific predictions of fluxes, but it has been useful to see if a certain perturbation is likely to cause severe growth problems even though the optimal FBA solution seems fine (for an example, see [143]).

ROOM (Regulatory On/Off Minimization) has been proposed as an alternative that does not suffer from these issues, as it is limited (by design) to turning fluxes "on" or "off" in an attempt to mimick regulatory activity in the cell [144]. However, it too does not see extensive use (its ample citations are largely from technology-oriented papers, a quirk shared by many other systems biology papers), perhaps because an attempt to study a mathematical representation (FBA model) of a biological system using another biological model (the ROOM extension, or MOMA for that matter) seems too far-fetched compared to solving a mathematical issue using mathematics.

In this light, it is also interesting to mention CoPE-FBA (Comprehensive Polyhedra Enumeration FBA), which is an example of the latter approach, and enables more general statements about the optimal flux space [145]. Specifically, CoPE-FBA has shown that in many biological systems that seem to have highly variable solution spaces, this variability is due to a rather small (indeed: manageable!) number of components. Components that might otherwise have been lost in a forest of flux values can then be followed up on in an attempt to understand their biological relevance to the system.

Clearly, these computational analyses of metabolic networks cannot be performed by hand, and Microsoft Excel — the biologist's staple — does not scale to the genome level. This need for community-wide standards and specialist software has resulted in the development of a standard language for describing metabolic models (SBML [146]), as well as in software facilitating their analysis (e.g. PySCeS [147] and the COBRA Toolbox [148]). We have expanded on this work by producing several tools that (further) facilitate the manipulation and visualization of genome-scale models. As this is described in detail in Chapters 4–5 of this thesis, we will not go into the topic of software here.

FBA predicts rates only through yield maximization

If one does want to optimize a model to obtain specific flux distributions, however, one requires a sensible objective function, and here some confusion and controversy has arisen in the literature. In a recent study, different objective functions were tested for the extent to which they could predict actual flux states under different conditions [149]. The study demonstrated that different

objective functions were needed to describe the flux states under different conditions. Notably, under energy (or carbon) limitation, optimization of biomass yield appeared to be the best objective function. This is in line with earlier studies in which the biomass formation function was taken as objective to predict functional states [150].

However, many microorganisms and cancer types display overflow metabolism, which is a wasteful lifestyle in terms of ATP generation and consequently, biomass yield. This behavior cannot be predicted by FBA, because it predicts rates through optimal yields. It could only be described by including additional capacity constraints on the oxidative phosphorylation pathways in the corresponding metabolic networks [151, 152].

It is important to fully appreciate the point that FBA predicts rates through optimal yields, and not rates directly. If we take our example from Figure 1.5, and suppose we want to maximize the production rate of P1 from S, we may formulate the FBA problem as:

$$\begin{aligned} &\max v_2 \\ &\textit{given}: \\ &\begin{cases} \text{mass balance constraint:} \\ v_1 - v_2 - v_3 = 0 \\ \text{capacity constraints:} \\ 0 \leq v_1 \leq 10 \\ -\infty \leq v_2 \leq \infty \\ 0 \leq v_3 \leq \infty \end{cases} \end{aligned} \tag{1.7}$$

The solution is easy in this case: v_1 should be 10, v_2 should be 10, and v_3 should be 0. But note that the rate of v_2 is fully dictated by the constraint on v1. We could write the rate of v_2 as:

$$v_2 = v_1 \cdot Y_{P1,S} \tag{1.8}$$

where $Y_{P1,S}$ is the yield of P1 on S (1 in this case). Since v_1 is fixed by the capacity constraint, the only way to maximize v2 is to maximize the yield; this is also fixed in this simple case, but this is not so in larger systems where the capacity constraint is always required to bound the problem. For larger, real genome-scale models, this argument therefore remains [88]: Since the solution space is bounded by an input flux, FBA finds an optimal objective *flux* by optimizing the *yield* on the incoming substrate. Hence, if in a model there

are two options to make ATP from glucose, fermentation (low ATP yield) and respiration (high ATP yield), optimization of ATP production rate will necessarily be achieved by respiration. Therefore, when applying biomass optimization in FBA, the underlying biological assumption is that *biomass yield maximization* was the strategy through which the organism has reached its fitness. It is clear that there are also other strategies that lead to fitness, and hence, the validity of FBA, using biomass yield as objective function, is organism and condition specific [149, 153, 154].

Applying genome-scale models to the study of cancer

Before we can fully appreciate the value and validity of biomass equations in genome-scale applications, we must first assess the value and validity of these applications in general. The use of GSMMs in the study of cancer got underway only fairly recently (in 2007) with the publication of a first reconstruction ("Recon 1") of human metabolism [155], although the state of technology had been ready for it for a while at the time. Other genome-scale models of human metabolism also appeared around this time, e.g. the Edinburg Human Metabolic Network (EHMN) reconstruction [156], and a bit later the liver-specific HepatoNet1 reconstruction [157]. While their usefulness goes undisputed, they never attained the popularity of Recon 1.

HepatoNet1 did mark a trend towards the generation of tissue-specific metabolic models. Ironically, while HepatoNet was extensively manually curated, subsequent studies of tissue-specific models employed automated approaches to create models and make statements about the function of specific cell types (e.g. INIT [158], mCADRE [159], and Recon 2 [50], the successor to Recon 1). The value of these automatically generated models turned out limited outside the scope within which they were created, and they saw limited to no reuse after publication. Nevertheless, the availability of "complete" models of human metabolism enabled a multitude of studies of human cellular physiology health and disease, and cancer was an obvious research topic this new generation of models could be used for.

For instance, Recon 1 was combined with information on targets of existing drugs, to investigate combinations of drugs that could specifically target cancer cells based on the disease's metabolic properties [160]. Other studies delved into the properties of the Warburg effect at a scale that used to be infeasible, by looking at the concentrations of external metabolites and the implications they could have on the internal flux distribution [45]. With some extensions to the FBA framework, even strictly metabolic explanations of the Warburg effect

were proposed [161, 162]. Though interesting in their own right as additions to the repertoire of FBA-related techniques, results of this approach have not been universally accepted as conclusive answers to the Warburg conundrum. One could question whether their authors even intended them to be: As we discussed above, it is likely that the workings of the Warburg phenotype include both metabolic as well as signaling components.

In any case, the approaches outlined above are illustrative of the broad array of research questions that is included in the scope of genome-scale stoichiometric modeling in the context of cancer. But tacitly, an unasked question underlies the current GSMM cancer modeling effort. The choice of biomass production as an objective function largely goes uncontested [163], but for some reason, so do the contents of the biomass equation. While there are often no suitable alternatives to the former [163], the latter is subject to a peculiar custom where, when an experimentally determined biomass equation is lacking, an amalgam of partial biomass compositions from various (related) species is employed. Examples of this can be found in the biomass equations that were included in various (and otherwise very helpful) reconstructions of human metabolism [50, 164]. In these models, data from human, mouse, rat, and dog were used in various combinations. Other models of human metabolism omit the biomass equation altogether, citing a lack of data and/or a lack of necessity [155,158]. To start ending this impasse, we set out experimentally determine of the biomass composition of human leukocytes and leukemia cells. Armed with this first "fully human" biomass equation, we can then assess the quality of earlier FBA results, and determine the relative importance of the accuracy of the biomass objective function. The results of this endeavor are described in Chapter 6 of this thesis.

Aim and outline of this thesis

It should be clear by now that the biology of cancer has many interconnected facets. In systems biology, we aim to take a step back and study the bigger picture, to see how it all fits together, but typically end up studying only a single subpart in detail.

In this thesis, I instead describe the study of several (sub)systems that contribute to cancer. By starting at microRNA-mediated regulation, a high level of abstraction, and moving via signaling to metabolism and growth, the eventual goal of many signaling processes in cancer, we encounter examples of most of the systems that make the difference between health and disease.

Specifically, in the case of microRNAs, we looked at the properties and implications of the miR-21 isoform landscape, which we characterized using next-generation sequencing (**Chapter 2**). In **Chapter 3**, we then apply a systems biology approach to a more traditional signaling challenge: elucidating chemokine signaling pathways that are thought to govern cancer progression on the regulatory level.

The step between these systems and growth is metabolism, but to adequately study metabolism and growth from a genome-wide perspective, adequate tools for dealing with genome-scale models must first be available. In **Chapter 4**, we describe the development of FAME, a web-based tool that allows the manipulation, simulation, and visualization of these large models. In **Chapter 5**, we present a comprehensive, interactive, and extensible map of metabolism, the availability of which is instrumental in the interpretation of metabolism-related data – in our case, modeling results.

Then, in **Chapter 6**, we can apply these tools to study the impact of the composition of human biomass to the modeling of human (cancer) cells. Our experimental determination of the biomass composition of human leukemia cells allows an unprecedented insight into the consequences of this systems biological staple on modeling results, and on their relationship to the system under study. The thesis is concluded by a General Discussion, in which these chapters are viewed in the context of the current state of the art in systems biology, and in the context of the path towards a better understanding of cellular function in health and disease.

Chapter 2

PAPD5-mediated 3′ adenylation and subsequent degradation of miR-21 is disrupted in proliferative disease

In collaboration with Helena Persson, Jay W. Shin, Yuri Ishizu, Inga Newie, Rolf Søkilde, Shannon M. Hawkins, Cristian Coarfa, Kazuhiro Ikeda, Ken-Ichi Takayama, Kuniko Horie-Inoue, Yoshinari Ando, Alexander M. Burroughs, Chihiro Sasaki, Chizuru Suzuki, Mizuho Sakai, Shintaro Aoki, Ayumi Ogawa, Akira Hasegawa, Marina Lizio, Kaoru Kaida, Bas Teusink, Piero Carninci, Harukazu Suzuki, Satoshi Inoue, Preethi Gunaratne, Carlos Rovira, Yoshihide Hayashizaki, and Michiel J.L. de Hoon. Adapted from: *Proc Natl Acad Sci U S A* (2014), **111**(31):11467–11472.

Background

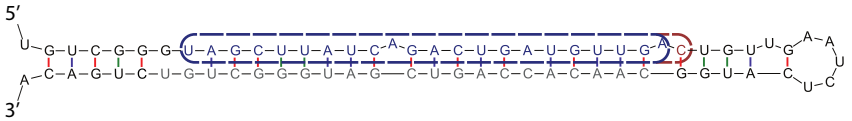
The advent and subsequently widespread use of next-generation sequencing technology has enabled the in-depth examination of the small RNA complement of cells. Among these small RNAs are microRNAs (miRNAs), a class of ~22-nt RNAs that was discovered about twenty years ago [165]. An impressive amount of miRNA sequencing data is now publicly available, which has aided the discovery of novel miRNAs, miRNA function, and their biogenesis [166].

In the canonical miRNA generation pathway, RNA polymerase II or III produces the primary miRNA transcript from which the pre-miRNA hairpin is excised by DROSHA. The pre-miRNA is exported to the cytoplasm where it is cut by DICER1 to release a double-stranded RNA duplex consisting of the mature miRNA and the passenger strand. Whereas the passenger strand is usually rapidly degraded, the mature miRNA is loaded as a single-stranded RNA into an Argonaute protein to form the RNA-induced silencing complex (RISC) [167, 168]. Within RISC the miRNA binds to partially complementary target sites in messenger RNAs, located preferentially in the 3′ UTR, to inhibit translation and induce degradation of the target mRNA. Base pairing to nucleotides two to eight of the miRNA, the so-called seed sequence, is particularly important for target gene recognition.

In recent years, deep-sequencing libraries of small RNAs have revealed the existence of a large number of sequence variants both in pre-miRNAs and in mature miRNAs. In pre-miRNAs, sequence variations associated with adenosine-to-inosine (A-to-I) editing have been shown to affect their processing [169, 170]. Similarly, polyuridylation mediated by the uridylyl transferase ZCCHC11 of the 3′ ends of pre-miRNAs of the let-7 miRNA family can suppress their maturation [171–174].

Sequence variations in mature miRNAs are referred to as isomiRs [175–177]. IsomiRs may be produced by differential excision by DICER1, which can lead to variations in the 5′ and 3′ ends of the mature miRNA [178]. This can alter their affinity for Argonaute proteins [179] and also, in case of 5′ end variations, their target specificity [180].

IsomiRs can also be generated by non-templated addition of A and U nucleotides to their 3′ end [175, 181]. Although several enzymes have been found to govern 3′ nucleotide addition to mature miRNAs [182], for most mature miRNAs 3′ adenylation is mediated by PAP associated domain containing 4 (PAPD4, also known as GLD2) [181]. Whereas mature miR-122 is selectively stabilized by PAPD4-mediated adenylation [183], it remains unclear whether



miR-21: UAGCUUAUCAGACUGAUGUUGA

miR-21+C: UAGCUUAUCAGACUGAUGUUGAC

miR-21+CA: UAGCUUAUCAGACUGAUGUUGACA

Passenger strand: CAACACCAGUCGAUGGGCUGU

$$\text{adenylation ratio} = \frac{\text{count of miR-21+CA}}{\text{count of miR-21+C}}$$

$$\text{degradation ratio} = \frac{\text{count of miR-21}}{\text{count of miR-21+C}}$$

Figure 2.1: The upper image shows the structure of pre-miR-21 as predicted by mfold [192]. The canonical miR-21 and miR-21+C isomiRs are highlighted in blue and burgundy, respectively. The definitions used in this chapter appear below the structure.

in general adenylation of the 3′ end of mature miRNAs increases their stability [181].

MicroRNAs are involved in the regulation of many cellular processes including developmental and oncogenic pathways [167, 184]. The oncogenic microRNA (oncomiR) miR-21 (Figure 2.1) has been of particular interest because of its importance in human diseases such as cancer [185–187], cardiovascular disease [186, 188], and inflammatory skin disease [189], with a role in cell cycle control [190], apoptosis [187] and metastasis [185, 191]. Small RNA sequencing data from The Cancer Genome Atlas (TCGA) (http://cancergenome.nih.gov/) across a panel of 10 cancer types showed that miR-21 is the most highly expressed miRNA in cancer (Figure S1A), with an expression level significantly elevated in tumor samples compared with the corresponding normal samples (Figure S1B). Due to its biological and clinical importance, miR-21 is by far the most widely studied miRNA (Figure S1C). Despite this, its isomiR composition and the functional significance of specific miR-21 isomiRs remain poorly understood.

Across a broad collection of small RNA deep-sequencing libraries, we found that two mature isomiRs of miR-21 are highly expressed, with their relative expression ratio strongly cell-type dependent. We show that the longer of the two isomiRs is the primary product of DICER1 cleavage, and that the shorter isomiR is produced by 3′-to-5′ degradation of miR-21 by a tailing-and-trimming process involving the nucleotidyl transferase PAP associated domain containing 5 (PAPD5) and the poly(A)-specific ribonuclease PARN. We then show that this degradation pathway is disrupted in most cancers, as well as in the non-cancerous proliferative skin disease psoriasis. This pro-

vides evidence for a pathway in which PAPD5, a tumor suppressor previously shown to stabilize tumor protein 53 (TP53) mRNA [193], also regulates oncomiR degradation.

Results

MicroRNA-21 has two prominent isoforms

We produced a set of small RNA libraries from MCF7 breast cancer cells (SI Materials and Methods and Figure S2) and found that miR-21 accounted for more than 39% of total miRNA expression in this cell line (Figure S3). Almost half of the overall miR-21 expression was due to a 23-nucleotide isomiR, which we refer to as miR-21+C, that has an additional cytosine at the 3′ end compared with the 22-nucleotide canonical miR-21 isomiR registered in miRBase [194] (Figures 2.1, S3). As this cytosine is encoded in the genome, miR-21+C may be produced directly by DICER1 processing of the pre-miR-21 hairpin structure. Reanalyzing small RNA data previously produced at our laboratory from THP1 monocytic leukemia cells [195] revealed abundant expression of both isomiRs also in this cell line (Figure S3). Previous Northern blotting experiments for miR-21 in RNA obtained from MCF7, HeLa, and HT29 cells [178] confirm the existence of both a 22 nucleotide and a 23 nucleotide mature miRNA in these cells.

MicroRNA-21+C is the primary product of pre-miR-21 cleavage by DICER1

To understand the mechanism by which these two isomiRs are produced, we first considered the possibility that DICER1 itself is able to generate both isomiRs by alternative excision from the pre-miRNA hairpin. However, previously only a single 23 nucleotide mature miRNA was observed after in vitro cleavage by DICER1 [196, 197], suggesting that at least in vitro DICER1 does not produce the 22 nucleotide canonical miR-21 isomiR. Furthermore, although we would expect such alternative excision to be accompanied by a variation in the 5′ starting nucleotide of the miR-21 passenger strand sequence, we detected only a single dominant miR-21 passenger strand sequence both in our MCF7 as well as in our THP1 sequencing data (Figures 2.1, S4). Sequencing reads of the loop section of the pre-miR-21 hairpin structure are also consistent with excision by DICER1 of miR-21+C and do not show evidence of direct excision of the 22-nucleotide canonical miR-21 isomiR (Figure S4).

Association of human DICER1 with its partner protein TAR (HIV-1) RNA binding protein 2 (TARBP2, also known as TRBP) has previously been shown to alter its pre-miRNA cleavage pattern in vitro [196]. However, Northern blots revealed only a single mature miRNA band of the same size after in vitro cleavage of pre-miR-21 by either DICER1 or the TARBP2-DICER1 complex [196], indicating that the existence of two distinct miR-21 isomiRs is not due to association of DICER1 with TARBP2.

Finally, we note that the 23-nucleotide isomiR miR-21+C has a 2-nt overhang at its 3′ side with respect to the 5′ end of the passenger strand (Figures 2.1, S4), which is biochemically favorable and is typically seen in mature miRNA excision by DICER1 [198]; in contrast, excision of the 22 nucleotide canonical miR-21 isomiR would yield a single nucleotide overhang. We conclude that the miR-21+C isomiR is the primary product of DICER1 cleavage of pre-miR-21 both in vitro and in cell lines.

MicroRNA-21+C is subject to cell-type specific 3′ adenylation

As DICER1 processing of pre-miR-21 only produces the 23 nt miR-21+C isomiR, we hypothesized that the canonical 22 nt mature miR-21 isomiR may instead be produced by 3′-to-5′ trimming of the 23 nt miR-21+C isomiR. Previously, stimulation of miR-21 degradation in HeLa cells by transfection of miR-21-specific antagomiRs was shown to not only result in 3′-to-5′ trimming of this miRNA, but also in its 3′ tailing by one or more nucleotides [199]. In most cases, a single adenosine was added to the 3′ end of miR-21+C, giving rise to a 24 nt miR-21+CA isomiR. This adenosine is not encoded in the genome and is therefore likely to have been added posttranscriptionally to miR-21+C. In agreement with these previous results, we found that the miR-21+CA isomiR comprised between 1% and 10% of the miR-21 sequences in our MCF7 and THP1 sequencing data (Figure S3).

To gauge whether these three isomiR forms are a general property of miR-21 expression in different cells and organisms, we examined publicly available sequencing data sets from human, rhesus macaque, mouse, rat, cow, platypus and Japanese flounder (Table S1). The examined data included both healthy and diseased samples, as well as primary tissue and cell lines, and between them cover approximately 450 million years of evolution. Remarkably, miR-21, miR-21+C, and miR-21+CA were found in each of these data sets, with their abundances and relative counts varying between cell types (Figure S3).

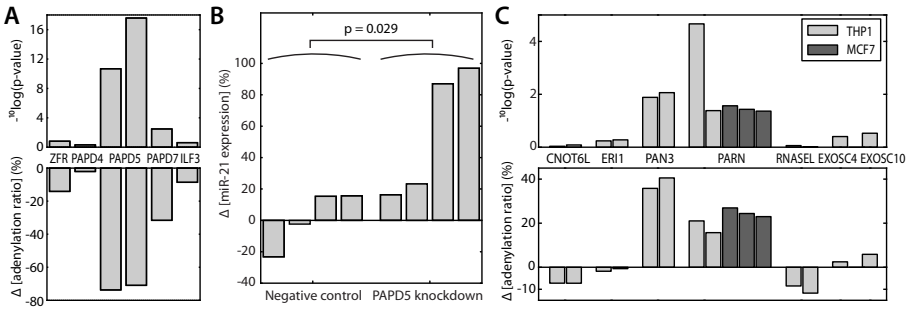


Figure 2.2: (A) Deep-sequencing data of small RNAs from THP1 cells upon knockdown of candidate adenylating enzymes showed a large and highly significant decrease in the adenylation ratio of miR-21+C after knockdown of PAPD5, which was confirmed in a replicate PAPD5 knockdown experiment. (B) The expression level of miR-21, as measured by qPCR and normalized against the expression of the spliceosomal RNA U6, increased significantly upon knockdown of PAPD5 in THP1 cells. (C) Upon knockdown of exoribonuclease PARN, the miR-21+C adenylation ratio increased significantly compared with NC both in THP1 (two replicates) and in MCF7 (three replicates). In THP1, the miR-21+C adenylation ratio also increased significantly upon knockdown of PAN3.

PAPD5 adenylates the 3′ end of miR-21

Previously, we investigated nontemplated 3′ nucleotide additions to miRNAs by analyzing sequencing data from THP1 cultures in which the nucleotidyl transferases ZFR, PAPD4, PAPD5, PAPD7, and ILF3 were knocked down by RNA interference [181]. We re-analyzed these data and found that knocking down PAPD5 dramatically decreases the adenylation ratio of miR-21+C (by 74%, p = 2.1e-11; Figure 2.2A). This finding surprised us, as previously most miRNA 3′ adenylation events appeared to be mediated by PAPD4 and few, if any, by PAPD5 [181]. We therefore repeated the PAPD5 knockdown experiment, which again showed a significant decrease in the miR-21+C adenylation ratio (by 71%; p = 2.6e-18; Figure 2.2A). MiR-142-3p was the only other miRNA that consistently showed a significant decrease in adenylation ratio in both replicates of the PAPD5 knockdown experiment (Figure S5).

In vitro, PAPD5 has previously been shown to be an RNA-specific nucleotidyl transferase that can add adenosine to the 3′ end of a variety of oligoribonucleotides [200]. To evaluate if also in cells PAPD5 acts directly on miR-21, we analyzed previously published photoactivatable-ribonucleoside-enhanced cross-linking and immunoprecipitation (PAR-CLIP) sequencing data obtained from HEK293 human embryonic kidney cells [200, 201]. These data revealed a statistically significant fraction of miR-21 cross-linking to PAPD5 in two in-

dependent replicates (p = 3.7e-4 and p = 0.0072, respectively; SI Materials and Methods and Table S2), suggesting that the decrease in miR-21 adenylation observed in the PAPD5 knockdown libraries is a direct effect.

MicroRNA-21+C is trimmed at the 3′ end following its adenylation

In human and murine liver cells, 3′ adenylation mediated by the nucleotidyl transferase PAPD4 has been shown to stabilize miR-122 [183]. In contrast, PAPD5 was recently found to aid in the maturation of SNORA63 and other small nucleolar RNAs in a pathway where oligoadenylation by PAPD5 is followed by 3′-to-5′ trimming by the exonuclease PARN [202,203] (Figure S6). In all sequencing data sets shown in Figure S3, we found miR-21+C transcripts with multiple adenosines added to their 3′ end. We hypothesized that miR-21+CA is similarly subject to 3′-to-5′ trimming, producing the 22-nt canonical miR-21 isomiR, which may subsequently be trimmed further in the 3′-to-5′ direction as part of a miR-21 degradation pathway. In all libraries, the vast majority of sequenced miR-21 reads shorter than the mature miRNA aligned to the 5′ end of miR-21, indicating that degradation of miR-21 indeed predominantly proceeds in the 3′-to-5′ direction. We regarded such 5′-matching reads shorter than 23 nucleotides as putative miR-21 degradation products. In both PAPD5 knockdown libraries, we found a significant decrease in the counts of these degradation products compared with negative control (p = 1.2e-4 and p = 1.0e-5), suggesting that PAPD5-mediated adenylation of miR-21+C indeed promotes its 3′-to-5′ degradation (Figure S7A). In agreement with such a degradation pathway, quantitative RT-PCR (qRT-PCR) showed a significantly higher expression level of miR-21 in THP1 cells upon PAPD5 knockdown compared with NC (p = 0.029; Figure 2.2B).

To assess whether expression of PAPD5 is sufficient to stimulate adenylation and trimming of miR-21, we overexpressed the protein in THP1 and MCF7 cells using a lentiviral vector (SI Materials and Methods; Figure S8A) and profiled their small RNAs by deep-sequencing. Both in THP1 and in MCF7 cells, we found an increase in the abundance of miR-21 degradation products two weeks post PAPD5 transduction compared with NC (Figure S7B, S7C). Confirming this observation we detected higher levels of miR-21 relative to miR-21+C (Figure S7D). However, we failed to confirm the accumulation of degradation products using RNA samples extracted at an earlier time point posttransduction (Figure S7E).

The exoribonuclease PARN mediates miR-21 degradation

To identify the enzyme responsible for 3′-to-5′ trimming of miR-21, we selected seven known exoribonucleases (CNOT6L, ERI1, PAN3 [poly(A) specific ribonuclease subunit], PARN, RNASEL, EXOSC4, and EXOSC10), knocked down their expression in THP1 cells using RNA interference (Figure S8B and Table S3), and sequenced their small RNA fraction. The sequencing data indeed revealed a significant increase in the adenylation ratio of miR-21+C upon PARN knockdown ($p = 2.2e\text{-}5$ and $p = 0.041$; Figure 2.2C) as well as upon PAN3 knockdown ($p = 0.013$ and $p = 0.0086$; Figure 2.2C).

Regarding 3′-to-5′ trimming of miR-21 as a degradation process, we define the ratio of the miR-21 count to the miR-21+C count as the degradation ratio (Figure 2.1). Although both replicates of the PARN knockdown showed a decrease in the degradation ratio (Figure S7F), we did not observe a decrease of this ratio upon PAN3 knockdown in either replicate (Figure S7G). This suggests that PARN, rather than PAN3, is the main exoribonuclease mediating 3′-to-5′ trimming and further degradation of miR-21. We further note that in both replicates of the PAN3 knockdown, on average, the adenylation ratio increased across all miRNAs ($p = 6.2e\text{-}4$ and $p = 1.6e\text{-}5$), suggesting that PAN3 acts as a deadenylating enzyme on a wide variety of miRNAs. No such effect was found for PARN, indicating that it acts specifically on miR-21.

To verify this role of PARN in a different cell type, we repeated its knockdown in MCF7 cells. Each of the three replicates again showed a significant increase in the miR-21 adenylation ratio ($p = 0.027$, $p = 0.037$, and $p = 0.043$; Figure 2.2C), as well as a decrease in the degradation ratio (Figure S7H).

PAPD5-mediated adenylation affects regulation by miR-21

To determine the downstream regulatory consequences of PAPD5-mediated adenylation and PARN-mediated degradation of miR-21+C, we analyzed gene expression in the PAPD4 and PAPD5 knockdown samples in THP1 cells using mRNA microarrays. Knockdown of PAPD4 and PAPD5 was confirmed both by qPCR and by the microarray expression profiling itself (Figure S8C, S8D). Target genes for miR-21 predicted by TargetScan [204, 205] showed significant down-regulation in the PAPD5 knockdown samples compared with the NC (Figure 2.3A; $p = 0.002$). We also find a down-regulation of target genes upon PAPD5 knockdown if we restrict ourselves to confident target predictions by requiring a TargetScan context score lower than -0.2 (Figure 2.3A; $p = 0.014$). In addition, the reduction in the expression of genes with at least two predicted

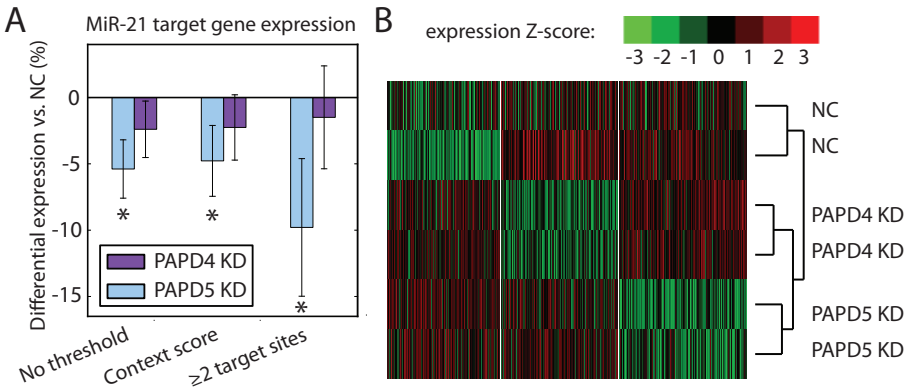


Figure 2.3: Gene expression profiling in THP1 cells was performed using microarrays for two biological replicates each for the NC, the PAPD4 knockdown condition, and the PAPD5 knockdown condition. (A) Compared with NC, predicted target genes of miR-21 were significantly downregulated upon knockdown of PAPD5, but not upon knockdown of PAPD4. This observation holds if more stringent selection criteria for miR-21 target genes are applied, such as requiring a context score lower than -0.2 or requiring multiple target sites for the same. Asterisks indicate statistically significant differential expression. Error bars represent the SDs of the estimated mean percentage change between the knockdown condition and the NC. (B) A heat map view of the expression data confirms that the profiled genes cluster by experimental condition.

miR-21 target sites was about twice as large compared with that of all predicted miR-21 target genes (Figure 2.3A; $p = 0.029$). None of these sets of genes showed a statistically significant down-regulation upon PAPD4 knockdown, demonstrating that the effect on miR-21 target genes is specific for PAPD5. Hierarchical clustering of the microarray expression data with respect to the knockdown conditions reveals distinct clustering into an NC cluster, a PAPD4 knockdown cluster, and a PAPD5 knockdown cluster (Figure 2.3B).

Dysregulation of the miR-21 degradation pathway in proliferative disease

To evaluate the relevance of the miR-21 degradation pathway in disease, we analyzed RNA sequencing and small RNA sequencing data from clinical samples from TCGA (http://cancergenome.nih.gov/). We applied our analysis to the 10 cancer types for which the TCGA data are currently publicly available, and for which both tumor and normal samples have been profiled (Figure 2.4). Across most cancer types, both the miR-21 degradation ratio (Figure 2.4A) and the miR-21 adenylation ratio (Figure 2.4B) were significantly lower in the tu-

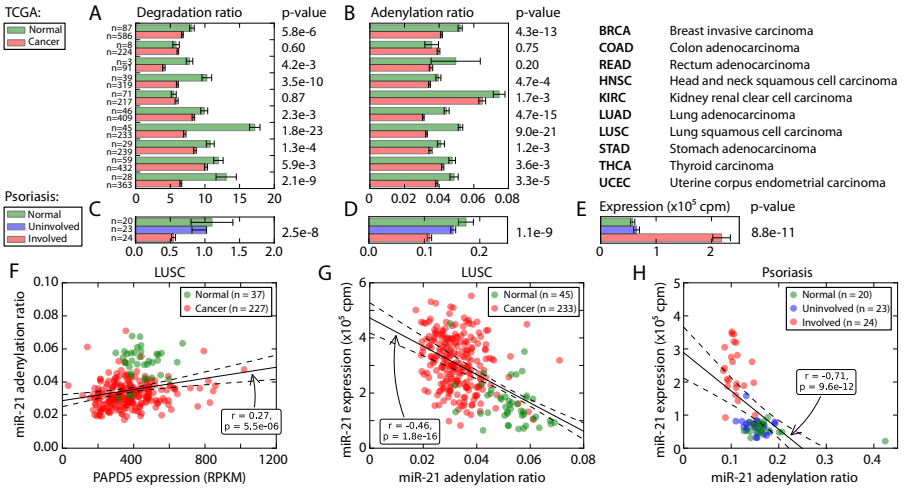


Figure 2.4: Across a wide variety of cancers, the degradation ratio (A) and the adenylation ratio (B) of oncomir miR-21+C is significantly lower in tumor samples compared with normal samples. A reduced degradation ratio (C) and adenylation ratio (D) is also observed in psoriasis, which is similarly characterized by elevated expression levels of miR-21 (E). Error bars indicate the SD of the estimated mean; Kruskal-Wallis p-values of the difference between healthy and disease samples are shown for each disease type. (F) As shown here for lung squamous cell carcinoma (LUSC), in most cancers the PAPD5 expression level is positively correlated with the miR-21 adenylation ratio, which in turn is negatively correlated with the miR-21 expression level (G). Similarly, in psoriasis the miR-21 adenylation ratio is negatively correlated with the miR-21 expression level (H). "Involved" refers to psoriatic skin samples, "uninvolved" to samples of skin from psoriatic patients that are not affected by psoriasis, and "normal" to skin from healthy donors. Dashed lines indicate the 95% confidence interval of the regression line. The Spearman correlation and its corresponding p-value are shown. MicroRNA-21 expression levels are normalized to the overall miRNA expression level in each sample; PAPD5 expression levels are evaluated as reads per kilobase of transcript per million mapped reads (RPKMs). cpm, counts per million miRNAs.

mor samples compared with the corresponding normal samples; there were no significantly elevated degradation ratios or adenylation ratios in any of the cancer types. Consistent with PAPD5-mediated adenylation of miR-21 leading to its degradation, across most cancer types we found a statistically significant positive correlation between PAPD5 expression level and adenylation ratio (Figures 2.4F, S9A), and a significant negative correlation between the adenylation ratio of miR-21 and its expression level (Figures 2.4G, S9B). The remaining correlation values were not significantly different from zero. Combined, the above indicates that disruptions in the degradation pathway of miR-21 are a hallmark of cancer across a wide range of tissues.

We then asked whether disruption of the miR-21 degradation pathway

is relevant in non-cancerous proliferative diseases such as psoriasis, a skin disease characterized by increased proliferation of basal cells in the epidermis [206]. Small RNA-sequencing data consisting of 67 patient samples of psoriatic and normal skin previously revealed [189] that at the molecular level psoriasis, like cancer, is characterized by elevated expression levels of miR-21 (Kruskal-Wallis p = 8.8e-11; Figure 2.4E). Analyzing the miR-21 isomiR counts across these samples showed that the adenylation ratio is negatively correlated with the miR-21 expression level (r = -0.71; p = 9.6e-12; Figure 2.4H), with the miR-21+C degradation ratio and adenylation ratio significantly lower in psoriasis samples compared with unaffected skin (Kruskal-Wallis p = 2.5e-8 and p = 1.1e-9, respectively; Figure 2.4C, 2.4D). Hence, we conclude that dysregulation of the miR-21 degradation pathway is not limited to cancer, but can also occur in other types of proliferative disease.

Discussion

In recent years, several posttranscriptional mechanisms have been uncovered that enable the fine-grained regulation of individual miRNAs beyond the canonical maturation pathway [167, 207, 208]. Next-generation sequencing of miRNAs has revealed a wide variety of miRNA variants that may be produced by differential excision by DICER1 or nucleotide additions to the 3′ ends of miRNAs or their precursors [175, 209]. Multiple nucleotidyl transferases, including PAPD4 and PAPD5, can mediate 3′ adenylation of mature miRNAs in human [181, 182]; these 3′ additions have been shown to be a physiological phenomenon rather than an artifact of library preparation and are specific for different miRNAs [182]. However, the biogenesis and biological significance of miRNA 3′ modifications as well as isomiRs with distinct 3′ ends are generally poorly understood.

Whereas mature miRNAs generally are stable and tend to persist even if the processing enzymes are depleted by RNAi [210, 211], examples are known in which specific miRNAs are degraded rapidly [212], and sequence-specific tailing and trimming was recently reported as a mechanism of miR-27a/b depletion in mouse [213]. Our results show that 3′ adenylation of oncomiR miR-21 by the nucleotidyl transferase PAPD5 may similarly lead to its degradation by targeting it for 3′ end trimming by exoribonuclease PARN. This enzyme thus is a human ortholog of the exonuclease Nibbler, which was recently discovered to trim the 3′ ends of miRNAs in *Drosophila* [207, 208]. PARN was recently also shown to trim the 3′ arm of human pre-miR-451 in the 3′-to-5′

direction as the final step in the maturation pathway of this miRNA [214].

In liver cells, an additional biological function of PAPD5 was recently discovered in a regulatory pathway affecting the tumor suppressor TP53 (also known as p53) [193]. In this pathway, PAPD4 stabilizes the microRNA miR-122 by 3′ adenylation [183], thereby stimulating translational repression by miR-122 of the cytoplasmic polyadenylation element-binding protein CPEB. This protein binds to the 3′ UTR of TP53 mRNA and recruits PAPD5, which modulates the stability and translation of the TP53 mRNA by polyadenylating its 3′ end. Hence, PAPD5 acts as a tumor suppressor by regulating translation of TP53 and thus promoting cellular senescence (Figure 2.5; [193]). Interestingly, PARN was recently shown to trim polyadenylated TP53 mRNA [215], which suggests a combined role for PAPD5 and PARN both in the regulation of TP53 abundance and in regulation of miR-21 after DICER1 cleavage.

Enzymatic addition of adenosine or uracil nucleotides to the 3′ end of miRNAs may in general modulate their stability and thereby play a role in regulating their abundance [184]. Whereas miR-122 was previously shown to be stabilized by PAPD4-mediated adenylation of its 3′ end [183, 184], we find instead that 3′ terminal addition by PAPD5 to miR-21 stimulates its degradation. This suggests the existence of a regulatory network of 3′ nucleotidyl transferases and exoribonucleases targeting specific mature miRNAs to enhance or prevent their degradation. As evidenced by our analysis of small RNA-sequencing data from clinical samples, the degradation pathway of miR-21 we identified is frequently disrupted across a wide variety of proliferative diseases, underscoring the importance of this regulatory network in human health and disease.

Materials and methods

A detailed description of the materials, data sources, and experimental and analysis methods is provided in SI Materials and Methods.

Analysis of miR-21 isoforms in sequencing data

For human, the mature miR-21 sequence listed in miRBase [194] is the common 22 nt isoform UAGCUUAUCAGACUGAUGUUGA. This 22 nt sequence and its 3′ downstream sequence are conserved in all species we analyzed, except for Japanese flounder [219], for which the mature sequence is UAGCUUAUCAGACUGGUGUUGG but the downstream sequence is the same as for the other species.

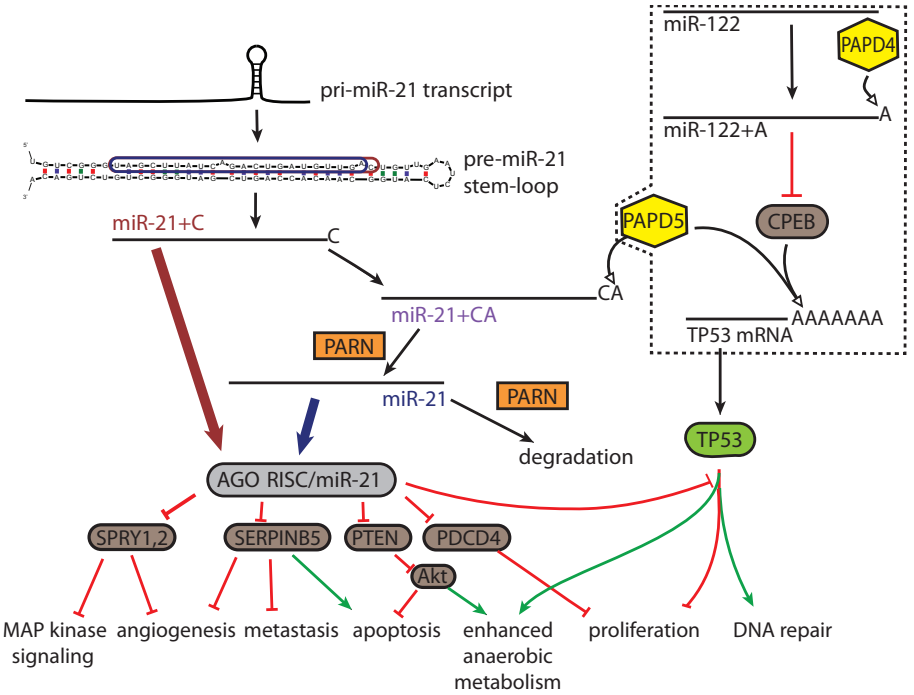


Figure 2.5: PAPD5 prolongs the life of the mRNA encoding tumor suppressor TP53 by polyadenylation. By adenylating miR-21+C and thus inducing its degradation by PARN, PAPD5 prevents the oncomiR from repressing other tumor suppressors. Once associated with an Argonaute protein, miR-21 isomiRs engage in the promotion of a variety of classical cancer hallmarks [14], such as angiogenesis, the avoidance of apoptosis, altered signaling, metastasis, enhanced anaerobic metabolism, and enhanced proliferation. The miR-21 targets shown in this figure have all been experimentally verified (PTEN: [63], PDCD4 [216], SPRY1 [217], SERPINB5: [218]). Nucleotidyl transferases are represented as yellow hexagons; TP53, a key transcription factor involved in tumor suppression, is represented by a green rounded rectangle; the exoribonuclease PARN is shown in orange; and regulatory interactions are represented by green arrows (up-regulation or stimulation) or red "inhibition" arrows (down-regulation or inhibition). The uptake of canonical miR-21 and miR-21+C into the RISC complex is denoted by arrows. Adenylated miR-21+C undergoes degradation by the 3′ exoribonuclease PARN, which yields the 22-nt miR-21 and shorter degradation products.

Statistical analysis of adenylation/degradation rates and overrepresentation of degradation products

The enzyme knockdown conditions were analyzed separately by comparing the sequence counts for isomiRs in the knockdown condition to their sequence counts in the negative control to estimate the common dispersion of the negative binomial distribution [220] using the maximum likelihood method. We then used the likelihood ratio test [221] to calculate the statistical significance of differences in adenylation ratios between each knockdown condition and the negative control for each miRNA.

Authors' contributions

MdH, SI, and YH designed research; JS, KT, KI, KH, YA, KK, CSa, CSu, MS, SA, HP, CR, YI, AO, HS, AH, ML performed research; JB, MdH, and CC analyzed data. JB and MdH wrote the paper, with AB, CR, HP, PC, HS, SMH, PG, and BT providing helpful comments.

Supplementary information

Supplementary material pertaining to this chapter can be accessed at http://www.joostboele.nl/thesis/.

Acknowledgments

The authors would like to acknowledge RIKEN GeNAS for the sequencing of the MCF7 small RNA libraries using the Illumina Genome Analyzer, and of the THP1 and MCF7 libraries on the Illumina HiSeq 2000. The authors would like to thank Fumi Hori for technical assistance, and Noriko Nakanishi for technical assistance and help with the manuscript. The authors would also like to acknowledge the TCGA Research Network for providing access to the data of The Cancer Genome Atlas. This work was supported by a research grant for RIKEN OSC from MEXT to YH, a grant of the Genome Network Project from MEXT to YH, a grant to the Cell Innovation Project (SI) from MEXT to YH, a Research Grant from MEXT to RIKEN CLST, a research grant from MEXT to RIKEN PMI, a grant to JB from the Hendrik Muller Foundation (Netherlands), and a grant from the Japan Society for the Promotion of Science to MdH.

Chapter 3

High-throughput phospho-proteome screening reveals differential signaling upon stimulation of the CXCR4/CXCR7/CXCL12-axis in breast cancer cells

In collaboration with Azra Mujic-Delic, Danny Scholten, Prahlad T. Ram, Frank J. Bruggeman, Bas Teusink, and Martine J. Smit.

Introduction

Chemokines are small peptides that are secreted by immune cells to modulate the immune response and immune surveillance of the body [222]. Even though most chemokines are released strictly in response to inflammation or injury, expression of some chemokines is constitutive and limited to specific cells and organs in order to sustain acquired immunity [223]. In case of inflammation, chemokines are released at the site of inflammation to form a gradient, which attracts leukocytes that express the cognate receptor the chemokine binds to.

C-X-C motif chemokine 12 (CXCL12) is a chemokine that mediates tissue homeostasis and immune surveillance in healthy tissue [224]. In cancer, CXCL12 is overexpressed by carcinoma-associated fibroblasts, which contributes to proliferation of the primary tumor [225]. Alternatively, it is secreted by distant organs, forming a gradient and promoting migration of primary tumor cells (metastasis) towards these organs [225].

For long, CXCL12 was thought to bind and activate only the chemokine receptor CXCR4. Yet, in the last decade CXCL12 was found to bind and activate the chemokine receptor CXCR7 as well [226]. In general, chemokine receptors are G protein-coupled receptors (GPCRs) that signal through activation of G proteins ($G\alpha_i$, $G\alpha_{12/13}$), initiating Ca^{2+}-release and migration [227]. This rule applies to CXCR4, but intriguingly, CXCR7 signals in a β-arrestin dependent manner [228], which can in turn lead to activation of the MAP kinase pathway as well as other prominent signaling pathways [229]. For this reason, CXCR7 has recently been designated as an atypical receptor (Atypical ChemoKine Receptor 3, ACKR3) [230]. Both receptors are overexpressed in various cancer types, and inhibition of either of them leads to decreased tumorigenesis *in vitro* and *in vivo* [231].

The CXCR4/CXCL12 signaling axis has been extensively studied, and there is ample knowledge on how this ligand-receptor pair interacts and which pathways it activates. However, as noted above, CXCL12 also binds to CXCR7, a chemokine receptor that was more recently discovered [232, 233], and thus the effects of the CXCR4/CXCL12 axis may have been confounded by CXCR7 and need to be revisited.

The complexity of this joint CXCR4/CXCR7/CXCL12 signaling axis has only recently begun to be appreciated. For example, it was reported for several cell types that CXCR4 and CXCR7 are able to form heterodimers both *in vitro* and *in vivo*, which alters G protein-dependent calcium responses, causes constitutive recruitment of β-arrestin, and enhances cell migration [234–236].

However, there are also reports of cell types in which one of the receptors is expressed, but not the other. In glioblastoma cells CXCR7 expression seems to predominate [237], while in gastric cancer, CXCR4 overexpression is well-established [238–242] but CXCR7 does not seem to be highly expressed [243]. The importance of CXCR4 in mammary tumor growth and metastasis is well-established [244]. Interestingly, CXCR7 is known to promote proliferation in this tissue as well as scavenge CXCL12 by internalization and recycling [245–247]. Moreover, expression of CXCR7 can lead to tumor growth and metastasis of CXCR4 positive cells [246].

We aimed to elucidate the signaling machinery that is activated by CXCL12 when both receptors are endogenously expressed. In the human breast cancer cell line BT474, both CXCR4 and CXCR7 are expressed, which makes it a suitable model system for the study of this network. In addition, the cells do not secrete CXCL12, a confounding trait some other cell types do display. Using receptor-specific inhibitors enabled us to study the downstream effects of the two receptors in isolation.

To be able to scan a broad panel of (phospho-)proteins for their involvement in CXCR4/7-mediated signaling, we evaluated our samples using a Reverse Phase Protein Array (RPPA). RPPA is an established, high-throughput and reproducible proteomics approach for detection of more than 100 validated antibodies [248]. Moreover, reliable methods have been developed for appropriate data processing, ensuring the quality of obtained results [249]. In view of the growing appreciation for the complexity of GPCR signaling, applying a systems biology approach is an increasingly attractive method for studying these systems, and RPPA is a suitable method in the exploratory phase of the model-building process [250].

Measuring the response of BT474 cells to CXCL12 at several time points after stimulation revealed changes in the phosphorylation status of proteins. Inhibition of CXCR4 or CXCR7 showed distinct effects on these responses. To mitigate observer bias, we applied unsupervised correlation-based analyses to the data, which confirmed the occurrence of several changes that appear to be specifically mediated through CXCR4 or CXCR7. Our findings underline that CXCR4 and CXCR7 have distinct signaling mechanisms and suggest their involvement in various cancer hallmarks.

Results

The BT474 breast cancer cell line expresses CXCR4 and CXCR7 but not CXCL12

First, we tested the BT474 breast carcinoma cell line for expression of CXCR4, CXCR7, and CXCL12. Since CXCR7 is known to also bind to CXCL11 [232], a characteristic it shares with CXCR3, we checked for the presence of this receptor and ligand as well (Figure 3.1A). Interestingly, while CXCR4 was mostly expressed on the cell surface, CXCR7 was found both on the cell surface as well as intracellularly (Figure 3.1B). Although there was expression of CXCL12 on the mRNA level, when we assayed for CXCL12 secretion in culture medium, we did not detect CXCL12 secretion on the protein level. The cells did not express CXCR3 or CXCL11 mRNA, making them a suitably "clean" system for the study of CXCL12-induced signaling through CXCR4 and/or CXCR7.

IT1t and VUF11207 are selective inhibitors of CXCR4 and CXCR7, respectively

Stimulation with CXCL12 leads to binding of the chemokine to CXCR4 and/or CXCR7, and activation of the respective signaling cascades (Figure 3.1C). At several time points after stimulation, we took samples for RPPA to study the effects of CXCL12 on the downstream machinery over time. To study the individual roles of CXCR4 and CXCR7 in this response, we also pre-incubated with the CXCR4-specific inhibitor IT1t and the CXCR7-specific inhibitor VUF11207, one hour prior to CXCL12 stimulation.

VUF11207 displays a high affinity for CXCR7 in a CXCL12 displacement assay (Figure 3.2A), and is able to recruit β-arrestin2 to the receptor in HEK293T cells transiently expressing CXCR7 (Figure 3.2B). Thus, it acts as a functional antagonist by reducing the capacity of CXCR7 that is available for binding CXCL12 [251].

CXCR4 inhibitor IT1t [252] has a higher affinity for CXCR4 compared to the more frequently studied inhibitor AMD3100 (Figure 3.2C), while showing equal potency in inhibiting CXCL12-induced G protein-dependent activation of CRE transcription factor in HEK293T cells transiently expressing CXCR4 (Figure 3.2D). Moreover, in contrast to IT1t, AMD3100 as well as the CXCR4 inhibitor TC14012 have been reported to activate CXCR7 at high concentrations [253, 254], an undesirable property that has not been associated with IT1t. Therefore, we selected IT1t for use in this study.

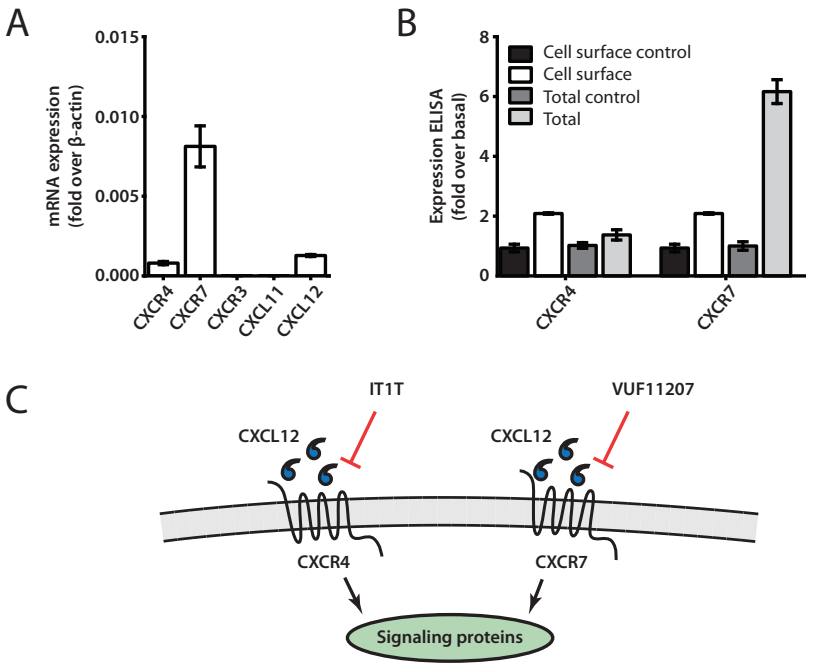


Figure 3.1: Characteristics of CXCR4/CXCR7/CXCL12 expression in BT474 cells. (A) qPCR showed that CXCR4, CXCR7, and CXCL12 are all expressed. (B) This was confirmed by an ELISA, which also shows the levels and locations of the proteins. (C) Schematic representation of the signaling system and its inhibitors. CXCR4 and CXCR7 are both activated by the chemokine CXCL12.

For each of the conditions (only CXCL12, CXCL12 + 1 μM IT1t, and CXCL12 + 1 μM VUF11207), we assayed by RPPA the relative levels of 161 antibody probes, at six time points (0, 2, 5, 10, 30, and 60 minutes after stimulation), making for a total of 3864 data points.

The signaling response to CXCL12 stimulation is due to both CXCR4 and CXCR7

The output of a reverse-phase protein array (RPPA) can be thought of as that of a western blot, performed on, in our case, 161 probes at once. A common way to initially organize this large quantity of data, after normalization, is to scale all relative expression values to a color bar and cluster the probes and samples, which yields a heatmap view. In the clustered heatmap view of

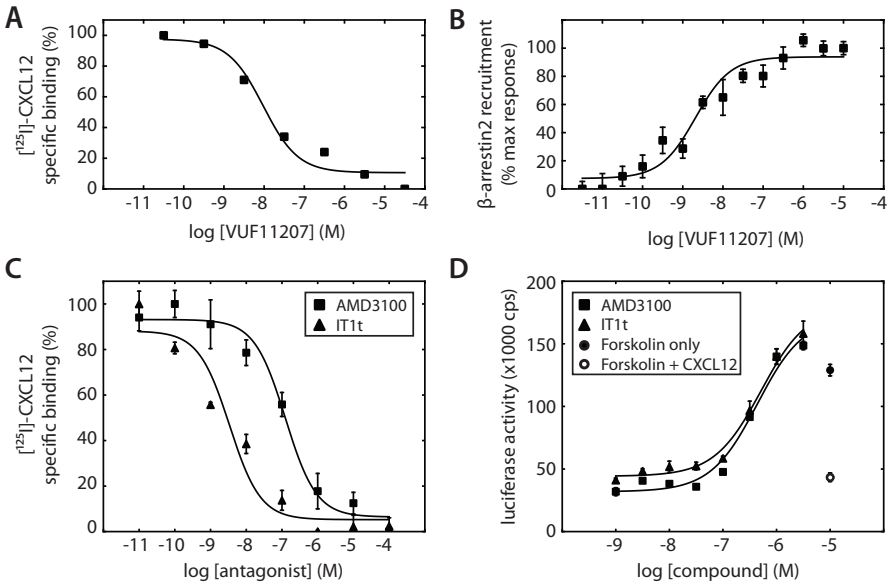


Figure 3.2: Binding and functional properties of CXCR4 and CXCR7 inhibitors in HEK293T cells. (A) VUF11207 displaces CXCL12 in HEK293T cells transiently expressing CXCR7 with high affinity (IC_{50} = 9.4 nM). (B) A bioluminescence resonance energy transfer (BRET) assay shows that VUF11207 is also able to recruit β-arrestin2 to CXCR7 with a high potency (EC_{50} = 2 nM). (C) IT1t displaces CXCL12 from CXCR4 with a higher affinity (IC_{50} = 3.4 nM) than AMD3100 (IC_{50} = 124 nM) in CXCR4-expressing HEK293 membranes. (D) IT1t and AMD3100 are able to inhibit CRE transcription factor activation with similar potency in HEK293T cells transiently expressing CXCR4, as shown by CRE reporter gene assay. In panels B–D, error bars represent the standard error of the mean.

our experiment, the different pre-incubation conditions (no inhibitor, IT1t, or VUF11207) show expression patterns that are distinct from one another, but similar among samples in the same treatment time course, indicating that the signaling responses to CXCR4 and CXCR7 can be distinguished on a molecular level (Figure 3.3A).

The signaling response to CXCL12 is diverse as well as receptor-specific. We show several examples of changes in the phosphorylation status of proteins in Figure 3.3B. For instance, phosphorylation of Akt at Thr308 is strongly increased upon CXCL12 stimulation with a peak at 30 minutes. This increase seems to be due to CXCR4 activation, since pre-incubation with its inhibitor IT1t abolishes the signal, while CXCR7 inhibitor VUF11207 is not able to attenuate the increase in phosphorylation. The opposite applies to phosphorylation

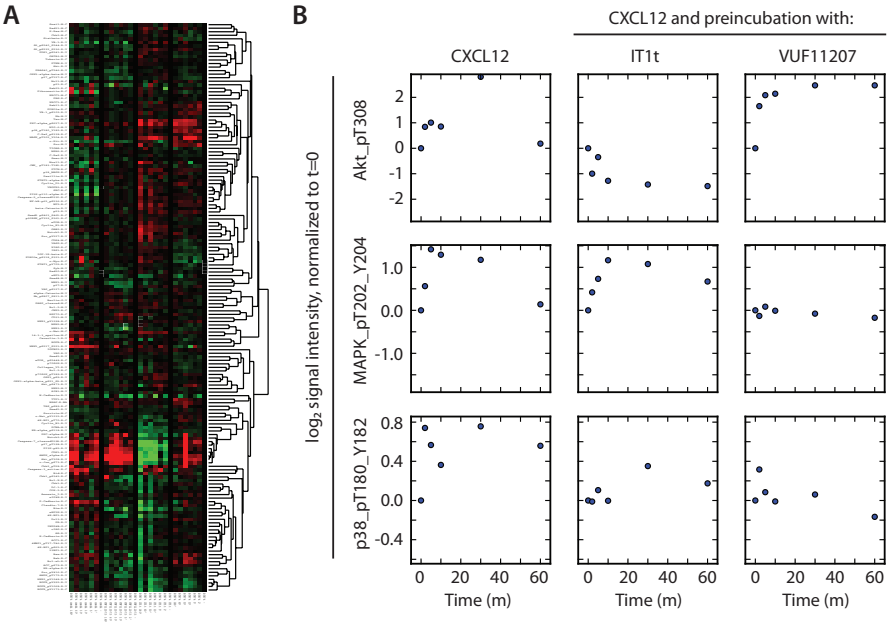


Figure 3.3: Overview of the RPPA results presented in this study. Pre-incubation with the CXCR4 inhibitor IT1t or the CXCR7 inhibitor VUF11207 in BT474 cells allowed us to study each receptor's downstream response in isolation. (A) Heatmap representation of the RPPA results. Red color indicates increased (phospho-)protein levels compared to basal (unstimulated cells), while green indicates decreased phospho-protein levels. (B) Changes in probe intensity of three phospho-proteins upon stimulation with CXCL12 alone, or stimulation with CXCL12 after pre-incubation with IT1t (CXCR4 inhibitor) or VUF11207 (CXCR7 inhibitor).

of p42/44 MAPK at Thr202/Tyr204, where CXCR7 inhibition leads to decreased phosphorylation while inhibition of CXCR4 has no effect (Figure 3.3B). Furthermore, we find proteins where CXCL12-induced phosphorylation seems to be mediated by both CXCR4 and CXCR7, as is the case for p38 MAPK phosphorylation at Thr180/Tyr182 (Figure 3.3B).

Differential clustering analysis reveals CXCR4- and CXCR7-mediated differential signaling

Although visual inspection alone already gives insights into some signaling proteins that are potentially activated downstream of CXCR4 and CXCR7, it is likely that the highly intertwined nature of signaling networks obscures more

subtle effects of the activation of these receptors by CXCL12. Therefore, we performed a computational analysis aimed at identifying probes that were differentially responsive to CXCL12 under various combinations of inhibitor pre-incubation conditions. For instance, one can preset biologically relevant questions such as *"which probes have response A in CXCL12-only and CXCL12 + IT1t, but the opposite response in CXCL12 + VUF11207?"*, while minimizing the amount of time spent on less informative correlations in the data (e.g., probes that respond to all treatments correlate strongly with themselves and similar probes, but the biological meaning of this is less relevant). Supplementing visual inspection with a computational analysis has the added benefit of helping to mitigate the risk of confirmation bias in this exploratory phase of the study. Details about the specific approach we used can be found in Materials and Methods, but we provide a summary in Figure 3.4A.

The results of an unsupervised computational analysis must always be manually pruned. In our analysis, an initial number of 34 probes of (potential) interest was reported. After applying cut-offs to eliminate probes that showed moderate or low differential expression, or erratic correlations, 18 were left, whose expression patterns were then manually inspected. The probes identified as MAPK p42/44 pT202/Y204, c-Jun pS73, and p70S6K pT389 were flagged by the computational analysis and withstood the subsequent manual curation step. As their responses to CXCL12 are dependent on either CXCR4 or CXCR7 according to the RPPA results (Figure 3.4B), we selected these proteins for further experimental validation.

CXCR7 mediates signaling through MAPK p42/44

To validate the differential phosphorylation of MAPK p42/44, we performed western blots to verify its abundance at several time points after CXCL12 stimulation, with and without pre-incubation with CXCR4/7 inhibitors. As indicated by the RPPA data we discussed above (Figure 3.4B), indeed, the level of MAPK p42/44 phospho-T202/Y204 increases after CXCL12 stimulation in a CXCR7-dependent manner (Figure 3.5A). The western blot results also refute the reverse (CXCR4-dependent signaling through activation of MAPK p42/44 phospho-T202/Y204), as pre-incubation with IT1t has no effect on phosphorylation.

Western blotting did not confirm CXCR4-dependent phosphorylation of c-Jun or CXCR7-dependent phosphorylation of p70S6K (Figure 3.5B-C), although their relative expression patterns on the RPPA resemble specific CXCR4 and CXCR7 responses (respectively), and thus warrant their selection in the

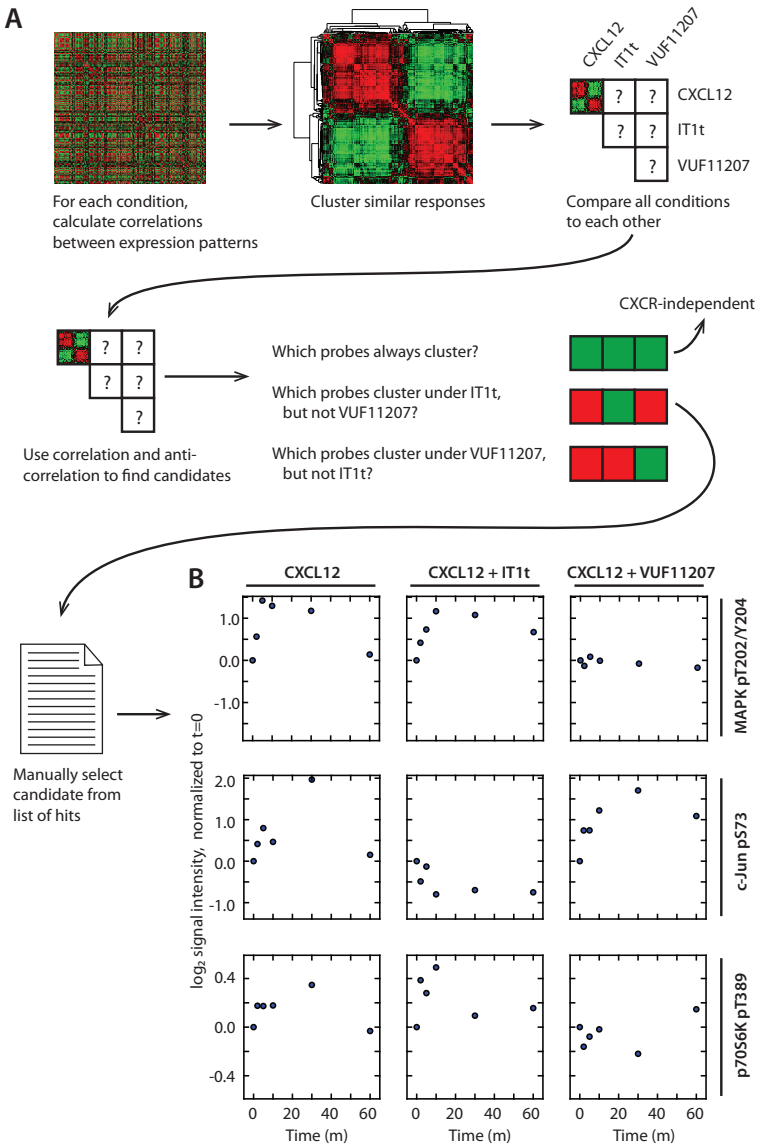


Figure 3.4: Summary of differential clustering analysis. (A) By comparing to each other all time course expression patterns, a matrix of correlation values is obtained. We can then cluster the probes according to their similarity under all combinations of inhibitor conditions, and use questions like *"which probes cluster under CXCR4 inhibition, but not under CXCR7 inhibition?"* to obtain lists of leads. (B) The list of ~20 leads is manually inspected. Phosphorylated MAPK p42/44, c-Jun, and p70S6K were reported by the analysis and stood up to scrutiny, and were selected for further investigation.

lead discovery stage (Figure 3.4B). Still, as MAPK p42/44 is a potent and well-known regulatory component, these findings underline the importance of the CXCR4/CXCR7/CXCL12 axis in the broader context of signaling in in breast cancer cells (Figure 3.5D).

Discussion

Chemokine receptor signaling is essential to the immune response, and it is often deregulated in diseases like cancer. Specifically, in breast tumors, CXCR4 signaling plays an important role in proliferation and vascularization of the primary tumor, as well as in migration and metastasis to distant sites [244]. Interestingly, CXCR4- and CXCR7-mediated signaling show distinct patterns in breast cancer.

For instance, a recent study where the rat mammary adenocarcinoma cell line MTLn3 was used to overexpress these two receptors showed that CXCR7 expression increases chemotaxis of MTLn3 cells towards CXCL12 compared to cells that only express CXCR4, and that binding of CXCL12 to CXCR4 was responsible for this response [245]. Also, both *in vitro* and *in vivo*, CXCL12-induced invasion was decreased when cells expressed both CXCR4 and CXCR7, compared to cells that only expressed CXCR4. Additionally, cells expressing CXCR4 had increased cell motility within the primary tumor as well as increased intravasation compared to non-CXCR4-expressing cells, while CXCR7-expressing cells enhanced primary tumor growth.

The BT474 breast cancer cell line expresses both CXCR4 as well as CXCR7 endogenously, which enabled us to study the contribution of both receptors to its signaling phenotype. Upon inhibition of CXCR4 or CXCR7 in BT474 cells, we observed distinct phosphorylation patterns of signaling proteins on a reverse-phase protein array.

Reverse-phase protein arrays (RPPA) were designed to enable the probing of samples of interest with (phospho-)protein-specific antibodies in a fast and reproducible manner. This method has been used widely in studies to gain insight into complex signaling networks. For example, the 60 cancer cell lines in the NCI-60 panel were characterized using RPPA to assess which signaling networks are most often affected in cancer, and to gain insight into sensitivity to standard anti-cancer therapy [260]. Furthermore, RPPA has been used to make a molecular distinction between different cell populations in acute myeloid leukemia [261].

Because RPPA is an antibody-based assay, it lends itself to a wide variety

Figure 3.5: (A) Western blotting con rms that phosphorylation of MAPK 42 /44 at T202/Y204 decreases upon CXCR7 inhibition. This nding con rms our computational prediction of CXCR7-mediated signaling through MAPK, and is consistent with the earlier RPPA results (Figure 3.4B). Conversely, inhibition of CXCR4 by IT1t has no discernable e ect on CXCL12-induced MAPK phosphorylation. CXCR4-mediated c-Jun phosphorylation (B) and CXCR7-mediated phosphorylation of p70S6K (C) were not con rmed by the western blot results. (D) The network context of MAPK, p70S6K, and c-Jun with respect to important signaling pathways and upstream stimulation. MAPK's namesake MAP kinase pathway is implicated in cell growth and frequently deregulated in cancer. The PI3K/Akt axis, shown partially in the center of the network, is known to respond to CXCR4 activation [255], and various means of cross-talk exist between the MAP kinase and PI3K/Akt pathways [256–259]. For each (phospho-)protein that was assayed by RPPA, its CXCL12 response pro le according to the computational analysis (CXCR4-associated, CXCR7-associated, and non-speci c response), as well as the results of the subsequent manual curation step ("selected upon manual inspection"), are denoted by colored circles.

Once generated, results can be visualized, but they are also always presented as a human-readable table (which includes reduced costs for each reaction) and as a machine-readable, tab-separated format file that can be imported in e.g. Excel (Figure 4.2D).

Visualization

The visualization module generates images in SVG format (Scalable Vector Graphics), based on the analysis results returned by PySCeS. The advantages of using SVG are manifold, some of the more notable being image scalability and ease of editing using third party software. Depending on the web browser used, users may need to download a (free) plug-in to view the images.

For each selected pathway an interactive KEGG-like image is drawn (Figure 4.2C), on which the run results are superimposed. To the biologist, this readily recognizable representation is an improvement over unsupervised visualization algorithms (e.g. in [285]), and while this approach to data visualization was already applied some years ago [286], to our knowledge, FAME is the first web-based application that both generates data and automatically visualizes results.

Many elements in the results images are clickable (another advantage of the SVG format), to make more information available more conveniently. For instance, clicking a metabolite will display an overview of reactions producing or consuming it, along with the KEGG information page for the metabolite, while clicking a reaction will display that reaction's KEGG information page.

Conclusions

With FAME, we present the community with an easy to use web-based "one stop shop" for the manipulation and execution of stoichiometric models. It enables biologists to create or import models, edit them, run them at the click of a button, and visualize the results from the browser window. We expect that its install-free integration of execution and visualization will appeal to investigators and educators alike.

Future releases of FAME will feature integration with web-based annotation services, gene associations, and further analysis options. Finally, the novel SOAP interface to PySCeS-CBM will facilitate the creation of user-friendly interfaces based on PySCeS that will uncover powerful modeling functions that may otherwise remain hidden behind the ever-enigmatic command line cursor.

Availability and Requirements

FAME is intended and offered as a web service, but can also be installed locally, as source code will be made available upon request. FAME can be accessed online at http://f-a-m-e.org/, where a full user manual and guided tutorial are available. PySCeS-CBM and Mariner are also open source, and can be downloaded from http://pysces.sourceforge.net/cbm. FAME and PySCeS/Mariner are covered by their own respective BSD-style licenses, which can be found on the respective web pages and, in short, entail that they are open-source and free to use for both academic and non-academic users.

Authors' contributions

JB created FAME. BGO created Mariner and PySCeS. JB, BGO, and BT wrote the paper. All authors read and approved the final manuscript.

Acknowledgments

The authors thank Filipe Santos and Anisha Goel for helping improve FAME by test-driving it within their daily modeling practice. BGO is supported by NWO Computational Life Science Grant 635-100-021. BT and JB acknowledge the Centre for Integrative Bioinformatics VU (IBIVU) and the Amsterdam Institute for Molecules, Medicines and Systems (AIMMS) for support.

Tool name	Ref.	Visualization			Run				Oper. mode		Model handling			
		Supervised	Unsupervised	User-supplied	FBA	FVA	MOMA	Gene/rxn	Stand-alone	Web-based	Build	Edit	Import SBML	Export SBML
FAME		•		•	•	•				•	•	•	•	•
Model SEED	[276]	•			•			○[1]		•	•	○[2]		•
COBRA Toolbox	[277]			•	•	•	•	•			•	•	○[3]	○[3]
OptFlux	[278]			•	•	•	•	•	•				•	•
CellNetAnalyzer	[279]			•	•	•							•	•
PySCeS	[147]				•	•			○[4]			•	•	•
YANASquare	[285]		•	•	•				•		•	•	•	
MEGU/Pathway Projector	[286]	•								•				
BioMet Toolbox	[141]				•			•	○[4]	•				
Cytoscape	[287]		•	•				○[1]	•		•	•		

Table 4.4: Comparison of the features offered in FAME and existing alternatives. A • indicates a feature that is present; an open circle (○) denotes a partial implementation of the feature in question. Under visualization, "supervised" means the application uses predefined network topologies, "unsupervised" means networks are drawn on the fly, as graphs, and "user-supplied" means that visualization is performed on user-supplied network topology maps. Most packages feature additional analyses besides the ones listed in this table. As many of these tools are custom-built for these analyses, or vice versa, we do not list those analysis options in this table. The tools whose functionalities most resemble FAME's have no web interface or have inferior visualization capabilities. On the other hand, tools featuring (or focusing on) superior visualization lack the powerful analysis features unlocked by PySCeS-CBM.

Notes:

1. Gene/reaction associations are used to build models, but no analyses are performed on them.
2. Model SEED has only limited editing capabilities once a model has been produced.
3. The COBRA Toolbox uses an SBML dialect specific to this tool.
4. Certain dependencies must be installed first.

Chapter 5

An interactive graphical map of metabolism

In collaboration with Timo R. Maarleveld, Remco C. Jongkind, Frank J. Bruggeman, and Bas Teusink. Adapted from: *Plant Physiology* (2014), **164**: 1111–1121. Copyright American Society of Plant Biologists.

Introduction

High-throughput experimental technologies have seen spectacular advances in recent years, and as a result, previously unimaginable amounts of biological data are now available. This "omics data explosion" includes transcriptomics (sequencing), metabolomics, and fluxomics data [288–291], and the analysis and integration of these data types can reveal important information about the system-level functioning of organisms [292]. Computational modeling of cellular behavior is a tried-and-tested method for the interpretation of "big data", and with genome-scale models now a reality, they are currently a staple of systems biology (Chapter 1).

The extensive body of existing biochemical knowledge of metabolism, combined with the relative ease with which relevant experiments can be performed and the central role of metabolism to the functioning of the cell, make genome-scale reconstructions of metabolism an obvious starting point for the interpretation of genome-scale data sets. Based on the contents of an organism's genome, these reconstructions list the metabolic reactions it can perform [124, 275]. However, even though many metabolic enzymes are well-characterized, for most reactions we still lack knowledge of the enzyme kinetics that are involved. To bypass this issue, genome-scale reconstructions are solely based on the reaction stoichiometry, and they are therefore generally referred to as genome-scale stoichiometric models (GSSMs).

To leverage the biological information that is included in GSSMs and predict physiological properties, various stoichiometric network analysis tools are available [293, 294]. For instance, Flux Balance Analysis (FBA) and Flux Variability Analysis (FVA) can be used to predict the internal flux distribution and its variability while optimizing for cellular growth yield [138, 139, 295], a typical objective in the study of micro-organisms or cancer cells. Alternatively, tools like OptKnock [140] can be used to predict which gene knockout strategies would result in increased production of metabolites of interest, such as biofuels.

Unfortunately, the results of analyses performed on GSSMs come in the form of lists of (reaction) identifiers and (flux) values, which often have thousands of entries. The sheer size of these lists makes their interpretation a tedious and human-unfriendly process, yet to be of any biological value, interpretation of e.g. FBA results in terms of biological functions, modules, and pathways is essential. To achieve this, adequate visualization is paramount, and attempts to visualize genome-scale data have indeed been

made. These approaches can be classified into supervised (human-driven) and non-supervised (fully automatic) ones. The latter approach, fully automated mapping of genome-scale models, results in graphs (in the mathematical sense) that are even more difficult to interpret than the raw data one set out to visualize on them, so we will not discuss them in detail here.

Human-driven visualization methods, on the other hand, are considerably more popular, as hand-drawn maps are markedly easier to interpret. Tools like KEGG [281], Cytoscape [296], CellDesigner [280], and the COBRA Toolbox [277] each offer their own particular take on modeling and visualization. However, since none of them was originally designed for visualizing genome-scale models or genome-scale data, they are of limited use for this purpose. This becomes particularly clear when one attempts to share, extend, or adapt an existing map. Commonly, users who attempt this quickly find themselves fazed by the need for special software, by incompatible identifiers, or simply by not being able to access an editable version of the desired image. Even iPath [297], which was designed with genome-scale applications in mind, suffers from these problems, and thus far, usability issues have prevented the wide-spread communal use of visualization aids, as well as their use in combination with computational results. However, the increased availability of high-throughput data makes the lack of adequate graphical interpretation tools more dearly felt than ever.

We address this issue by presenting the first comprehensive genome-scale data visualization tool that is both extensible and open-source: A graphical map of metabolism that can display any type of data that is associated with reaction identifiers, gene identifiers, or metabolite identifiers. The map is fully integrated with FAME, the web-based Flux Analysis and Modeling Environment [298], but it can also be downloaded and used to display results from other analysis tools. To enable advanced visualization options, we also provide a stand-alone Python console application named VoNDA, which can be used to take advantage of the functionality of (other) existing constraint-based modeling software. For instance, using VoNDA in combination with PySCeS-CBM [147] or COBRApy [299] allows for a fully Python-based modeling experience.

As a demonstration of how the generic map can be instantiated to suit a specific genome-scale model, we also present a graphical map that is specific to *Synechocystis PCC 6803* (hereafter simply *Synechocystis*) [300–303]. This cyanobacterium can directly convert carbon dioxide (CO_2) into carbon compounds, which makes it of great interest for the development of production

techniques for third-generation biofuels [304–306].

At http://f-a-m-e.org/genericmapdemo/, the generic map is pre-loaded into FAME, along with a model of *Escherichia coli* to demonstrate the map's use in analysis and visualization in an interactive manner. Similarly, the *Synechocystis* map and model are made available at http://f-a-m-e.org/synechocystis/. VoNDa can be downloaded from http://vonda.sf.net/.

Results and Discussion

An overview of the generic metabolic map

Which reactions should one draw on a "generic" metabolic map? Drawing all imaginable reactions would be very labor-intensive process that would, ironically, render the map cluttered and useless. Selecting reactions by hand is a reliable method to include biologically relevant pathways, but it introduces a bias towards familiar or favorite organisms. Therefore, we selected about 50% of the reactions on our map by hand, and performed a computational analysis on all genome-scale models in the MetaNetX database [307] to select the other reactions based on their prevalence in existing models (see Materials and Methods).

The resulting generic metabolic map that we present here contains 1605 reactions (excluding exchange and transport reactions). The metabolites and reactions on it are represented internally in the MetaNetX namespace [307]. MetaNetX is designed to be very comprehensive and comes with cross-referencing tables that aid the integration of the map with models from a variety of sources (see Table 5.1 for an example). Users who favor an alternative naming scheme can either convert their model to MetaNetX format (a tool for this is provided on http://metanetx.org/), or the map to their model's format. The latter process is automated in FAME, to maximize the map's immediate utility and minimize the conversion effort on the user's part.

The map can be used to visualize reaction-based results, such as those of a FBA or FVA, by color-coding reactions based on their flux or variability value (Figure 5.1).

Each reaction group also includes three small boxes that can be used to display information about the genes that are associated with that reaction (see Materials and Methods). These rectangles are normally hidden from view, but can be shown when gene expression information is available.

Namespace		“Enolase”	“ATP”
BiGG	[282]	R_ENO	M_ATP
BioPath	[308]	RXN01306	Adenosine-5-prime-triphosphate
Brenda	[96]	BR3384	BG1395
KEGG	[281]	R00658	C00002
MetaCyc	[309]	2PGADEHYDRAT-RXN	ATP
MetaNetX	[307]	MNXR1898	MNXM3
SEED	[310]	rxn00459	cpd00002
UniPathway	[311]	UCR00658/UER00187	UPC00002

Table 5.1: Example showing the identifiers of a reaction (2-phosphoglycerate to phosphoenylpyruvate (PEP), catalyzed by the metabolic enzyme enolase, Figure 5.3) and a metabolite (ATP) in various namespaces.

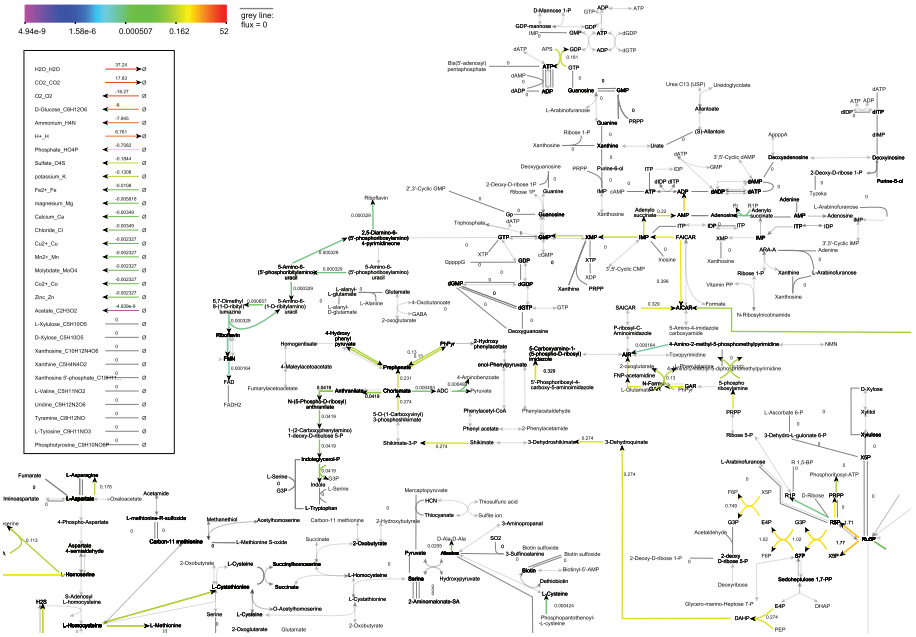


Figure 5.1: A summary image showing the map’s flux analysis visualization features. To prevent the figure from becoming illegible, this example covers only <10% of the entire generic metabolic map. This image shows FBA results of the iAF1260 model of *E. coli* being visualized by color-coding the reaction lines. Exchange reactions in this model have been automatically added to the map by FAME, and sorted by absolute flux magnitude (partially shown in inlay on left hand side of figure).

Using the generic metabolic map to visualize arbitrary organisms

Although the map can be downloaded separately and adapted to the user's needs, we have also integrated it in FAME, which can directly load it and tailor it to the user's model. The link between a model and the generic map is achieved by ensuring that the identifiers of the reactions, metabolites, and genes in the underlying model match those on the map.

FAME offers several tools that help make sure the model and map are optimally linked. In addition to converting the map to the namespace the model is composed in, it can hide reactions that are not in the model, and it can automatically draw reactions that are defined in the model, but that are not yet on the generic map. The user can then drag these into place to customize the map to their specific needs, and use this customized map instead of the generic one. When used in conjunction with a model that uses the MetaNetX namespace, the map can also be directly loaded in VoNDa to execute more complex visualization queries from the command line.

It follows from the above that the generic map we present here should be seen as a starting point, both for individual visualization needs as well as for a comprehensive generic visualization solution for the community. While we provide various tools that help tailor the map to the user's needs based on the contents of the model, and this is likely to be sufficient for incidental use, workflows that rely heavily on graphical interpretation will require that the map be customized to a greater extent. In most cases, this means expanding the map with additional reactions. As each reaction needs to be manually added in the desired location, this is necessarily a rather labor-intensive process, but we have attempted to minimize the time investment that is required to customize the map.

Besides using an open source image format, we do this by using the MetaNetX namespace, whose identifiers can be easily migrated to cognate identifiers in alternative nomenclature systems (e.g. [282, 312–314]). FAME can migrate the map's identifiers to a variety of namespaces to provide a starting point for expansion of the map. Additions to the map can be drawn by hand or in whichever other way the user deems most effective or most pretty, as long as each new element is associated with the correct identifier. We provide an online SVG-editor in FAME, which allows for easy manipulation of the metabolic map, including any results superimposed on it.

While the quality of any customized map will be strongly dependent on the knowledge and skill of whoever modified it, as well as the quality of the model it represents, the present work presents the community with a solid ba-

sis from which to start the process of gaining biological insight through graphical representations of large data sets. The base files for the map are hosted on GitHub (https://joostboele.github.io/Metabolic-Map/), which encourages users to share their modifications for others to (re)use (e.g. by further adapting them and sharing the result) [315].

Positioning the map in the spectrum of visualization approaches

Obviously, the challenge our map addresses is not a novel one, and consequently, alternative solutions have been proposed. Broadly, these can be categorized into solutions that use the existing metabolic map from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [316], those that build upon the simulation framework offered by the COBRA Toolbox [277] or directly on Matlab, and all other solutions. While each approach has its merits, we decided against plugging into an existing framework, since existing solutions either impose too many restrictions on form and functionality (as in the case of KEGG or WikiPathways), or because of limited support for (open-source) graphical applications (as in the case of the COBRA Toolbox, which instead focuses on simulation). Also, importantly, many otherwise adequate visualization tools have no analysis functionality; instead, they rely on the user supplying data from an external source.

In Tables 5.2 and 5.3, we provide general and technical overviews of the various extant visualization solutions, and briefly describe their approach to the visualization of FBA-type results. It should be noted that the capabilities of the various tools vary with their aims and intended use, and that Tables 5.2 and 5.3 solely assess their usefulness for genome-scale metabolic applications, and not their general usefulness or quality. For instance, CellDesigner [317] is very commonly used for a broad range of visualization tasks, but finds application mostly in the realm of small to medium sized signaling models.

As our visualization solution was developed specifically for genome-scale applications, it should come as no surprise that the combination of FAME and VoNDa covers the foreseeable user requirements in this area. Thus, regardless of any modifications that may be required to tailor the map we present here to a specific purpose, this first generic, comprehensive, extensible, and human- and computer-friendly map of metabolism will surely be helpful in the interpretation of systems biology results. By offering this resource not only as a flat file, but also as an integrated part of an interactive online analysis environment, we expect it to be of even greater use to the community.

Tool name	Ref.	Short description
KEGG-based		
KEGG	[316]	The well-known KEGG metabolic map has been made available in SVG format. Very comprehensive mapping effort. Detailed information is hidden for reasons of perspicuity, which makes detailed interpretation of run results difficult. KEGG cannot perform analyses, but it can superimpose data on maps.
CytoScape	[318]	Cytoscape is a widely used open-source visualization platform. It cannot perform analyses on metabolic networks, but has an intuitive editor for network images and can visualize a variety of data. Uses fixed layout elements. The KEGG map of metabolism is available in Cytoscape format.
Pathway Projector	[126]	Pathway Projector is a web-based pathway browser that can be used to view, edit and visualize data on the KEGG pathway maps. It is based on the Google Maps API, and does not itself perform analyses. Images can be exported in XML or Biopax format.
iPath 2.0	[297]	iPath is web-based and can visualize KEGG maps of metabolic and regulatory networks. It can export images in various formats, including SVG, but it cannot change the basic layout of elements or perform analyses.
COBRA/Matlab-based		
COBRA Toolbox 2.0	[277]	Well-known Matlab-based modeling suite. Requires the use of the Simpheny image definition format and uses a fixed layout scheme. Can superimpose data on image, but visualization capabilities and options to gain a deeper understanding in the data or the model are limited.
Paint4Net	[319]	Plug-in for the COBRA Toolbox which adds detailed information about run results to visualization capabilities. Can be used to make new maps, but this is based on unsupervised graph algorithms.
(Cy)FluxViz	[320]	Plug-in for the COBRA Toolbox (CyFluxViz) and Cytoscape (FluxViz) designed to specifically facilitate the visualization of flux distributions. Relies on the host program for analysis features (COBRA Toolbox only) and requires a map to be supplied by the user.
RAVEN Toolbox	[321]	While primarily presented as an aid in the metabolic reconstruction process, the RAVEN Toolbox also has visualization functionality. For this, it requires pre-drawn maps in the CellDesigner format.
FAME (current)	[298]	Web-based, open-source. Can analyze models and visualize results on SVG maps, or visualize arbitrary data. Features an image editor, and can hide unused features of the image or automatically expand it with additional reactions. Is indifferent to the layout of maps as long as elements have the right IDs.
VoNDa	[322]	Stand-alone supplement to the above. Has visualization features similar to FAME, but offers additional control over the parameters of visualization and can include or exclude reactions based on the pathway they are in, their flux value, or the metabolites that participate in them.
Other		
WikiPathways	[323]	This community-driven project centers around pathways (unsurprisingly), and currently includes over 1700 species-specific pathway maps. The project's strength lies in the extensive annotation of all elements on all maps, which ensures that information can be reused in the future. The task of combining the various maps into single genome-scale visualization aids has yet to be undertaken.
CellDesigner	[317]	CellDesigner is a graphically oriented stand-alone modeling tool that can be used for both simulation and visualization. Its feature set is impressive, and mostly geared towards kinetic models. Consequently, it has mainly seen (published) use in the study of regulatory networks.

Table 5.2: Overview of extant visualization aids for biological systems.

Tool name	Ref.	Accessibility			Functionality				Automatic drawing	
		OS	WB	SA	ED	Analysis	Overlay	CLI	Leave out	Add
KEGG-based										
KEGG	[316]		•				•			
CytoScape	[318]	•		•	•		•	○		
Pathway Projector	[126]	•	•		•		•			
iPath 2.0	[297]		•				•			
COBRA/Matlab-based										
COBRA Toolbox 2.0	[277]	○		•		•	•	•		
Paint4Net	[319]	•					•	○		•
(Cy)FluxViz	[320]	•					•	○		
RAVEN Toolbox	[321]	○		•		•	•	•		
FAME (current)	[298]	•	•		•	•	•	○	•	•
VoNDa	[322]	•		•		○	•	•		
Other										
Wikipathways	[323]	•	•		•					
CellDesigner	[317]			•	•	•	•			○

Table 5.3: Comparison of of the technical capabilities of visualization tools. Only tools that can visualize data are shown here. The tools are grouped depending on their "base" drawing; such drawings include the KEGG map of metabolism, the COBRA Toolbox drawing philosophy, and the map we present here. SBGN and SBML Layout are not shown here, as they are graphical frameworks but not tools. Interestingly, various popular visualization tools cannot perform analyses and rely on the user supplying data instead. OS: Tool and image formats are open source (•); if non-open-source dependencies exist ○ is displayed instead. WB: Web-based. SA: Stand-alone. ED: Ability to edit the structure of the graphical map. Analysis: FBA, FVA, etc. can be executed from within tool. Overlay: Results of analysis or external data sets can be visualized on map. CLI: Command line interface for advanced analyses (○ is displayed for CLIs that can be accessed through additional software). Leave out: Unused reactions can be hidden from map. Add: Reactions that are not drawn can be automatically or manually drawn (•) or manually added as necessary (○).

Materials and Methods

Selection of reactions for the generic graphical map

Previously, we developed a graphical map of *Synechocystis* metabolism, which included 902 reactions (including exchange and transport reactions) [322]. We started by selecting core metabolic reactions from this set, and converting them from the BiGG namespace to MetaNetX (see Table 5.1 for an example of such conversions). As *Synechocystis* is a rather specialized organism, it lacks many reactions that are common in other parts of the tree of life, but rather than just hand-picking these, we applied a computer-guided selection procedure to avoid introducing a bias in the map's content.

To do this, we assessed the prevalence of each reaction by determining in which fraction of the models in the MetaNetX database it occurs. The MetaNetX database contains models from a variety of sources; some of these are built and curated by hand

(e.g. the models from the BiGG database), whereas others were generated computationally, and are barely curated (e.g. the models from The SEED). We took this into account when assessing a reaction's "prevalence in existing models", by not only determining the prevalence of each reaction in all models, but also by restricting our scope to curated models in the database. The results of these analyses were highly comparable, presumably because the information the automatically constructed models are based on is generally provided by the available manually curated models.

By sorting the list of reactions in order of descending prevalence, we could determine how much of each model we could visualize if we would create a map with the first n reactions. By varying the value of n, and the models whose coverage we assess (all models, or only curated models), we then determined that approximately 2000 reactions would be an optimal number for our map, as adding reactions beyond that number is no longer effective towards generating a "generic" map (Figure 5.2). However, additional reactions can still be added to tailor the map specifically to a certain (group of) organism(s).

We manually pruned the list of suggested reactions for consistency and biological sense, and combined with the reactions we carried over from *Synechocystis* (many of which were among the selected ~2000 reactions), the first draft of our generic map contains 1605 reactions. This number is lower than the 2000 quoted above, because exchange and transport reactions are not included in the generic map by default. Instead, we included an option in FAME to automatically draw these reactions based on the model the map is used with.

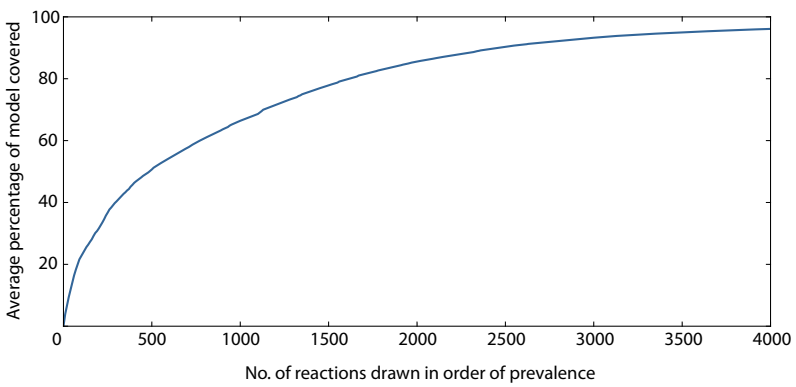


Figure 5.2: After ~2000 reactions, adding more reactions to the generic map results in diminishing returns on the map's coverage. In determining how many reactions to include in the generic map, we selected approximately half of the reactions manually, and the other half based on their prevalence across the models in the MetaNetX database. This plot shows that adding more than ~2000 reactions to the map would have a relatively low return on investment for a generic map, although it may still be a good idea when one is interested in a specific organism.

Construction of the graphical metabolic map in SVG

The map is represented in the Scalable Vector Graphics (SVG) format to maximize its potential for reuse by the community as well as its accessibility by both humans and computers. Each reaction is represented by a group node whose `id` attribute is the same as the reaction's identifier in the MetaNetX namespace. This way, the entire group can be hidden or removed if the reaction is not in the model that is being visualized. Each reaction's group includes the metabolites participating in the reaction, a path that connects these metabolites, rectangles that can be used to visualize e.g. gene expression data, and a hidden text node with the string `ReactionValue` followed by the reaction ID (Figure 5.3B). Because each reaction's group contains all metabolites participating in it (with the exception of some abundant co-factors), removing a reaction from the middle of a pathway does not affect adjacent reactions that use the same metabolites: If the metabolite also participates in an adjoining reaction, that reaction will have a copy of the metabolite in its group as well.

Using the SVG format allowed us to make the map interactive and dynamic. Specifically, it includes hyperlinks that can direct the user towards sources of additional information (e.g. KEGG) or, when FAME is used to generate simulation results and display them on the map, to details about the run that generated the displayed results. If we had utilized one of the budding visualization standards like the Systems Biology Graphical Notation (SBGN) [324] or SBML layout [325], including similar functionality would require the user to install additional third-party software simply to view the image, which would be undesirable. The SVG format, on the other hand, is open source, and it is commonly used and widely supported on the web.

Importantly, popular image manipulation software (such as InkScape or Adobe Illustrator) can import SVG files, which ensures that the map can be used as base material for publication-grade figures. As SVG is also computer-friendly, it is conducive to the automation of tasks like the extraction of subnetworks from maps, or the development of a new metabolic map for a specific organism using the generic map as a blueprint.

The map's features and their relation to biology

The map includes multiple instances of common metabolites, such as ATP, ADP, NAD(P)(H), and other co-factors. These metabolites are included on the map in various locations, which greatly reduces the clutter that would arise if only one instance of each were drawn (graph based visualization solutions tend to do this). Other helpful features are hidden when they are not in use, but deserve a brief explanation here; an overview of these features is given in Figure 5.3.

For instance, space has been reserved on the map to display flux values for reactions and the direction in which the reaction takes place (Figure 5.3B). Furthermore, when displaying run results generated in FAME, metabolite names link to a page with a list of fluxes that produce and consume the respective metabolite. Finally, when display-

ing run results, the lines that represent reactions can be color-coded to represent the magnitude of the flux through them (or the variability of this flux, in the case of FVA results), and an explanatory color bar is displayed at the top of the image (Figure 5.3C).

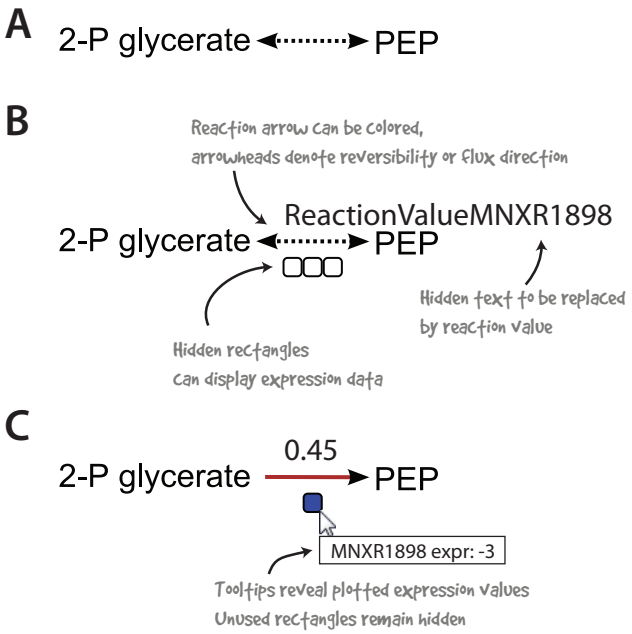


Figure 5.3: Overview of the "hidden" features on the map. The example reaction is catalyzed by enolase, a highly conserved enzyme that was also used as an example in Table 5.1. (A) When viewing the map without visualizing data, only the reaction's main reactants are shown, as well as the reaction's reversibility. (B) Exposing the hidden features reveals a text string that serves as a placeholder for flux data ("ReactionValue" plus the ID of the reaction) as well as three rectangles that can be used to display gene expression data. (C) Example of all of these features in use simultaneously. Hovering the mouse pointer over a rectangle reveals a tooltip with the plotted expression value. When viewing FBA results, clicking a metabolite opens a page that shows which reactions consume and produce the metabolite.

Integration of the map with FAME

To demonstrate the functionality of the map as well as facilitate its use, we have added a customized section to FAME, the Flux Analysis and Modeling Environment [298]. Visiting http://f-a-m-e.org/genericmapdemo causes the generic map to be loaded, along with the iAN1260 *E. coli* model from the BiGG database [282]. After converting the map to the BiGG namespace, which is a one click operation, the user is able to view, edit, and run the model, and have analysis results visualized on the map we present in this article. Though the ability to visualize results on user-supplied (custom) maps

was already present in FAME, the URL described above pre-loads the model and map and eliminates the need for loading several files separately.

When the map is loaded, FAME displays a new input box for gene expression data, allowing for visualization of these data on the map for interpretation in conjunction with flux analysis results. In addition, we have expanded FAME with an in-browser image editor. Because this image editor runs on the client machine as a JavaScript applet, its effectivity at editing the map is dependent on this machine's specifications. However, users can always download an editor-friendly version of the map for offline editing.

Authors' contributions

JB, TRM, and RCJ developed the map. JB, TRM, and BT wrote the chapter, with FJB providing helpful comments.

Acknowledgments

The authors would like to thank Brett Olivier for helpful discussions and tips for developing the data integration and visualization tool.

Chapter 6

The biomass composition of human cells

In collaboration with Petra Krumpochova, Brett Olivier, Dick-Paul Kloos, Martin Giera, and Bas Teusink.

Background

Genome-scale stoichiometric models have become a highly valued tool for the study of metabolism [326]. But the scale of these models comes at a price: To allow the effective study of these large models, reaction kinetics must be (largely) ignored, even when they are known. In the study of human cells, a similar agnosticism occurs with respect to the objective function, but this disinterest is not justifiable. The choice of objective function has a profound impact on simulation results, yet strangely, it does not usually receive much attention, although historically there have been notable exceptions [163, 327].

In the study of microorganisms, there is agreement in literature that cell growth is an appropriate choice for an objective function (see e.g. [328–330]), and to an extent, this also applies to cancer (e.g. [160, 161]). However, healthy cells from multicellular organisms are a different matter, on which the jury is still out. For instance, Gille et al. tested a liver model with 442 different objectives, none of which amounted to biomass production [157], and Shlomi et al. predicted metabolic capabilities based on gene expression data, forgoing the use of an objective function altogether [331].

Alternative types of objective functions have been proposed (see e.g. [110, 149, 332]), but none seem to appeal to researchers on an intuitive level the way biomass production [163] does. Regardless, once one decides that biomass production is the objective function to go with, there is a second issue: what is "biomass" made of? Classical biochemistry has been used to answer that question for various organisms. For *Escherichia coli* [333], *Lactococcus lactis* [328], and *Lactobacillus plantarum* [329], the biomass composition was determined directly, whereas for *Saccharomyces cerevisiae*, the commonly used "Yeast 5" [330] and iTO977 [334] models contain biomass equations that are based on information from literature. In yeast, it is known that growth conditions (carbon limitation, nitrogen limitation) have a significant impact on biomass composition, and in the two latter references, the authors employ computational methods (e.g. sensitivity analysis [334]) to circumvent this issue.

The situation is more complex for multicellular organisms. Healthy tissue does not strive (solely) for the production of *more* healthy tissue, and the composition of human biomass is expected to vary by tissue type [164]. Also, acquiring sufficient biological material of which to determine the biomass composition can be problematic. The study of cancer cells does not suffer from these limitations, as cancer cell lines can be cultured *in vitro* and few people object to parting with primary tumor cells. In addition, their hyperproliferative

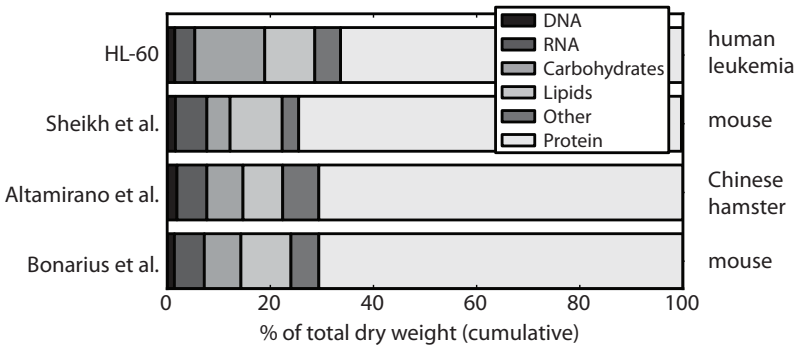


Figure 6.1: The human biomass composition described in this chapter (top), along with three popular biomass compositions from literature. While all three other biomass equations are widely used in genome-scale modeling, none of them is derived solely from human data.

phenotype [14] justifies the use of biomass production as an objective function in the study of their metabolic properties. But these can hardly be called new insights. In 1996, mouse hybridoma cells were studied using a flux balance analysis (FBA) approach [46], and their biomass composition was determined as a matter of course. A similar study, also on murine hybridoma cells, was performed about a decade later by Sheikh et al. [335], and Altamirano et al. determined the biomass composition of Chinese hamster ovary two years later [336] (Figure 6.1).

In this light, it is highly surprising that so far, no genome-scale study of human metabolism has included an experimentally determined biomass composition. The "Recon 1" reconstruction [155] does not include a biomass reaction. Derivative works add to this model a biomass reaction that is based on literature from a variety of eukaryote species [161], or refer to the "pluriformity of human biomass" and omit it altogether [158]. "Recon 2", the successor to Recon 1, does include a biomass reaction, but this reaction is based directly on the murine hybridoma work of Bonarius et al. [46, 50] and therefore is of limited value to the study of human metabolism (we also discovered some other technical shortcomings, which we will discuss in later sections).

The above illustrates that thus far, in spite of some notable exceptions [157], many FBA-derived conclusions about human metabolism [164] and disease [160, 161] have ultimately been based on arcane combinations of mouse, rat, dog, and human biomass samples. Therefore, we set out to experimentally determine the first (fully) human biomass composition, and to examine the

effect of this novel biomass equation on modeling results. Because there can be all sorts of differences between different cell types, we did not aim to determine the be-all and end-all of human biomass compositions. Instead, we focused on the HL-60 leukemia cell line [337], which has several properties that make it especially suitable for this kind of study. For instance, (neutrophilic) HL-60 cells can be induced to differentiate into granulocytes by adding specific chemicals to the culture medium [337], and healthy counterparts to the cancer cells can be easily and safely acquired. Most importantly, however, the cells can be grown in suspension, which makes it easier to accumulate the required amount of material for our biomass determination.

So what is "biomass" made of? Broadly, it consists of five major fractions. These are protein, lipids, carbohydrates, RNA, and DNA. In addition, there is a category of "other" contributors to biomass, which is usually taken to include ions and vitamins [163]. Often, the exact contribution of these other components to biomass is difficult to determine, but they are sometimes explicitly included as biomass components to make sure the pathways leading up to them carry a (small) flux, mimicking the biological situation. Biomass equations typically also include an energy requirement in the form of ATP hydrolysis. This represents the part of the (measured) energy expenditure that is associated with maintenance and/or the production of new cells, but that is not accounted for by known anabolic processes [329, 338].

For most of these main categories, there are assays that enable their quantification. The exceptions are "other", which by definition cannot itself be quantified, and the energy requirement, which is usually taken from literature [338].

The subcomposition of the protein, lipid, carbohydrate, RNA, and DNA components is then determined in various ways. The contributions of the (deoxy)nucleotides to RNA and DNA follow from their prevalence in ribosomal RNA and genomic DNA, respectively. Likewise, the amino acids that make up the protein fraction can be inferred from protein sequences, but as protein expression patterns are known to vary greatly across cells and cell types, the preferred method is to experimentally quantify them using gas or liquid chromatographic methods. For the carbohydrate fraction, a (phospho-)hexose is usually selected as a representative for glycogen and included in the biomass equation, after quantifying total hexoses/pentoses in biomass. Models of human cells therefore necessarily overlook their inclusion in glycoproteins, but since these cannot yet be accurately modeled, it does not make much sense to include them in such detail anyway [339]. The lipid subcomposition is often

estimated or taken from literature (if it is modeled in any detail at all), as the modest contributions of its subcomponents to total biomass are usually not deemed sufficient to warrant research into their quantities.

In this chapter, we describe the first fully human biomass composition, which we derived from HL-60 cells. Insertion of this biomass reaction into Recon 2.1 [339] uncovered some fundamental issues with the currently most popular model of human metabolism, including elemental imbalances. After remedying those, an evaluation of the effects of our more complex new biomass equation on modeling results showed that reactions from various pathways in the underlying network, which had laid "dormant" before, were now active. Our analysis also indicated the need for several essential fatty acids to be supplied in the growth medium. Although the standard specification of the medium does not include these compounds, we experimentally confirmed their presence in complete, serum-containing growth medium. The data and improved human metabolic model that result from this project are made available to the community, and represent what we hope to be an important step on the path to a full understanding of human molecular physiology.

Results

The HL-60 biomass composition

To determine the biomass composition of HL-60 leukemia cells, we first experimentally assessed the degree to which each major component (DNA, RNA, lipids, carbohydrates, protein) contributes to total dry weight. We then experimentally determined the subcomposition of the lipid and protein fractions, estimated the DNA and RNA subcompositions from literature, and appointed glucose as a representative of the carbohydrate fraction. For the "other" category, of which the subcomponents are uncertain (and if they are known, require individual assays for each subcomponent to measure), we assumed a fraction of 5% of total biomass, which is in line with previous estimations (Figure 6.1). The total lipid fraction, whose direct measurement proved elusive, was estimated by subtracting the other fractions from one. The resulting HL-60 biomass composition is shown in Table 6.1 and Figure 6.2, along with the biomass reaction that is included in the Recon 2 reconstruction for comparison [50].

Table 6.1 shows that to an extent, the contributions of the main biomass fractions are comparable between the HL-60 biomass equation and the one

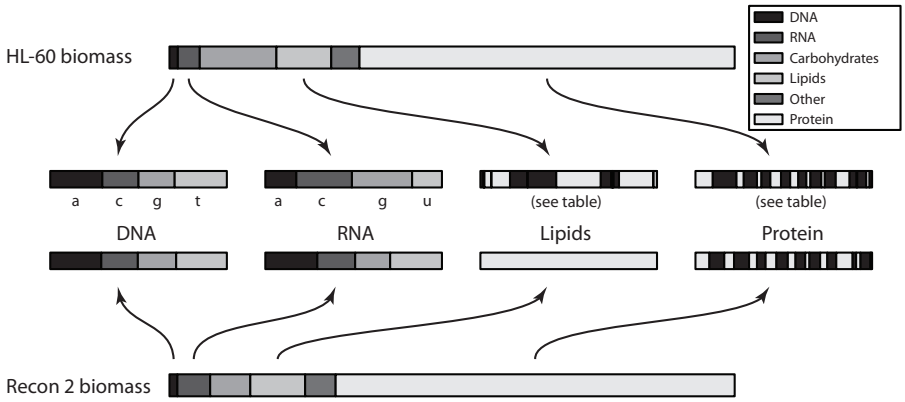


Figure 6.2: Graphical comparison of the new biomass composition (top) with the one that is included in Recon 2 (bottom). For the RNA subcomposition, the differences are likely due to the sequence source used (i.e. in our case, rRNA; for Recon 2, it is unspecified but likely based on the full genome sequence [50, 124]). The protein subcomposition in our biomass equation was determined experimentally, which (partially) explains the slight differences between the two bars. Although the individual amino acids are not annotated in the figure, the order and coloring is the same. Since the subcomponents of DNA, RNA, and protein are the same in both models, these differences are unlikely to have profound consequences for modeling results. However, the new lipid subcomposition is clearly more complex, especially when the R-group components in the Recon 2 biomass equation are disregarded, as shown.

in Recon 2. However, a disparity is apparent when the subcompositions of the protein and lipids fractions are considered (Tables 6.2 and 6.3). For the amino acid subcomposition, this could be due to the cell type (or organism) under study, but for the lipid fraction, the difference is a product of our more extensive subcharacterization.

	Fraction of biomass (%)	
Component	HL-60	Recon 2
Total protein	66.38	70.6
Total lipids	9.7	9.7
RNA	3.88	5.8
DNA	1.48	1.4
Carbohydrates	13.56	7.1
Other	5.0	5.4

Table 6.1: Biomass composition of HL-60 cells, compared to the biomass composition used in Recon 2 [46]. The values represent percentages of total dry weight. Details of the conversion from weight fractions to stoichiometric coefficients are given in Materials and Methods.

Component	Amino acid subcomposition (mmol of AA per g protein)	
	HL-60	Recon 2
Alanine	0.92	0.53
Arginine	0.48	0.36
Asparagine	0.43	0.43
Aspartic acid	0.43	0.81
Cysteine	0.18	0.57
Glutamine	0.58	0.22
Glutamic acid	0.58	0.21
Glycine	1.26	0.48
Histidine	0.37	0.75
Isoleucine	0.24	0.06
Methionine	0.22	0.49
Phenylalanine	0.19	0.17
Proline	0.49	0.54
Serine	0.57	0.74
Threonine	0.47	0.45
Tryptophan	0.13	0.48
Tyrosine	0.03	0.02
Valine	0.39	0.38

Table 6.2: Amino acid subcomposition of HL-60 cells (determined in this chapter) and the biomass equation in the Recon 2 model. Values represent millimoles of each amino acid per gram of protein in the biomass equation.

Towards a new biomass equation

A set of measurements does not a biomass equation make. Even after we converted the measured data to a biomass equation (see Materials and Methods for details), inserting it into Recon 2 did not make for a working model.

Recon 2 [50] is currently the latest, most comprehensive, and most popular reconstruction of human metabolism. Following its initial release, it was soon revised by its authors to address some problems with its use, including carbon imbalances (i.e. reactions with differing numbers of carbon atoms on the reactant and product sides). The resulting reconstruction is known as version 2.03, but it was only made available in a Matlab-based format, severely limiting its accessibility.

More importantly, both Recon 2 and Recon 2.03 are reconstructions rather than models. This means that although they look and behave like stoichiometric models, they should be interpreted as curated collections of reaction information that can be used as starting points for building models. A model, then, uses this network to simulate biological behavior; for instance, by specifying

Component	Species ID	% w/w	Stoich. coefficient
C10	M_dcacoa_c	0.01	0.0003
C11	M_undcoa_m	0.19	0.0101
C12	M_ddcacoa_c	0.08	0.0038
C13	M_tridcoa_m	0.01	0.0004
C14:0	M_tdcoa_c	3.82	0.1325
C15	M_ptdcacoa_m	0.22	0.0092
C16:0	M_pmt_c	9.31	0.3629
C16:1	M_hd_c	8.76	0.3444
C17	M_hpdca_c	0.18	0.0067
C18	M_st_c	15.97	0.5615
C18:1 (cis)	M_ocdcea_c	24.80	0.8779
C18:1 (trans)	M_elaid_c	5.86	0.2074
C18:2	M_lnlc_c	0.63	0.0226
C18:3 (α-linolenate)	M_lnlnca_c	0.07	0.0024
C18:3 (γ-linolenate)	M_lnlncg_c	0.07	0.0027
C20	M_arach_c	0.10	0.0031
C20:3	M_dlnlcg_c	0.44	0.0144
C20:4	M_eicostet_c	3.25	0.1067
C20:5 (EPA)	M_5eipencoa_m	20.61	0.6814
C22	M_docosac_c	0.03	0.0009
C22:6	M_crvnc_c	2.17	0.0661
C24:1	M_nrvnc_c	0.05	0.0013
Other fatty acids	—	0.88	—
Cholesterol	M_chsterol_c	2.50	0.0647
Other sterols	—	BDL	—

Table 6.3: Subcomposition of the lipid content of HL-60. For each lipid component, its contribution to the total lipid fraction is shown as a weight percentage. We also show the species ID in the model that corresponds to each lipid subcomponent, as well as its stoichiometric coefficient in the lipids sub-reaction (= mmol per gram of lipids). Recon 2 is not shown for comparison here: after removing the metabolites that only participate in unbalanced reactions, it only includes cholesterol, and no fatty acids. BDL: below detection limit.

flux constraints based on external conditions like the nutrient concentrations in the medium, or the availability of certain reactions. Being a reconstruction rather than a model, when Recon 2 is simulated as-is, it produces biomass, but upon closer inspection, it turns out that all biomass components can be freely imported into the cell, and there is no need for the rest of metabolism to work.

Elemental balancing

Soon after its initial release, one of the authors of the original Recon 2 reconstruction proposed Recon 2.1 as a carbon-balanced successor to Recon 2 [339]. Version 2.1 fixes the carbon balance and some other issues Recon 2 suffered

from, but the other elements (H, O, N, P, S) may still have imbalances due to the use of generic metabolites (metabolites with an R-group that symbolizes an unspecified tail of carbohydrates, acyls, etc.). Smallbone, the author of Recon 2.1, realized this and proposed Recon 2.1x as a solution to this in the same paper. He resolved the issue by enumerating (almost) all possible instantiations of generic metabolites, but while the resulting reconstruction is elementally balanced, at a whopping 71159 reactions, it is also too large to handle using current-day software/hardware. Because we wanted to be able to simulate the model after adding the new biomass equation, we had to find another solution.

Since the elemental composition of each metabolite is provided in the metadata of the reconstruction, we can identify imbalanced reactions in Recon 2.1 by comparing the elemental composition of each reaction's left-hand side and right-hand side. We detected 339 unbalanced reactions using this method, and we disabled them in the model. Importantly, this operation revealed that the Recon 2 biomass equation included seven metabolites with R-groups. By removing them from the biomass reaction (instead of removing the biomass reaction altogether; see Materials and Methods for details), we ensured we could compare balanced models with functional biomass equations.

Medium specifications imposed by the biomass equation

Recon 2.1 has 1290 exchange reactions, which allow the model to consume or produce 645 metabolites (import and export are split into separate reactions). By default, they can all carry flux, which essentially means any biomass equation will "work", because all components can just be imported. To begin to make statements about the modeling implications of using a different biomass equation, it is important to first determine which metabolites are essential, meaning they must be supplied externally, and which metabolites can be produced by the cell from these external inputs.

Using a mixed-integer linear programming approach (see Materials and Methods), we determined the smallest possible set of metabolites that is needed to produce biomass, given Recon 2.1 with either its original biomass equation (without R-group metabolites) or our novel one. As expected, the resulting minimal media (Table 6.4) include essential amino acids, a source of phosphate, and, in the case of the new biomass equation, a subset of the fatty acids in the lipid subcomposition. We confirmed that the other lipids could either be produced from these fatty acids, or from other metabolites.

The predicted minimal media do not include ions or vitamins, as they are not included in the old or new biomass equation.

Both minimal media also include the peptide kinetensin, which is an uncommon find in this context. Kinetensin is a peptide whose amino acid sequence is Ile-Ala-Arg-Arg-His-Pro-Tyr-Phe-Leu [340]. This makes it an efficient source of essential amino acids, as importing kinetensin means that phenylalanine and leucine need not be imported separately anymore. It is also a very rich carbon and nitrogen source. Therefore, the presence of kinetensin in the computationally determined minimal media is a consequence of the method used to determine said media, rather than of a biological property of the metabolic network.

Component	Recon 2	New biomass eqn
Histidine	✓	✓
Isoleucine	✓	✓
Lysine	✓	✓
Methionine	✓	✓
Threonine	✓	✓
Tryptophan	✓	✓
Valine	✓	✓
Kinetensin	✓	✓
Phosphate	✓	✓
Margaric acid (C17)		✓
Linoleic acid (C18:2)		✓
α-linolenate (C18:3)		✓
Behenic acid (C22)		✓
Nervonic acid (C24:1)		✓

Table 6.4: Minimal media requirements of the Recon 2 model with its original biomass reaction and with the new biomass equation. If a metabolite is in the set of minimal media requirements for a model with a particular biomass equation, this is indicated by a check mark. The new biomass equation adds requirements on lipid intake, which were absent in the original model.

Comparing simulation outcomes using both biomass equations

Although we can also simulate a kinetensin-free medium by adding the other essential amino acids, since both minimal media contain the peptide, we first compared the implications of each biomass equation on the metabolic network that lies between the consumption of metabolites and the production of biomass. Since we already know that the model will grow when supplied with its minimal medium, we can proceed directly to a flux variability analy-

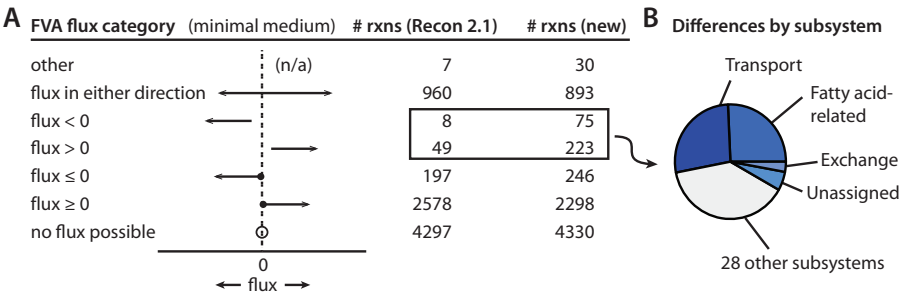


Figure 6.3: Summary of FVA results when simulating a minimal medium and using either the Recon 2 biomass equation or our new one. (A) Based on their minimum and maximum flux values, reactions can be categorized into those that cannot carry flux, those that can optionally carry flux in one direction, those that *must* carry flux in a certain direction, those that can (optionally) carry flux in either direction, and reactions for which the analysis is infeasible of nonsensical. For each of these classes, the number of reactions that belongs to them in both versions of the model is shown. (B) For those reactions that *must* carry a nonzero flux when using one biomass equation, but not when using the other, we evaluated the subsystem this reaction was annotated with. The model with the new biomass equation requires almost five times as many reactions to be active compared to the old biomass composition, and fatty acid oxidation and synthesis and intracellular transport reactions cover about half of these differences.

sis (FVA) and find the minimum and maximum flux for each reaction, given optimal biomass production. For each reaction, this now gives an indication of its importance to the network: can it carry flux in the first place? If it can, in which direction, and is this mandatory? Maybe it can carry flux in both directions (Figure 6.3A)?

We categorized reactions based on these properties. Among the resulting categories, the most interesting ones are those containing reactions that *must* carry flux in some direction. Interestingly, this number is much larger when using the new biomass equation, compared to the old one (Figure 6.3A, highlighted values).

We examined the annotations of these reactions, and unsurprisingly, found that the subsystems associated with fatty acid oxidation and synthesis were strongly represented among the reactions that must carry flux when using the new biomass equation, but not when using the old one (Figure 6.3B). Therefore, we can conclude that using the more complex new biomass equation has an impact on the flux distribution of the underlying network, particularly in subsystems related to exchange, transport, and fatty acid oxidation and synthesis.

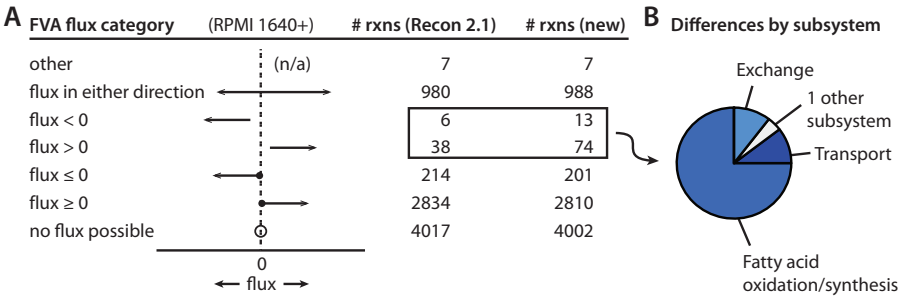


Figure 6.4: Summary of FVA results when simulating RPMI 1640 medium, augmented with essential fatty acids (RPMI 1640+) and using either the Recon 2 biomass equation or our new one. (A) Based on their minimum and maximum flux values, reactions can be categorized into those that cannot carry flux, those that can optionally carry flux in one direction, those that *must* carry flux in a certain direction, those that can (optionally) carry flux in either direction, and reactions for which the analysis is infeasible or nonsensical. For each of these classes, the number of reactions that belongs to them in both versions of the model is shown. (B) For those reactions that *must* carry a nonzero flux when using one biomass equation, but not when using the other, we evaluated the subsystem this reaction was annotated with. Increased use of fatty acid oxidation and synthesis pathways, due to the increased complexity of the lipid fraction in our new biomass equation, explains the vast majority of these differences.

Simulating a more realistic medium reduces the number of differentially activated subsystems

As we stated above, simulations with kinetensin as a medium component are unlikely to represent a realistic situation. HL-60 cells are normally cultured in RPMI 1640 medium, supplemented with fetal bovine serum. While it is not known exactly what is in the serum part of the medium, the base formulation of RPMI 1640 is common knowledge: in a nutshell, it contains amino acids (including all essential ones), glucose, and various vitamins, cofactors, and salts.

We repeated the flux variability analysis described above with RPMI 1640 instead of an artificial minimal medium. The fatty acid requirements our new biomass equation imposes were met by supplementing the simulated RPMI 1640 with five fatty acids that we determined as essential to growth of the (new) model (Table 6.4).

When using our modified RPMI 1640 medium ("RPMI 1640+"), the results of an FVA comparison between models with the old and new biomass equations are largely comparable to the situation with the computationally determined minimal medium (Figure 6.4A). As expected, the more complete

(base) formulation of RPMI 1640 translates to an overall need for fewer flux-carrying reactions.

However, the reactions that *must* carry flux when using the new biomass equation but not when using the old one show a much clearer distribution across annotated subsystems. The vast majority of the affected reactions is now associated with fatty acid oxidation or synthesis, with a minor role for intracellular transporters (Figure 6.4B).

Moreover, the number of different subsystems the affected reactions are classified in when using the RPMI 1640+ medium is vastly smaller than in the simulations with a minimal, kinetensin-containing medium. In the former case, only a single subsystem was not exchange-, transport-, or fatty acid-related (the exception was "valine, leucine and isoleucine metabolism", which contained 2 reactions), whereas with the minimal medium 28 other subsystems contained 92 reactions.

Because of their essentiality for growth in simulations, we hypothesized that the essential fatty acids we added to the simulated medium are included in real-world growth medium through the addition of serum. Indeed, when we sampled complete culture medium and determined its lipid subcomposition, we found substantial concentrations of each of the five essential fatty acids. The amounts of the fatty acids in the medium, as a weight percentage of total lipids, ranged from 22% to 83% of the relative amount in the HL-60 samples, indicating the cells selectively that them up during the course of their growth. This finding justifies their inclusion in these simulations, but also shows the potential of a biomass-based approach for determining the molecular requirements of growth.

Combined with the results of our simulations using the minimal media, this result illustrates the importance of a detailed biomass equation as well as a realistically defined medium: it is easy to overlook entire pathways if there is no demand for them to be functional, and conversely, an overly minimal medium may lose touch with biology by requiring unrealistic activity in exotic pathways.

Discussion

Biomass requires balance

An understanding of cell growth and hyperproliferative disease is contingent on an understanding of what cells are made of. In an effort to bring this goal

closer, we have presented a more detailed human biomass composition than any other that we were aware of. Perhaps unexpectedly, its mere inclusion in Recon 2.1 already revealed places in which our current model of metabolism diverges from physiological reality.

A major culprit is the elemental imbalance in the Recon 2 reconstruction. Although this issue has been identified before and a solution proposed [339], our solution of removing all unbalanced reactions yields the first Recon 2-derived model that can be used in practice.

Unfortunately, the imbalance also extended to Recon 2's biomass equation, as it too includes R-group metabolites. This property seems to have been introduced in a Recon 1-based model published by Folger et al. [160], and carried over to later versions of the human metabolic reconstruction [50]. The specific composition of these glycophospholipid biomass components is an important piece of information that genome-scale models would clearly benefit from, but as Smallbone demonstrated [339], the combinatorial aspect their specification entails is an adversary to be reckoned with: the almost-complete specification of these compounds grew the model to ~71000 reactions.

Even so, we expect that further increasing our understanding of the specifics of human biomass is a great way to improve the predictive power of human metabolic models, by shedding light on the state of various pathways that would go unused if these biomass components are ignored.

Growing awareness of medium in simulations

Besides biomass production, an output variable, the other major determinant of the outcome of model simulations is which metabolites the model can consume, on the input side. In other words, to which medium components should the simulated cells have access?

In microorganisms, this question is commonly asked, as FBA is often used to either optimize growth or the production of a biotechnologically interesting metabolite given some environment [329,341]. In contrast, for reconstructions and models of human metabolism, a variety of approaches are found in recent literature.

Recon 1, 2, and 2.1 are reconstructions rather than models, so they contain no predefined medium, and all exchange reactions are available [50,155,339]. Bordbar et al. produced cell-type specific reconstructions that were studied with respect to each other, but not so much with respect to the environment [164]. Nevertheless, although the authors do not describe the medium in detail, the number of exchange reactions in the models is kept low and seems

well thought out. Interestingly, they include exchange reactions for many of the fatty acids our analysis identified as essential for biomass production in Recon 2.1.

Shlomi et al., on the other hand, took a structure-driven approach, and predicted exchange fluxes based on the activity of enzymes in the network [331]. Ågren et al. automatically constructed tissue-specific models, similar to the Recon 2 approach, but did not include a biomass equation and therefore forwent the medium specification as well [158]. Jain et al., in constrast, focused primarily on the production and consumption of medium components [45]. The resulting data set spans sixty cell lines, and is of great value in the context of tissue-specific modeling. The leukocyte-based biomass equation we present here can be viewed as a contribution in the context of this tissue-specific future; however, cell type-specific models based on Recon 2.1 have yet to be developed. Ideally, these will include cell type-specific biomass equations.

Folger et al. also specify the medium they simulated [160]. The model presented in their paper, uses RPMI 1640 (as well as the mouse hybridoma biomas equation that found its way into Recon 2 [46, 50]), but in contrast to our approach, it only specifies the components of the medium and not their quantities relative to each other.

With our analysis, we aimed for a hybrid computational/experimental approach. By first looking at the model structure to discover what the model needs to grow, and then comparing this to the metabolites we know are in the medium, we could deduce which other, unseen medium components exist, and confirm this experimentally. We expect that medium and biomass-based approaches will be of increasing importance in the future, as models of human metabolism become more detailed and therefore more dependent on knowledge about their exchanges with the environment.

Going from reconstructions to cell-type specific models

The biomass composition we present here was determined in the HL-60 promyelocytic leukemia cell line [337]. Although we focused on this as a primary objective, an obvious direction for future work is to compare these cells to, say, carcinomas, differentiated HL-60 cells, or healthy leukocyte samples. The experimental challenge this presents is quite surmountable, but to be able to properly compare cell types, cell type-specific models must also be developed. Various methodologies have been proposed to do this [50, 158, 164, 342]. Most are somehow based on gene expression data; however, there is no consensus as to which method is the best.

Whichever methodology ends up finding broad adoption by the community, we expect that the structure of a cell type-specific network, combined with a demand-driven approach like the one we present will yield valuable insights in the metabolic properties of various tissues. Data like the biomass composition we present here, as well as the metabolite profiles produced by Jain et al. [45], are instrumental in this endeavor, as we will sooner or later reach the limitations of what can be done using structural analyses. In the end, there is no substitute for biochemical experimental knowledge, just like the structural findings we present here would have been meaningless without a fully human biomass equation.

It is evident that the first fully human biomass equation we present here is a resource upon which the community can build in a wide variety of future projects. We have used it to check, improve, and instantiate Recon 2.1, the current state of the art in modeling human metabolism, and have made the resulting model available for download at http://www.joostboele.nl/thesis/. Together, the two resources can be built upon to optimize the predictive power of human genome-scale metabolic modeling approaches, and improve our understanding of human molecular physiology.

Materials and Methods

Determination of the HL-60 biomass composition

Cell culture, cell count, and dry weight determination

HL-60 human promyelocytic leukemia cells were obtained from the European Collection of Cell Cultures, and maintained in RPMI-1640 medium (PAA) with 8% fetal bovine serum (PAA) and 1% Penicillin/Streptomycin (PAA).

The cell density of cultures was determined using a Bio-Rad TC-20 Cell Counter. To determine the amount of dry weight per volume of culture, a known volume of sample was filtered using 1μm nitrocellulose filters (Whatman), and the weight of the filters before and after filtration compared using a Sartorius 4503 Micro scale.

Total protein content and protein hydrolysis

The cells were lysed with 1% Triton X-100, 50 mM HEPES, pH 7.4, 150 mM NaCl, 1.5 mM $MgCl_2$, 1mM EGTA, 100 mM NaF, 10 mM Na pyrophosphate, 1 mM Na_3VO_4 and 10% glycerol, containing freshly added protease and phosphatase inhibitors. Total protein content was then determined colorimetrically using the bicinchoninic acid (BCA) assay (Pierce).

The contribution of the individual amino acids was determined by first hydrolysing the protein by keeping the samples in 6M HCl at 115°C for 24h, and then measuring the amino acid concentrations in the hydrolyzed samples using the GC/MS method described below.

Amino acid subcomposition GC/MS analysis

Extraction and derivatization of the amino acids were performed with a commercially available kit according to the instructions of the manufacturer [343]. Briefly, 100μL of sample including ^{13}C internal standards was mixed with Reagent 1 in order to adjust pH and add the internal standard (norvaline). The mixture was slowly passed through sorbent tip and washed with Reagent 2. Then, 200μL of freshly prepared eluting medium was used to elute sample and sorbent from the tip. In the next step, derivatization reagent (Reagent 4) was added to the mixture and the sample was vigorously vortexed. Finally, 100μL of Reagent 5 containing isooctane was added to sample in order to separate different phases after vortexing, and we collected the top (organic) phase.

The derivatized samples were analyzed with a GC-2010 gas chromatograph (Shimadzu), equipped with injector AOC-20i (Shimadzu) and autosampler AOC-20s (Shimadzu) which was coupled to a GCMS-QP2010 Plus with an electron ionization source (Shimadzu). One microliter of derivatized sample was injected onto a ZB-AAA column (10 m x 0.25 mm ID with a film thickness of 0.25 μm). The injection port, equipped with a GC liner for Shimadzu 17B (deactivated, Phenomenex), was set to split ratio 30 and held at a temperature of 250°C. The column oven temperature was initially set at 110°C and then raised to 320°C at a rate of 30°C/min. Helium was used as carrier gas with a gas flow of 2 mL/min. The interface (GC to MS) was at temperature 320°C and the ion source at 240°C. Electron ionization was operated with 70 eV. For quantitative measurements, the MS was used in selected ion monitoring (SIM) mode. All analyses were performed in duplicate. ^{13}C-^{15}N labeled amino acid standards were added as internal standard to both samples and calibration standards for IDMS-based quantification.

The results included concentrations for most amino acids, but did not include arginine, cysteine, tryptophan. For these amino acids, their relative contributions to total protein were estimated based on protein sequence data. Also, due to the HCl treatment, asparagine in the original samples was measured as aspartic acid, and glutamine as glutamic acid, so the measured concentration of aspartic acid was split evenly between asparagine and aspartic acid, and of glutamic acid between glutamine and glutamic acid.

Carbohydrates

Since we are interested in the total of all carbohydrates in the cell, commonly used enzymatic assays are too specific for our purposes. Instead, we determined the car-

bohydrate fraction using the colorimetric phenol-sulphuric acid method, similarly to previous biomass composition determinations in prokaryotes [329]. We scaled down the assay to economize reagents, used the reagent proportions described by Masuko et al. [344], and performed the assay in glass labware to minimize the background signal. We measured the samples at 488 nm; because the absorption spectra of various carbohydrates overlap, measuring at this wavelength gives a fairly complete overview of the total abundance of carbohydrates in the sample.

RNA and DNA

RNA was isolated from samples using the acid phenol/chloroform extraction protocol protocol, treated with DNAse-I and then quantified using a NanoDrop ND-1000 (Thermo Scientific), which yields an RNA concentration in ng/μl. The RNA subcomposition was then determined by assessing the contribution of each nucleotide (A, C, G, and U) to known human ribosomal RNA sequences (28S, 18S, 5.8S, 5S; accession numbers NR_003287.2, NR_003286.2, NR_003285.2, NR_023379.1, respectively) relative to the total number of nucleotides in these sequences. We used ribosomal RNA, as this class of RNA composes the vast majority of the total RNA fraction in proliferating cells.

DNA was isolated using the phenol/chloroform extraction protocol, and then quantified using a NanoDrop ND-1000 (Thermo Scientific), which yields a DNA concentration in ng/μl. The DNA subcomposition was then determined by assessing the contribution of each nucleotide (A, C, G, and T) to the hg19 assembly of the human genome sequence, and averaging the contributions of A and T, and C and G, to account for basepairing.

For both RNA and DNA, of each (deoxy)nucleotide, the triphosphate form is included on the reactants side of the biomass equation. The diphosphate that is released when a DNA or RNA molecule is elongated by one nucleotide is included on the product side. When determining the molecular weight of DNA and RNA in the biomass composition, the molecular weight of all nucleotides was reduced by the weight of one diphosphate.

Total lipid content and lipid subcomposition

Due to difficulties directly measuring the lipid fraction in total biomass, it was calculated by subtracting all other fractions from one. This yielded a lipid fraction of 9.7%, which is consistent with earlier biomass determinations in mammals. While taking samples for the lipid assays, we took care to avoid plastic disposables to minimize leeching of lipid compounds to/from them.

To determine the lipid subcomposition, we spun down $\sim 2 \cdot 10^6$ cells and washed the pellet twice with PBS. The sample was then diluted 4x with ice cold methanol and stored at -80°C until use.

The fatty acid and sterol quantification was performed as follows. First, 2.5 μL of sample was mixed with 180 μL of MeOH and 2 0μL of water. 70 μL of 6 N HCl was added, after which 10 μL internal standard (IS) was added. Lipid extraction was carried out by adding 1 mL of *n*-hexane. After vigorously shaking the sample for 30 s, the *n*-hexane layer was transferred to a 1.5 mL glass vial and concentrated to dryness under a gentle stream of nitrogen. The procedure was repeated and the organic extracts were combined.

Dried samples were derivatized using 25 μL of M*t*BSTFA for 10 min at room temperature. Subsequently, 25 μL of BSTFA (1% TMCS) and 2.5 μL of pyridine were added and the sample was heated for 15 min to 50°C. Next 947.5 μL of *n*-hexane, containing 10 μg/mL octadecane (C18) as system monitoring component, was added.

A Scion TQ GC-MS system from Bruker (Bremen, Germany), equipped with a Bruker BR5-MS 15 m × 0.25 mm × 0.25 μm column was used. The injector was held at 280°C. The injection volume was 1 μL for selected ion monitoring (SIM) and 2 μL for isotopologue analysis (splitless). The temperature program was kept at 90°C for 0.5 min, then ramped with 30°C/min to 180°C, then to 250°C with 10°C/min, then to 266°C with 2°C/min, and finally to 300°C with 120°C/min, kept for 2 min. Helium (99.999%) was used as carrier gas at a flow rate of 1.2 mL/min. All analyses were carried out using EI at 70 eV. For validation and quantification purposes, SIM analysis was applied. For isotopologue analysis, a scan window from the monoisotopic SIM value up to +15 *m/z* units was measured in a window of ± 0.3 min around the retention time of the target component.

Composing the biomass equation

To compose a biomass equation for use in models that use (milli)moles as a unit, and not "percentage of dry weight" (which would indeed be of very limited use in genome-scale models), we processed the data as follows.

To reflect that the biomass equation is initially determined by quantifying its components in broad categories, the biomass equation itself simply specifies these categories as fictitious metabolites, with their stoichiometric coefficients equal to their respective contribution to a gram of biomass. Thus, the biomass equation simply states that a gram of biomass consists of e.g. 0.036 grams of RNA, etc.

For this to work in a model that uses millimoles as a primary unit, a transformation to this unit is required, and this takes place at the level of the broadly categorized biomass components. As the subcomponents of these components (and their relative quantities and molecular weights) are known, we can calculate the molecular weight of 1 (milli)mole of this component. For instance, RNA is composed of adenine, cytosine, guanine, and uridine, and their relative contributions to the RNA fraction are known. Thus, we can calculate the weight of "1 RNA" by multiplying these relative contributions by the molecular weight of each nucleotide. By scaling the result of this step to 1

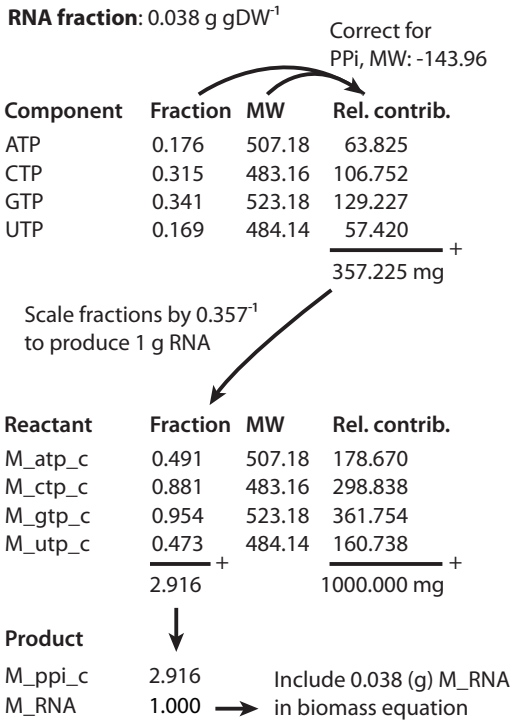


Figure 6.5: Conversion of the (known) composition of a biomass fraction to one gram of that fraction for inclusion in the biomass reaction. In the case of RNA, for each nucleotide, we first multiply its relative contribution to total RNA (in this case, determined from rRNA sequence) by its molecular weight. Since polymerization of nucleotides into RNA results in a pyrophosphate (PPi) being expelled, the molecular weight of PPi must be subtracted from the nucleotide's before multiplication. The sum of these products is the weight of one fictitious millimole of RNA. When all fractions are scaled by the inverse of this number and the calculation is repeated, the result will be 1 gram of RNA. In our case, we know we need 0.038 gram of RNA per gram of biomass, so we can simply include 0.038 "RNA" in the biomass reaction. The subcompositions of the other biomass components were converted and modeled in a similar fashion.

gram, we can intuitively add fractions of this "gram of RNA" to total biomass, and do the same for all other components (Figure 6.5).

For biomass components that are polymers (proteins are polymers of amino acids, RNA and DNA are polymers of NTPs and dNTPs, respectively), the production of these polymers results in the exclusion of water or diphosphate from the participating monomers. As a result, the molecular weight of a monomer is reduced by the weight of the excluded parts once it is included in the polymer, and we took this into account when performing the transformation from biomass components to biomass.

Including the energy requirements for maintenance and proliferation

To account for the energy cells expend for maintenance and for the anabolism of biomass components, biomass equations typically include an ATP requirement. The 1969 paper on this topic by Kilburn et al. is a prominent resource for this information [338], as evidenced by its pervasive use in genome-scale modeling. Kilburn et al. report an ATP requirement of 1.6 x 10^{-11} mole per cell for proliferating cells, which includes both maintenance and proliferation. By multiplying this number by the cell count per mL of culture, and then dividing by the measured dry weight (in grams) per mL of culture, we arrive at an ATP requirement of 154 mmol gDW^{-1}. This figure, although it may seem large at first, is consistent with what is used in previous biomass equations, as the ratios between it and the stoichiometric coefficients of other biomass components are comparable.

Inclusion of the biomass equation in Recon 2.1

We included the new biomass equation in the Recon 2.1 model in the form of five subreactions (DNA, RNA, lipids, protein, and carbohydrates) and one biomass reaction in which they and the ATP requirement for maintenance and proliferation come together. Each subreaction produces a gram of the respective biomass component (e.g. RNA) from its subcomponents (in the case of RNA, ribonucleotides; see Figure 6.5 for an example). This way, the fictitious metabolite ("RNA") can then be included in the final biomass reaction in the desired quantity (in grams).

For the purposes of our study, the model's objective was changed to maximization of flux through the new biomass reaction, and constraints on it and its subreactions were set to make them irreversible. To ensure the more detailed new lipid requirements for biomass productions could be met, five essential fatty acids were additionally specified to be in the medium (that is, the constraints on their exchange reactions were set to allow importation of these lipids). These fatty acids were margaric acid, linoleic acid, α-linolenate, behenic acid, and nervonic acid (Table 6.4). The resulting model, which we used for the remainder of the analyses described in this chapter, comprises 8095 reactions and 9209 metabolites. We have made it available for download at http://www.joostboele.nl/thesis/.

Modeling analyses using the modified human model

To compare the effects of using a new, more detailed biomass equation to using the old one, we took the following approach. All simulations were performed using the Python Simulator for Cellular Systems (PySCeS) [147]. First, we disabled all reactions that were not elementally balanced, and removed seven metabolites from the old biomass equation that had become dead ends because of this process.

Identifier	Description in model	Chemical formula in model
M_clpn_hs_c	Cardiolipin	$C_9H_{16}O_9P_2(RCO_2)_4$
M_pail_hs_c	1-phosphatidyl-1D-myo-inositol(1-)	$C_9H_{16}O_9P(RCO_2)_2$
M_pchol_hs_c	Phosphatidylcholine	$C_8H_{18}NO_4P(RCO_2)_2$
M_pe_hs_c	Phosphatidylethanolamine	$C_5H_{12}NO_4P(RCO_2)_2$
M_pglyc_hs_c	Phosphatidylglycerol(1-)	$C_6H_{12}O_6P(RCO_2)_2$
M_ps_hs_c	Phosphatidylserine	$C_6H_{11}NO_6P(RCO_2)_2$
M_sphmyln_hs_c	Sphingomyelin betaine	$C_{23}H_{48}N_2O_5P(RCO)$

Table 6.5: R-group metabolites in the original Recon 2 biomass equation. As all reactions these metabolites participate in are unbalanced by definition, they were removed from the biomass reaction in order to perform the analyses described in this chapter. Metabolite identifiers, descriptions and chemical formulae are taken from the Recon 2 SBML file. For readability, in the chemical formulae, "FULLRCO2" and "FULLR2CO2" are shortened to "RCO_2", and FULLRCO is shortened to "RCO".

Then, to be able to compare growing models, we determined a minimal medium that would sustain growth with the selected biomass equation as the objective function. We then performed a comparative FVA as described below.

We repeated the entire analysis using RPMI 1640 augmented with five essential fatty acids (RPMI 1640+) as the growth medium, instead of the bare-bones minimal medium we determined earlier.

Removal of elementally unbalanced reactions

Using the `fbc:chemicalFormula` attribute of each species (which in turn derived from the `FORMULA` field in the `notes` nodes of Recon 2) and the stoichiometric coefficients in each reaction, we determined the number of atoms of each element on each side of the reaction. If for any element the number of atoms was not the same on both sides of the reaction, the reaction was flagged as unbalanced and disabled by setting its upper and lower bounds to zero.

We did not disable the original biomass reaction (which strictly speaking was also unbalanced), because we wanted to compare our new biomass equation to it. Instead, we removed from its reactants side the seven R-group metabolites which caused the imbalance, since no non-R-group equivalents to these molecules could be found in the model. The metabolites removed from the biomass equation are shown in Table 6.5. It should be noted that removing these lipids from the biomass equation does not substantially affect the analyses described here or the results displayed in Figures 6.3 and 6.4, as these metabolites can only be produced using unbalanced reactions (i.e., they can be made "out of thin air") or directly imported from the environment (which does not affect the flux distribution, as it bypasses metabolism). The resulting biomass reaction was functional and balanced, albeit slightly leaner than before. In total, 339 (out of 8095) reactions were disabled because of elemental imbalances.

Determination of the minimal medium for a given model

A "minimal medium" is one that has the fewest components. However, in order to properly model this we need to first quantify "fewer". One classical approach, generally applied to internal fluxes, is that of "flux minimization" which assumes that a cell has to produce more protein (enzyme) in order to carry a larger flux through a reaction. Therefore for a cell to be efficient, for any particular optimal objective function (e.g. max growth), there exists a solution which minimizes the maximal flux of each reaction needed to achieve that optimum. Practically, this is implemented by minimizing the sum of absolute fluxes, which allows this to be applied to reversible reactions without bias towards the reverse reaction. Generally, flux minimizations are relatively simple to calculate using a linear program (LP), and the procedure is implemented in most CBM/FBA software.

Modifying this method slightly to only minimize a selected set of transporters (e.g. 300 potential uptake reactions) does result in a minimal medium; however, we observed that the algorithm resulted in a significant number of transporters each carrying a small flux. If one considers that, in general, transporters are large, multimeric protein complexes that are expressed in relative low numbers (relative to metabolic enzymes), and that they are not easily saturated with substrate, then it seems highly unlikely that a "large number of low flux carrying transporters" strategy would be adopted by the cell.

As an alternative, we propose that it is more likely that the cell minimizes the number of active transporters used to achieve a specific objective rather than the combined flux passing through them. We therefore define a minimal medium as a medium which requires the smallest number of active transporters (uptake reactions) needed to satisfy a particular objective (in our case, biomass production).

We can algorithmically generate such a minimal medium by using a mixed integer linear program (MILP). In addition to the standard FBA LP formulation (constraints and flux bounds, cf. Equation 1.7), we introduce binary variables (z_i) to control whether a flux is on (a non-zero flux) or off, and a constraint that fixes the objective to a specific target value ($J_{opt} \geq optarg$). This MILP is described in Equation 6.1 and is analogous to an approach used in [345].

$$\begin{aligned}
&\text{minimize } \textstyle\sum_1^i z_i \\
&\text{such that:} \\
&\begin{cases} NJ = 0 \\ J_{opt} \geq optarg \\ z_i = 0 \rightarrow J_i = 0 \\ lb_n \leq J_n \leq ub_n \\ z_i \in \{0 : 1\} \end{cases}
\end{aligned} \tag{6.1}$$

The solution space of this MILP includes exchange reactions that must always be

part of the minimal medium, and/or sets of exchange reactions that can be substituted for one another. An example of the latter is the requirement for a phosphate source: the cell can consume inorganic phosphate, but ATP and many other phosphate-containing molecules would be adequate as well. In such cases, we selected the metabolite that made the most sense from a biological point of view to be part of the minimal medium.

Comparative flux variability analysis

With a minimal medium determined for models with the old or new biomass equation, we then performed an FVA on each of these models, with as input flux bounds only their respective minimal medium. Based on the results, each reaction was then categorized into one of the following classes: reactions that cannot carry flux under these conditions, reactions with flux ≤ 0, reactions with flux ≥ 0, reactions with flux < 0, reactions with flux > 0, reactions with lower bound < 0 *and* upper bound > 0, and reactions that either returned an infeasible status when performing the FVA or whose minimum and maximum fluxes were moot (e.g. the biomass equations and their subreactions).

For the categories of reactions that must carry flux in one direction or the other (flux < 0 or > 0), we then determined which reactions had this property in the FVA results of one model but not of the other. Of these reactions, we looked up the annotated subsystem in the model, to get an indication of the structural differences between the models with different biomass equations.

Since the minimal medium we determined was indeed quite minimal, and contained a peptide that is rarely on the menu for human cells in culture (kinetensin), we repeated the analysis after setting the exchange reaction constraints to mimick RPMI 1640 medium. To do so, the ingredients of RPMI 1640, as reported by the manufacturer in g/l, were converted to mM, and the essential lipids our earlier analysis identified were added to the simulated medium to ensure both biomass reactions would be supported.

Authors' contributions

JB, PK, MG, and DK performed experiments. JB and BGO analyzed data. JB and BT wrote the chapter, with MG providing helpful comments.

Acknowledgments

The work presented here was carried out on the Dutch national e-infrastructure with the support of SURF Foundation.

Chapter 7

General Discussion

The state of the art in systems biology

The chapters in this thesis tell of microRNAs, of sequencing, signaling, and stoichiometric modeling, of biomass and tools. Despite their obvious differences, these topics have one thing in common: the systems they tie into are relevant to cellular function (or dysfunction). But if we zoom out to look at the bigger picture, they also tie into each other. In this final chapter, we will revisit some of these earlier topics, and view them in the context of systems. . . and cancer.

Let us begin our recapitulation of the subfields of systems biology in this thesis on familiar ground. Since its first application, the analysis of genome-scale metabolic models (GSMMs) has led to valuable insights for a wide variety of species (see Chapter 1 and e.g. [151, 153, 164, 284, 334]). Of course, *Homo sapiens* is among the species for which GSMMs have been reconstructed [50, 155–157], and studies like those by Folger and Shlomi [160, 161] have successfully harnessed existing knowledge to facilitate the generation of new knowledge.

The insights gained from genome-scale modeling include suggestions on how to use new combinations of existing pharmaceuticals to combat cancer [160], but as we described in Chapter 6, to date, all genome-scale modeling approaches to human metabolism have ultimately used a mixture of human and non-human data from various sources to make up the model's final objective: the production of biomass. While the use of biomass as the be-all and end-all of objectives is certainly debatable (see e.g. [163]), there can be no doubt that the availability of any non-composite fully-human biomass equation will enhance our understanding of human metabolism.

When we inserted our new biomass equation into the Recon 2 reconstruction and tried to import or produce all components, we already ran into some issues. Notably, there were elemental imbalances in many reactions in Recon 2, as well as in Recon 2.1. Perhaps surprisingly, these issues were known, but the recently proposed solution, Recon 2.1x, while technically correct, made the reconstruction so large it became unwieldy, and ended up replacing one problem with another [339]. An alternative, crowd-driven curation method was recently proposed by the IMPROVER project team [346], and such ongoing initiatives are more likely to be a future-proof solution to the technical issues we encountered.

Model usability aside, the choice of biomass equation influences modeling results on a more biologically fundamental level. Using a more detailed

biomass equation requires the activation of various biosynthetic pathways, in addition to those the old biomass equation required. This is not surprising, as the more complex biomass equation just places more demands on the underlying network, but it is important that we gain a quantitative understanding of these differences and their consequences. These differences can be studied by characterizing more biomass components (e.g. of glycophospholipids) and their connections to the underlying networks [339]; their consequences become visible through gene or metabolite profiling (e.g. [45]).

In recognition of the pivotal position that abundant high-throughput data have recently acquired in systems biology, I have mostly adopted a top-down perspective on metabolism throughout this thesis. Data of this kind have found their way into current-day metabolic models, for instance, in the form of gene-protein-reaction (GPR) relations [124]. GPR relations codify which gene produces a given protein, and which reaction this protein can perform (either by itself or as part of a complex), so that genotype can be linked directly to phenotype. The practical applicability of this information has so far been limited, primarily because the availability of data is not yet at the point that the genotype of individuals can be easily linked to their metabolic phenotype.

The GPR framework could extend to models of signaling/miRNA regulation, as these models describe what happens between **g**enes and **p**roteins, a conversion that is usually taken for granted in metabolic models (the step from **p**roteins to **r**eactions, on the other hand, has been explicitly modeled [136]). Unfortunately, our knowledge of the topology of the cell's regulatory networks is lacking in some places, and our knowledge of the dynamics of these networks is lacking in many places, especially where crosstalk enters the picture. To make matters more complicated, we do not know exactly where our knowledge lacks. Therefore, creating a model that includes transcription, translation, and the reaction that is eventually performed, as well as the regulatory processes that influence these phenomena, is currently still a utopic vision.

But even if we were already there, the GPR framework would still require a(n arbitrary) preprocessing step to go from expression data to something that would fit the **g**ene layer. The OME-matrix, the prokaryotic alternative, explicitly models this **e**xpression dimension, and has so far come the closest to providing a link between the dynamics of regulation and the relatively static structure of the metabolic network. It does so by providing a framework in which interactions between **o**perons, **m**etabolism, and **e**xpression can be formalized [136].

Regulation is more complex in mammals, though. The well-known central dogma of molecular biology holds that "DNA becomes RNA becomes protein" [347] (Figure 7.1), and we have known for decades that regulatory proteins can influence the transcription of DNA to RNA [348]. So far this could fit into an OME-like framework, but more recent discoveries have indicated that small RNAs can perform similar roles, and that these small RNAs, like proteins, are themselves subject to regulation [349, 350] (Figure 7.1). While this does not violate the central dogma, the two-way nature of these checks and balances does preclude them from being properly represented using the GPR or OME frameworks without major data (pre)processing.

Traditional frameworks are definitely not prepared for the layer of microRNA regulation. In Chapter 2, we describe a novel mechanism by which the abundance of microRNA-21 is regulated in an isoform-specific manner. Since miRNAs are generally thought of as stable molecules [351], this finding sheds new light on the regulation of the regulators, and it is fair to ask what this means for systems biology. The answer is deceptively simple. Simple, because if there is an additional layer of regulation, then it should (eventually) be added to our model of cellular function. *Deceptively* simple, because we do not yet have any model that can encompass all the "old" modes of regulation — at least not all at once.

But now that we know of this additional layer's existence, we will want to somehow determine its significance, either through experimentation or through computational estimation (i.e. systems biological methods). Given the modest effect on gene expression we observed (see Figure 2.4A), it is unlikely that the addition of this extra layer to regulatory or metabolic models would make a difference to predictions that were based on them in the past. Still, we can only be sure of this once we elucidate the full extent to which this mechanism influences regulation.

With all this talk of new mechanisms, it is easy to overlook the ones we are already familiar with. The "conventional" high-throughput approaches like microarray profiling and sequencing, which we dealt with above, focus on the sequences of proteins and other targets, but we have long known that there is more to proteins than their amino acid sequence. The phosphorylation state of enzymes matters greatly to their function, both in metabolism as well as in signaling (see Chapter 1 and e.g. [3, 22, 74, 81–86]). Therefore, it is important to quantify phosphorylated protein isoforms in addition to total protein or

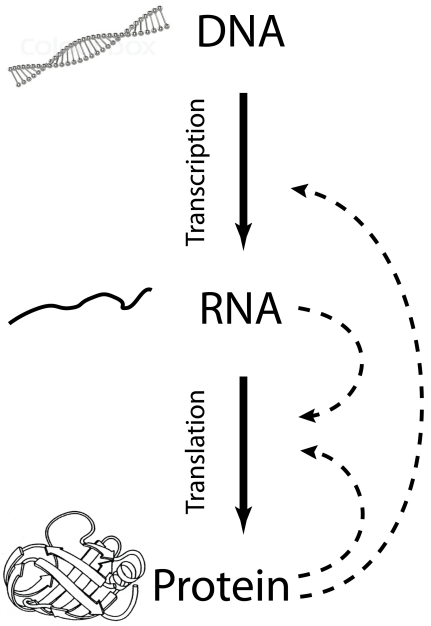


Figure 7.1: The central dogma of molecular biology. Proteins are produced by translating messenger RNA (mRNA) molecules into amino acid sequences. The mRNA, in turn, is transcribed from the genome, which is made from DNA. This flow of information is highly conserved throughout evolution (retroviruses are an exception, as they add one step in the inverse direction), but the products of transcription and translation (RNA and protein) can exert influence on transcription and translation. The dashed arrows represent RNA interference (RNA to translation), transcriptional regulation (protein to transcription) and (post-)translational regulation (protein to translation). As in these processes no RNA is produced from protein, and no DNA is produced from RNA, they do not violate the central dogma, but they add a layer of complexity to the simplistic "DNA becomes RNA becomes protein" mantra.

transcript levels, and microarray or sequencing approaches are not the tools of choice for this.

Instead, the field traditionally relied on knowledge-driven research and western blotting, but the recent implementation of reverse-phase protein arrays (RPPA) has brought big data into the picture [248]. As a result, signaling network dynamics can now be estimated based on high-throughput results, rather than determined through specific, targeted experimentation. This has the potential to accelerate knowledge generation in this field, and the approach has been shown to work in practice [352].

The work on the CXCR4/CXCR7/CXCL12 signaling axis described in Chap-

ter 3 is another example of such a data-driven systems biological study of a signaling network. It illustrates the application of high-throughput technology in the identification of qualitative connections, or broadly speaking, the network's structure. Quantitative statements about the dynamics of networks and interactions require even larger amounts of data, in addition to knowledge of the network structure. Thus, in the case of CXCR7-mediated signaling, a system whose network structure is poorly understood at present, we cannot understand the dynamics before we understand the structure, and this is what we contributed to in Chapter 3.

The methodological side of systems biology

In the section above, the analysis step that lies between data and knowledge was omitted for readability, but the importance of appropriate methods for analyzing large quantities of data is uncontested. Indeed, the availability of methods that help make sense of data and models is an important determining factor of the success of research. For example, the mathematical basis of stoichiometric modeling and flux balance analysis was laid more than two decades ago, but convenient tools to leverage this technique had yet to be developed at that time [353]. Interestingly, the strain that FBA places on computational resources is quite limited (solving large models would have been feasible two decades ago — genome-scale models just weren't there yet, because there was no genome-scale data), but the use of the technique only accelerated when larger models and adequate tools became available. Notable among the latter for their widespread adoption by the community were PySCeS [147] and the COBRA toolbox [148], but while these tools are highly versatile, they have a steep learning curve which, for many users, already starts at the installation step. The methodological work I have presented in this thesis can be seen as an effort to make existing mathematical techniques accessible to a broader audience of investigators and educators (Chapters 4–5).

In addition to facilitating the mathematical side of modeling, with FAME, we broke ground by offering so-called "supervised" results visualization functionality. Users can now superimpose results on a graphical map of metabolism, without first having to supply their own hand-drawn map. Prior to this, visualization had mostly been an afterthought. FAME's initial visualization capabilities were limited by the availability of high-quality base material (specifically, the KEGG maps [281]), but this changed with our development of a "universal map": One that is both human and computer friendly, and that

can be used for a multitude of species.

It is difficult to quantify the impact of a graphical representation of something on our way of thinking, but it is clear that it is there. In linguistics, a related notion is known as "linguistic relativity": The language you speak influences the way you think [354,355]. Similarly, most biologists will agree that the impact of the "standard" two-dimensional representations of metabolic pathways on our way of thinking about them has been huge. For instance, virtually all biologists with basic knowledge of metabolism will schematically draw core metabolism, and particularly glycolysis, the same way.

This has profound implications for the way we approach the questions we face as scientists, which in turn suggests that if there were a zero-effort way of visualizing our thoughts, data, and networks differently, or more comprehensively, then this would directly influence the way we generate hypotheses and approach problems. The universal map we describe in Chapter 5 is a first step upon this path, but in contrast to the WikiPathways and Recon 2 projects [50, 323], which are of an encyclopedic nature, expansion of this map does not require encyclopedic knowledge from individual contributors — they can just draw whatever they think is pretty or useful. Still, we feel that a comprehensive map is of vital importance to the study of metabolism, and its expansion will require a community effort. We have strived to make this as seamless as possible by using the tried-and-tested Git collaboration architecture [315].

The many aspects of systems biology

Systems biology is as multifaceted a field as cancer is a disease. Successfully applying a systems approach to, well, any kind of biology requires the simultaneous consideration of many of these facets. At the same time, many of these individual aspects are mature research areas in their own right, and taking them for granted would not do them justice. Therefore, in this section, we will explore several facets of systems biology in isolation, before reappraising the bigger picture.

Of systems and software

A key feature of scientifically relevant work, in any field, is that it gives rise to new ideas or inspires new research. Computational tools aid this process, but to those interested in applied bioinformatics, also pose questions and challenges of their own. For instance, spending time to create user-friendly

software, rather than plain raw functional software, is often frowned upon in scientific software development (the installation procedure of the COBRA Toolbox [148] is a testament to this). There seems to be an unspoken resignation that installation procedures need only be marginally documented, and that aspiring users just have to invest some of their time to figuring it out before they can use their peers' work to go do science.[1]

After all, scientists (generally) are smart people. But imagine the number of extra (wo)man-hours these smart people could spend on asking and answering interesting scientific questions if they would not have to figure out complex installation procedures or troubleshoot poorly documented analysis subroutines! Add to this the number of extra (wo)men that will be able to carry out the analyses in the first place if adequately human-friendly software is available, and the number of students that can be instructed in procedures that were previously too complex, and it becomes clear that more time should be spent improving the availability of computational methods, and less time should be spent ridiculing those who cannot operate "good old-fashioned academic software" (a euphemism for "software that may or may not work as advertized").

Fortunately, in spite of the harsh-yet-hopeful words above, awareness of a more user-centered perspective on scientific software has recently been building [356,357]. Thus, the future of systems biology, from a technology perspective, is one in which this growing field of scientists appreciate what software can do for them, *as well as* what they can do for software.

Of systems and sequencing

In the field of sequence analysis, software is a dire necessity, but as we can now generate data faster than we can interpret them, software matters recede into the background once the raw data has been processed and the analysis can begin. The technology-driven chapters of this thesis (Chapters 2 and 3) illustrate this point, as there, too, the "bio" takes precedence over the "informatics".

The biology-driven analyses, in turn, lead to new knowledge and more questions, which new (high throughput) experiments can then be designed for, leading to analysis and new knowledge and questions, etc. The intertwinement of systems and sequencing make "systems bioinformatics" more than just a catchphrase, it is the art of asking the right questions. And with

[1] This is not only a problem with scientific software. In scientific literature, materials and methods sections are usually insufficiently detailed to reproduce experiments and confirm conclusions.

regard to the opportunities RPPA and smallRNA sequencing present, we are not nearly out of good questions.

RPPA offers an unprecedented high throughput insight into the levels and, more importantly, the phosphorylation states of proteins [248]. The questions we asked and answered in Chapter 3 were of the obvious but necessary kind, as the study of any system becomes much more meaningful once the system's structure is known. But since RPPA is an antibody-based assay that can specifically target phosphorylated isoforms, we can think bigger.

How about simultaneously sampling the levels and phosphorylation states of glycolytic enzymes, as well as metabolite levels, and using the resulting data to refine our already fine-grained understanding of glycolytic kinetics [98]? This could even be combined with an evaluation of the cells' response to some kind of stimulus, such as activation of CXCR4 or CXCR7 by CXCL12.

Likewise, our analysis of sequencing data leads to some interesting follow-up questions, ranging from the obvious to the more far-reaching. As for the former, does a loss of PAPD5, PARN, or other components of the tailing-and-trimming machinery cause phenotypical alterations at the organism level? This would speak to their significance, but more importantly, it could shed light on idiosyncratic disease phenomena, which could in turn lead to therapy development. Of course, homing in on the specific contribution of miRNA tailing-and-trimming is complicated by the fact that PAPD5 and PARN also have roles in the maturation of snoRNA-63 [202] and other small RNAs (e.g. [214]), as well as by the involvement of other enzymes, like GLD2, in the same processes [183, 193, 193].

Further removed from the work presented in this thesis, but no less interesting, is the question of how sequencing data can help form connections in our knowledge of metabolism and signaling on a broader scale. For instance, single nucleotide polymorphisms (SNPs) were recently shown to have the potential to influence various metabolic processes [358]. In chemokine signaling, too, the influence of SNPs on phenotypical phenomena has long been known and appreciated [359], and mutations that are relevant to human health have been described for CXCR4 as well as CXCL12 [360, 361]. The combination of this kind of genetic/phenotypic knowledge with our expanding knowledge of the structures of the systems inside the cell will herald a new kind of insight, one which will surely help form and reinforce our (mental) models of the link between genotype and phenotype.

Of systems and signal transduction

In signal transduction, the function of a protein is almost always ambiguous, if its function is even known to begin with. This is a consequence of the way signaling networks work, and the biochemical requirements this puts on receptors and downstream effectors. Like metabolism [362, 363], the immune system [364], and even the internet [365], signaling is often schematically compared to a bow tie structure [366] (Figure 7.2A). There are many inputs (receptors for different environmental stimuli), a relatively small number of enzymes that integrate these many signals (the signal transduction network), and a large number of executive outputs that can tell the cell what to do (mostly transcription factors) [367].

CXCR4 and CXCR7 are receptors and obviously reside on the input side of this imaginary bow tie, but the results of the RPPA experiment described in Chapter 3 illustrate that despite the relatively small number of proteins in the middle section of the bow tie, the complexity of this section far exceeds that of the input and output sides (Figure 7.2B). Just like a real-world bow tie hides an intricate knot in the center, so does the (woefully overused) biological analog.

So while Figure 7.2A may look elegant, it does not do justice to the complex biology it describes. The signaling response to CXCL12 stimulation is an illustration of this. The two chemokine receptors that bind this ligand elicit different signaling responses, but exhibit crosstalk with each other and with other receptor networks [368–370]. In part, this is due to the atypical β-arrestin dependent signaling that CXCR7 has been associated with [228]. However, it is fair to ask what the functional consequences are of having CXCR4, CXCR7, both receptors, or none of them on the cell surface, and this is an important future line of research.

It is tempting to think that CXCR4 is the major contributor to the CXCL12 response, with CXCR7 as a (minor) backup, but as it stands, the only reason for this speculation would be our meager understanding of the response to CXCR7 activation. It requires no explanation that this kind of reasoning cannot stand on its own, so instead, in Chapter 3, we described the discovery of the involvement of MAPK in the downstream signaling network of CXCR7 in BT474 breast cancer cells.

Connections like these lay a qualitative foundation upon which quantitative models can be based [371]. Ample examples of these systems approaches to signaling exist and have yielded valuable results in the past [372–374], so not only is the elucidation of the CXCR4/CXCR7 signaling network "an important future line of research", it is also a feasible one. Just like in metabolism,

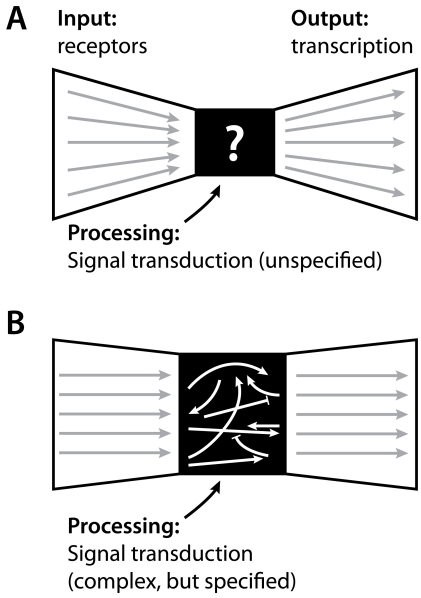


Figure 7.2: The bow tie representation of regulation. (A) In its classical form, the bow tie representation of regulation depicts many different inputs (stimuli), many different outputs (responses), and a relatively compact core that sets in motion a specific (set of) output(s) when triggered by a specific (combination of) input(s). The core is viewed as a black box with relatively few actors, but the highly complex internals are never specified. (B) Those who venture inside the black box invariably learn that the complexity of the interactions in the signal transduction core is vast, and to reflect this, I propose the "poorly knotted bow tie" model of regulation. In this poorly knotted bow tie, the knot is enlarged and its internals spill out into view, in an effort to more accurately represent the complex and important contribution of the signal transduction machinery to the integration of signals into actions by the cell.

however, stringing together multiple kinetic models of signaling is a tricky endeavor, but just like for metabolism, an answer to this complexity lies in abstraction. Petri nets are an implementation of one such answer [375–377], and even though the first genome-scale Petri net model of signaling has yet to be created, this future larger-scale signaling model would be mathematically compatible with existing metabolic Petri net reconstructions [378].

Of systems and stoichiometric modeling

For metabolic models, stoichiometric models have been used to leverage structural information even in the absence of complete kinetic knowledge [326,379].

As the structural properties of models of human metabolism are unambiguous and seldom form a source of disagreement, one would expect the resulting models to be used to generate predictions that can subsequently be tested. However, several issues have thus far stood in the way of this.

For one, Recon 1 and Recon 2 are reconstructions, and not models [50, 155]. The authors do indicate this (hence the "Recon" nomenclature), but the inclusion in Recon 2 of a non-human biomass equation inadvertently reinforces the misunderstanding that it is a model. Another issue is that the format in which derivative models were made available (if they even were) was not standardized, even though a suitable and widely ccepted standard was available in the form of SBML [146].

This made existing work difficult to reuse. For instance, a Recon 1-based model published by Folger et al. [160] requires extensive parsing to get it into SBML format. Nevertheless, and perhaps surprisingly, information from the model was reused: the biomass equation in Recon 2 was taken straight from the Folger et al. model, which in turn had adapted it from (mouse) literature [46]. The fact that this biomass equation was not elementally balanced could easily go unnoticed in the review process, as neither tools nor policies for checking this type of supplementary information were generally implemented in the review process.

Fortunately, this situation is on the verge of being remedied, as publishers like PLoS are adapting their policies to mandate the deposition of mathematical models into repositories, just like similar policies were developed for protein structures and sequencing data in the past. This development will also make it easier for reviewers to check for consistency between manuscript text and modeling results.

To underscore the importance of readers and reviewers being able to easily reproduce modeling results, we implemented an innovative solution of our own when we recently published a graphical map of Synechocystis metabolism [322]. The web-based environment in which we made the map and model available (http://f-a-m-e.org/synechocystis/) supports the reproduction of all modeling results and figures in the article, and we encourage and offer to facilitate all researchers who wish to follow this example when publishing models and results. Additionally, with the biomass equation we present in Chapter 6, modelers now have a starting point for making their work (truly) specific to human cells. But between that and detailed reconstructions like Recon 2.1 [339], where do the opportunities lie for future work?

The long term goal, of course, is integration of the metabolic domain with

knowledge and data from neighboring fields, like the regulatory machinery. A way to start, without requiring *a priori* knowledge of the combined system's full structure, is to assess what we know about the coincidence of a certain regulatory state of the cell with certain metabolic states (or vice versa). Given enough data points, such information may be used to elucidate the structure of the network at the interface of regulation (in a broad sense) and metabolism. Structural knowledge, in turn, can be leveraged by detailed models that include a regulatory layer, much like was done for *Escherichia coli* by Thiele et al. [136].

Until we get to that point, however, there are things we can do now with the tools, knowledge, and data that are available. These things form up three broad categories. Firstly, tissue-specific *models* can be generated by taking subsets of the generic human metabolic *reconstruction*, based on e.g. gene expression data. Bordbar et al. [164] exemplify this approach, and while their subsequent analysis incorporated only the structure of the networks, their methods are transferable.

Secondly, tissue-specific models can be used for tissue-specific simulations by using tissue-specific data quantifying metabolite consumption and production. The metabolite profiling study by Jain et al. produced such data for the NCI-60 panel (sixty cell lines), which makes this data set useful for the instantiation of single models as well as for semi-automated comparisons and lead discovery [45].

Finally, we can leverage existing knowledge about the structure of the human metabolic network and infer new knowledge or ideas from it. This is the approach taken by Folger et al. and Li et al. [160, 380], who used drug-reaction interactions in combination with the Recon 1 network structure and predicted novel drug targeting strategies and synergies. It is to be expected that this approach will soon be applied to the next generation of human metabolic models (i.e. Recon 2.1/2.1x [339]), and our knowledge base increases further, at very modest expense.

Of systems and cancer

All subfields of systems biology come together in the study of disease in human cells. Cancer is notable among human diseases, as it lends itself to being studied from the genome-scale metabolic perspective because of its distinctive metabolic phenotype [1]. Given the many different arguments for either a metabolic or a regulatory explanation, it is likely that the Warburg effect is the result of a combination the two. Indeed, when viewed within

an integrative mental framework, the many interconnections between the two realms argue for an integrative (but yet unknown) explanation. We have made a case for this in Chapter 1, and the integrative context of these systems has also been reviewed by others, who have reached similar conclusions [53].

There are potentially profound implications for the treatment of cancer if this systems view becomes widely adopted and finds its way into the clinic. Of course, most of the work in this thesis is of a very fundamental nature, and even a corny speculative "bench to bedside" prediction would seem like a stretch here. But it is also difficult to predict the shape an integrative approach to cancer treatment would have to take, since the currently dominant paradigm of evidence-based medicine [381] relies heavily on randomized controlled trials to supply the evidence that give it its name.

Systems biology concerns itself with the (biochemical) mechanisms upon which these trials are later based, but not so much with the trials or drugs themselves. It is, in the clinical sense, a hypothesis-generating link in the chain from bench to bedside (there, I did it); even as it generates and tests its own biochemical hypotheses, the clinical implications are usually uncertain. In other words, systems biology can be used to discover connections, that may then later be exploited clinically.

Therefore, the onus is now on the systems biology community to not only study individual systems, but to connect them, and to develop the tools needed to do this. Working from the bottom up, while useful in the study of individual systems, is not feasible in this realm, and loses out to high-throughput approaches.

This grand vision of integration, first and foremost, means that a model of how the regulatory (= signaling, transcription) and executive (= metabolism, motility, etc.) systems interact will have to be developed. To maximize the clinical impact of this understanding, the regulation of regulation (by e.g. microRNAs) will also have to be modeled, as it is highly likely that these phenomena, too, influence cancer on various levels. For instance, even in the nascent and fundamental field of miRNA modification, 3′ adenylation of a miRNA has already been shown to influence the translation of p53 mRNA [193].

Although it cannot exist without building upon previous work in individual systems, the importance of an integrative view in the quest for a cure to cancer cannot be overstated. It is in the context of this integrative framework that the work that is described in this thesis will have to be assessed, and hopefully prove useful in understanding the physiology of healthy and cancerous cells. Ultimately, both the scientific community, who will be required

to team up and share credit, as well the general public, who will have to fund them, will need to be convinced of the merits of this approach to make this happen [382].

But the best summary of our current position in the fight against cancer is perhaps in a quote from an ancient Chinese general, more than 2,000 years before there even was a fight against cancer — and he did not have to worry about funding.

> *Know yourself and know your enemy.*
> *You will be safe in every battle.*
> *You may know yourself but not know your enemy.*
> *You will then lose a battle for every battle you win.*
> *You may know neither yourself nor your enemy.*
> *You will then lose every battle.*
>
> — Sun Tzu, *The Art of War*

References

[1] Warburg O: **On respiratory impairment in cancer cells**. *Science* 1956, **124**(3215):269–270.

[2] Vander Heiden MG, Locasale JW, Swanson KD, Sharfi H, Heffron GJ, Amador-Noguez D, Christofk HR, Wagner G, Rabinowitz JD, Asara JM, Cantley LC: **Evidence for an alternative glycolytic pathway in rapidly proliferating cells**. *Science* 2010, **329**(5998):1492–1499.

[3] Gao X, Wang H, Yang JJ, Liu X, Liu ZR: **Pyruvate kinase M2 regulates gene transcription by acting as a protein kinase**. *Mol. Cell* 2012, **45**(5):598–609.

[4] Christofk HR, Vander Heiden MG, Harris MH, Ramanathan A, Gerszten RE, Wei R, Fleming MD, Schreiber SL, Cantley LC: **The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth**. *Nature* 2008, **452**(7184):230–233.

[5] Lu H, Forbes RA, Verma A: **Hypoxia-inducible factor 1 activation by aerobic glycolysis implicates the Warburg effect in carcinogenesis**. *J. Biol. Chem.* 2002, **277**(26):23111–23115.

[6] Matsumoto K, Obara N, Ema M, Horie M, Naka A, Takahashi S, Imagawa S: **Antitumor effects of 2-oxoglutarate through inhibition of angiogenesis in a murine tumor model**. *Cancer Sci.* 2009, **100**(9):1639–1647.

[7] Duran RV, Mackenzie ED, Boulahbel H, Frezza C, Heiserich L, Tardito S, Bussolati O, Rocha S, Hall MN, Gottlieb E: **HIF-independent role of prolyl hydroxylases in the cellular response to amino acids**. *Oncogene* 2012.

[8] Hanahan D, Weinberg RA: **The hallmarks of cancer**. *Cell* 2000, **100**:57–70.

[9] Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR: **A census of human cancer genes**. *Nat. Rev. Cancer* 2004, **4**(3):177–183.

[10] Soussi T, Ishioka C, Claustres M, Beroud C: **Locus-specific mutation databases: pitfalls and good practice based on the p53 experience**. *Nat. Rev. Cancer* 2006, **6**:83–90.

[11] Vogelstein B, Lane D, Levine AJ: **Surfing the p53 network**. *Nature* 2000, **408**(6810):307–310.

[12] Ross JS, Slodkowska EA, Symmans WF, Pusztai L, Ravdin PM, Hortobagyi GN: **The HER-2 receptor and breast cancer: ten years of targeted anti-HER-2 therapy and personalized medicine**. *Oncologist* 2009, **14**(4):320–368.

[13] De Vita F, Giuliani F, Silvestris N, Catalano G, Ciardiello F, Orditura M: **Human epidermal growth factor receptor 2 (HER2) in gastric cancer: a new therapeutic target**. *Cancer Treat. Rev.* 2010, **36 Suppl 3**:S11–15.

[14] Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation**. *Cell* 2011, **144**(5):646–674.

[15] Gatenby RA, Gillies RJ: **Why do cancers have high aerobic glycolysis?** *Nat. Rev. Cancer* 2004, **4**(11):891–899.

[16] Warburg O: **On the origin of cancer cells**. *Science* 1956, **123**(3191):309–314.

[17] Hsu PP, Sabatini DM: **Cancer cell metabolism: Warburg and beyond**. *Cell* 2008, **134**(5):703–707.

[18] van Dijken JP, Weusthuis RA, Pronk JT: **Kinetics of growth and sugar consumption in yeasts**. *Antonie Van Leeuwenhoek* 1993, **63**(3-4):343–352.

[19] Vemuri GN, Altman E, Sangurdekar DP, Khodursky AB, Eiteman MA: **Overflow metabolism in Escherichia coli during steady-state growth: transcriptional regulation and effect of the redox ratio**. *Appl. Environ. Microbiol.* 2006, **72**(5):3653–3661.

[20] Mazurek S, Boschek CB, Hugo F, Eigenbrodt E: **Pyruvate kinase type M2 and its role in tumor growth and spreading**. *Semin. Cancer Biol.* 2005, **15**(4):300–308.

[21] Bluemlein K, Gruning NM, Feichtinger RG, Lehrach H, Kofler B, Ralser M: **No evidence for a shift in pyruvate kinase PKM1 to PKM2 expression during tumorigenesis**. *Oncotarget* 2011, **2**(5):393–400.

[22] Hitosugi T, Kang S, Vander Heiden MG, Chung TW, Elf S, Lythgoe K, Dong S, Lonial S, Wang X, Chen GZ, Xie J, Gu TL, Polakiewicz RD, Roesel JL, Boggon TJ, Khuri FR, Gilliland DG, Cantley LC, Kaufman J, Chen J: **Tyrosine phosphorylation inhibits PKM2 to promote the Warburg effect and tumor growth**. *Sci Signal* 2009, **2**(97):ra73.

[23] Fell DA: **Metabolic control analysis: a survey of its theoretical and experimental development**. *Biochem. J.* 1992, **286 (Pt 2)**:313–330.

[24] Hofmeyr JH, Rohwer JM: **Supply-demand analysis a framework for exploring the regulatory design of metabolism**. *Meth. Enzymol.* 2011, **500**:533–554.

[25] Kochanowski K, Volkmer B, Gerosa L, Haverkorn van Rijsewijk BR, Schmidt A, Heinemann M: **Functioning of a metabolic flux sensor in Escherichia coli**. *Proc. Natl. Acad. Sci. U.S.A.* 2013, **110**(3):1130–1135.

[26] Hitosugi T, Zhou L, Elf S, Fan J, Kang HB, Seo JH, Shan C, Dai Q, Zhang L, Xie J, Gu TL, Jin P, Ale?kovi? M, LeRoy G, Kang Y, Sudderth JA, DeBerardinis RJ, Luan CH, Chen GZ, Muller S, Shin DM, Owonikoko TK, Lonial S, Arellano ML, Khoury HJ, Khuri FR, Lee BH, Ye K, Boggon TJ, Kang S, He C, Chen J: **Phosphoglycerate mutase 1 coordinates glycolysis and biosynthesis to promote tumor growth**. *Cancer Cell* 2012, **22**(5):585–600.

[27] Moellering RE, Cravatt BF: **Functional lysine modification by an intrinsically reactive primary glycolytic metabolite**. *Science* 2013, **341**(6145):549–553.

[28] Huang Z, Zhu L, Cao Y, Wu G, Liu X, Chen Y, Wang Q, Shi T, Zhao Y, Wang Y, Li W, Li Y, Chen H, Chen G, Zhang J: **ASD: a comprehensive database of allosteric proteins and modulators**. *Nucleic Acids Res.* 2011, **39**(Database issue):D663–669.

[29] Sugden MC, Holness MJ: **Recent advances in mechanisms regulating glucose oxidation at the level of the pyruvate dehydrogenase complex by PDKs**. *Am. J. Physiol. Endocrinol. Metab.* 2003, **284**(5):E855–862.

[30] Abbot EL, McCormack JG, Reynet C, Hassall DG, Buchan KW, Yeaman SJ: **Diverging regulation of pyruvate dehydrogenase kinase isoform gene expression in cultured human muscle cells**. *FEBS J.* 2005, **272**(12):3004–3014.

[31] Hitosugi T, Fan J, Chung TW, Lythgoe K, Wang X, Xie J, Ge Q, Gu TL, Polakiewicz RD, Roesel JL, Chen GZ, Boggon TJ, Lonial S, Fu H, Khuri FR, Kang S, Chen J: **Tyrosine phosphorylation of mitochondrial pyruvate dehydrogenase kinase 1 is important for cancer metabolism**. *Mol. Cell* 2011, **44**(6):864–877.

[32] Possemato R, Marks KM, Shaul YD, Pacold ME, Kim D, Birsoy K, Sethumadhavan S, Woo HK, Jang HG, Jha AK, Chen WW, Barrett FG, Stransky N, Tsun ZY, Cowley GS, Barretina J, Kalaany NY, Hsu PP, Ottina K, Chan AM, Yuan B, Garraway LA, Root DE, Mino-Kenudson M, Brachtel EF, Driggers EM, Sabatini DM: **Functional genomics reveal that the serine synthesis pathway is essential in breast cancer**. *Nature* 2011, **476**(7360):346–350.

[33] DeBerardinis RJ, Cheng T: **Q's next: the diverse functions of glutamine in metabolism, cell biology and cancer**. *Oncogene* 2010, **29**(3):313–324.

[34] Metallo CM, Gameiro PA, Bell EL, Mattaini KR, Yang J, Hiller K, Jewell CM, Johnson ZR, Irvine DJ, Guarente L, Kelleher JK, Vander Heiden MG, Iliopoulos O, Stephanopoulos G: **Reductive glutamine metabolism by IDH1 mediates lipogenesis under hypoxia**. *Nature* 2012, **481**(7381):380–384.

[35] Scott DA, Richardson AD, Filipp FV, Knutzen CA, Chiang GG, Ronai ZA, Osterman AL, Smith JW: **Comparative metabolic flux profiling of melanoma cell lines: beyond the Warburg effect**. *J. Biol. Chem.* 2011, **286**(49):42626–42634.

[36] Yoo H, Antoniewicz MR, Stephanopoulos G, Kelleher JK: **Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line**. *J. Biol. Chem.* 2008, **283**(30):20621–20627.

[37] Reich JG, Selkov EE: *Energy metabolism of the cell : a theoretical treatise*. New York: Academic Press 1981.

[38] Kacser H, Burns JA: **The control of flux**. *Biochem. Soc. Trans.* 1995, **23**(2):341–366.

[39] Fell DA, Thomas S: **Physiological control of metabolic flux: the requirement for multisite modulation**. *Biochem. J.* 1995, **311 (Pt 1)**:35–39.

[40] Semenza GL: **Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics**. *Oncogene* 2010, **29**(5):625–634.

[41] Neermann J, Wagner R: **Comparative analysis of glucose and glutamine metabolism in transformed mammalian cell lines, insect and primary liver cells**. *J. Cell. Physiol.* 1996, **166**:152–169.

[42] Ziegler A, von Kienlin M, Decorps M, Remy C: **High glycolytic activity in rat glioma demonstrated in vivo by correlation peak 1H magnetic resonance imaging**. *Cancer Res.* 2001, **61**(14):5595–5600.

[43] Snell K, Fell DA: **Metabolic control analysis of mammalian serine metabolism**. *Adv. Enzyme Regul.* 1990, **30**:13–32.

[44] Chaneton B, Hillmann P, Zheng L, Martin AC, Maddocks OD, Chokkathukalam A, Coyle JE, Jankevics A, Holding FP, Vousden KH, Frezza C, O'Reilly M, Gottlieb E: **Serine is a natural ligand and allosteric activator of pyruvate kinase M2**. *Nature* 2012, **491**(7424):458–462.

[45] Jain M, Nilsson R, Sharma S, Madhusudhan N, Kitami T, Souza AL, Kafri R, Kirschner MW, Clish CB, Mootha VK: **Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation**. *Science* 2012, **336**(6084):1040–1044.

[46] Bonarius HP, Hatzimanikatis V, Meesters KP, de Gooijer CD, Schmid G, Tramper J: **Metabolic flux analysis of hybridoma cells in different culture media using mass balances**. *Biotechnol. Bioeng.* 1996, **50**(3):299–318.

[47] Maier K, Hofmann U, Reuss M, Mauch K: **Identification of metabolic fluxes in hepatic cells from transient 13C-labeling experiments: Part II. Flux estimation**. *Biotechnol. Bioeng.* 2008, **100**(2):355–370.

[48] Schlisio S: **Neuronal apoptosis by prolyl hydroxylation: implication in nervous system tumours and the Warburg conundrum**. *J. Cell. Mol. Med.* 2009, **13**(10):4104–4112.

[49] Keith B, Johnson RS, Simon MC: **HIF1α and HIF2α: sibling rivalry in hypoxic tumour growth and progression**. *Nat. Rev. Cancer* 2012, **12**:9–22.

[50] Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, Haraldsdottir H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bolling C, Bordel S, Chavali AK, Dobson P, Dunn WB, Endler L, Hala D, Hucka M, Hull D, Jameson D, Jamshidi N, Jonsson JJ, Juty N, Keating S, Nookaew I, Le Novere N, Malys N, Mazein A, Papin JA, Price ND, Selkov E, Sigurdsson MI, Simeonidis E, Sonnenschein N, Smallbone K, Sorokin A, van Beek JH, Weichart D, Goryanin I, Nielsen J, Westerhoff HV, Kell DB, Mendes P, Palsson BO: **A community-driven global reconstruction of human metabolism**. *Nat. Biotechnol.* 2013, **31**(5):419–425.

[51] Vander Heiden MG, Cantley LC, Thompson CB: **Understanding the Warburg effect: the metabolic requirements of cell proliferation**. *Science* 2009, **324**(5930):1029–1033.

[52] Levine AJ, Puzio-Kuter AM: **The control of the metabolic switch in cancers by oncogenes and tumor suppressor genes**. *Science* 2010, **330**(6009):1340–1344.

[53] Cairns RA, Harris IS, Mak TW: **Regulation of cancer cell metabolism**. *Nat. Rev. Cancer* 2011, **11**(2):85–95.

[54] Munoz-Pinedo C, El Mjiyad N, Ricci JE: **Cancer metabolism: current perspectives and future directions**. *Cell Death Dis* 2012, **3**:e248.

[55] Semenza GL: **HIF-1: upstream and downstream of cancer metabolism**. *Curr. Opin. Genet. Dev.* 2010, **20**:51–56.

[56] Laughner E, Taghavi P, Chiles K, Mahon PC, Semenza GL: **HER2 (neu) signaling increases the rate of hypoxia-inducible factor 1alpha (HIF-1alpha) synthesis: novel mechanism for HIF-1-mediated vascular endothelial growth factor expression**. *Mol. Cell. Biol.* 2001, **21**(12):3995–4004.

[57] Iyer NV, Kotch LE, Agani F, Leung SW, Laughner E, Wenger RH, Gassmann M, Gearhart JD, Lawler AM, Yu AY, Semenza GL: **Cellular and developmental control of O2 homeostasis by hypoxia-inducible factor 1 alpha**. *Genes Dev.* 1998, **12**(2):149–162.

[58] Dang L, White DW, Gross S, Bennett BD, Bittinger MA, Driggers EM, Fantin VR, Jang HG, Jin S, Keenan MC, Marks KM, Prins RM, Ward PS, Yen KE, Liau LM, Rabinowitz JD, Cantley LC, Thompson CB, Vander Heiden MG, Su SM: **Cancer-associated IDH1 mutations produce 2-hydroxyglutarate**. *Nature* 2009, **462**(7274):739–744.

[59] Losman JA, Looper RE, Koivunen P, Lee S, Schneider RK, McMahon C, Cowley GS, Root DE, Ebert BL, Kaelin WG: **(R)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible**. *Science* 2013, **339**(6127):1621–1625.

[60] Ke Q, Costa M: **Hypoxia-inducible factor-1 (HIF-1)**. *Mol. Pharmacol.* 2006, **70**(5):1469–1480.

[61] Yuan G, Nanduri J, Khan S, Semenza GL, Prabhakar NR: **Induction of HIF-1alpha expression by intermittent hypoxia: involvement of NADPH oxidase, Ca2+ signaling, prolyl hydroxylases, and mTOR**. *J. Cell. Physiol.* 2008, **217**(3):674–685.

[62] Song MS, Salmena L, Pandolfi PP: **The functions and regulation of the PTEN tumour suppressor**. *Nat. Rev. Mol. Cell Biol.* 2012, **13**(5):283–296.

[63] Meng F, Henson R, Wehbe-Janek H, Ghoshal K, Jacob ST, Patel T: **MicroRNA-21 regulates expression of the PTEN tumor suppressor gene in human hepatocellular cancer**. *Gastroenterology* 2007, **133**(2):647–658.

[64] He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM: **A microRNA polycistron as a potential human oncogene**. *Nature* 2005, **435**(7043):828–833.

[65] Huang X, Le QT, Giaccia AJ: **MiR-210–micromanager of the hypoxia pathway**. *Trends Mol Med* 2010, **16**(5):230–237.

[66] Tarnowski M, Grymula K, Reca R, Jankowski K, Maksym R, Tarnowska J, Przybylski G, Barr FG, Kucia M, Ratajczak MZ: **Regulation of expression of stromal-derived factor-1 receptors: CXCR4 and CXCR7 in human rhabdomyosarcomas**. *Mol. Cancer Res.* 2010, **8**:1–14.

[67] Ishikawa T, Nakashiro K, Klosek SK, Goda H, Hara S, Uchida D, Hamakawa H: **Hypoxia enhances CXCR4 expression by activating HIF-1 in oral squamous cell carcinoma**. *Oncol. Rep.* 2009, **21**(3):707–712.

[68] Dang CV, Le A, Gao P: **MYC-induced cancer cell energy metabolism and therapeutic opportunities**. *Clin. Cancer Res.* 2009, **15**(21):6479–6483.

[69] Osthus RC, Shim H, Kim S, Li Q, Reddy R, Mukherjee M, Xu Y, Wonsey D, Lee LA, Dang CV: **Deregulation of glucose transporter 1 and glycolytic gene expression by c-Myc**. *J. Biol. Chem.* 2000, **275**(29):21797–21800.

[70] Gao P, Tchernyshyov I, Chang TC, Lee YS, Kita K, Ochi T, Zeller KI, De Marzo AM, Van Eyk JE, Mendell JT, Dang CV: **c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism**. *Nature* 2009, **458**(7239):762–765.

[71] Dang CV, Kim JW, Gao P, Yustein J: **The interplay between MYC and HIF in cancer**. *Nat. Rev. Cancer* 2008, **8**:51–56.

[72] Shaw RJ, Kosmatka M, Bardeesy N, Hurley RL, Witters LA, DePinho RA, Cantley LC: **The tumor suppressor LKB1 kinase directly activates AMP-activated kinase and regulates apoptosis in response to energy stress**. *Proc. Natl. Acad. Sci. U.S.A.* 2004, **101**(10):3329–3335.

[73] Mihaylova MM, Shaw RJ: **The AMPK signalling pathway coordinates cell growth, autophagy and metabolism**. *Nat. Cell Biol.* 2011, **13**(9):1016–1023.

[74] Shackelford DB, Shaw RJ: **The LKB1-AMPK pathway: metabolism and growth control in tumour suppression**. *Nat. Rev. Cancer* 2009, **9**(8):563–575.

[75] Koivunen P, Lee S, Duncan CG, Lopez G, Lu G, Ramkissoon S, Losman JA, Joensuu P, Bergmann U, Gross S, Travins J, Weiss S, Looper R, Ligon KL, Verhaak RG, Yan H, Kaelin WG: **Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation**. *Nature* 2012, **483**(7390):484–488.

[76] Zhao Y, Coloff JL, Ferguson EC, Jacobs SR, Cui K, Rathmell JC: **Glucose metabolism attenuates p53 and Puma-dependent cell death upon growth factor deprivation**. *J. Biol. Chem.* 2008, **283**(52):36344–36353.

[77] De Saedeleer CJ, Copetti T, Porporato PE, Verrax J, Feron O, Sonveaux P: **Lactate activates HIF-1 in oxidative but not in Warburg-phenotype human tumor cells**. *PLoS ONE* 2012, **7**(10):e46571.

[78] Blad CC, Tang C, Offermanns S: **G protein-coupled receptors for energy metabolites as new therapeutic targets**. *Nat Rev Drug Discov* 2012, **11**(8):603–619.

[79] Kaneda M, Takeuchi K, Inoue K, Umeda M: **Localization of the phosphatidylserine-binding site of glyceraldehyde-3-phosphate dehydrogenase responsible for membrane fusion**. *J. Biochem.* 1997, **122**(6):1233–1240.

[80] Seweryn E, Pietkiewicz J, Bednarz-Misa IS, Ceremuga I, Saczko J, Kulbacka J, Gamian A: **Localization of enolase in the subfractions of a breast cancer cell line**. *Z. Naturforsch., C, J. Biosci.* 2009, **64**(9-10):754–758.

[81] Schulze A, Downward J: **Flicking the Warburg switch-tyrosine phosphorylation of pyruvate dehydrogenase kinase regulates mitochondrial activity in cancer cells**. *Mol. Cell* 2011, **44**(6):846–848.

[82] Fan J, Hitosugi T, Chung TW, Xie J, Ge Q, Gu TL, Polakiewicz RD, Chen GZ, Boggon TJ, Lonial S, Khuri FR, Kang S, Chen J: **Tyrosine phosphorylation of lactate dehydrogenase A is important for NADH/NAD(+) redox homeostasis in cancer cells**. *Mol. Cell. Biol.* 2011, **31**(24):4938–4950.

[83] Fernandez-Garcia P, Pelaez R, Herrero P, Moreno F: **Phosphorylation of yeast hexokinase 2 regulates its nucleocytoplasmic shuttling**. *J. Biol. Chem.* 2012, **287**(50):42151–42164.

[84] Dieni CA, Storey KB: **Regulation of hexokinase by reversible phosphorylation in skeletal muscle of a freeze-tolerant frog**. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* 2011, **159**(4):236–243.

[85] Pilkis SJ, Claus TH, Kurland IJ, Lange AJ: **6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase: a metabolic signaling enzyme**. *Annu. Rev. Biochem.* 1995, **64**:799–835.

[86] Marsin AS, Bouzin C, Bertrand L, Hue L: **The stimulation of glycolysis by hypoxia in activated monocytes is mediated by AMP-activated protein kinase and inducible 6-phosphofructo-2-kinase**. *J. Biol. Chem.* 2002, **277**(34):30778–30783.

[87] Heinrich R, S S: *The regulation of cellular systems*. New York: Chapman & Hall 1996.

[88] Schuster S, Fell DA, Dandekar T: **A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks**. *Nat. Biotechnol.* 2000, **18**(3):326–332.

[89] Fell D: *Understanding the control of metabolism*. London: Portland Press 1997.

[90] Veening JW, Smits WK, Kuipers OP: **Bistability, epigenetics, and bet-hedging in bacteria**. *Annu. Rev. Microbiol.* 2008, **62**:193–210.

[91] Goldbeter A: *Biochemical Oscillations And Cellular Rhythms - The Molecular Bases Of Periodic And Chaotic Behaviour*. Cambridge: Cambridge University Press 1997.

[92] Kitano H: **Systems biology: a brief overview**. *Science* 2002, **295**(5560):1662–1664.

[93] Kitano H: **Towards a theory of biological robustness**. *Mol. Syst. Biol.* 2007, **3**:137.

[94] Bakker BM, van Eunen K, Jeneson JA, van Riel NA, Bruggeman FJ, Teusink B: **Systems biology from micro-organisms to human metabolic diseases: the role of detailed kinetic models**. *Biochem. Soc. Trans.* 2010, **38**(5):1294–1301.

[95] Hoefnagel MH, Starrenburg MJ, Martens DE, Hugenholtz J, Kleerebezem M, Van Swam II, Bongers R, Westerhoff HV, Snoep JL: **Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis**. *Microbiology (Reading, Engl.)* 2002, **148**(Pt 4):1003–1013.

[96] Barthelmes J, Ebeling C, Chang A, Schomburg I, Schomburg D: **BRENDA, AMENDA and FRENDA: the enzyme information system in 2007**. *Nucleic Acids Res.* 2007, **35**(Database issue):D511–514.

[97] Rojas I, Golebiewski M, Kania R, Krebs O, Mir S, Weidemann A, Wittig U: **Storing and annotating of kinetic data**. *In Silico Biol. (Gedrukt)* 2007, **7**(2 Suppl):37–44.

[98] Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, Walsh MC, Bakker BM, van Dam K, Westerhoff HV, Snoep JL: **Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry**. *Eur. J. Biochem.* 2000, **267**(17):5313–5329.

[99] Mashego MR, van Gulik WM, Heijnen JJ: **Metabolome dynamic responses of Saccharomyces cerevisiae to simultaneous rapid perturbations in external electron acceptor and electron donor**. *FEMS Yeast Res.* 2007, **7**:48–66.

[100] Theobald U, Mailinger W, Reuss M, Rizzi M: **In vivo analysis of glucose-induced fast changes in yeast adenine nucleotide pool applying a rapid sampling technique**. *Anal. Biochem.* 1993, **214**:31–37.

[101] Visser D, van Zuylen GA, van Dam JC, Eman MR, Proll A, Ras C, Wu L, van Gulik WM, Heijnen JJ: **Analysis of in vivo kinetics of glycolysis in aerobic Saccharomyces cerevisiae by application of glucose and ethanol pulses**. *Biotechnol. Bioeng.* 2004, **88**(2):157–167.

[102] Banga JR, Balsa-Canto E: **Parameter estimation and optimal experimental design**. *Essays Biochem.* 2008, **45**:195–209.

[103] Liebermeister W, Baur U, Klipp E: **Biochemical network models simplified by balanced truncation**. *FEBS J.* 2005, **272**(16):4034–4043.

[104] Liebermeister W, Klipp E: **Biochemical networks with uncertain parameters**. *Syst Biol (Stevenage)* 2005, **152**(3):97–107.

[105] Liebermeister W, Klipp E: **Bringing metabolic networks to life: convenience rate law and thermodynamic constraints**. *Theor Biol Med Model* 2006, **3**:41.

[106] Steuer R, Gross T, Selbig J, Blasius B: **Structural kinetic modeling of metabolic networks**. *Proc. Natl. Acad. Sci. U.S.A.* 2006, **103**(32):11868–11873.

[107] Wang L, Hatzimanikatis V: **Metabolic engineering under uncertainty. I: framework development**. *Metab. Eng.* 2006, **8**(2):133–141.

[108] Jamshidi N, Palsson BO: **Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models**. *Biophys. J.* 2010, **98**(2):175–185.

[109] Resendis-Antonio O: **Filling kinetic gaps: dynamic modeling of metabolism where detailed kinetic information is lacking**. *PLoS ONE* 2009, **4**(3):e4967.

[110] Price ND, Reed JL, Palsson BO: **Genome-scale models of microbial cells: evaluating the consequences of constraints**. *Nat. Rev. Microbiol.* 2004, **2**(11):886–897.

[111] Papin JA, Stelling J, Price ND, Klamt S, Schuster S, Palsson BO: **Comparison of network-based pathway analysis methods**. *Trends Biotechnol.* 2004, **22**(8):400–405.

[112] Hoppe A, Hoffmann S, Holzhutter HG: **Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks**. *BMC Syst Biol* 2007, **1**:23.

[113] Kummel A, Panke S, Heinemann M: **Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data**. *Mol. Syst. Biol.* 2006, **2**:2006.0034.

[114] Kummel A, Panke S, Heinemann M: **Systematic assignment of thermodynamic constraints in metabolic network models**. *BMC Bioinformatics* 2006, **7**:512.

[115] Teusink B, Smid EJ: **Modelling strategies for the industrial exploitation of lactic acid bacteria**. *Nat. Rev. Microbiol.* 2006, **4**:46–56.

[116] Lipton P: **Testing hypotheses: prediction and prejudice**. *Science* 2005, **307**(5707):219–221.

[117] Kell DB: **Metabolomics and systems biology: making sense of the soup**. *Curr. Opin. Microbiol.* 2004, **7**(3):296–307.

[118] van der Werf MJ: **Towards replacing closed with open target selection strategies**. *Trends Biotechnol.* 2005, **23**:11–16.

[119] Knijnenburg TA, de Winde JH, Daran JM, Daran-Lapujade P, Pronk JT, Reinders MJ, Wessels LF: **Exploiting combinatorial cultivation conditions to infer transcriptional regulation**. *BMC Genomics* 2007, **8**:25.

[120] Tai SL, Boer VM, Daran-Lapujade P, Walsh MC, de Winde JH, Daran JM, Pronk JT: **Two-dimensional transcriptome analysis in chemostat cultures. Combinatorial effects of oxygen availability and macronutrient limitation in Saccharomyces cerevisiae**. *J. Biol. Chem.* 2005, **280**:437–447.

[121] Kim C, Paik S: **Gene-expression-based prognostic assays for breast cancer**. *Nat Rev Clin Oncol* 2010, **7**(6):340–347.

[122] Francke C, Siezen RJ, Teusink B: **Reconstructing the metabolic network of a bacterium from its genome**. *Trends Microbiol.* 2005, **13**(11):550–558.

[123] Reed JL, Famili I, Thiele I, Palsson BO: **Towards multidimensional genome annotation**. *Nat. Rev. Genet.* 2006, **7**(2):130–141.

[124] Thiele I, Palsson BO: **A protocol for generating a high-quality genome-scale metabolic reconstruction**. *Nat Protoc* 2010, **5**:93–121.

[125] Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D, Gavin AC: **Visualization of omics data for systems biology**. *Nat. Methods* 2010, **7**(3 Suppl):56–68.

[126] Kono N, Arakawa K, Ogawa R, Kido N, Oshita K, Ikegami K, Tamaki S, Tomita M: **Pathway projector: web-based zoomable pathway browser using KEGG atlas and Google Maps API**. *PLoS ONE* 2009, **4**(11):e7710.

[127] Stevens MJ, Wiersma A, de Vos WM, Kuipers OP, Smid EJ, Molenaar D, Kleerebezem M: **Improvement of Lactobacillus plantarum aerobic growth as directed by comprehensive transcriptome analysis**. *Appl. Environ. Microbiol.* 2008, **74**(15):4776–4778.

[128] Patil KR, Nielsen J: **Uncovering transcriptional regulation of metabolism by using metabolic network topology**. *Proc. Natl. Acad. Sci. U.S.A.* 2005, **102**(8):2685–2689.

[129] Cakir T, Patil KR, Onsan Zi, Ulgen KO, Kirdar B, Nielsen J: **Integration of metabolome data with metabolic networks reveals reporter reactions**. *Mol. Syst. Biol.* 2006, **2**:50.

[130] Burgard AP, Nikolaev EV, Schilling CH, Maranas CD: **Flux coupling analysis of genome-scale metabolic network reconstructions**. *Genome Res.* 2004, **14**(2):301–312.

[131] Notebaart RA, Kensche PR, Huynen MA, Dutilh BE: **Asymmetric relationships between proteins shape genome evolution**. *Genome Biol.* 2009, **10**(2):R19.

[132] Notebaart RA, Teusink B, Siezen RJ, Papp B: **Co-regulation of metabolic genes is better explained by flux coupling than by network distance**. *PLoS Comput. Biol.* 2008, **4**:e26.

[133] Palsson B: **Two-dimensional annotation of genomes**. *Nat. Biotechnol.* 2004, **22**(10):1218–1219.

[134] Herrgard MJ, Covert MW, Palsson BO: **Reconstruction of microbial transcriptional regulatory networks**. *Curr. Opin. Biotechnol.* 2004, **15**:70–77.

[135] Hyduke DR, Palsson BO: **Towards genome-scale signalling network reconstructions**. *Nat. Rev. Genet.* 2010, **11**(4):297–307.

[136] Thiele I, Jamshidi N, Fleming RM, Palsson BO: **Genome-scale reconstruction of Escherichia coli's transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization**. *PLoS Comput. Biol.* 2009, **5**(3):e1000312.

[137] Gianchandani EP, Chavali AK, Papin JA: **The application of flux balance analysis in systems biology**. *Wiley Interdiscip Rev Syst Biol Med* 2010, **2**(3):372–382.

[138] Orth JD, Thiele I, Palsson BO: **What is flux balance analysis?** *Nat. Biotechnol.* 2010, **28**(3):245–248.

[139] Mahadevan R, Schilling CH: **The effects of alternate optimal solutions in constraint-based genome-scale metabolic models**. *Metab. Eng.* 2003, **5**(4):264–276.

[140] Burgard AP, Pharkya P, Maranas CD: **Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization**. *Biotechnol. Bioeng.* 2003, **84**(6):647–657.

[141] Cvijovic M, Olivares-Hernandez R, Agren R, Dahr N, Vongsangnak W, Nookaew I, Patil KR, Nielsen J: **BioMet Toolbox: genome-wide analysis of metabolism**. *Nucleic Acids Res.* 2010, **38**(Web Server issue):W144–149.

[142] Segre D, Vitkup D, Church GM: **Analysis of optimality in natural and perturbed metabolic networks**. *Proc. Natl. Acad. Sci. U.S.A.* 2002, **99**(23):15112–15117.

[143] Teusink B, Wiersma A, Jacobs L, Notebaart RA, Smid EJ: **Understanding the adaptive growth strategy of Lactobacillus plantarum by in silico optimisation**. *PLoS Comput. Biol.* 2009, **5**(6):e1000410.

[144] Shlomi T, Berkman O, Ruppin E: **Regulatory on/off minimization of metabolic flux changes after genetic perturbations**. *Proc. Natl. Acad. Sci. U.S.A.* 2005, **102**(21):7695–7700.

[145] Kelk SM, Olivier BG, Stougie L, Bruggeman FJ: **Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks**. *Sci Rep* 2012, **2**:580.

[146] Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J: **The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models**. *Bioinformatics* 2003, **19**(4):524–531.

[147] Olivier BG, Rohwer JM, Hofmeyr JH: **Modelling cellular systems with PySCeS**. *Bioinformatics* 2005, **21**(4):560–561.

[148] Becker SA, Feist AM, Mo ML, Hannum G, Palsson B, Herrgard MJ: **Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox**. *Nat Protoc* 2007, **2**(3):727–738.

[149] Schuetz R, Kuepfer L, Sauer U: **Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli**. *Mol. Syst. Biol.* 2007, **3**:119.

[150] Edwards JS, Ibarra RU, Palsson BO: **In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data**. *Nat. Biotechnol.* 2001, **19**(2):125–130.

[151] Famili I, Forster J, Nielsen J, Palsson BO: **Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network**. *Proc. Natl. Acad. Sci. U.S.A.* 2003, **100**(23):13134–13139.

[152] Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110**. *Appl. Environ. Microbiol.* 1994, **60**(10):3724–3731.

[153] Fong SS, Marciniak JY, Palsson BO: **Description and interpretation of adaptive evolution of Escherichia coli K-12 MG1655 by using a genome-scale in silico metabolic model**. *J. Bacteriol.* 2003, **185**(21):6400–6408.

[154] Ibarra RU, Edwards JS, Palsson BO: **Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth**. *Nature* 2002, **420**(6912):186–189.

[155] Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO: **Global reconstruction of the human metabolic network based on genomic and bibliomic data**. *Proc. Natl. Acad. Sci. U.S.A.* 2007, **104**(6):1777–1782.

[156] Ma H, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I: **The Edinburgh human metabolic network reconstruction and its functional analysis**. *Mol. Syst. Biol.* 2007, **3**:135.

[157] Gille C, Bolling C, Hoppe A, Bulik S, Hoffmann S, Hubner K, Karlstadt A, Ganeshan R, Konig M, Rother K, Weidlich M, Behre J, Holzhutter HG: **HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology**. *Mol. Syst. Biol.* 2010, **6**:411.

[158] Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J: **Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT**. *PLoS Comput. Biol.* 2012, **8**(5):e1002518.

[159] Wang Y, Eddy JA, Price ND: **Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE**. *BMC Syst Biol* 2012, **6**:153.

[160] Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T: **Predicting selective drug targets in cancer through metabolic networks**. *Mol. Syst. Biol.* 2011, **7**:501.

[161] Shlomi T, Benyamini T, Gottlieb E, Sharan R, Ruppin E: **Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect**. *PLoS Comput. Biol.* 2011, **7**(3):e1002018.

[162] Resendis-Antonio O, Checa A, Encarnacion S: **Modeling core metabolism in cancer cells: surveying the topology underlying the Warburg effect**. *PLoS ONE* 2010, **5**(8):e12383.

[163] Feist AM, Palsson BO: **The biomass objective function**. *Curr. Opin. Microbiol.* 2010, **13**(3):344–349.

[164] Bordbar A, Feist AM, Usaite-Black R, Woodcock J, Palsson BO, Famili I: **A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology**. *BMC Syst Biol* 2011, **5**:180.

[165] Lee RC, Feinbaum RL, Ambros V: **The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14**. *Cell* 1993, **75**(5):843–854.

[166] Lee LW, Zhang S, Etheridge A, Ma L, Martin D, Galas D, Wang K: **Complexity of the microRNA repertoire revealed by next-generation sequencing**. *RNA* 2010, **16**(11):2170–2180.

[167] Winter J, Jung S, Keller S, Gregory RI, Diederichs S: **Many roads to maturity: microRNA biogenesis pathways and their regulation**. *Nat. Cell Biol.* 2009, **11**(3):228–234.

[168] Hutvagner G, Simard MJ: **Argonaute proteins: key players in RNA silencing**. *Nat. Rev. Mol. Cell Biol.* 2008, **9**:22–32.

[169] Kawahara Y, Zinshteyn B, Chendrimada TP, Shiekhattar R, Nishikura K: **RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex**. *EMBO Rep.* 2007, **8**(8):763–769.

[170] Kawahara Y, Megraw M, Kreider E, Iizasa H, Valente L, Hatzigeorgiou AG, Nishikura K: **Frequency and fate of microRNA editing in human brain**. *Nucleic Acids Res.* 2008, **36**(16):5270–5280.

[171] Viswanathan SR, Daley GQ, Gregory RI: **Selective blockade of microRNA processing by Lin28**. *Science* 2008, **320**(5872):97–100.

[172] Heo I, Joo C, Cho J, Ha M, Han J, Kim VN: **Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA**. *Mol. Cell* 2008, **32**(2):276–284.

[173] Heo I, Joo C, Kim YK, Ha M, Yoon MJ, Cho J, Yeom KH, Han J, Kim VN: **TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation**. *Cell* 2009, **138**(4):696–708.

[174] Hagan JP, Piskounova E, Gregory RI: **Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells**. *Nat. Struct. Mol. Biol.* 2009, **16**(10):1021–1025.

[175] Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foa R, Schliwka J, Fuchs U, Novosel A, Muller RU, Schermer B, Bissels U, Inman J, Phan Q, Chien M, Weir DB, Choksi R, De Vita G, Frezzetti D, Trompeter HI, Hornung V, Teng G, Hartmann G, Palkovits M, Di Lauro R, Wernet P, Macino G, Rogler CE, Nagle JW, Ju J, Papavasiliou FN, Benzing T, Lichter P, Tam W, Brownstein MJ, Bosio A, Borkhardt A, Russo JJ, Sander C, Zavolan M, Tuschl T: **A mammalian microRNA expression atlas based on small RNA library sequencing**. *Cell* 2007, **129**(7):1401–1414.

[176] Kuchenbauer F, Morin RD, Argiropoulos B, Petriv OI, Griffith M, Heuser M, Yung E, Piper J, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Hansen CL, Marra MA, Humphries RK: **In-depth characterization of the microRNA transcriptome in a leukemia progression model**. *Genome Res.* 2008, **18**(11):1787–1797.

[177] Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA: **Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells**. *Genome Res.* 2008, **18**(4):610–621.

[178] Starega-Roslan J, Krol J, Koscianska E, Kozlowski P, Szlachcic WJ, Sobczak K, Krzyzosiak WJ: **Structural basis of microRNA length variety**. *Nucleic Acids Res.* 2011, **39**:257–268.

[179] Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Suzuki H, Hayashizaki Y, Daub CO: **Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin**. *RNA Biol* 2011, **8**:158–177.

[180] Azuma-Mukai A, Oguri H, Mituyama T, Qian ZR, Asai K, Siomi H, Siomi MC: **Characterization of endogenous human Argonautes and their miRNA partners in RNA silencing**. *Proc. Natl. Acad. Sci. U.S.A.* 2008, **105**(23):7964–7969.

[181] Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Nishibu T, Ukekawa R, Funakoshi T, Kurokawa T, Suzuki H, Hayashizaki Y, Daub CO: **A comprehensive survey of 3′ animal miRNA modification events and a possible role for 3′ adenylation in modulating miRNA targeting effectiveness**. *Genome Res.* 2010, **20**(10):1398–1410.

[182] Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, Krouse MA, Webster PJ, Tewari M: **Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity**. *Genome Res.* 2011, **21**(9):1450–1461.

[183] Katoh T, Sakaguchi Y, Miyauchi K, Suzuki T, Kashiwabara S, Baba T, Suzuki T: **Selective stabilization of mammalian microRNAs by 3′ adenylation mediated by the cytoplasmic poly(A) polymerase GLD-2**. *Genes Dev.* 2009, **23**(4):433–438.

[184] Krol J, Loedige I, Filipowicz W: **The widespread regulation of microRNA biogenesis, function and decay**. *Nat. Rev. Genet.* 2010, **11**(9):597–610.

[185] Yang CH, Yue J, Pfeffer SR, Handorf CR, Pfeffer LM: **MicroRNA miR-21 regulates the metastatic behavior of B16 melanoma cells**. *J. Biol. Chem.* 2011, **286**(45):39172–39178.

[186] Jazbutyte V, Thum T: **MicroRNA-21: from cancer to cardiovascular disease**. *Curr Drug Targets* 2010, **11**(8):926–935.

[187] Medina PP, Nolde M, Slack FJ: **OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma**. *Nature* 2010, **467**(7311):86–90.

[188] Cheng Y, Zhang C: **MicroRNA-21 in cardiovascular disease**. *J Cardiovasc Transl Res* 2010, **3**(3):251–255.

[189] Joyce CE, Zhou X, Xia J, Ryan C, Thrash B, Menter A, Zhang W, Bowcock AM: **Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome**. *Hum. Mol. Genet.* 2011, **20**(20):4025–4040.

[190] Liu M, Wu H, Liu T, Li Y, Wang F, Wan H, Li X, Tang H: **Regulation of the cell cycle gene, BTG2, by miR-21 in human laryngeal carcinoma**. *Cell Res.* 2009, **19**(7):828–837.

[191] Niu J, Shi Y, Tan G, Yang CH, Fan M, Pfeffer LM, Wu ZH: **DNA damage induces NF-κB-dependent microRNA-21 up-regulation and promotes breast cancer cell invasion**. *J. Biol. Chem.* 2012, **287**(26):21783–21795.

[192] Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction**. *Nucleic Acids Res.* 2003, **31**(13):3406–3415.

[193] Burns DM, D'Ambrogio A, Nottrott S, Richter JD: **CPEB and two poly(A) polymerases control miR-122 stability and p53 mRNA translation**. *Nature* 2011, **473**(7345):105–108.

[194] Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ: **miRBase: microRNA sequences, targets and gene nomenclature**. *Nucleic Acids Res.* 2006, **34**(Database issue):D140–144.

[195] Taft RJ, Glazov EA, Cloonan N, Simons C, Stephen S, Faulkner GJ, Lassmann T, Forrest AR, Grimmond SM, Schroder K, Irvine K, Arakawa T, Nakamura M, Kubosaki A, Hayashida K, Kawazu C, Murata M, Nishiyori H, Fukuda S, Kawai J, Daub CO, Hume DA, Suzuki H, Orlando V, Carninci P, Hayashizaki Y, Mattick JS: **Tiny RNAs associated with transcription start sites in animals**. *Nat. Genet.* 2009, **41**(5):572–578.

[196] Lee HY, Doudna JA: **TRBP alters human precursor microRNA processing in vitro**. *RNA* 2012, **18**(11):2012–2019.

[197] Ando Y, Maida Y, Morinaga A, Burroughs AM, Kimura R, Chiba J, Suzuki H, Masutomi K, Hayashizaki Y: **Two-step cleavage of hairpin RNA with 5′ overhangs by human DICER**. *BMC Mol. Biol.* 2011, **12**:6.

[198] Zhang H, Kolb FA, Jaskiewicz L, Westhof E, Filipowicz W: **Single processing center models for human Dicer and bacterial RNase III**. *Cell* 2004, **118**:57–68.

[199] Ameres SL, Horwich MD, Hung JH, Xu J, Ghildiyal M, Weng Z, Zamore PD: **Target RNA-directed trimming and tailing of small silencing RNAs**. *Science* 2010, **328**(5985):1534–1539.

[200] Rammelt C, Bilen B, Zavolan M, Keller W: **PAPD5, a noncanonical poly(A) polymerase with an unusual RNA-binding motif**. *RNA* 2011, **17**(9):1737–1746.

[201] Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T: **Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP**. *Cell* 2010, **141**:129–141.

[202] Berndt H, Harnisch C, Rammelt C, Stohr N, Zirkel A, Dohm JC, Himmelbauer H, Tavanez JP, Huttelmaier S, Wahle E: **Maturation of mammalian H/ACA box snoRNAs: PAPD5-dependent adenylation and PARN-dependent trimming**. *RNA* 2012, **18**(5):958–972.

[203] Virtanen A, Henriksson N, Nilsson P, Nissbeck M: **Poly(A)-specific ribonuclease (PARN): an allosterically regulated, processive and mRNA cap-interacting deadenylase**. *Crit. Rev. Biochem. Mol. Biol.* 2013, **48**(2):192–209.

[204] Friedman RC, Farh KK, Burge CB, Bartel DP: **Most mammalian mRNAs are conserved targets of microRNAs**. *Genome Res.* 2009, **19**:92–105.

[205] Lewis BP, Shih IH, Jones-Rhoades MW, Bartel DP, Burge CB: **Prediction of mammalian microRNA targets**. *Cell* 2003, **115**(7):787–798.

[206] Roberson ED, Bowcock AM: **Psoriasis genetics: breaking the barrier**. *Trends Genet.* 2010, **26**(9):415–423.

[207] Liu N, Abe M, Sabin LR, Hendriks GJ, Naqvi AS, Yu Z, Cherry S, Bonini NM: **The exoribonuclease Nibbler controls 3′ end processing of microRNAs in Drosophila**. *Curr. Biol.* 2011, **21**(22):1888–1893.

[208] Han BW, Hung JH, Weng Z, Zamore PD, Ameres SL: **The 3′-to-5′ exoribonuclease Nibbler shapes the 3′ ends of microRNAs bound to Drosophila Argonaute1**. *Curr. Biol.* 2011, **21**(22):1878–1887.

[209] Burroughs AM, Kawano M, Ando Y, Daub CO, Hayashizaki Y: **pre-miRNA profiles obtained through application of locked nucleic acids and deep sequencing reveals complex 5′/3′ arm variation including concomitant cleavage and polyuridylation patterns**. *Nucleic Acids Res.* 2012, **40**(4):1424–1437.

[210] Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN: **The nuclear RNase III Drosha initiates microRNA processing**. *Nature* 2003, **425**(6956):415–419.

[211] Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R: **The Microprocessor complex mediates the genesis of microRNAs**. *Nature* 2004, **432**(7014):235–240.

[212] Pedersen IM, Cheng G, Wieland S, Volinia S, Croce CM, Chisari FV, David M: **Interferon modulation of cellular microRNAs as an antiviral mechanism**. *Nature* 2007, **449**(7164):919–922.

[213] Marcinowski L, Tanguy M, Krmpotic A, Radle B, Lisni? VJ, Tuddenham L, Chane-Woon-Ming B, Ruzsics Z, Erhard F, Benkartek C, Babic M, Zimmer R, Trgovcich J, Koszinowski UH, Jonjic S, Pfeffer S, Dolken L: **Degradation of cellular mir-27 by a novel, highly abundant viral transcript is important for efficient virus replication in vivo**. *PLoS Pathog.* 2012, **8**(2):e1002510.

[214] Yoda M, Cifuentes D, Izumi N, Sakaguchi Y, Suzuki T, Giraldez AJ, Tomari Y: **Poly(A)-specific ribonuclease mediates 3′-end trimming of Argonaute2-cleaved precursor microRNAs**. *Cell Rep* 2013, **5**(3):715–726.

[215] Devany E, Zhang X, Park JY, Tian B, Kleiman FE: **Positive and negative feedback loops in the p53 and mRNA 3′ processing pathways**. *Proc. Natl. Acad. Sci. U.S.A.* 2013, **110**(9):3351–3356.

[216] Asangani IA, Rasheed SA, Nikolova DA, Leupold JH, Colburn NH, Post S, Allgayer H: **MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer**. *Oncogene* 2008, **27**(15):2128–2136.

[217] Thum T, Gross C, Fiedler J, Fischer T, Kissler S, Bussen M, Galuppo P, Just S, Rottbauer W, Frantz S, Castoldi M, Soutschek J, Koteliansky V, Rosenwald A, Basson MA, Licht JD, Pena JT, Rouhanifard SH, Muckenthaler MU, Tuschl T, Martin GR, Bauersachs J, Engelhardt S: **MicroRNA-21 contributes to myocardial disease by stimulating MAP kinase signalling in fibroblasts**. *Nature* 2008, **456**(7224):980–984.

[218] Zhu S, Wu H, Wu F, Nie D, Sheng S, Mo YY: **MicroRNA-21 targets tumor suppressor genes in invasion and metastasis**. *Cell Res.* 2008, **18**(3):350–359.

[219] Fu Y, Shi Z, Wu M, Zhang J, Jia L, Chen X: **Identification and differential expression of microRNAs during metamorphosis of the Japanese flounder (Paralichthys olivaceus)**. *PLoS ONE* 2011, **6**(7):e22957.

[220] Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinformatics* 2010, **26**:139–140.

[221] Wilks SS: **The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses**. *Ann Math Statist.* 1938, **9**:60–62.

[222] Viola A, Luster AD: **Chemokines and their receptors: drug targets in immunity and inflammation**. *Annu. Rev. Pharmacol. Toxicol.* 2008, **48**:171–197.

[223] Zlotnik A, Yoshie O: **The chemokine superfamily revisited**. *Immunity* 2012, **36**(5):705–716.

[224] Karin N: **The multiple faces of CXCL12 (SDF-1alpha) in the regulation of immunity during health and disease**. *J. Leukoc. Biol.* 2010, **88**(3):463–473.

[225] Burger JA, Kipps TJ: **CXCR4: a key receptor in the crosstalk between tumor cells and their microenvironment**. *Blood* 2006, **107**(5):1761–1767.

[226] Duda DG, Kozin SV, Kirkpatrick ND, Xu L, Fukumura D, Jain RK: **CXCL12 (SDF1alpha)-CXCR4/CXCR7 pathway inhibition: an emerging sensitizer for anticancer therapies?** *Clin. Cancer Res.* 2011, **17**(8):2074–2080.

[227] O'Hayre M, Salanga CL, Handel TM, Allen SJ: **Chemokines and cancer: migration, intracellular signalling and intercellular communication in the microenvironment**. *Biochem. J.* 2008, **409**(3):635–649.

[228] Rajagopal S, Kim J, Ahn S, Craig S, Lam CM, Gerard NP, Gerard C, Lefkowitz RJ: **Beta-arrestin- but not G protein-mediated signaling by the "decoy" receptor CXCR7**. *Proc. Natl. Acad. Sci. U.S.A.* 2010, **107**(2):628–632.

[229] DeWire SM, Ahn S, Lefkowitz RJ, Shenoy SK: **Beta-arrestins and cell signaling**. *Annu. Rev. Physiol.* 2007, **69**:483–510.

[230] Bachelerie F, Ben-Baruch A, Burkhardt AM, Combadiere C, Farber JM, Graham GJ, Horuk R, Sparre-Ulrich AH, Locati M, Luster AD, Mantovani A, Matsushima K, Murphy PM, Nibbs R, Nomiyama H, Power CA, Proudfoot AE, Rosenkilde MM, Rot A, Sozzani S, Thelen M, Yoshie O, Zlotnik A: **International Union of Pharmacology. LXXXIX. Update on the extended family of chemokine receptors and introducing a new nomenclature for atypical chemokine receptors**. *Pharmacol. Rev.* 2014, **66**:1–79.

[231] Hattermann K, Mentlein R: **An infernal trio: the chemokine CXCL12 and its receptors CXCR4 and CXCR7 in tumor biology**. *Ann. Anat.* 2013, **195**(2):103–110.

[232] Burns JM, Summers BC, Wang Y, Melikian A, Berahovich R, Miao Z, Penfold ME, Sunshine MJ, Littman DR, Kuo CJ, Wei K, McMaster BE, Wright K, Howard MC, Schall TJ: **A novel chemokine receptor for SDF-1 and I-TAC involved in cell survival, cell adhesion, and tumor development**. *J. Exp. Med.* 2006, **203**(9):2201–2213.

[233] Balabanian K, Lagane B, Infantino S, Chow KY, Harriague J, Moepps B, Arenzana-Seisdedos F, Thelen M, Bachelerie F: **The chemokine SDF-1/CXCL12 binds to and signals through the orphan receptor RDC1 in T lymphocytes**. *J. Biol. Chem.* 2005, **280**(42):35760–35766.

[234] Decaillot FM, Kazmi MA, Lin Y, Ray-Saha S, Sakmar TP, Sachdev P: **CXCR7/CXCR4 heterodimer constitutively recruits beta-arrestin to enhance cell migration**. *J. Biol. Chem.* 2011, **286**(37):32188–32197.

[235] Levoye A, Balabanian K, Baleux F, Bachelerie F, Lagane B: **CXCR7 heterodimerizes with CXCR4 and regulates CXCL12-mediated G protein signaling**. *Blood* 2009, **113**(24):6085–6093.

[236] Luker KE, Gupta M, Luker GD: **Imaging chemokine receptor dimerization with firefly luciferase complementation**. *FASEB J.* 2009, **23**(3):823–834.

[237] Hattermann K, Held-Feindt J, Lucius R, Muerkoster SS, Penfold ME, Schall TJ, Mentlein R: **The chemokine receptor CXCR7 is highly expressed in human glioma cells and mediates antiapoptotic effects**. *Cancer Res.* 2010, **70**(8):3299–3308.

[238] Schimanski CC, Schwald S, Simiantonaki N, Jayasinghe C, Gonner U, Wilsberg V, Junginger T, Berger MR, Galle PR, Moehler M: **Effect of chemokine receptors CXCR4 and CCR7 on the metastatic behavior of human colorectal cancer**. *Clin. Cancer Res.* 2005, **11**(5):1743–1750.

[239] He H, Wang C, Shen Z, Fang Y, Wang X, Chen W, Liu F, Qin X, Sun Y: **Upregulated expression of C-X-C chemokine receptor 4 is an independent prognostic predictor for patients with gastric cancer**. *PLoS ONE* 2013, **8**(8):e71864.

[240] Oh YS, Kim HY, Song IC, Yun HJ, Jo DY, Kim S, Lee HJ: **Hypoxia induces CXCR4 expression and biological activity in gastric cancer cells through activation of hypoxia-inducible factor-1**α. *Oncol. Rep.* 2012, **28**(6):2239–2246.

[241] Zhu S, Hong J, Tripathi MK, Sehdev V, Belkhiri A, El-Rifai W: **Regulation of CXCR4-mediated invasion by DARPP-32 in gastric cancer cells**. *Mol. Cancer Res.* 2013, **11**:86–94.

[242] Lee HJ, Jo DY: **The role of the CXCR4/CXCL12 axis and its clinical implications in gastric cancer**. *Histol. Histopathol.* 2012, **27**(9):1155–1161.

[243] Maussang D, Mujic-Delic A, Descamps FJ, Stortelers C, Vanlandschoot P, Stigter-van Walsum M, Vischer HF, van Roy M, Vosjan M, Gonzalez-Pajuelo M, van Dongen GA, Merchiers P, van Rompaey P, Smit MJ: **Llama-derived single variable domains (nanobodies) directed against chemokine receptor CXCR7 reduce head and neck cancer cell growth in vivo**. *J. Biol. Chem.* 2013, **288**(41):29562–29572.

[244] Mukherjee D, Zhao J: **The Role of chemokine receptor CXCR4 in breast cancer metastasis**. *Am J Cancer Res* 2013, **3**:46–57.

[245] Hernandez L, Magalhaes MA, Coniglio SJ, Condeelis JS, Segall JE: **Opposing roles of CXCR4 and CXCR7 in breast cancer metastasis**. *Breast Cancer Res.* 2011, **13**(6):R128.

[246] Luker KE, Lewin SA, Mihalko LA, Schmidt BT, Winkler JS, Coggins NL, Thomas DG, Luker GD: **Scavenging of CXCL12 by CXCR7 promotes tumor growth and metastasis of CXCR4-positive breast cancer cells**. *Oncogene* 2012, **31**(45):4750–4758.

[247] Miao Z, Luker KE, Summers BC, Berahovich R, Bhojani MS, Rehemtulla A, Kleer CG, Essner JJ, Nasevicius A, Luker GD, Howard MC, Schall TJ: **CXCR7 (RDC1) promotes breast and lung tumor growth in vivo and is expressed on tumor-associated vasculature**. *Proc. Natl. Acad. Sci. U.S.A.* 2007, **104**(40):15735–15740.

[248] Tibes R, Qiu Y, Lu Y, Hennessy B, Andreeff M, Mills GB, Kornblau SM: **Reverse phase protein array: validation of a novel proteomic technology and utility for analysis of primary leukemia specimens and hematopoietic stem cells**. *Mol. Cancer Ther.* 2006, **5**(10):2512–2521.

[249] Zhang L, Wei Q, Mao L, Liu W, Mills GB, Coombes K: **Serial dilution curve: a new method for analysis of reverse phase protein array data**. *Bioinformatics* 2009, **25**(5):650–654.

[250] Heitzler D, Crepieux P, Poupon A, Clement F, Fages F, Reiter E: **Towards a systems biology approach of-G protein-coupled receptor signalling: challenges and expectations**. *C. R. Biol.* 2009, **332**(11):947–957.

[251] Wijtmans M, Maussang D, Sirci F, Scholten DJ, Canals M, Muji?-Deli? A, Chong M, Chatalic KL, Custers H, Janssen E, de Graaf C, Smit MJ, de Esch IJ, Leurs R: **Synthesis, modeling and functional activity of substituted styrene-amides as small-molecule CXCR7 agonists**. *Eur J Med Chem* 2012, **51**:184–192.

[252] Thoma G, Streiff MB, Kovarik J, Glickman F, Wagner T, Beerli C, Zerwes HG: **Orally bioavailable isothioureas block function of the chemokine receptor CXCR4 in vitro and in vivo**. *J. Med. Chem.* 2008, **51**(24):7915–7920.

[253] Gravel S, Malouf C, Boulais PE, Berchiche YA, Oishi S, Fujii N, Leduc R, Sinnett D, Heveker N: **The peptidomimetic CXCR4 antagonist TC14012 recruits beta-arrestin to CXCR7: roles of receptor domains**. *J. Biol. Chem.* 2010, **285**(49):37939–37943.

[254] Kalatskaya I, Berchiche YA, Gravel S, Limberg BJ, Rosenbaum JS, Heveker N: **AMD3100 is a CXCR7 ligand with allosteric agonist properties**. *Mol. Pharmacol.* 2009, **75**(5):1240–1247.

[255] Delgado-Martin C, Escribano C, Pablos JL, Riol-Blanco L, Rodriguez-Fernandez JL: **Chemokine CXCL12 uses CXCR4 and a signaling core formed by bifunctional Akt, extracellular signal-regulated kinase (ERK)1/2, and mammalian target of rapamycin complex 1 (mTORC1) proteins to control chemotaxis and survival simultaneously in mature dendritic cells**. *J. Biol. Chem.* 2011, **286**(43):37222–37236.

[256] Aksamitiene E, Kiyatkin A, Kholodenko BN: **Cross-talk between mitogenic Ras/MAPK and survival PI3K/Akt pathways: a fine balance**. *Biochem. Soc. Trans.* 2012, **40**:139–146.

[257] Carracedo A, Ma L, Teruya-Feldstein J, Rojo F, Salmena L, Alimonti A, Egia A, Sasaki AT, Thomas G, Kozma SC, Papa A, Nardella C, Cantley LC, Baselga J, Pandolfi PP: **Inhibition of mTORC1 leads to MAPK pathway activation through a PI3K-dependent feedback loop in human cancer**. *J. Clin. Invest.* 2008, **118**(9):3065–3074.

[258] Moelling K, Schad K, Bosse M, Zimmermann S, Schweneker M: **Regulation of Raf-Akt Cross-talk**. *J. Biol. Chem.* 2002, **277**(34):31099–31106.

[259] Kinkade CW, Castillo-Martin M, Puzio-Kuter A, Yan J, Foster TH, Gao H, Sun Y, Ouyang X, Gerald WL, Cordon-Cardo C, Abate-Shen C: **Targeting AKT/mTOR and ERK MAPK signaling inhibits hormone-refractory prostate cancer in a preclinical mouse model**. *J. Clin. Invest.* 2008, **118**(9):3051–3064.

[260] Park ES, Rabinovsky R, Carey M, Hennessy BT, Agarwal R, Liu W, Ju Z, Deng W, Lu Y, Woo HG, Kim SB, Cheong JH, Garraway LA, Weinstein JN, Mills GB, Lee JS, Davies MA: **Integrative analysis of proteomic signatures, mutations, and drug responsiveness in the NCI 60 cancer cell line set**. *Mol. Cancer Ther.* 2010, **9**(2):257–267.

[261] Kornblau SM, Qutub A, Yao H, York H, Qiu YH, Graber D, Ravandi F, Cortes J, Andreeff M, Zhang N, Coombes KR: **Proteomic profiling identifies distinct protein patterns in acute myelogenous leukemia CD34+CD38- stem-like cells**. *PLoS ONE* 2013, **8**(10):e78453.

[262] Furusato B, Mohamed A, Uhlen M, Rhim JS: **CXCR4 and cancer**. *Pathol. Int.* 2010, **60**(7):497–505.

[263] Lee E, Han J, Kim K, Choi H, Cho EG, Lee TR: **CXCR7 mediates SDF1-induced melanocyte migration**. *Pigment Cell Melanoma Res* 2013, **26**:58–66.

[264] Odemis V, Lipfert J, Kraft R, Hajek P, Abraham G, Hattermann K, Mentlein R, Engele J: **The presumed atypical chemokine receptor CXCR7 signals through G(i/o) proteins in primary rodent astrocytes and human glioma cells**. *Glia* 2012, **60**(3):372–381.

[265] Kumar R, Tripathi V, Ahmad M, Nath N, Mir RA, Chauhan SS, Luthra K: **CXCR7 mediated Giα independent activation of ERK and Akt promotes cell survival and chemotaxis in T cells**. *Cell. Immunol.* 2012, **272**(2):230–241.

[266] Steen A, Schwartz TW, Rosenkilde MM: **Targeting CXCR4 in HIV cell-entry inhibition**. *Mini Rev Med Chem* 2009, **9**(14):1605–1621.

[267] Heckmann D, Maier P, Laufs S, Wenz F, Zeller WJ, Fruehauf S, Allgayer H: **CXCR4 Expression and Treatment with SDF-1α or Plerixafor Modulate Proliferation and Chemosensitivity of Colon Cancer Cells**. *Transl Oncol* 2013, **6**(2):124–132.

[268] Chen FH, Fu SY, Yang YC, Wang CC, Chiang CS, Hong JH: **Combination of vessel-targeting agents and fractionated radiation therapy: the role of the SDF-1/CXCR4 pathway**. *Int. J. Radiat. Oncol. Biol. Phys.* 2013, **86**(4):777–784.

[269] Welschinger R, Liedtke F, Basnett J, Dela Pena A, Juarez JG, Bradstock KF, Bendall LJ: **Plerixafor (AMD3100) induces prolonged mobilization of acute lymphoblastic leukemia cells and increases the proportion of cycling cells in the blood in mice**. *Exp. Hematol.* 2013, **41**(3):293–302.

[270] Jahnichen S, Blanchetot C, Maussang D, Gonzalez-Pajuelo M, Chow KY, Bosch L, De Vrieze S, Serruys B, Ulrichts H, Vandevelde W, Saunders M, De Haard HJ, Schols D, Leurs R, Vanlandschoot P, Verrips T, Smit MJ: **CXCR4 nanobodies (VHH-based single variable domains) potently inhibit chemotaxis and HIV-1 replication and mobilize stem cells**. *Proc. Natl. Acad. Sci. U.S.A.* 2010, **107**(47):20565–20570.

[271] Wang J, Shiozawa Y, Wang J, Wang Y, Jung Y, Pienta KJ, Mehra R, Loberg R, Taichman RS: **The role of CXCR7/RDC1 as a chemokine receptor for CXCL12/SDF-1 in prostate cancer**. *J. Biol. Chem.* 2008, **283**(7):4283–4294.

[272] Muller A, Homey B, Soto H, Ge N, Catron D, Buchanan ME, McClanahan T, Murphy E, Yuan W, Wagner SN, Barrera JL, Mohar A, Verastegui E, Zlotnik A: **Involvement of chemokine receptors in breast cancer metastasis**. *Nature* 2001, **410**(6824):50–56.

[273] Phillips RJ, Burdick MD, Lutz M, Belperio JA, Keane MP, Strieter RM: **The stromal derived factor-1/CXCL12-CXC chemokine receptor 4 biological axis in non-small cell lung cancer metastases**. *Am. J. Respir. Crit. Care Med.* 2003, **167**(12):1676–1686.

[274] Verzijl D, Storelli S, Scholten DJ, Bosch L, Reinhart TA, Streblow DN, Tensen CP, Fitzsimons CP, Zaman GJ, Pease JE, de Esch IJ, Smit MJ, Leurs R: **Noncompetitive antagonism and inverse agonism as mechanism of action of nonpeptidergic antagonists at primate and rodent CXCR3 chemokine receptors**. *J. Pharmacol. Exp. Ther.* 2008, **325**(2):544–555.

[275] Oberhardt MA, Palsson BO, Papin JA: **Applications of genome-scale metabolic reconstructions**. *Mol. Syst. Biol.* 2009, **5**:320.

[276] Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: **High-throughput generation, optimization and analysis of genome-scale metabolic models**. *Nat. Biotechnol.* 2010, **28**(9):977–982.

[277] Schellenberger J, Que R, Fleming RM, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BO: **Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0**. *Nat Protoc* 2011, **6**(9):1290–1307.

[278] Rocha I, Maia P, Evangelista P, Vilaca P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC, Rocha M: **OptFlux: an open-source software platform for in silico metabolic engineering**. *BMC Syst Biol* 2010, **4**:45.

[279] Klamt S, Saez-Rodriguez J, Gilles ED: **Structural and functional analysis of cellular networks with CellNetAnalyzer**. *BMC Syst Biol* 2007, **1**:2.

[280] Funahashi A, Tanimura N, Morohashi M, Kitano H: **CellDesigner: a process diagram editor for gene-regulatory and biochemical networks**. *BIOSILICO* 2003, **1**:159–162.

[281] Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes**. *Nucleic Acids Res.* 2000, **28**:27–30.

[282] Schellenberger J, Park JO, Conrad TM, Palsson BO: **BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions**. *BMC Bioinformatics* 2010, **11**:213.

[283] **Main Page - SBML.org** [http://sbml.org/Documents/Specifications/SBML_Level_3/Packages/Flux_Balance_Constraints_(flux)].

[284] Duarte NC, Herrgard MJ, Palsson BO: **Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model**. *Genome Res.* 2004, **14**(7):1298–1309.

[285] Schwarz R, Liang C, Kaleta C, Kuhnel M, Hoffmann E, Kuznetsov S, Hecker M, Griffiths G, Schuster S, Dandekar T: **Integrated network reconstruction, visualization and analysis using YANAsquare**. *BMC Bioinformatics* 2007, **8**:313.

[286] Kono N, Arakawa K, Tomita M: **MEGU: pathway mapping web-service based on KEGG and SVG**. *In Silico Biol. (Gedrukt)* 2006, **6**(6):621–625.

[287] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome Res.* 2003, **13**(11):2498–2504.

[288] Gstaiger M, Aebersold R: **Applying mass spectrometry-based proteomics to genetics, genomics and network biology**. *Nat. Rev. Genet.* 2009, **10**(9):617–627.

[289] Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics**. *Nat. Rev. Genet.* 2009, **10**:57–63.

[290] Kahn SD: **On the future of genomic data**. *Science* 2011, **331**(6018):728–729.

[291] Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A: **Full-length transcriptome assembly from RNA-Seq data without a reference genome**. *Nat. Biotechnol.* 2011, **29**(7):644–652.

[292] Berger B, Peng J, Singh M: **Computational solutions for omics data**. *Nat. Rev. Genet.* 2013, **14**(5):333–346.

[293] Price ND, Papin JA, Schilling CH, Palsson BO: **Genome-scale microbial in silico models: the constraints-based approach**. *Trends Biotechnol.* 2003, **21**(4):162–169.

[294] Maarleveld TR, Khandelwal RA, Olivier BG, Teusink B, Bruggeman FJ: **Basic concepts and principles of stoichiometric modeling of metabolic networks**. *Biotechnol J* 2013, **8**(9):997–1008.

[295] McCloskey D, Palsson BO, Feist AM: **Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli**. *Mol. Syst. Biol.* 2013, **9**:661.

[296] Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T: **Cytoscape 2.8: new features for data integration and network visualization**. *Bioinformatics* 2011, **27**(3):431–432.

[297] Yamada T, Letunic I, Okuda S, Kanehisa M, Bork P: **iPath2.0: interactive pathway explorer**. *Nucleic Acids Res.* 2011, **39**(Web Server issue):W412–415.

[298] Boele J, Olivier BG, Teusink B: **FAME, the Flux Analysis and Modeling Environment**. *BMC Syst Biol* 2012, **6**:8.

[299] Ebrahim A, Lerman JA, Palsson BO, Hyduke DR: **COBRApy: COnstraints-Based Reconstruction and Analysis for Python**. *BMC Syst Biol* 2013, **7**:74.

[300] Shastri AA, Morgan JA: **Flux balance analysis of photoautotrophic metabolism**. *Biotechnol. Prog.* 2005, **21**(6):1617–1626.

[301] Knoop H, Zilliges Y, Lockau W, Steuer R: **The metabolic network of Synechocystis sp. PCC 6803: systemic properties of autotrophic growth**. *Plant Physiol.* 2010, **154**:410–422.

[302] Nogales J, Gudmundsson S, Knight EM, Palsson BO, Thiele I: **Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis**. *Proc. Natl. Acad. Sci. U.S.A.* 2012, **109**(7):2678–2683.

[303] Saha R, Verseput AT, Berla BM, Mueller TJ, Pakrasi HB, Maranas CD: **Reconstruction and comparison of the metabolic potential of cyanobacteria Cyanothece sp. ATCC 51142 and Synechocystis sp. PCC 6803**. *PLoS ONE* 2012, **7**(10):e48285.

[304] Chisti Y: **Biodiesel from microalgae**. *Biotechnol. Adv.* 2007, **25**(3):294–306.

[305] Ducat DC, Way JC, Silver PA: **Engineering cyanobacteria to generate high-value products**. *Trends Biotechnol.* 2011, **29**(2):95–103.

[306] Angermayr SA, Paszota M, Hellingwerf KJ: **Engineering a cyanobacterial cell factory for production of lactic acid**. *Appl. Environ. Microbiol.* 2012, **78**(19):7098–7106.

[307] Ganter M, Bernard T, Moretti S, Stelling J, Pagni M: **MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks**. *Bioinformatics* 2013, **29**(6):815–816.

[308] Schreiber F: **High quality visualization of biochemical pathways in BioPath**. *In Silico Biol. (Gedrukt)* 2002, **2**(2):59–73.

[309] Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, Karp PD: **The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases**. *Nucleic Acids Res.* 2012, **40**(Database issue):D742–753.

[310] Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Ruckert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V: **The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes**. *Nucleic Acids Res.* 2005, **33**(17):5691–5702.

[311] Morgat A, Coissac E, Coudert E, Axelsen KB, Keller G, Bairoch A, Bridge A, Bougueleret L, Xenarios I, Viari A: **UniPathway: a resource for the exploration and annotation of metabolic pathways**. *Nucleic Acids Res.* 2012, **40**(Database issue):D761–769.

[312] Lang M, Stelzer M, Schomburg D: **BKM-react, an integrated biochemical reaction database**. *BMC Biochem.* 2011, **12**:42.

[313] Bernard T, Bridge A, Morgat A, Moretti S, Xenarios I, Pagni M: **Reconciliation of metabolites and biochemical reactions for metabolic networks**. *Brief. Bioinformatics* 2012.

[314] Kumar A, Suthers PF, Maranas CD: **MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases**. *BMC Bioinformatics* 2012, **13**:6.

[315] Ram K: **Git can facilitate greater reproducibility and increased transparency in science**. *Source Code Biol Med* 2013, **8**:7.

[316] Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, Bork P, Goto S, Kanehisa M: **KEGG Atlas mapping for global analysis of metabolic pathways**. *Nucleic Acids Res.* 2008, **36**(Web Server issue):W423–426.

[317] Kitano H, Funahashi A, Matsuoka Y, Oda K: **Using process diagrams for the graphical representation of biological networks**. *Nat. Biotechnol.* 2005, **23**(8):961–966.

[318] Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD: **Integration of biological networks and gene expression data using Cytoscape**. *Nat Protoc* 2007, **2**(10):2366–2382.

[319] Kostromins A, Stalidzans E: **Paint4Net: COBRA Toolbox extension for visualization of stoichiometric models of metabolism**. *BioSystems* 2012, **109**(2):233–239.

[320] Konig M, Holzhutter HG: **Fluxviz - Cytoscape plug-in for visualization of flux distributions in networks**. *Genome Inform* 2010, **24**:96–103.

[321] Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J: **The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum**. *PLoS Comput. Biol.* 2013, **9**(3):e1002980.

[322] Maarleveld TR, Boele J, Bruggeman FJ, Teusink B: **A Data Integration and Visualization Resource for the Metabolic Network of Synechocystis sp. PCC 6803**. *Plant Physiol.* 2014.

[323] Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C: **WikiPathways: pathway editing for the people**. *PLoS Biol.* 2008, **6**(7):e184.

[324] Le Novere N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villeger A, Boyd SE, Calzone L, Courtot M, Dogrusoz U, Freeman TC, Funahashi A, Ghosh S, Jouraku A, Kim S, Kolpakov F, Luna A, Sahle S, Schmidt E, Watterson S, Wu G, Goryanin I, Kell DB, Sander C, Sauro H, Snoep JL, Kohn K, Kitano H: **The Systems Biology Graphical Notation**. *Nat. Biotechnol.* 2009, **27**(8):735–741.

[325] Gauges R, Rost U, Sahle S, Wegner K: **A model diagram layout extension for SBML**. *Bioinformatics* 2006, **22**(15):1879–1885.

[326] Santos F, Boele J, Teusink B: **A practical guide to genome-scale metabolic models and their analysis**. *Meth. Enzymol.* 2011, **500**:509–532.

[327] Thiele I, Price ND, Vo TD, Palsson B: **Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet**. *J. Biol. Chem.* 2005, **280**(12):11683–11695.

[328] Oliveira AP, Nielsen J, Forster J: **Modeling Lactococcus lactis using a genome-scale flux model**. *BMC Microbiol.* 2005, **5**:39.

[329] Teusink B, Wiersma A, Molenaar D, Francke C, de Vos WM, Siezen RJ, Smid EJ: **Analysis of growth of Lactobacillus plantarum WCFS1 on a complex medium using a genome-scale metabolic model**. *J. Biol. Chem.* 2006, **281**(52):40041–40048.

[330] Herrgard MJ, Swainston N, Dobson P, Dunn WB, Arga KY, Arvas M, Bluthgen N, Borger S, Costenoble R, Heinemann M, Hucka M, Le Novere N, Li P, Liebermeister W, Mo ML, Oliveira AP, Petranovic D, Pettifer S, Simeonidis E, Smallbone K, Spasi? I, Weichart D, Brent R, Broomhead DS, Westerhoff HV, Kirdar B, Penttila M, Klipp E, Palsson BO, Sauer U, Oliver SG, Mendes P, Nielsen J, Kell DB: **A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology**. *Nat. Biotechnol.* 2008, **26**(10):1155–1160.

[331] Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E: **Network-based prediction of human tissue-specific metabolism**. *Nat. Biotechnol.* 2008, **26**(9):1003–1010.

[332] Holzhutter HG: **The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks**. *Eur. J. Biochem.* 2004, **271**(14):2905–2922.

[333] Neidhardt FC: **Chemical composition of Escherichia coli.** In *Escherichia coli and Salmonella typhimurium: Cellular and molecular biology*. Edited by Neidhardt FC, Washington, D.C.: American Society for Microbiology 1987.

[334] Osterlund T, Nookaew I, Bordel S, Nielsen J: **Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling**. *BMC Syst Biol* 2013, **7**:36.

[335] Sheikh K, Forster J, Nielsen LK: **Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of Mus musculus**. *Biotechnol. Prog.* 2005, **21**:112–121.

[336] Altamirano C, Illanes A, Casablancas A, Gamez X, Cairo JJ, Godia C: **Analysis of CHO cells metabolic redistribution in a glutamate-based defined medium in continuous culture**. *Biotechnol. Prog.* 2001, **17**(6):1032–1041.

[337] Gallagher R, Collins S, Trujillo J, McCredie K, Ahearn M, Tsai S, Metzgar R, Aulakh G, Ting R, Ruscetti F, Gallo R: **Characterization of the continuous, differentiating myeloid cell line (HL-60) from a patient with acute promyelocytic leukemia**. *Blood* 1979, **54**(3):713–733.

[338] Kilburn DG, Lilly MD, Webb FC: **The energetics of mammalian cell growth**. *J. Cell. Sci.* 1969, **4**(3):645–654.

[339] Smallbone K: **Striking a balance with Recon 2.1**. *arXiv* 2013, :1311.5696 [q–bio.MN].

[340] Mogard MH, Kobayashi R, Chen CF, Lee TD, Reeve JR, Shively JE, Walsh JH: **The amino acid sequence of kinetensin, a novel peptide isolated from pepsin-treated human plasma: homology with human serum albumin, neurotensin and angiotensin**. *Biochem. Biophys. Res. Commun.* 1986, **136**(3):983–988.

[341] Raman K, Chandra N: **Flux balance analysis of biological systems: applications and challenges**. *Brief. Bioinformatics* 2009, **10**(4):435–449.

[342] Jerby L, Shlomi T, Ruppin E: **Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism**. *Mol. Syst. Biol.* 2010, **6**:401.

[343] Husek P: **Amino acid derivatization and analysis in five minutes**. *FEBS Lett.* 1991, **280**(2):354–356.

[344] Masuko T, Minami A, Iwasaki N, Majima T, Nishimura S, Lee YC: **Carbohydrate analysis by a phenol-sulfuric acid method in microplate format**. *Anal. Biochem.* 2005, **339**:69–72.

[345] Klitgord N, Segre D: **Environments that induce synthetic microbial ecosystems**. *PLoS Comput. Biol.* 2010, **6**(11):e1001002.

[346] Ansari S, Binder J, Boue S, Di Fabio A, Hayes W, Hoeng J, Iskandar A, Kleiman R, Norel R, O'Neel B, Peitsch MC, Poussin C, Pratt D, Rhrissorrakrai K, Schlage WK, Stolovitzky G, Talikka M: **On Crowd-verification of Biological Networks**. *Bioinform Biol Insights* 2013, **7**:307–325.

[347] Crick F: **Central dogma of molecular biology**. *Nature* 1970, **227**(5258):561–563.

[348] Britten RJ, Davidson EH: **Gene regulation for higher cells: a theory**. *Science* 1969, **165**(3891):349–357.

[349] Ghildiyal M, Zamore PD: **Small silencing RNAs: an expanding universe**. *Nat. Rev. Genet.* 2009, **10**(2):94–108.

[350] Liu Q, Paroo Z: **Biochemical principles of small RNA pathways**. *Annu. Rev. Biochem.* 2010, **79**:295–319.

[351] Zhang Z, Qin YW, Brewer G, Jing Q: **MicroRNA degradation and turnover: regulating the regulators**. *Wiley Interdiscip Rev RNA* 2012, **3**(4):593–600.

[352] Iadevaia S, Lu Y, Morales FC, Mills GB, Ram PT: **Identification of optimal drug combinations targeting cellular networks: integrating phospho-proteomics and computational network analysis**. *Cancer Res.* 2010, **70**(17):6704–6714.

[353] Fell DA, Small JR: **Fat synthesis in adipose tissue. An examination of stoichiometric constraints**. *Biochem. J.* 1986, **238**(3):781–786.

[354] Hill J, Mannheim B: **Language and Worldview**. *Annual Reviews in Anthropology* 1992, **21**:381–406.

[355] Hoijer H: **The Sapir-Whorf hypothesis**. In *Language in culture: Conference on the interrelations of language and other aspects of culture*. Edited by Hoijer H, Chicago: University of Chicago Press 1954:92–105.

[356] Joppa LN, McInerny G, Harper R, Salido L, Takeda K, O'Hara K, Gavaghan D, Emmott S: **Computational science. Troubling trends in scientific software use**. *Science* 2013, **340**(6134):814–815.

[357] Sanders R, Kelly D: **Dealing with Risk in Scientific Software Development**. *IEEE Software* 2008, **25**(4):21–28.

[358] Illig T, Gieger C, Zhai G, Romisch-Margl W, Wang-Sattler R, Prehn C, Altmaier E, Kastenmuller G, Kato BS, Mewes HW, Meitinger T, de Angelis MH, Kronenberg F, Soranzo N, Wichmann HE, Spector TD, Adamski J, Suhre K: **A genome-wide perspective of genetic variation in human metabolism**. *Nat. Genet.* 2010, **42**(2):137–141.

[359] Geraghty DE, Daza R, Williams LM, Vu Q, Ishitani A: **Genetics of the immune response: identifying immune variation within the MHC and throughout the genome**. *Immunol. Rev.* 2002, **190**:69–85.

[360] Hernandez PA, Gorlin RJ, Lukens JN, Taniuchi S, Bohinjec J, Francois F, Klotman ME, Diaz GA: **Mutations in the chemokine receptor gene CXCR4 are associated with WHIM syndrome, a combined immunodeficiency disease**. *Nat. Genet.* 2003, **34**:70–74.

[361] Petersen DC, Glashoff RH, Shrestha S, Bergeron J, Laten A, Gold B, van Rensburg EJ, Dean M, Hayes VM: **Risk for HIV-1 infection associated with a common CXCL12 (SDF1) polymorphism and CXCR4 variation in an African population**. *J. Acquir. Immune Defic. Syndr.* 2005, **40**(5):521–526.

[362] Csete M, Doyle J: **Bow ties, metabolism and disease**. *Trends Biotechnol.* 2004, **22**(9):446–450.

[363] Zhao J, Yu H, Luo JH, Cao ZW, Li YX: **Hierarchical modularity of nested bow-ties in metabolic networks**. *BMC Bioinformatics* 2006, **7**:386.

[364] Tieri P, Grignolio A, Zaikin A, Mishto M, Remondini D, Castellani GC, Franceschi C: **Network, degeneracy and bow tie integrating paradigms and architectures to grasp the complexity of the immune system**. *Theor Biol Med Model* 2010, **7**:32.

[365] Butler D: **Souped-up search engines**. *Nature* 2000, **405**(6783):112–115.

[366] Oda K, Kitano H: **A comprehensive map of the toll-like receptor signaling network**. *Mol. Syst. Biol.* 2006, **2**:2006.0015.

[367] Citri A, Yarden Y: **EGF-ERBB signalling: towards the systems level**. *Nat. Rev. Mol. Cell Biol.* 2006, **7**(7):505–516.

[368] Cabioglu N, Summy J, Miller C, Parikh NU, Sahin AA, Tuzlali S, Pumiglia K, Gallick GE, Price JE: **CXCL-12/stromal cell-derived factor-1alpha transactivates HER2-neu in breast cancer cells by a novel pathway involving Src kinase activation**. *Cancer Res.* 2005, **65**(15):6493–6497.

[369] Hartmann TN, Grabovsky V, Pasvolsky R, Shulman Z, Buss EC, Spiegel A, Nagler A, Lapidot T, Thelen M, Alon R: **A crosstalk between intracellular CXCR7 and CXCR4 involved in rapid CXCL12-triggered integrin activation but not in chemokine-triggered motility of human T lymphocytes and CD34+ cells**. *J. Leukoc. Biol.* 2008, **84**(4):1130–1140.

[370] Singh AK, Arya RK, Trivedi AK, Sanyal S, Baral R, Dormond O, Briscoe DM, Datta D: **Chemokine receptor trio: CXCR3, CXCR4 and CXCR7 crosstalk via CXCL11 and CXCL12**. *Cytokine Growth Factor Rev.* 2013, **24**:41–49.

[371] Kholodenko BN, Kiyatkin A, Bruggeman FJ, Sontag E, Westerhoff HV, Hoek JB: **Untangling the wires: a strategy to trace functional interactions in signaling and gene networks**. *Proc. Natl. Acad. Sci. U.S.A.* 2002, **99**(20):12841–12846.

[372] Raia V, Schilling M, Bohm M, Hahn B, Kowarsch A, Raue A, Sticht C, Bohl S, Saile M, Moller P, Gretz N, Timmer J, Theis F, Lehmann WD, Lichter P, Klingmuller U: **Dynamic mathematical modeling of IL13-induced signaling in Hodgkin and primary mediastinal B-cell lymphoma allows prediction of therapeutic targets**. *Cancer Res.* 2011, **71**(3):693–704.

[373] Schilling M, Maiwald T, Hengl S, Winter D, Kreutz C, Kolch W, Lehmann WD, Timmer J, Klingmuller U: **Theoretical and experimental analysis links isoform-specific ERK signalling to cell fate decisions**. *Mol. Syst. Biol.* 2009, **5**:334.

[374] Kolodkin A, Sahin N, Phillips A, Hood SR, Bruggeman FJ, Westerhoff HV, Plant N: **Optimization of stress response through the nuclear receptor-mediated cortisol signalling network**. *Nat Commun* 2013, **4**:1792.

[375] Sackmann A, Heiner M, Koch I: **Application of Petri net based analysis techniques to signal transduction pathways**. *BMC Bioinformatics* 2006, **7**:482.

[376] Chaouiya C: **Petri net modelling of biological networks**. *Brief. Bioinformatics* 2007, **8**(4):210–219.

[377] Breitling R, Gilbert D, Heiner M, Orton R: **A structured approach for the engineering of biochemical network models, illustrated for signalling pathways**. *Brief. Bioinformatics* 2008, **9**(5):404–421.

[378] Chen M, Hofestadt R: **Quantitative Petri net model of gene regulated metabolic networks in the cell**. *In Silico Biol. (Gedrukt)* 2003, **3**(3):347–365.

[379] Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO: **Reconstruction of biochemical networks in microorganisms**. *Nat. Rev. Microbiol.* 2009, **7**(2):129–143.

[380] Li L, Zhou X, Ching WK, Wang P: **Predicting enzyme targets for cancer drugs by profiling human metabolic reactions in NCI-60 cell lines**. *BMC Bioinformatics* 2010, **11**:501.

[381] Dawes M, Summerskill W, Glasziou P, Cartabellotta A, Martin J, Hopayian K, Porzsolt F, Burls A, Osborne J: **Sicily statement on evidence-based practice**. *BMC Med Educ* 2005, **5**:1.

[382] Swierstra T, Vermeulen N, Braeckman J, van Driel R: **Rethinking the life sciences. To better serve society, biomedical research has to regain its trust and get organized to tackle larger projects**. *EMBO Rep.* 2013, **14**(4):310–314.

Nederlandse samenvatting

Over systemen en kanker

Al sinds de jaren '30 van de vorige eeuw weten we dat het metabolisme van kankercellen afwijkt van dat van normale cellen. De bekendste uiting hiervan is het Warburg-effect, de observatie dat kankercellen glucose op inefficiënte wijze omzetten in (2) ATP en melkzuur, ook als er voldoende zuurstof aanwezig is om het via oxidatieve phosphorylering voor ~30 ATP te verbranden. Voorgestelde verklaringen voor het typische kanker-fenotype zetten meestal in op de signaleringsmachinerie of op het metabolisme, maar het ontbreken van dé verklaring voor kanker is wellicht zelf al de beste illustratie van de behoefte aan een systeembenadering die beide velden verenigt. In dit proefschrift belicht ik enkele onderwerpen die bijdragen tot het afwijkende metabole fenotype van kanker, en doe ik een poging die te begrijpen binnen het grotere raamwerk van de cel als geheel van systemen.

Cellulaire signalering wordt vaak beschouwd als het mechanisme dat aan de hand van stimuli uit de omgeving het lot van de cel bepaalt. Binnen deze regulatiemachinerie hebben microRNAs zich in recente jaren een plaats verworven, en ook in de etiologie van kanker laten ze zich niet onbetuigd. Een van de archetypische oncomiRs, miR-21, is in deze laatste context geen onbekende, en keer op keer blijkt dit microRNA bij allerlei typische kankerkenmerken betrokken te zijn. In borstkankercellen, waarin overexpressie van miR-21 eerder regel dan uitzondering is, vonden we indicaties dat bepaalde 3′ isovormen van miR-21 het doelwit zijn van een enzymatisch degradatiemechanisme. Dit mechanisme, dat we daarna ook in allerlei andere celtypen en organismen terugvonden, werkt doordat de tumor suppressor PAPD5 miR-21+C eerst adenyleert, waardoor het microRNA vervolgens vatbaar wordt

voor degradatie door een exoribonuclease, waarschijnlijk PARN (Hoofdstuk 2).

Het onderzoek naar de bijdrage van individuele microRNAs aan het algehele (kanker)fenotype maakt deel uit van een relatief jong en dynamisch veld, waarin nog veel onbekend is. De rol die de "gewone" signaleringsmachinerie speelt wordt al langer onderkend, en het napluizen van de interacties tussen signaleringseiwitten en de pathways die zij samen vormen heeft al vele nuttige inzichten opgeleverd. In dit kader wordt behalve enzymkinetiek tegenwoordig ook de samenhang tussen eiwitten en pathways op systeembiologische wijze onderzocht.

In dit verband bestudeerden wij het netwerk van reacties dat het gevolg is van de activatie van de chemokinereceptoren CXCR4 en CXCR7 door hun ligand, de chemokine CXCL12. De betrokkenheid van deze receptoren bij het verloop van kanker is bekend, maar wat hun precieze bijdrage is is nog onduidelijk, vooral in het geval van CXCR7. Met behulp van Reverse Phase Protein Arrays, een op antilichamen gebaseerde technologie, konden we in een tijdsreeks niet alleen de relatieve hoeveelheid van allerlei signaleringseiwitten bepalen, maar ook een onderscheid maken tussen de gefosforyleerde en niet-gefosforyleerde varianten van deze eiwitten. Door de tijdsreeks te herhalen met toevoeging van inhibitors van CXCR4, CXCR7, of beide receptoren, konden we vervolgens de bijdragen van de twee eiwitten onderscheiden, en vaststellen dat p42/44 MAPK als gevolg van CXCR7-activatie gefosforyleerd wordt, hetgeen we vervolgens met Western blots bevestigden (Hoofdstuk 3).

In contrast tot de signaleringsmachinerie staat het metabolisme. Metabole enzymen bewerken in de regel kleine moleculen in plaats van andere eiwitten, en hierdoor is voor het op grote (genoom)schaal bestuderen van het metabolisme een heel andere aanpak nodig dan voor onderzoek naar signalering. Zelfs de op kinetiek gebaseerde aanpak die voor een kleinschaliger bestudering van het metabolisme in het verleden zeer nuttig is gebleken, staat buitenspel wanneer inzicht in het grotere plaatje het doel is.

Het antwoord op deze uitdaging wordt gevormd zogenoemde genome-scale metabolic models (GSMM's), waarin weliswaar informatie over regulering en kinetiek expres wordt weggelaten, maar die toch op basis van de omgeving van de cel en de structuur van een complex en groot netwerk van metabole reacties licht kunnen werpen op de eigenschappen van het systeem. Door de omvang van de hiervoor gebruikte modellen was het lastig om deze modellen op een handige manier samen te stellen en te 'ondervragen', nog

voordat men aan de oorspronkelijke biologische uitdaging toekwam. Om deze barrière te slechten voor zowel onderzoekers (inclusief ondergetekende) en onderwijzers hebben we FAME ontwikkeld, een grafische online tool voor het makkelijker werken met GSMM's, voordat we aan het 'echte' modelleerwerk begonnen (Hoofdstuk 4).

Een tweede uitdaging bij het werken met GSMM's wordt gevormd door de grote hoeveelheden data die resulteren uit simulaties. Het overgrote deel van deze data is helemaal niet interessant, maar zonder op maat gemaakte analyseprogrammatuur was het zoeken naar de biologische betekenis van resultaten een haast Herculeaanse onderneming. Om dit analyseproces mensvriendelijker te maken hebben we grote metabole landkaarten ontworpen die zowel door computers te lezen en bewerken zijn, als door mensen te interpreteren zijn. Door hergebruik van bestaande technologie hebben we deze kaarten vervolgens geïntegreerd in FAME, om ze voor alle onderzoekers van nut te laten zijn (Hoofdstuk 5).

Het uiteindelijke doel, het bestuderen van de eigenschappen van het menselijk metabolisme in kankercellen, was hiermee in zicht. Een zeer volledige reconstructie van het metabole netwerk dat gecodeerd wordt door het gehele menselijk genoom was hiervoor beschikbaar in de vorm van 'Recon2'. Het uiteindelijke doel van de simulatie van kankermetabolisme, de productie van nog meer kankercellen, is echter zeer afhankelijk van de in het gebruikte model aanwezige biomassavergelijking. Omdat in het geval van Recon2 sprake was van een uit allerlei verschillende (niet humane) diersoorten opgebouwde biomassareactie, hebben we de eerste experimenteel verkregen op GSMM's gerichte biomassavergelijking bepaald. Na het als 'objective function' opnemen van deze op de HL-60 leukemiecellijn gebaseerde biomassareactie in Recon2 bepaalden we vervolgens wat met het oog op de structuur van het metabole netwerk de minimale (medium)vereisten zijn voor de groei van deze kankercellen. Opvallend hierbij was dat de resultaten de aandacht vestigden op verscheidene vaak over het hoofd geziene vetzuren. Daar deze vetzuren niet aanwezig zijn in het gedefinieerde gedeelte van RPMI1640-medium, moesten deze dus wel afkomstig zijn geweest uit het serumgedeelte van dit groeimedium; we konden dit inderdaad experimenteel bevestigen (Hoofdstuk 6).

English summary

It has long been known that cancer has aberrant metabolic properties compared to healthy cells, including, famously, the tendency to forgo oxidative phosphorylation and produce lactate instead, even when oxygen is present. In the study of these properties, traditional metabolism-only and regulation-only approaches have shed some light on the matter, but no single explanation has covered all aspects of the Warburg effect — let alone all hallmarks of cancer. This thesis attempts to view the various systems that contribute to cancer's reprogramming of energy metabolism in a broader context, to shed light on the mechanisms that underlie cellular function in health and disease.

Regulation is traditionally viewed as a "governing mechanism" in cellular function, as the system that decides cell's responses to its environment as well as its eventual fate. MicroRNAs have relatively recently secured their position in our concept of the regulatory landscape, as well as in the study of cancer etiology. MiR-21, as one of the archetypical oncomiRs, is quite famous in the latter context, and its involvement in various cancer hallmarks (including cancer metabolism) has been demonstrated in the past. In breast cancer, a cancer type in which miR-21 is known to be strongly overexpressed, we discovered that an enzymatic degradation mechanism specifically operates on certain 3′ isoforms of miR-21. In this mechanism, which was also found in a variety of other cell types and species, miR-21+C is first adenylated by the tumor suppressor PAPD5, which renders the microRNA susceptible to a subsequent degradation step by an exoribonuclease, probably PARN (Chapter 2).

While the elucidation of the contribution of individual microRNAs to the cancer phenotype is a budding field of research, "regular" signaling enzymes have had their role in cancer appreciated for quite some time, and in the past,

the identification of the roles of individual proteins within signaling pathways and the roles of these pathways in cancer has proven fruitful on many occasions. We studied the downstream network of the chemokine receptors CXCR4 and CXCR7, which are both activated by the chemokine CXCL12, using a time course of reverse-phase protein array (RPPA) assays. The involvement of chemokine receptors in cancer is well known, but the precise extent of this involvement is unclear, especially in the case of CXCR7. As RPPA is an antibody-based technology, it can distinguish between phosphorylated and non-phosphorylated isoforms of the same signaling enzyme, and their relative levels can be tracked over a time course of CXCL12 stimulation after inhibition of either CXCR4 or CXCR7, or of both. Using RPPA data, we identified p42/44 MAPK phosphorylation as an event downstream of CXCR7 activation, and we then confirmed this in separate Western blotting experiments (Chapter 3).

In contrast to the signaling machinery, metabolic enzymes primarily act on small molecules rather than on other enzymes. Therefore, the tools used to study metabolism on a genome-wide scale are vastly different from those used to study signaling, and even from those traditionally used to study metabolism on a smaller scale (namely, kinetic models).

Genome-scale metabolic models (GSMMs), although a coarse-grained approach that purposely ignores various regulatory and kinetic aspects, can shed light on the properties of a large metabolic system based on its structure and on the specifications of its environment. However, their size and format made working with them an exercise in software installation and troubleshooting in addition to the biological challenge at hand. To help overcome this impediment for both researchers (including yours truly) and educators, we developed FAME, an install-free graphical online tool for working with GSMMs, before doing our modeling work (Chapter 4).

A subsequent secondary challenge was the interpretation of the vast amounts of results that GSMM analyses produce. Without custom-made computational support, finding the biological phenomenon of interest if one does not know where to look beforehand is an almost Herculean task. Thus, to make the analysis process more human-friendly, we developed hand-drawn maps of metabolism that can be read by computers as well as intuitively interpreted by researchers. By repurposing existing technology to make the map adapt to the model under study as well as integrate with FAME, the results interpretation step of GSMM research should be greatly facilitated (Chapter 5).

We then proceeded to study the properties of human metabolism in cancer cells. A very comprehensive reconstruction of human metabolism ("Recon2") was available for just this purpose, but its biomass equation being derived from a combination of non-human species, we first experimentally determined the first human biomass equation for use in genome-scale models. After including the biomass composition of HL-60 human leukemia cells as an objective function in Recon2, we determined the requirements for growth that this biomass equation places on the medium. We discovered that several fatty acids are essential for HL-60 growth, and since the defined part of RPMI1640 medium does not contain them, we separately confirmed their presence in serum-containing medium (Chapter 6).

Acknowledgments (dankwoord)

This page intentionally left blank in this version.

This page intentionally left blank in this version.

This page intentionally left blank in this version.

List of publications

Santos F., **Boele J.**, Teusink B. (2011) A practical guide to genome-scale metabolic models and their analysis. *Methods in Enzymology* **500**: 509-532. doi:10.1016/B978-0-12-385118-5.00024-4.

Boele J., Olivier B.G., Teusink B. (2012) FAME, the Flux Analysis and Modeling Environment. *BMC Systems Biology* **6**: 8. doi:10.1186/1752-0509-6-8.

Maarleveld T.R., **Boele J.**, Bruggeman F.J., Teusink B. (2014) A Data Integration and Visualization Resource for the Metabolic Network of Synechocystis sp. PCC 6803. *Plant Physiology* **164**: 1111–1121. doi:10.1104/pp.113.224394.

Yamaga R., Ikeda K., **Boele J.**, Horie-Inoue K., Takayama K., Urano T., Kaida K., Carninci P., Kawai J., Hayashizaki Y., Ouchi Y., de Hoon M., Inoue S. (2014) Systemic identification of estrogen-regulated genes in breast cancer cells through cap analysis of gene expression mapping. *Biochemical and Biophysical Research Communications* **447**: 531–536. doi:10.1016/j.bbrc.2014.04.033.

Boele J., Persson H., Shin J.W. Ishizu Y., Newie I.S., Søkilde R., Hawkins S.M., Coarfa C., Ikeda K., Takayama K., Horie-Inoue K., Ando Y., Burroughs A.M., Sasaki C., Suzuki C., Sakai M., Aoki S., Ogawa A., Hasegawa A., Lizio M., Kaida K., Teusink B., Carninci P., Suzuki H., Inoue S., Gunaratne P.H., Rovira C., Hayashizaki Y., de Hoon M.J. (2014) PAPD5-mediated 3′ adenylation and subsequent degradation of miR-21 is disrupted in proliferative disease. *Proceedings of the National Academy of Science of the United States of America* **111**(31): 11467–11472. doi:10.1073/pnas.1317751111.

Boele J., Krumpochova P., Olivier B.G., Kloos D., Giera M., Teusink B. (2014) An experimental determination of the biomass composition of HL-60 human leukemia cells has implications for genome-scale metabolic modeling of cancer physiology. *Manuscript in preparation*.

Boele J., Mujic-Delic A., Scholten D., Ram P.T., Bruggeman F.J., Teusink B., Smit M.J. (2014) High-throughput phospho-proteome screening reveals differential signaling upon stimulation of the CXCR4/CXCR7/CXCL12-axis in breast cancer cells. *Manuscript in preparation*.

Boele J., Maarleveld T.R., Jongkind R.C., Bruggeman F.J., Teusink B. (2014) An interactive multi-organism graphical map of metabolism that facilitates the analysis of genome-scale modeling results. *Manuscript in preparation*.

Boele J., Bank E., Santos F., Teusink B. (2014) FiJo, a semi-automated approach to the reconstruction of genome-scale metabolic models, and its application to modeling a pathogenic bacterium. *Manuscript in preparation*.