

Lessons from a genetic network about the causes of dominance

Invited contribution to “*Biology of Dominance*”

By Veitia, R. (ed.)

Andreas Wagner

University of New Mexico and The Santa Fe Institute

University of New Mexico

Department of Biology

167A Castetter Hall

Albuquerque, NM 817131-1091

Phone: +505-277-2021

FAX: +505-277-0304

Email: wagnera@unm.edu

Abstract. I review recent findings from the stoichiometric analysis of metabolic networks. These findings show that the physiological theory of dominance, which explains dominance in metabolic genes from kinetic properties of biochemical pathways and redundant allelic functions is incomplete. Stoichiometric analysis of metabolic networks indicates that a mix of causes – local allele redundancy and global network properties – is responsible for dominance in metabolic genes.

Dominance is a special case of a genetic system's robustness against mutations. It can be defined as robustness to a 50% change in gene dosage, and is caused by the redundant functions of two alleles at a diploid gene locus. While robustness to a 50% change in gene dosage is remarkable, it is also worth pointing out that robustness against more drastic genetic change is widespread. For example, more than one third of all synthetic null (gene knockout) mutations in the yeast *Saccharomyces cerevisiae* have weak or no phenotypic effects under standard laboratory conditions.^{1,2} These are mutations where either two copies of the gene in a diploid cell or one copy in a haploid cell are eliminated. They completely eliminate a gene from the genome – a 100% change in gene dosage. One might think that weak phenotypic effects of gene knockout mutations are due to redundant gene functions, not among alleles at one locus, but among duplicate genes. However, between 40 and 77% of yeast genes with weak knockout effects are single copy genes.^{3,4} This translates into hundreds to thousands of single copy genes with weak phenotypic effects. Similar numbers exist in the nematode worm *Caenorhabditis elegans* (Conant and Wagner, unpubl. manuscript). These findings do not negate the role of gene duplications and gene redundancy in mutational robustness – both are undoubtedly important.^{3,5} However, they indicate that simple redundancy of parts can not be the only source of mutational robustness.

Arguably, most knockout mutations with no phenotypic effects are also examples of the wild-type gene's dominance, where a 50% reduction of gene dosage would have no phenotypic effect. And if many such cases of gene knockout mutations involve single copy genes, it follows that gene redundancy cannot be the only source of dominance. Unfortunately, the reasons why a single copy gene may have no phenotypic effect are generally poorly understood, with one exception: a growing body of work on genetic perturbations in metabolic networks. I will here review some relevant background, as well as implications of this work for our understanding of mutational robustness in general and dominance in particular.

The mechanistic reasons for dominance are best understood for metabolic genes acting in metabolic pathways, where dominance results from the cooperation of multiple enzymes in producing the output of a pathway and the flux of matter and energy through the pathway⁶. However, this framework predicts that complete elimination of a gene from a pathway—when considered in isolation—would have grave phenotypic effects, because it would block the entire pathway. In contrast, the work I review below suggests that considering pathways in the larger context of metabolic networks changes this picture. Blocking individual pathways may have little phenotypic effects, because a metabolic network as a whole may be able to reroute metabolic flux around the pathway. This suggests that an improved understanding of genetic interactions in genetic networks – whether metabolic or regulatory – can teach us much about the mechanistic causes of dominance.

Background of stoichiometric network analysis. Metabolic pathways are but figments of the complex reaction networks that sustain the living. Ultimately, to understand robustness in such networks will require understanding the functioning of large metabolic networks. A quantitative theory such as metabolic control analysis would allow such an understanding.⁷ However, such understanding also has a price, namely that the theory requires much quantitative information, in particular about kinetic rate constants of enzymatic reactions. Unfortunately, with the exception of a few simple model systems, such as the metabolism of human red blood cells,⁸ such information is unavailable. This holds even for the best studied metabolic networks, such as that of *Escherichia coli*. Although most if not all chemical reactions catalyzed by this bacterium are known, kinetic information is available only for a small fraction of these enzymes. Thus, even if the substantial mathematical problems of applying metabolic control analysis to large reaction networks

were solved, insufficient information would be available to apply the theory to large networks.

This problem raises the question: is it possible to characterize large chemical reaction networks and their robustness even though information about the reaction rates of many enzymes is unavailable? The answer is yes. It can be done by merely examining stoichiometric properties of metabolic networks, using related approaches such as flux balance analysis⁹ or elementary mode analysis.^{10, 11} I will briefly outline the foundation of these approaches and some of the insights they provide.

Consider the following simple reaction scheme (Fig. 1), where some substrate S external to the cell is imported and/or converted into some metabolite m_1 , which can then be converted through a reversible reaction (indicated by a double-headed arrow) into metabolite m_2 . m_2 , in turn, is a precursor to some product P , which might be a biomass component essential for cellular growth, or it might be secreted from the cell. Alternatively to its direct interconversion with m_2 , m_1 can also be converted into m_2 through a chain of two irreversible chemical reactions involving a metabolite m_3 .

The most important variables in a stoichiometric analysis of chemical reaction networks are the amounts of matter that flow through each of the chemical reactions indicated by the arrows in Figure 1. These are referred to as metabolic fluxes v_i , where the subscript i corresponds to the name or index of the chemical reaction, as written above the arrows in Figure 1. The changes in concentrations of the metabolites m_i , dm_i/dt are simple functions of these fluxes. For the example of Figure 1, these are

$$\begin{aligned}\frac{dm_1}{dt} &= v_s - v_1 - v_2 \\ \frac{dm_2}{dt} &= v_1 + v_3 - v_p \\ \frac{dm_3}{dt} &= v_2 - v\end{aligned}\tag{1}$$

Metabolic "inputs" like S and "outputs" like P are usually referred to as "external" metabolites. The changes in their concentrations are not explicitly modeled and are distinguished from "internal" metabolites m_i , whose interconversion constitutes metabolism proper. The internal metabolites are subject to conservation of mass, which is reflected in the fact that by summing all derivatives dm_i/dt the internal fluxes v_i cancel. That is, one is left only with fluxes from and to external metabolites

$$\frac{dm_1}{dt} + \frac{dm_2}{dt} + \frac{dm_3}{dt} = v_s - v_p$$

Under constant environmental conditions, the changes in metabolite concentrations dm_i/dt must approach zero, otherwise some metabolites would disappear completely, whereas the concentration of others would approach infinity. Thus, a metabolic network approaches a steady state, in which individual metabolite concentrations do not change. This steady state is dynamic, in the sense that constant metabolite concentrations are maintained by ongoing interconversions of metabolites. Restricting the analysis only to steady states is motivated by two further observations. First, whereas transient changes in metabolite concentrations occur, for example when an environment changes, a new steady state is reached rapidly, typically within minutes.¹² Secondly, even if metabolite concentrations sometimes show more complex behavior, such as sustained oscillations, the

time-averaged metabolite concentrations are constant, and thus, effectively, a steady-state is reached.

The steady state condition simplifies the mathematical treatment of a metabolic network considerably, as the following simple example shows. I begin by writing (1) in a more compact form, i.e.,

$$\frac{d\vec{m}}{dt} = \mathbf{S}\vec{v} \quad (2)$$

This is now a linear matrix differential equation, where $\vec{m} = (m_1, m_2, m_3)$ is the vector of internal metabolites. $\vec{v} = (v_s, v_1, v_2, v_3, v_p)$ is the vector of all fluxes, and \mathbf{S} is a matrix that contains the coefficients of this differential equation, which are simply the stoichiometric coefficients of the chemical reactions for each metabolite m_i . For the simple reaction scheme of Figure 1, this matrix has the following structure.

$$\mathbf{S} = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix} \quad (3)$$

Its columns correspond to the reactions s, 1,2,3, and p of Figure 1. Its rows correspond to metabolites $m_1, m_2,$ and m_3 . Positive and negative signs of the respective entries indicate whether a metabolite occur in the left or right side of a chemical reaction in Figure 1b. For example $S_{32} = 1$ because metabolite m_3 is a product (occurs on the right-hand side) of reaction 2. In the notation of (2), the steady state condition (no changes in internal metabolite concentrations) can be written as

$$\frac{d\vec{m}}{dt} = 0$$

which is equivalent to

$$\mathbf{S}\vec{v} = 0 \quad (4)$$

For our example, this condition is equivalent to

$$\begin{aligned} v_s - v_1 - v_2 &= 0 \\ v_1 + v_3 - v_p &= 0 \\ v_2 - v_3 &= 0 \end{aligned} \quad (5)$$

As a mere matter of convention, the fluxes corresponding to external metabolites are sometimes written on the right-hand side of (5), such that one obtains

$$\begin{aligned} -v_1 - v_2 &= -v_s \\ v_1 + v_3 &= v_p \\ v_2 - v_3 &= 0 \end{aligned} \quad (6)$$

This can be written again in matrix form

$$\mathbf{S}'\vec{v}' = \vec{b} \quad (7)$$

where the stoichiometry matrix \mathbf{S}' is

$$\mathbf{S}' = \begin{pmatrix} -1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

and the vector $\vec{v}' = (v_1, v_2, v_3)$, and $\vec{b} = (-v_s, v_p, 0)$. The equivalence of (5) and (6) shows that the two matrix formulations also have equivalent solutions. However, there is a conceptual difference. In (4), all fluxes, including those from and to the external metabolites, v_s and v_p , are treated as variables, and the steady state flux distribution is obtained by solving for \vec{v} . In (7), these external fluxes are absorbed into the vector \vec{b} , and are thus treated as constants. (One can think of them as availabilities of a substrate in the environment, or as export rates of a product.) One thus solves only for the three internal fluxes.

I note that the above stoichiometry matrices \mathbf{S} and \mathbf{S}' are unusual in several respects. First, and trivially, they contain many fewer reactions than are encountered in any network of realistic complexity. Second, the reaction scheme of Fig. 1 contains only monomolecular reaction, i.e., reactions where each reaction has only one educt and only one product. Much more frequent are bimolecular reactions. Such reactions are easily incorporated into a stoichiometry matrix \mathbf{S} . Each column of \mathbf{S} (corresponding to one reaction) can carry as many negative entries as the reaction has educts, and as many positive entries as it has products. Third, \mathbf{S} in this example contains only nonzero entries (+1) and (-1), but many reactions do not convert molecules in equimolar proportions. This complication is also easily incorporated. For example, if reaction 3 needed two molecules of m_3 to produce one molecule of m_2 , then S_{33} would be equal to -2 instead of -1.

In addition to the flux-balance condition (4) in the steady state, any metabolic reaction network has to fulfill several additional constraints. First, fluxes can not become arbitrarily small or large, such that they need to be bounded between some real values. The reason is that only a limited amount of any one enzyme can be present, and that enzymes catalyze reactions at finite rates. Second, some reactions are irreversible and can proceed in only one direction. By convention, the respective flux v_i can not be negative. The same principles hold for the import of an external metabolite, such as a carbon source, or for the export of a metabolic end product. The respective flux has a maximal rate, which reflects factors such as the concentration of a metabolite in the environment and the transport mechanism of the metabolite.

Main tasks of stoichiometric network analysis. Stoichiometric analysis has two main tasks. First, it identifies the flux-vectors \vec{v} that fulfill all the constraints on a metabolic network. These include the steady-state condition (4), as well as the additional constraints on flux magnitudes and signs just listed. Flux vectors that meet these conditions are 'allowed' fluxes, that is, fluxes that a cell can realize. Importantly there is almost never just one unique allowed flux vector. The reason is that in most metabolic reaction networks with n internal metabolites and m chemical reactions, there are many more chemical reactions than

metabolites ($m \gg n$), or many fewer equations than variables, such that (4) is massively underdetermined. If no constraints other than (4) were operational, the set of allowable fluxes would be an $(n-m)$ -dimensional (vector) space, which is also called the null space of the stoichiometry matrix S . However, because fluxes are bounded, the allowable flux vectors occupy a bounded region in this space, as indicated in the three-dimensional caricature of Figure 2.

The second task of stoichiometric analysis is to identify regions within the set of allowable fluxes that maximize a desirable property. One example of such a property is cell growth. Consider a genetically heterogeneous population of single-celled organism that actively grow and divide. Each cell or genotype in this population may occupy a different position in the region of allowed metabolic fluxes, because its enzymes and their expression levels under any particular environmental condition differ from those of other cells. Those cells in the population that grow at a maximal rate will outgrow all other cells, and thus come to dominate the population. For such maximal cell growth, biosynthetic precursors such as amino acids need to be made in well-defined ratios. Similarly, high-energy phosphate bonds (ATP and related molecules) and redox potential (NADH and related molecules) need to be produced in balanced amounts. This means that the metabolic fluxes generating them must have particular values, which can be identified if the optimal proportions of biosynthetic precursors, energy carriers, and redox potential are known. In some well studied organisms, such as the bacterium *Escherichia coli*, these proportions are known from the biomass composition of the organism¹³. In trying to find the flux vectors \vec{v} that yield maximal growth under any one environmental condition, one tries to identify one (or more) points in the realizable region of the space of fluxes (Fig. 2) that maximize or minimize some function Z of the flux \vec{v} . In practice, linear functions of \vec{v} are most important, such that

$$Z(\vec{v}) = \sum_{i=1}^m c_i v_i \quad (8)$$

and one tries to find values of \vec{v} that maximize Z . (There may be many flux vectors \vec{v} that give the same maximal value of Z). As an aside, finding fluxes that ensure maximal growth is by far not the only application of this approach. Others include identification of fluxes that minimize ATP production – corresponding to energy-efficient growth –, or identification of fluxes that produce maximal amounts of an industrially important metabolite, such as an antibiotic. (The approach is thus of great relevance for metabolic engineering of organisms in industrial biotechnology.) In practice, the function Z is maximized by standard numerical techniques such as linear programming.

Note that even though a particular point in realizable flux space may be identified as optimal, it is by no means assured that a cell can achieve the desired metabolic fluxes. For example, when faced with certain carbon sources, a cell may not be able to express the required enzymes in the amounts that assure optimal growth. A case in point is the *E. coli* strain MG1655, which grows on glycerol as sole carbon source at a rate lower than predicted as optimal by this approach. However, within 40 days or 700 generations of evolution in the laboratory, the strain increases growth substantially, and the fluxes approach those predicted by theory as optimal¹⁴. This shows not only the power of this approach to predict optimal flux distributions, but also the power of evolution to achieve the appropriate patterns of gene regulation within short amounts of time.

To summarize, stoichiometric analysis of large enzymatic reaction networks can identify realizable metabolic fluxes, fluxes that fulfill certain boundary conditions. Within this set of all realizable fluxes, it can identify fluxes that endow a cell with properties of

interest. A condition of particular relevance for evolutionary studies is that of maximal cell growth, because it is a good correlate of fitness under conditions where cells actively divide.

Applications to robustness. The approach outlined above is suitable to analyze robustness of cell growth to changes in individual fluxes. That is, under conditions where cells grow maximally, one can ask: what are the effects of reducing only one individual flux drastically, e.g., by forcing it to assume a value of zero. Biologically, such a change would correspond to a loss-of-function mutation in an individual enzyme catalyzing a particular reaction, or in a loss of its expression. Because the approach is computational, it is easily possible to determine robustness to changing – one by one – all fluxes in a network, and one can do the same for all pairwise combinations of fluxes.

In a series of papers, Edwards and Palsson¹⁵⁻¹⁷ analyzed the robustness of the well-characterized chemical reaction networks in two prokaryotic organisms, *Escherichia coli* and *Haemophilus influenzae*. For *E. coli*, they assembled a reaction network comprising 436 metabolites and 736 reactions from the biochemical literature, genome sequence information, and metabolic databases.¹⁷ They determined the allowable steady-state fluxes under the constraints listed above. Within this allowable region of fluxes, they determined those fluxes for which growth on a minimal medium under aerobic conditions with glucose as sole carbon source was maximal. Growth is maximal for fluxes that produce the necessary metabolites in proportions that correspond to the (empirically known) biomass composition of *E. coli*. (This maximal growth flux distribution can be thought of as a single chemical reaction that converts biosynthetic precursors into biomass with the least possible wastage.) With this optimal flux distribution \vec{v} in hand, individual fluxes can be forced to a value of zero (corresponding to deletion of the respective enzyme-coding genes) and the resulting effect on growth can be studied. The parts of metabolism Edwards and Palsson analyzed in this way comprise 48 chemical reactions and include all of glycolysis, the pentose phosphate pathway, the tricarboxylic acid cycle, and respiration. Perhaps surprisingly, only seven of the 48 eliminated reactions turned out to be lethal. That is, they generated fluxes not allowable in steady state, or fluxes from which an essential biochemical precursor, such as an amino acid, is not produced. Two of the essential reactions, glyceraldehyde 3-phosphate dehydrogenase and phosphoglycerate kinase, are part of the three-carbon stage of glycolysis. Three others are the first reactions of the tricarboxylic acid cycle, namely citrate synthase, aconitase, and isocitrate dehydrogenase. The remaining two essential reactions are part of the pentose phosphate shunt. These reactions are ribose-5-phosphate isomerase and transketolase. Figure 3 shows the essential reactions as part of an overview of central metabolism.

Of the remaining 41 nonessential reactions, 32 reduced growth by less than 5%, and only 9 reduced growth by more than 5%. The authors compared their results for a variety of carbon sources to experimental data that had determined the effect of deletions in enzyme-coding genes of *E. coli*. They found that the experimental data and computational predictions on growth/no-growth coincided in 86% of the cases. Thus, the computational results show good agreement with experiment.

The surprising aspect of these results is the high fraction – about two-thirds – of complete loss-of-function mutations that reduce growth by less than 5%. It bears emphasizing that there is no redundancy in this system, in the sense that there are no two enzymes that can carry out the same chemical reactions. The explanation of these findings is that flux can be shuttled through parts of a metabolic network that are not affected by blocking a reaction. Here is an example.¹⁵

The pentose phosphate shunt diverts metabolites from glycolysis and serves two main purposes¹⁸. Its oxidative branch generates NADPH for biosyntheses. (This branch can only be entered via glucose-6-phosphate.) Its nonoxidative branch generates biosynthetic precursors such as ribose-5-phosphate. It can be entered from the oxidative branch, but also

from other glycolytic intermediates such as fructose 6-phosphate. Edwards and Palsson¹⁵ blocked the first reaction of the pentose phosphate shunt (reaction zwf in Fig. 3), glucose 6-phosphate dehydrogenase, which converts glucose 6-phosphate (G6P) into D-6-phosphate-glucone-lactone (6PGA). This reaction leads into the oxidative branch of the pentose phosphate shunt. Deletion of this reaction completely blocks the oxidative branch but affects metabolic output only minimally: wild-type growth is reduced by only one percent under aerobic conditions in glucose minimal medium. However, the deletion has profound systemic consequences on the flux in this network. Before the deletion (i.e., in the wild-type state), about two-thirds of the NADPH needed is produced by the pentose phosphate shunt. Most of the NADH is produced by the tricarboxylic acid cycle. (High energy phosphate bonds are largely generated via oxidative phosphorylation.) One of the major systemic reorganizations of metabolic flow after the mutation regards NADPH production. To compensate for the blocked oxidative branch of the pentose phosphate shunt, most of the NADPH is now produced through an increased flux in the tricarboxylic acid cycle, which generates NADH. This NADH is then transformed into NADPH via a massively increased flux through the transhydrogenase reaction. The nonoxidative branch of the pentose phosphate shunt can still be entered through other metabolites of glycolysis, and thus still serves to produce sufficient quantities of biosynthetic precursors. (Figure 4 contains an overview of the flux in this mutant.)

A further striking result of this analysis is that the network is robust even to substantial manipulations of flux through some of the seven essential reactions¹⁶. That is, although complete elimination of flux through the essential reactions is lethal, substantial quantitative reductions in flux may be neutral. A case in point is the essential transketolase reaction in the pentose phosphate shunt. As long as a mutation preserves more than 15% of the wild-type flux through this reaction, growth is greater than 99.2% of the wild-type growth rate. This absence of a phenotypic effect, however, camouflages profound systemic changes that have to take place to compensate for the reduction in flux. One of these changes is again an increased production of NADPH through the tricarboxylic acid cycle and through transhydrogenase. Another change is an increased flux through glycolytic reactions such as pyruvate kinase that absorb the reduced flux through the pentose phosphate shunt. As flux decreases below 15%, the reduced flux through transketolase limits the production of erythrose 4-phosphate, an essential precursor of aromatic amino acids. The result is a reduced growth rate.

The first three reactions of the tricarboxylic acid cycle, equally essential, also are quite robust to large quantitative changes in flux. For example, unless flux through the citrate synthase reaction falls below 18% of the wildtype, wild-type growth is essentially unchanged. However, as flux through this reaction is gradually reduced from the optimal wildtype level, a variety of systemic changes occur that allow wild-type growth to be sustained. They include increased flux through the pentose phosphate shunt, reduction and eventual complete elimination of flux through the pyruvate kinase reaction, (which is upstream of the blocked reaction), as well as reduction of cyclic flux through the tricarboxylic acid cycle. The cycle eventually ceases to function cyclically, and serves only to generate biosynthetic precursors. If the flux through citrate synthase falls below 18% of wild-type levels, the metabolic network can not produce sufficient α -ketoglutarate, an essential precursor of amino acids, to ensure maximal growth. This is the reason for the reduction in growth rate under these conditions.¹⁶

The two essential glycolytic reactions are the most restricted of the seven essential reactions. Their fluxes can not be reduced to less than 70% of the wild-type level without affecting growth substantially. Finally, it is worth mentioning that even the essential reactions can be quite insensitive to increases in flux. (I have not emphasized robustness to such "gain-

of-function" mutations, because most mutations are likely to reduce flux, as they reduce either enzyme activity or enzyme expression.) For example, flux through the citrate synthase reaction can increase to 160% of wild-type without affecting growth.¹⁶

Edwards and Palsson also studied the metabolic reaction network of the bacterium *Haemophilus influenzae*. Their results add additional facets to the studies in *E. coli*. They used physiological and genome information to construct a map of 488 metabolic reactions and 343 metabolites, and examined robustness of fluxes in this network to deletions in 36 central reactions. They found a larger fraction of essential genes than in *E. coli* (33% vs 14% in *E. coli*), and a smaller fraction of genes with no effects on growth when eliminated (42% vs 69% in *E. coli*). However, the networks are not straightforward to compare, because of their different sizes and features. (For example, glutamate was an essential amino acid for the *H. influenzae* network, whereas it can be produced by the *E. coli* network.)

Two aspects of the *H. influenzae* study add substantially to the *E. coli* work. First, the authors also examined the effects of multiple (double and triple) deletions of enzymatic reactions on growth. They found only 7 lethal double-mutants among 361 double mutations whose singly mutant constituents were not lethal. Similarly, among 5270 triple mutations, only 7 were lethal. That is, in large metabolic networks under defined conditions, it is quite feasible to eliminate multiple network components without destroying network function.

The second important aspect of this study is that the authors did not only analyze robustness in one environment, but they extended their analysis to several environments. The *H. influenzae* results cited above make specific assumptions about the availability and uptake rate of glutamate, an essential amino acid, and fructose, the sole carbon source. When both the availability of fructose and glutamate were varied over a range of values, the number of reactions that did not affect growth when deleted shrank from 14 to 9. When in addition the availability of oxygen was also varied, this number was further reduced to five. In other words, chemical reactions that do not affect growth in one environment may well do so in another environment.

A note of caution to all these results must be reiterated at this point. While the stoichiometric analysis reveals optimality criteria for cell growth, it does not guarantee that cells can attain the required fluxes. For example, a cell may not be able to express enzymes in the quantities necessary to ensure maximal growth. This holds in particular for unusual environmental or genetic conditions which a population has not encountered in its evolutionary history, and where no evolutionary pressure has forced an adaptive response. A possible case in point is the *E. coli* double mutant in the enzymes glucose-6-phosphate dehydrogenase and transhydrogenase. Its growth rate is predicted to be 92% of the wild-type.¹⁷ However, experimental results from a strain in which both genes encoding the respective enzymes were deleted show that the mutant grows only at 57% of the wild-type rate. Such quantitative discrepancies may well be due to the fact that the *E. coli* strain may never have undergone adaptive evolution in the double mutant condition.

Summary and conclusion. The results summarized above indicate that gene redundancy is by far not the only source of robustness in metabolic networks. For instance, among 48 emulated gene knockout mutations in central metabolism of *E. coli*, only 7 revealed essential reactions. Among these essential reactions, only two can not tolerate a flux reduction of more than 50% without affecting cell growth. In other words, the vast majority of loss-of-function mutations in metabolic networks may be recessive for reasons that have nothing to do with gene redundancy. Rather, such recessivity is a property of a network capable of reallocating metabolic flux to different pathways. The canonical explanation of dominance in metabolic pathways, which emerges from metabolic control theory, is only a part of the mechanistic explanation for dominance. Its emphasis on simple parts redundancy in individual pathways renders it incomplete. It will be instructive to see whether a similar

mix of causes – local gene redundancy and global network properties – accounts for robustness in other genetic networks that have come under increased scrutiny with the availability of functional genomic techniques.

Acknowledgments.

I would like to thank the NIH for its support through NIH grant GM063882-01 and the Santa Fe Institute for its continued support.

Figure Captions

Figure 1. A simple chemical reaction scheme where an external substrate S is transformed into a metabolic product P via a series of chemical reactions involving internal metabolites m_i . The scheme is represented in two equivalent forms.

Figure 2. A schematic representation of allowed steady state fluxes for a hypothetical set of three chemical reactions (not shown). The figure illustrates that the allowed fluxes do not form a vector space but instead a bounded subset (flux cone) of a vector space, the null space of a stoichiometric matrix S .

Figure 3. Central metabolic pathway reactions in *Escherichia coli*. Reactions shaded in grey are essential, that is, their removal eliminates growth according to stoichiometric analysis^{16, 17}.

Reactions: aceA, isocitrate lyase; aceB, malate synthase; aceEF, pyruvate dehydrogenase; ack, acetate kinase; acn, aconitase; adh, acetaldehyde dehydrogenase; eno, enolase; fba, fructose-1,6-bisphosphatase aldolase; fbp, fructose-1,6-bisphosphatase; frd, fumarate reductase; fum, fumarase; gap, glyceraldehyde-3-phosphate dehydrogenase; glk, glucokinase; glt, citrate synthase; gnd, 6-phosphogluconate dehydrogenase; gpm, phosphoglycerate mutase; icd, isocitrate dehydrogenase; ldh, lactate dehydrogenase; mae, malic enzyme; mdh, malate dehydrogenase; pck, phosphoenolpyruvate carboxykinase; pfk, phosphofructokinase; pfl, pyruvate formate lyase; pgi, phosphoglucose isomerase; pgk, phosphoglycerate kinase; pgl, 6-phosphogluconolactonase; ppc, phosphoenolpyruvate carboxylase; pps, phosphoenolpyruvate synthase; pts, phosphotransferase system; pyk, pyruvate kinase; rpe, ribulose phosphate 3-epimerase; rpi, ribose-5-phosphate isomerase; sdh, succinate dehydrogenase; sfc, malic enzyme; sucAB, 2-ketoglutarate dehydrogenase; sucCD, succinyl-CoA synthetase; tal, transaldolase; tkt, transketolase; tpi, triosphosphate isomerase; zwf, glucose 6-phosphate-1-dehydrogenase. Metabolites: 2PG, 2-phosphoglycerate; 3PG, 3-phosphoglycerate; 6PG, D-6-phosphate-gluconate; 6PGA, D-6-phosphate-glucono- γ -lactone; AC, acetate; AcCoA, Acetyl-CoA; R-KG, R -ketoglutarate; CIT, citrate; DHAP, dihydroxyacetone phosphate; DPG, 1,3-bis-phosphoglycerate; E4P, erythrose 4-phosphate; ETH, ethanol; F6P, fructose 6-phosphate; FDP, fructose 1,6-diphosphate; FOR, formate; FUM, fumarate; G6P, glucose 6-phosphate; GA3P, glyceraldehydes 3-phosphate; ICIT, isocitrate; LAC, lactate; MAL, malate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, ribose 5-phosphate; Ru5P, ribulose 5-phosphate; S7P, sedo-heptulose; SUCC, succinate; SuccCoA, succinyl CoA; X5P, dihydroxyacetone phosphate. Figure courtesy of Jeremy Edwards. Reprinted with permission from Edwards JS, Palsson BO. *Biotechnology Progress* 2000;16(6):927-939.

Figure 4. Rerouting of metabolic fluxes in a metabolic mutant. Growth-maximizing flux for a wild type metabolic genotype and in the *zwf* mutant. This mutation (indicated by the crossed arrow) eliminates the glucose 6-phosphate dehydrogenase reaction that leads into the pentose phosphate shunt. Biomass yield in this mutant is only one percent lower than that in the wild-type. Thickness of arrows is proportional to the flux through a reaction in the wild-type. The upper (lower) number next to each reaction indicates the metabolic flux in the wild-type (mutant) [substrate converted h^{-1} gram dry weight (DW) $^{-1}$] at a glucose uptake rate of 6.6 mmol glucose h^{-1} g DW $^{-1}$ and an oxygen uptake rate of 12.4 mmol oxygen h^{-1} g DW $^{-1}$. See legend to Figure 3 for an explanation of acronyms used. Figure courtesy of Jeremy Edwards. Reprinted with permission from Edwards JS, Palsson BO. *Proc. Natl. Acad. Sci. U.S.A.* 2000;97:5528-5533.

Literature Cited

1. Smith V, Chou KN, Lashkari D et al. Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* 1996; 274:2069-2074.
2. Giaever G, Chu A, Ni L et al. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 2002;418:387-391.
3. Gu Z, Steinmetz L, Gu X et al. Role of duplicate genes in genetic robustness against null mutations. *Nature* 2003; 421:63-66.
4. Wagner A. Mutational robustness in genetic networks of yeast. *Nature Genetics* 2000;24:355-361.
5. Conant GC, Wagner A. GenomeHistory: a software tool and its applications to fully sequenced genomes. *Nucleic Acids Research* 2002;30,1-10.
6. Kacser H, Burns JA. The molecular basis of dominance. *Genetics* 1981;97:639-666.
7. Fell D. Understanding the control of metabolism. 1997. Miami: Portland Press.
8. Palsson BO, Lee I-D. A comprehensive model of human erythrocyte metabolism: extensions to include pH effects. *Biomed. Biochim. Acta* 1991;49:771-789.
9. Edwards JS, Ramakrishna R, Schilling C, Palsson BO. Metabolic flux balance analysis. In: Lee S, Papoutsakis E, eds. *Metabolic engineering* New York: Marcel Dekker; 1999: 13-57.
10. Schuster S, Hilgetag C, Woods J, Fell D. Elementary modes of functioning in biochemical networks. In: Cuthbertson R, Holcombe M, Paton R, eds. *Computation in cellular and molecular biological systems*. Singapore: World Scientific; 1996:151-164.
11. Schuster S, Dandekar T, Fell DA. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology* 1999;17:53-60.
12. Palsson BO, Lightfoot EN. Mathematical-Modeling of Dynamics and Control in Metabolic Networks .1. On Michaelis-Menten Kinetics. *Journal of Theoretical Biology*. 1984;111:273-302.
13. Neidhardt FC, Umberger HE. chemical composition of *Escherichia coli*. In: Neidhardt FC, ed. *Escherichia coli and Salmonella: cellular and molecular biology*. Vol 1. Washington, D.C.: American Society of Microbiology Press; 1996:13-16.
14. Ibarra RU, Edwards JS, Palsson BO. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 2002;420:186-189.
15. Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *Journal of Biological Chemistry*. 1999;274:17410-17416.
16. Edwards JS, Palsson BO. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnology Progress* 2000;16(6):927-939.
17. Edwards JS, Palsson BO. The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. U.S.A.* 2000;97:5528-5533.
18. Stryer L. *Biochemistry*. 4th ed. New York: Freeman; 1995.

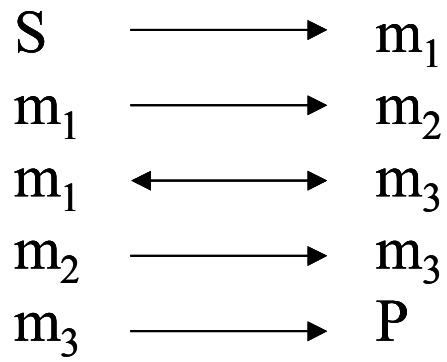
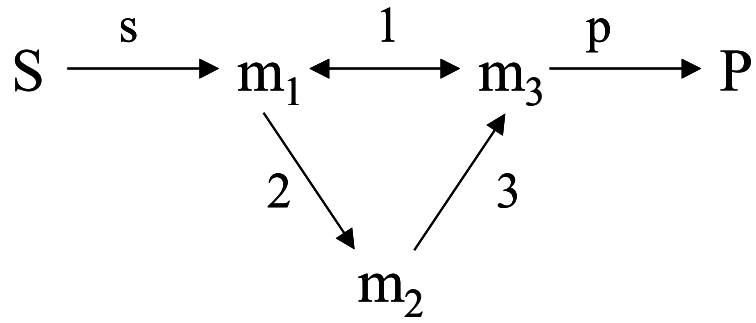


Figure 1

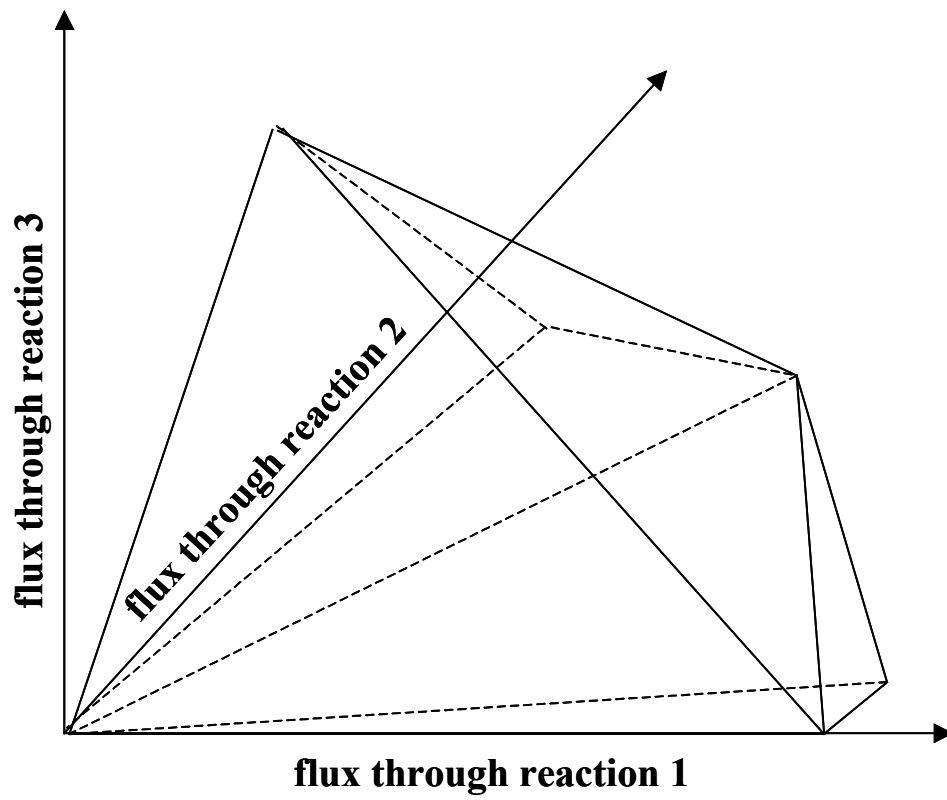


Figure 2

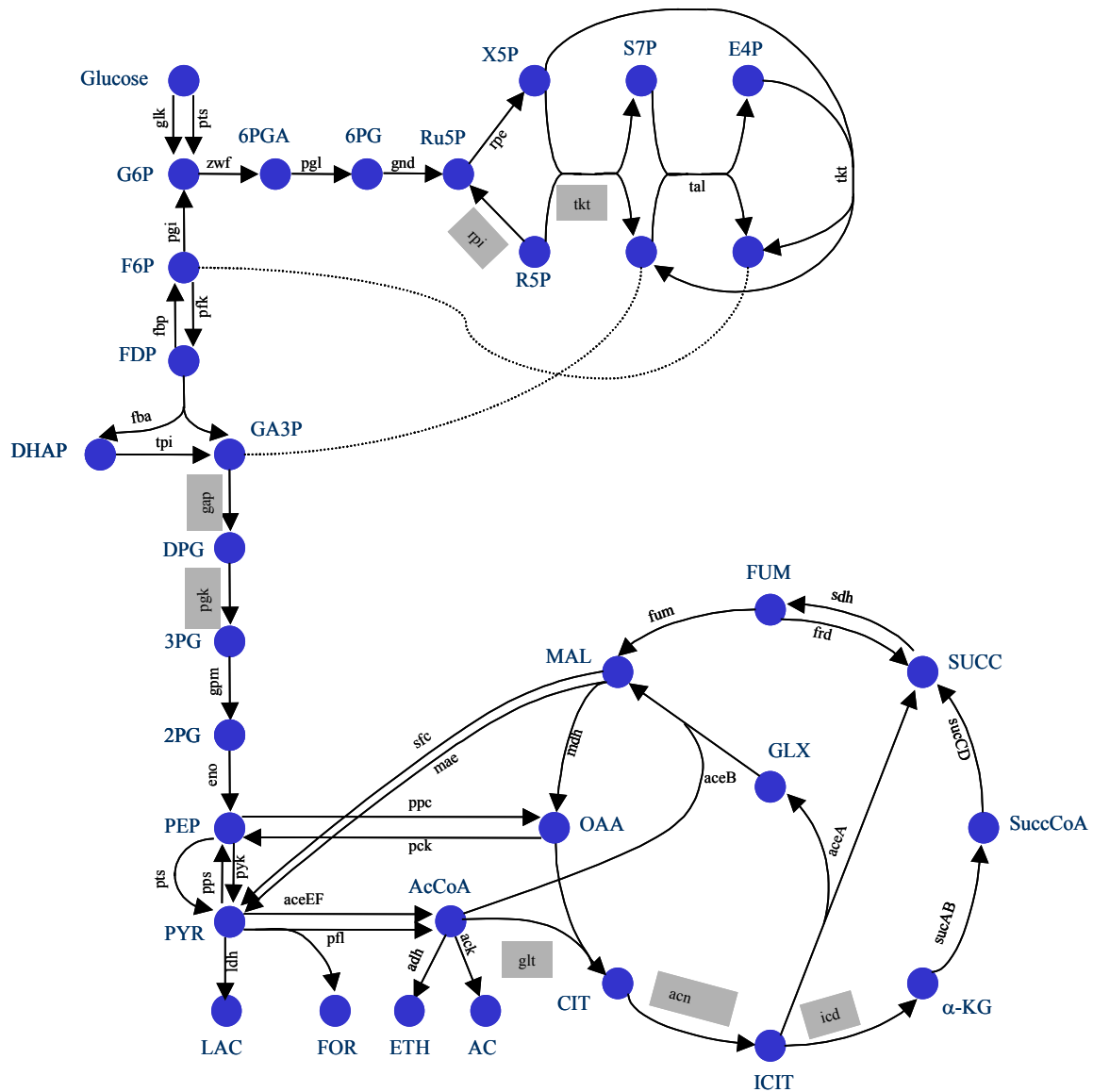


Figure 3

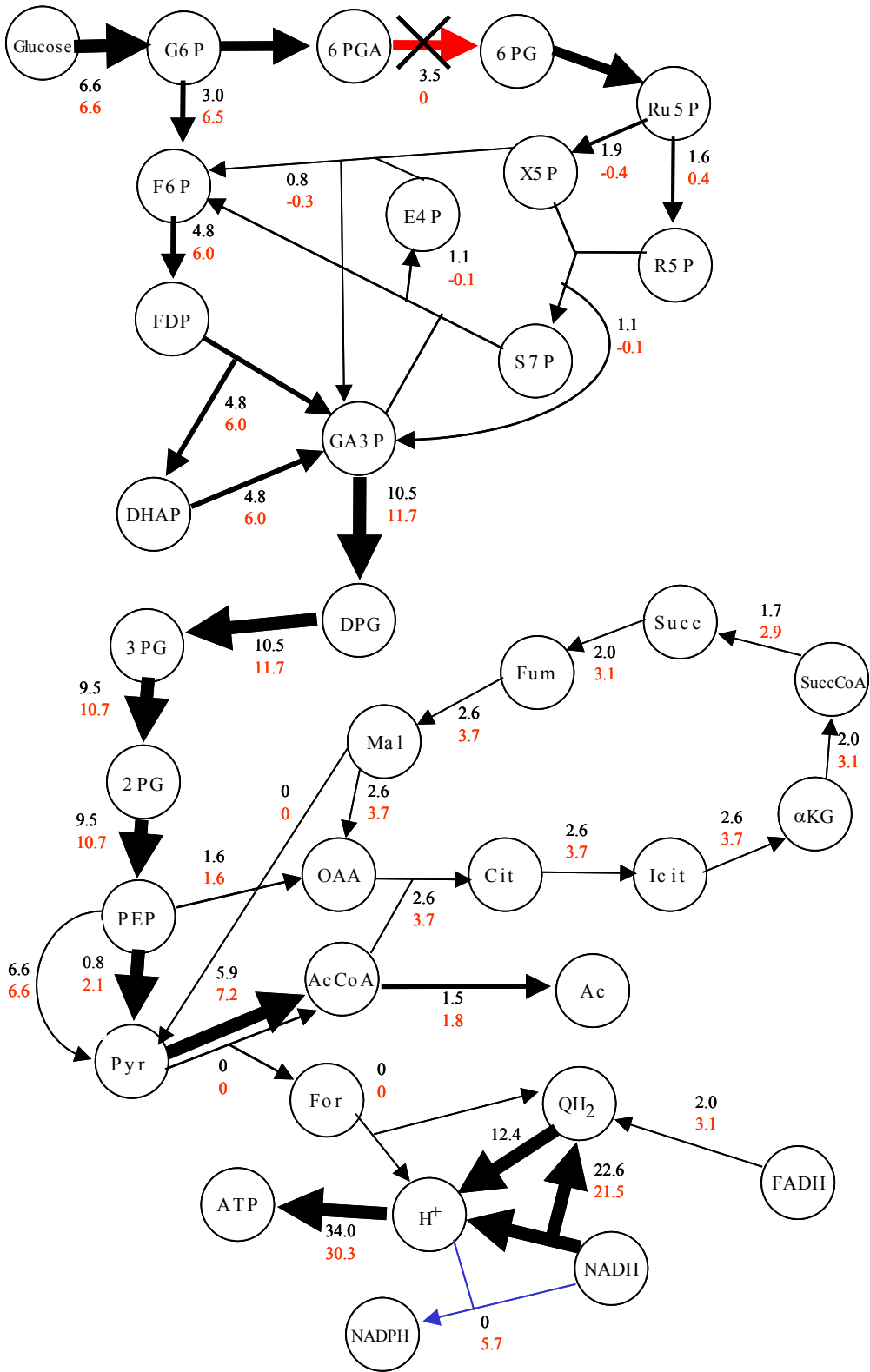


Figure 4

