# A mechanistic study of evolvability using the mitogen-activated protein kinase cascade

# H. F. Nijhout,\* A. M. Berg, and W. T. Gibson

Department of Biology, Duke University, Durham, NC 27708, USA \*Author for correspondence (e-mail: hfn@duke.edu)

**SUMMARY** Evolvability is a function of the way genetic variation interacts with the mechanisms that produce the phenotype. We explore an explicitly mechanistic way of studying the evolvability of phenotypes that are produced by a relatively simple genetic mechanism, the mitogen-activated protein kinase (MAPK) cascade. We developed a quantitative model of MAPK activation that can be used to study the effects of genetic variation on the various components of this signaling cascade. We show how some standard tools of applied mathematics, such as steady-state formulations and nondimensionalization, can be used to elucidate the relative importance of variation in each gene of this mechanism. We also give insights into non-intuitive patterns of dependence

# INTRODUCTION

The term evolvability has acquired several meanings. The original usage in statistical genetics refers to the ability of a population to respond to selection (Roff 1997; Lynch and Walsh 1998). It is equivalent to the heritability of a trait in the direction of selection. The term evolvability has also been used to refer to the genome's ability to produce adaptive variants (Wagner and Altenberg 1996), an organism's capacity to generate heritable phenotypic variation (Kirschner and Gerhart 1997), and the capacity of a lineage to evolve (Kirschner and Gerhart 1998). The interest in evolvability in evolutionary developmental biology emerges from an earlier preoccupation with developmental constraints and dissatisfaction with the fact that this negative concept did not easily lend itself to analysis. The various definitions of evolvability attempt to capture two quite different concepts: (a) the availability of the appropriate genetic variation so that there can be an evolutionary response to selection, and (b) the ability of a developmental-genetic mechanism to generate phenotypic variation that is potentially adaptive. The first defines the short-term evolvability of a system based on current genetic variation. This kind of evolvability cannot be projected far into the future because the heritability of a trait changes when gene frequencies change after selection (Faland trade-off among the genes. The mechanism produces several different phenotypes (ultrasensitivity to stimulation, switch-like behavior, amount of MAPK-PP delivered, persistence of MAPK-PP activity), each of which is sensitive to different (but partially overlapping) combinations of genes. We show that the mechanism imposes clear limitations on the evolvability of each of the different phenotypes of the pathway, even in the presence of genetic variation in the components of the mechanism. This approach to the study of evolvability is generally applicable and complements the traditional approach through statistical genetics by providing a mechanistic understanding of the genetic interactions that produce the phenotype.

coner and Mackay 1996). The second is *evolvability in principle*. It asks whether the developmental mechanism is in principle able to produce a particular phenotype, given the appropriate genetic variation, but does not ask whether the appropriate genetic variation is currently available. It is the latter type of evolvability we are concerned with here.

To examine the range of phenotypes a given developmental mechanism can produce, it is necessary to have an accurate description of the mechanism that can be translated into a mathematical model. The model can include roles for the effects of genetic, structural, and environmental factors and can be used to examine how variation in one or more of these factors affects the behavior of the mechanism and the resultant phenotype. The effects of genetic variation, and the results of selection on that variation, can then be simulated. Mitogen-activated protein kinase (MAPK) signaling is such a mechanism; it is a well-studied and widespread cellular signal transduction mechanism that plays a critical role in the regulation of a host of developmental processes. We therefore use the MAPK cascade as an exemplar to illustrate how knowledge of the developmental or genetic mechanism that underlies a particular phenotype can be used to study evolvability in a quantitative manner. In doing so we explore several approaches that allow us to focus on different aspects of the phenotype and its underlying genetic mechanism.

### MAPK CASCADE

MAPK cascades constitute one of a handful of signal transduction mechanisms that mediate the transmission of signals from the cell surface to the cytoplasm and nucleus. The MAPK cascade is a highly conserved signaling pathway and has been found in all major groups of eukaryotes. The cascade can be activated by growth factors, differentiation factors, or stress. The downstream members of the cascade can be involved in activating a variety of cytoplasmic regulatory proteins, cytoskeletal elements, or nuclear transcription factors. The pathway has a broad array of functions that range from the control of cell proliferation to cell differentiation, cytoskeletal rearrangement, and apoptosis in response to signals received from outside.

MAPK cascades typically have three members (Fig. 1). The terminal member of the cascade is the MAPK proper (e.g., ERK, JNK, p38, p42), which is activated by a MAPK kinase (MAPKK; e.g. MEK, MKK), which in turn is activated by a MAPKK kinase (MAPKKK; e.g. Raf, Mos). Activation of MAPK and MAPKK consists of a double phosphorylation at highly conserved phosphorylation sites (Huang and Ferrell 1996). The phosphorylated kinases are inactivated by specific phosphatases that remove one or both phosphates. Removal of one phosphate can completely or partially inactivate a doubly phosphorylated molecule. It should be noted that MAPK, MAPKK, and MAPKKK are used here as generic terms and do not refer to specific molecules. A wide range of species- and cell-specific gene products, with interesting evolutionary relationships to each other, serve these functions. For the purposes of this article we refer to the members of the cascade by their generic names only. For examples of the diversity of gene products that serve these functions, the reader is referred to the reviews by Herskowitz (1995), Robinson and Cobb (1997), Garrington and Johnson (1999), and Keyse (2000).

The MAPK cascade is activated by the binding of a ligand (a growth factor, for instance) to a cell surface receptor tyrosine kinase. Ligand binding to the receptor leads to phosphorylation of a protein complex associated with the receptor on the cytoplasmic side of the cell membrane, which in turn leads to the activation of the Ras protein (by binding to GTP), which then activates MAPKKK (often a member of the Raf gene family). The receptor–ligand complex then becomes internalized into vesicles where it may continue to be active or may be subject to degradation and recycling (for a theoretical study of the functional roles of receptor internalization, see Haugh and Lauffenburger 1998).

Depending on the system, two or three members of the cascade are attached to a scaffolding protein. This scaffolding ensures that successive members of the cascade are physically adjacent to each other. The functional reason for this is that a cell can contain many different MAPK pathways that are used to transduce different signals and produce different results (Schwartz and Baron 1999). To prevent crosstalk between these pathways, that is, to prevent a kinase from one pathway from activating a member of a different pathway, members of a particular pathway are held in physical proximity to each other by specific scaffolding proteins (Garrington and Johnson 1999). The scaffolding provides an interesting structural constraint, in that it restricts the number of molecules that can interact at any one time. Molecules that are attached to a scaffold presumably are active at equimolar concentrations.

The terminal member of the cascade (a MAPK such as ERK) is typically free to move through the cytoplasm, and a portion of the activated molecules may translocate to the nucleus where they regulate gene transcription. Whether or not the MAPK actually translocates to the nucleus depends on how long the cascade is activated. In the mammalian PC12 cell line, it has been shown that stimulation by nerve growth factor induces a brief activation of the MAPK pathway and results in cell differentiation, whereas stimulation by epidermal growth factor causes a much more prolonged activation of the pathway and results in stimulation of the cell cycle (Marshall 1995). It appears that prolonged activation of the pathway is a requirement for MAPK translocation to the nucleus (Marshall 1995; Brondello et al. 1997). This sensitivity to temporal regulation implies that the duration of activity of a cascade upon stimulation must be regulated. It has been suggested that this regulation may occur by down-regulation or degradation of the cell surface receptor soon after activation (Marshall 1995) or by negative feedback from a MAPK phosphatase, whose expression is stimulated by MAPK (Brondello et al. 1997; Keyse 2000; Asthagiri and Lauffenburger 2001).

# MATERIALS AND METHODS

### Simulation model

Our simulation model for a generic MAPK signaling cascade is derived from that of Huang and Ferrell (1996). Our simulation consists of three parts whose properties can be studied independently or together: (a) activation and inactivation of the cell surface receptor, (b) the phosphorylation cascade, and (c) negative feedback by the last member of the cascade. Receptor synthesis, activation, and inactivation are summarized as follows:

$$dR/dt = k_1 \quad \text{if } R < 1 \tag{1}$$
  
$$dR/dt = 0 \quad \text{if } R \ge 1$$

$$dRL/dt = k_2 \cdot L \cdot (R - RL) - k_3 \cdot RL$$
(2)

where R is the receptor, L is a ligand (a growth factor, for instance), RL is the activated growth factor–receptor complex, and k is a rate constant corresponding to labeled arrows in Figure 1. Equation (1) describes up-regulation of the receptor to a level normalized to 1. Equation (2) describes changes in the level of active ligand–receptor

Nijhout et al.



**Fig. 1.** Structure of the MAPK cascade. We assume that the active receptor–ligand complex ( $RL_{active}$ ) controls the phosphorylation of MAPKKK directly. The MAPKK and MAPK levels of the cascade are activated by double phosphorylation. Dotted lines indicate negative feedback by the activated MAPK. The rate constants for the reactions are designated as  $f_n$  for the forward (phosphorylation) reactions and  $r_n$  for the reverse (dephosphorylation) reactions.

complex (*RL*) as a fraction of the total amount of receptor. We assume that the total level of receptor begins to decline exponentially after initial stimulation by the ligand, and this down-regulation is given by  $dR/dt = -k_3 \cdot [R]$ , using the same rate constant as inactivation of the ligand–receptor complex.

The phosphorylation and dephosphorylation reactions of the MAPK cascade are represented by Eqs. (3)–(5), where we follow the active (phosphorylated) kinase as a fraction of the total kinase present. To simplify notation we omit the MAP prefix (so MAPKKK becomes KKK, etc.), and we indicate single and double phosphorylations by P and PP, respectively. We also assume that the rate constants for the first and second phosphorylation have the same value and are thus represented by the same parameter (Fig. 1).

$$dKKKP/dt = f_1 RL (N_1 - KKKP) - r_1 KKKP$$
(3)

$$dKKP/dt = f_2 KKKP (N_2 - KKP - KKPP) - r_2 KKP + r_2 KKPP - f_2 KKKP KKP (4.1) dKKPP/dt = f_2 KKKP KKP - r_2 KKPP (4.2)$$

$$dKP/dt = f_3 KKPP (N_3 - KP - KPP) - r_3 KP + r_3 KPP - f_3 KP KKPP$$
(5.1)

$$dKPP/dt = f_3 KP KKPP - r_3 KPP$$
(5.2)

Here N indicates the total amount of each kinase in the cascade. When N is set to unity, the values of the kinases represent the fraction of the total amount of kinase that is in the single- or doublephosphorylated form (the effect of different assumptions about the values of N is explored below [cf. Figs. 10 and 11]). f is the rate constant for the forward reaction (phosphorylation), and r is the rate constant for the reverse reaction (dephosphorylation). We assume that the source of phosphates is saturating, so all rate constants are in units of t<sup>-1</sup>. Equation (3) represents the activation of MAPKKK by the activated receptor/Ras component of the cascade. Equations (4) and (5) represent the double phosphorylation of the MAPKK and MAPK, respectively, with Eqs. (4.1) and (5.1) representing the first phosphorylation and Eqs. (4.2) and (5.2) the second phosphorylation. We note that this reaction scheme differs from that used by Huang and Ferrell (1996) in that we explicitly model mass action kinetics rather than Michaelis-Menten kinetics. This approach allows us to investigate the sensitivity of the system to each of the kinetic constants and to the total amount of each kinase independently rather than via their effect on composite parameters such as  $V_{max}$ and the Michaelis constant. Unless noted otherwise, the parameter values used in all simulations shown below are as follows:  $f_1 = f_2 =$  $f_3 = 1150, r_1 = r_2 = r_3 = 150, N_1 = N_2 = N_3 = 1.$ 

Feedback inhibition of MAPK occurs when active MAPK induces or enhances the synthesis and activation of a MAPK phosphatase, which dephosphorylates and thus inactivates MAPK. Feedback inhibition by the terminal product of the cascade (a MAPK-PP) is believed to occur at the first and last steps in the cascade (Fig. 1). This inhibition is simulated by using the MAPK-PP to enhance the dephosphorylation (phosphatase) rate in Eqs. (3), (5.1), and (5.2) above, giving the following set of equations that model the MAPK cascade with feedback:

$$dKKKP/dt = f_1 RL (N_1 - KKKP) - r_1 KKKP KPP$$
(6)

$$dKKP/dt = f_2 KKKP (N_2 - KKP - KKPP) - r_2 KKP$$

$$KKPP/dt = f_2 KKKP KKP - r_2 KKKP (7.1)$$

$$dKP/dt = f_3 KKPP (N_3 - KP - KPP) - r_3 KP KPP$$
  
+ r, KPP KPP - f, KP KKPP (8.1)

$$dKPP/dt = f_3 KP KKPP - r_3 KPP KPP$$
(8.2)

The MAPK cascade can also be subject to positive feedback from MAPK to MAPKKK in some cells, and this feedback mechanism can result in a bistable switch with sharp transitions between the active and inactive forms of MAPK and hysteresis (Ferrell and Xiong 2001). The consequences of this form of positive feedback on evolvability is not examined here.

### **RESULTS AND DISCUSSION**

d

### Kinetic behavior of the simulation model

We first examined the kinetic behavior of the basic model, that is, in the absence of up- and down-regulation of the receptor (setting R in Eq. (2) to 1) and without the negative feedback. The activation level of each member of the cascade in response to a range of stimulus strengths is shown in Figure 2. This figure illustrates an important property of the cascade, *ultrasensitivity* to stimulation, which was elucidated by Huang and Ferrell (1996; see also Ferrell 1996, 1997, 1999; Hoek and Kholodenko 1997). Ultrasensitivity describes the increasing steepness of the sigmoid "activity" curve for each successive member of the cascade. Huang and



**Fig. 2.** Response of the model to variation in stimulation by ligand (*L*). Progressively lower members of the cascade exhibit increasing sensitivity and an increasingly steep sigmoidal response. The terminal member of the cascade (MAPK-PP) is fully saturated at L = 0.1 and half-saturated at L = 0.008, using the parameter values shown in Materials and Methods and in the legend to Fig. 3.



**Fig. 3.** The effect of various components of the model on the time evolution of the MAPK-PP level. (A) Profile in the absence of feedback and down-regulation of stimulation by ligand; (B) profile with feedback only; (C) profile with down-regulation of ligand stimulation only; (D) profile with both feedback and ligand down-regulation. All plots use the following parameter values:  $f_1 = f_2 = f_3 = 1150$ ,  $r_1 = r_2 = r_3 = 150$ ,  $N_1 = N_2 = N_3 = 1$ .



**Fig. 4.** Experimentally obtained profile of p44<sup>MAPK</sup> activity (phosphorylation) in CCL39 cells after stimulation by serum. Redrawn from Brondello et al. (1997).

Ferrell (1996) showed that the increasing steepness of the sigmoid response is a consequence of the double phosphorylation steps. For the parameter values used here, the response to ligand (*L*) is at half saturation at L = 0.01 and is saturating at L > 0.05. As noted above, the effect of double phosphorylation of the middle and terminal members of the cascade is to increase the steepness of the switch (Huang and Ferrell 1996). The same effect could be obtained by adding additional steps to the cascade, though this solution does not appear to have evolved in nature.

The concentration profile of active MAPK that develops upon exposure to a ligand is shown in Figure 3A. Upon introduction of ligand the level of active MAPK (MAPK-PP) rises rapidly and then stabilizes at a steady state whose value is determined by the relative values of the phosphorylation and dephosphorylation rates. It is clear that Figure 3A does not reproduce the detailed temporal pattern of activation observed experimentally (Fig. 4). In CCL39 fibroblast cells, MAPK activity rises to a peak upon stimulation and then declines, first rapidly and then more slowly with an approximate half-life of 45 minutes (Brondello et al. 1997). Changing the relative values of the phosphorylation and dephosphorylation rates, in any combination, changes the level of the steady state but has no effect on the overall shape of the time profile, that is, it does not produce an initial peak of phosphorylation of MAPK (Fig. 4). The peak-like initial overshoot is readily simulated by the introduction of the negative feedback loop at the end of the cascade (Fig. 3B).

After the initial overshoot, the level of active MAPK settles down to a constant steady state, where it remains indefinitely. In life, however, the level of active MAPK gradually declines (Fig. 4). We found two ways to simulate this decline. The most straightforward way is by introducing downregulation of the receptor complex (Fig. 3C). The rate of receptor complex deactivation ( $k_3$  in Eq. (2)) determines the apparent half-life of active MAPK. We have already noted that different persistence times of MAPK activity can produce different physiological effects in the cell. As can be seen from Figure 3, down-regulation of the receptor, by itself, does not reproduce the initial sharp overshoot of MAPK activity. Hence, to produce the experimentally observed profiles of MAPK activity upon stimulation, it is necessary to have both receptor inactivation and terminal negative feedback (Fig. 3D). The second way of producing at least the appearance of a gradual decline in MAPK activity is by reducing the feedback inhibition, so that the rate of dephosphorylation of MAPK is, for a time at least, substantially slower than its rate of phos-



**Fig. 5.** Variation in the persistence time of MAPK-PP is affected by the rate of down-regulation of the growth factor-receptor complex (A) and by the rate of phosphatase activity (B). Graphs represent the time evolution of MAPK-PP levels after stimulation by ligand at time = 0. Numbers on the right side indicate the factor by which the relevant rate is reduced to obtain the profile indicated, where 1.00 represents the value used to obtain the dynamics in Figure 3D. Reducing either the rate of down-regulation of the receptor or the phosphatase activity increases the persistence time of active MAPK-PP.

phorylation. The results of such a simulation are shown in Figure 5. Changing the rate of receptor down-regulation (Fig. 5A) decreases the rate at which the level of MAPK-PP declines, whereas changing phosphatase activity (Fig. 5B) mainly elevates the peak level of MAPK-PP.

### Nondimensional parameters

The parameter values used to obtain the results in Figure 3 are shown in the legend. There are almost certainly an infinite number of combinations of parameter values that could produce the same results. For instance, the value of an equilibrium constant is the ratio of the values of the forward and backward rate constants, and the effect of increasing the value of the forward rate constant by an arbitrary factor can be offset by increasing the value of the backward rate constant by a similar factor. In systems of even moderate complexity, it is typically not possible to deduce the patterns of trade-off among parameters by inspection of the equations. It is possible, however, to express this system in terms of nondimensional parameters. Nondimensionalization has several advantages: It almost always reduces the number of parameters, and it clarifies the trade-off relationships among the original parameters (Edelstein-Keshet 1988).

We calculated the following dimensionless form of the MAPK cascade:

$$dKKKP/dt = A_1 RL (1 - KKKP) - KKKP KPP$$
(9)

$$dKKP/dt = A_2 KKKP (1 - 2 KKP - KKPP)$$

$$-A_{3} (KKP - KKPP)$$
(10.1)  
dKKPP/dt =  $A_{2} KKKP KKP - A_{3} KKPP$ (10.2)

$$dKP/dt = A_4 KKPP (1 - 2 KP - A_5 KPP)$$

$$+ A_6 \operatorname{KPP} (A_5 \operatorname{KPP} - \operatorname{KP})$$
(11.1)

$$dKPP/dt = A_4 KKPP KP/A_5 - A_6 KPP^2$$
(11.2)

where  $A_{1-6}$  are the new dimensionless parameters, whose definitions are as follows:

where  $k_3$  is the decay rate of the receptor–ligand complex (*RL*) from Eq. (2). In this case the nondimensionalization reduces the number of parameters only slightly (from 10 to 6). The dimensionless parameters are made up of combinations of the original parameters, and the trade-offs among them can now be deduced by inspection (keeping in mind that  $A_4$ ,  $A_5$ , and  $A_6$  interact as shown in Eq. (11)). Thus, if  $N_1$  is substantially smaller than  $N_2$  and  $N_3$ , this would require that  $f_2$  be substantially larger than  $f_1$  and  $f_3$ , all other things being equal to obtain the behavior shown in Figure 3. The constraints given by the definitions of  $A_{1-6}$  and the requirement to produce the dynamics of Figure 4 suggest that once the values of some of the parameters are known, others can be predicted to fall within rather narrow limits. Nondimensionalization

thus produces a unit-less system and a somewhat simplified notation that can facilitate certain types of analysis.

### Steady-state behavior

The ultimate role of the MAPK cascade is to deliver a quantity of activated MAPK to target sites in the cell. The amount of active MAPK varies over time in a pattern that is determined by the parameters of the mechanism. This dynamic temporal variation makes it difficult to describe and evaluate the relative significance of the various parameters. The effect of parameter variation can, however, be readily quantified by studying the steady-state behavior of the system. Eliminating receptor down-regulation and feedback allows us to study the timeindependent steady state of the cascade, which is a measure of the amount of active (doubly phosphorylated) MAPK that is deliverable. At steady state, the derivatives in Eqs. (3) through (5) are 0, and the system of four equations can be solved for [MAPK-PP] as a function of the eight reaction constants and the relative number of molecules at each step in the cascade. After some algebra, the solution can be expressed as follows:

$$\begin{aligned} \text{MAPK-PP}_{\text{ss}} &= L^4 f_1^4 f_2^4 f_3^2 N_1^4 N_2^2 N_3 / r_3^2 \left( (L \ f_1 + r_1)^2 \ r_2^2 \right. \\ &+ L \ f_1 \ (L \ f_1 + r_1) \ f_2 \ r_2 \ N_1 + L^2 \ f_1^2 \ f_2^2 \ N_1^2 \right)^2 \\ &+ L^2 \ f_1^2 \ f_2^2 \ f_3 \ r_3 \ N_1^2 \ N_2 \ \left( (L \ f_1 + r_1)^2 \ r_2^2 \right. \\ &+ L \ f_1 \ (L \ f_1 + r_1) \ f_2 \ r_2 \ N_1 + L^2 \ f_1^2 \ f_2^2 \ N_1^2 \right) \\ &+ L^4 \ f_1^4 \ f_2^4 \ f_3^2 \ N_1^4 \ N_2^2 \end{aligned} \tag{12}$$

Here MAPK-PP refers to the level of MAPK-PP as a fraction of the total MAPK available, L refers to the active receptor– ligand complex (RL elsewhere), and other parameters are as defined before. Equations (3)–(5) can also be expressed in nondimensional form as follows:

$$dKKKP/dt = RL - RL KKKP - B_1 KKKP$$
$$dKKP/dt = B_2 KKKP(1-2 KKP- KKPP)$$
$$- B_3 (KKP - KKPP)$$
$$dKKPP/dt = B_2 KKKP KKP-B_3 KKPP$$
$$dKP/dt = B_4 KKPP(1-2 KP)$$
$$+ B_5 KPP (B_6-B_4 KKPP) - B_6 KP$$
$$dKPP/dt = B_4 KKPP KP/B_5-B_6 KPP$$

using the following nondimensional parameters:

$$B_1 = r_1/f_1 \qquad B_2 = N_1 f_2/f_1 \qquad B_3 = r_2/f_1 B_4 = N_2 f_3/f_1 \qquad B_5 = f_3 / N_3 r_3 \qquad B_6 = r_3/f_1$$

This nondimensionalization requires a different choice of nondimensional parameters because the equations do not include decay of the receptor complex. The nondimensional steady state is found to be

$$\begin{aligned} \text{MAPK-PP}_{ss} &= (B_2^4 B_4^2 L^4) / (B_1^4 B_3^4 B_5^2 + 2 B_1^3 B_3^3 B_5^2 L (B_2 + 2 B_3) \\ &+ B_1^2 B_3^2 B_5 L^2 (6 B_2 B_3 B_5 + 6 B_3^2 B_5 \\ &+ B_2^2 (B_4 + 3 B_5)) + B_1 B_3 B_5 L^3 (B_2 + 2 B_3) \\ &(B_2^2 B_4 + 2 B_5 (B_2^2 + B_2 B_3 + B_3^2)) \\ &+ (B_2^4 B_4^2 + B_2^2 B_4 B_5 (B_2^2 + B_2 B_3 + B_3^2) \\ &+ (B_2^2 + B_2 B_3 + B_3^2)^2 B_5^2) L^4 \end{aligned}$$
(13)

### Nijhout et al.

Although it is possible to use Eq. (13) to study the steadystate behavior of the nondimensional system, for the present purposes we only examine the fully dimensional system represented by Eq. (12), because we are ultimately interested in the relative effects of variation of individual genes. How variation in one gene might interact with that of other genes can be deduced from the structure of the nondimensional system. The steady-state Eq. (12) is sufficiently complicated that the effect of parameter variation cannot be easily deduced by inspection. A graphical analysis (see below) provides some of the required insights. Before proceeding with the analysis of the steady state, it is important to have clarity about what is meant by genotype and phenotype in this system.

### What is the genotype?

The genetically determined properties of this system are the parameters of the simulation model. Mutations in the regulatory region of the MAPK genes will affect their level of expression, and this is modeled by variation in N. Mutations in coding regions will affect the kinase and phosphatase rate constants, the k of the model. Alternative alleles can thus be represented by alternative values of these parameters. We have already seen some ways in which parameter variation affects the behavior of this system. A systematic exploration of the sensitivity of this system to parameter variation should give us some insight into the sensitivity of this cascade to mutations and genetic variation in its various components. Evolvability is in effect a measure of the association between genetic variation and phenotypic variation. Mutations (or allelic variation) that have no effect on the phenotype will be selectively neutral and thus will not affect evolvability of the system. Thus, we would like to understand how sensitive the phenotype is to all the possible kinds of genetic variation in the underlying mechanism (Nijhout 2002).

### What is the phenotype?

Evolvability is defined with reference to a specific phenotype, so it is necessary to establish what the relevant phenotype is, whose variation and evolvability we wish to examine. Equations (12) and (13) were derived with the assumption that variation in the total deliverable amount of active MAPK is the phenotype of interest. Other variable phenotypes of the MAPK cascade are sensitivity to stimulation, amount of amplification of the signal, persistence of the MAPK signal, and robustness to parameter variation. All these phenotypes can presumably be under independent selection, and they should be able to evolve if the appropriate genetic variation is available. We first consider the evolvability of the phenotype described by Eq. (12) before addressing the other possible phenotypes.

# Evolvability of MAPK-PP delivery (the MAPK-PP phenotype)

The function of the MAPK cascade is to deliver activated MAPK to the nucleus upon receiving an appropriate signal

at the cell surface. The phenotypes of interest here are the total amount of MAPK-PP delivered and the amount of MAPK-PP delivered relative to the strength of the signal received at the cell surface. The steady-state Eq. (12) shows that the parameters interact with each other in relatively complicated ways, which makes it difficult to get a sense of the behavior of this system from simple univariate plots. Ideally one would like to explore the full 10-dimensional parameter space at once, but such a space cannot be depicted. Instead, we examine the steady-state properties of this system by means of selected bivariate plots. Although the dimensionality of such plots is still low relative to the dimensionality of the whole system, they nevertheless give an immediate indication of how genes interact in controlling the value of the phenotype. We begin by setting the parameters to values that produce the kinetics shown in Figure 3. This provides a baseline from which we can begin to explore the effects of parameter variation and hence the evolvability of this system. The effects of variation in the rate constants and expression level of the various kinases on the steady-state MAPK-PP phenotype are shown in Figures 6–12. We refer to these graphs as phenotypic surfaces (Rice 1998; Nijhout 2002). The sensitivity of a system to mutations in a given gene is proportional to the slope of the phenotypic surface in the direction parallel to the axis that represents that genetic property. Thus, in areas where the surface is flat and orthogonal to the phenotypic axis, the systems would be insensitive to mutations and genetic variation in that region would be effectively neutral, because changes in genetic values have no effect on the phenotype.

### Variation in kinase activity

We first examine the effects of variation in the rate constants. Figure 6 illustrates the sensitivity of the system to variation in the first and second kinase rate, under different strengths of stimulation by ligand. Here we vary the ligand concentration from about its half-saturation level to its full saturation level (cf. Fig. 2). At full ligand saturation, the steady-state behavior of the MAPK cascade is relatively insensitive to genetic variation over a large portion of the range of values of the rate constants considered.

At subsaturating levels of ligand, the system is very sensitive to genetic variation in the kinase rate constants. This is evidenced by the fact that there is no horizontal flat region in the phenotypic surfaces except at high kinase rates (Fig. 7). Thus, if the system typically works at saturating levels of ligand, then one would predict that all but the most severe hypomorphic (or null) mutations in the kinase rate constants would be neutral with respect to this phenotype. As a result, one would expect to find a lot of small-level genetic variation in the rate constants of the kinases, which would be detectable as sequence variation in the coding regions of the kinases.



**Fig. 6.** (A–C) The effect of variation in the three kinase rate constants  $(f_1, f_2, and f_3)$  on the responsiveness to ligand stimulation. The diagrams represent phenotypic surfaces that describe how the steady-state level of MAPK-PP depends on the values of two parameters. The steady-state concentration of MAPK-PP levels off at high ligand concentration and also at high values of each of the kinase rate constants, so in this region variation in either parameter has little or no effect on the phenotype. At a given level of stimulation by ligand, the steady-state level of MAPK-PP depends on the value of the kinase rate constants. To obtain high levels of MAPK-PP, it is necessary to have higher kinase rates for the downstream  $(f_3)$  than for the upstream  $(f_1)$  members of the cascade.



**Fig. 7.** Relative sensitivity to variation of the first  $(f_1)$  and second  $(f_2)$  kinase constants under different levels of stimulation by ligand (*L*). At and near saturating levels of ligand (A, L = 0.1; B, L = 0.05) the phenotypic surface has a large flat region where variation of the kinase rate constant has no effect on the MAPK-PP phenotype. Near half-saturation (C, L = 0.01) MAPK-PP levels are sensitive to a broad range of kinase activities.

It is worth asking whether ligands typically occur at supra- or subsaturation levels. If the principal function of the cascade is to produce a switch-like response, then one would expect stimulating concentration of ligand to generally be suprasaturation. If, on the other hand, it is necessary to dynamically control the level of MAPK-PP in response to a graded signal, then one would expect the ligand level to be centered roughly at half-saturation, because that is where one would obtain the most accurate and sensitive control. MAPK-PP levels typically fluctuate dynamically and are at peak level for only very brief periods, if at all (e.g., Brondello et al. 1997). These kinds of fluctuations could be obtained by very brief suprasaturation pulses of ligand (e.g., Fig. 3D), or they could come about through dynamic fluctuations in the level of ligand. In the latter case we would expect the ligand levels to remain at subsaturation levels.

We can see from Figure 7 that the level of ligand should have a profound influence on the evolvability of the system. If the ligand level is typically above saturation, then a situation like that depicted in Figure 7A should be obtained. In this case we would expect most variation in the rate constants to be neutral with respect to phenotypic variation, as we just noted. But if the ligand is typically below saturation, we would expect a situation like that depicted in Figure 7C, in which the level of MAPK-PP is very sensitive to genetic variation in one or another of the rate constants. The sensitivity of the phenotype to genetic variation is proportional to the slope of the genetic surface (Nijhout 2002), so we can examine the relative sensitivities to the two rate constants shown in Figure 7 by plotting the first derivatives with respect to the rate constants. Figure 8 illustrates the first derivatives with respect to each of the rate constants as a function of both rate constants. The system is most sensitive to variation in  $f_1$  whenever  $f_1 < f_2$ , and vice versa. We noted above that  $f_2$  is likely to be substantially larger than  $f_1$  when  $N_1$  is smaller than  $N_2$  and  $N_3$ . Under these conditions, the system should be more sensitive to variation in  $f_1$  than to variation in  $f_2$ . If this is true, then under stabilizing selection one would expect stronger selection against variation in f<sub>1</sub> than to variation in  $f_2$ . Recently, Riley et al. (in press) showed that in Drosophila there is little or no sequence variation in the upstream members of the Ras-mediated MAPK cascade, whereas the middle members of the cascade are much more variable. This suggests that the upstream members are subject to strong purifying selection and that selection on the middle members is much weaker. This indicates that the system is much less tolerant of genetic variation in MAPKKK than in MAPKK, which could be explained if  $f_1 < f_2$ .

### Variation in phosphatase activity

In general, one expects the phosphatase activity to be smaller than the kinase activity, so that the equilibrium upon stimulation would favor the accumulation of activated kinases. In



**Fig. 8.** The first derivatives of MAPK-PP with respect to  $f_1$  (light gray surface) and  $f_2$  (dark gray surface). The value of the first derivative indicates the slope of the phenotypic surface of Figure 7C in the direction of the nominal axis. The derivative with respect to  $f_1$  is largest wherever  $f_1 > f_2$ . The sensitivity of the phenotype to genetic variation is proportional to the slope of the phenotypic surface, so this implies that if  $f_1 > f_2$  the phenotype is more sensitive to genetic variation in the first kinase constant than to variation in the second one. Under stabilizing selection, the system will tend to tolerate less genetic variation in  $f_1$  than in  $f_2$ .

Figure 9 we illustrate the relative sensitivities of MAPK-PP delivery to variation in kinase and phosphatase activities at the first level of the cascade. The relative sensitivity to phosphatase appears to depend on whether stimulation by ligand is below or above saturation. At or above saturation the system is almost completely insensitive to variation in phosphatase activity, except at very low kinase rates, whereas at subsaturating levels of ligand, variation in phosphatase can have a dramatic effect on variation in deliverable MAPK-PP.

### Variation in expression level

In addition to genetic variation in kinetic activity, there can be genetic variation in expression level. Genetic variation in the expression level of a gene that is not due to genetic variation in its transcriptional activators or inhibitors must be due to variation in the regulatory region of the gene. Variation in the *expression level* of the kinases is simulated by variation in the value of N. This can have a large effect on the amount of deliverable MAPK-PP. Figures 10 and 11 show that variation in the expression level of MAPK generally has a large effect on deliverable MAPK-PP, which increases linearly with increasing levels of the kinase. By contrast, the effects of variation in the expression level of either MAPKK or MAPKKK is strong at low expression levels but saturates at higher levels. Thus, genetic variation in the regulatory region of MAPK would be expected to have a profound effect on this phenotype, over a broad range of values, and would



**Fig. 9.** The effect of variation in kinase rate  $(f_1)$  and phosphatase rate  $(r_1)$  of the first level of the cascade on the steady-state level of MAPK-PP, under two levels of stimulation by ligand (L). (A) L = 0.05 (saturating); (B) L = 0.01 (near half-saturating).

be expected to be under stronger selection than variation in the regulatory regions of the two upstream kinases.

### Evolvability of sensitivity to stimulation

Ultrasensitivity of the MAPK cascade emerges from two structural features: the multiplicity of steps in the cascade and multiple phosphorylation of some members of the cascade (Huang and Ferrell 1996) (Fig. 1). Hence, the principal ways of genetically changing the sensitivity of the system are to have genetic variation in the number of steps in the cascade or in the degree of phosphorylation of the relevant steps. Neither of these changes could be accomplished by mutational changes in the existing system but would require the intervention of new regulatory events. Hence, these ways of altering the sensitivity to stimulation should probably be thought of as differences in kind rather than degree. Adding phosphorylation sites, and adding levels to a cascade, would be



**Fig. 10.** (A–C) The effects of variation in the expression level (N) of the three kinases on the response of MAPK-PP to ligand (L) concentration. The expression level controls the number of molecules of each kinase that are present. Increasing expression of the first two levels of the cascade  $(N_1 \text{ and } N_2)$  does not result in an increase in the maximum steady-state of MAPK-PP, which depends only on the expression level of MAPK itself  $(N_3)$ . Variation in ligand concentration has little effect on the steady state of MAPK-PP except at very low levels of expression of the kinases.







GF



**Fig. 11.** (A–C) The relative effects of variation in the expression level of the three kinases on the level of MAPK-PP for ligand concentration L = 0.05. Each graph shows the effect of a different pair-wise combination of the three kinases. The phenotypic surfaces for the first two kinases are nearly flat, indicating that variation in their expression has little effect on the steady-state level of MAPK-PP. Therefore, selection for increased levels of MAPK-PP will not favor alleles that increase the expression level of the first two kinases.

**Fig. 12.** Cascades of different levels differ in their response to stimulation by ligand (L) and variation in the kinase rate of the first member of the cascade ( $f_1$ ). (A) One-level cascade; (B) two-level cascade; (C) three level cascade. z axes represent the steady-state responses of the last members of each cascade. These graphs show clearly that the shape of the phenotypic surface and the local slopes change when the mechanism changes. Adding members to a cascade makes the steady-state level of the last member increasingly insensitive to variation in ligand and the first kinase ( $f_1$ ).

considered to be evolutionary novelties at the molecular level. It would be interesting to know whether these kinds of changes are inherently less probable than mutational changes in the activities or the expression levels of the members of the cascade.

Mutational changes in rate constants and expression levels can alter the sensitivity to stimulation, but only hypomorphic mutations can do so, and these also diminish the overall level of deliverable MAPK-PP (Figs. 6 and 7). Hypermorphic mutations in general are expected to be without phenotypic effect, just as in the case of the MAPK-PP phenotype. Thus, mutational variation in the members of the cascade cannot increase sensitivity to stimulation but can, in many cases, decrease this sensitivity.

### Evolvability of signal persistence

Quantitative differences in the persistence of active MAPK can cause qualitative differences in the nature of the cellular response. In PC12 cells, a brief persistence triggers cell division, whereas a prolonged persistence causes the cell to begin differentiation (Marshall 1995). The rate of decay of active MAPK is governed almost entirely by the rate of inactivation of the receptor complex (Fig. 3). The strength of the negative feedback within the cascade can also affect persistence of the signal. The effect of variation in feedback can be modeled by varying the phosphatase rate constants ( $r_1$  and  $r_3$ ) through which the feedback inhibition takes place (Fig. 1). The effect of variation in these parameters is illustrated in Figure 5. Here signal persistence is accompanied by a slight increase in the peak level of MAPK-PP.

### **Evolvability of robustness**

Robustness refers to the ability of a system to withstand variation in those genetic or environmental factors that are essential for the generation of the phenotype (Nijhout 2002). The phenotype is robust to a given factor, say the value of a rate constant, if variation in that value has little or no effect on the phenotype. From the various bivariate plots shown above, it is clear that there are large regions of parameter space where the phenotypic surface is flat and normal to the phenotype (z) axis. In these regions, variation of the relevant parameters does not produce variation in the phenotype, and hence the phenotype is robust to that variation.

Robustness can thus evolve by mutations and selection that move genotypes into a region where the phenotypic surface is flat. Evolvability of robustness thus exists if a flat surface is within "reach" of genetic variation and mutation. For the steady state, simply increasing the values of the forward reaction constants can make the MAPK-PP level robust to their variation, because they in effect take the system to saturation. The flat horizontal regions of the phenotypic surfaces in Figures 6, 7, 10, and 11 represent regions where the phenotype is robust to variation in the relevant parameters. Robustness can also evolve, not by moving the genotype into a flat region of the surface but by altering the local shape of the surface through evolutionary changes in the underlying mechanism. As a result, the surface around a given genotype (or set of parameter values) can go from being steeply sloped to being nearly horizontal and flat. Figure 12 illustrates how the addition of members to the cascade has had such an effect. Figure 12 also illustrates the effect of sensitivity to ligand (e.g., growth factor), and of variation in the kinase rate of the first step in the cascade, on the steady state level of the last member of the cascade for a cascade of one, two, and three steps.

### CONCLUSIONS

In the absence of genetic variation there can be no evolution. Genetic variation by itself, however, does not produce evolvability if it does not cause phenotypic variability. Evolvability is therefore a function of the way genetic variation interacts with the molecular and developmental mechanisms that produce the phenotype. Here we have explored a way of studying the evolvability of phenotypes that are produced by a relatively simple genetic mechanism, the MAPK signaling cascade. Because the parameters in this system represent the activities and expression levels of genes, evolvability is equivalent to the sensitivity of the properties of this system to parameter variation. A convenient way of depicting this sensitivity is by means of phenotypic surfaces. The sensitivities are proportional to the slopes of these surfaces (Nijhout 2002). Gene interactions are typically nonlinear, and this causes the quantitative effects of genetic variation to be context dependent. The phenotypic surfaces provide a convenient way of visualizing the ways in which variation in one gene affects sensitivity to variation in another.

Phenotypic surfaces are graphical representations of a mathematical model of the phenotype, and their shape is therefore completely defined by the structure of the model. The axes of a particular plot represent a range of genetic variation. Individual genotypes can be represented by points on the surface, and populations with genetic variation would be represented by a distribution of points. Mutations move genotypes from one place to another on the surface, and selection on populations alters the distribution of genotypes. Given a certain amount of genetic variation, selection on the phenotype will move the population mean genotype only if the phenotypic surface is not flat and normal to the phenotypic axis. This is because in flat horizontal regions genetic variation has no effect on the value of the phenotype. The way in which allele frequencies respond to selection on the phenotype depends on the covariance between genetic variation and phenotypic variation. The covariance between genetic and phenotypic variation depends on two factors: the

### Nijhout et al.

slope of the surface and how broadly individuals are distributed on the surface (Nijhout and Paulsen 1997; Nijhout 2002). Formulas for how the correlations and covariances can be calculated from the slopes of the phenotypic surface and the genetic variances of a population are given by Nijhout (2002).

The model mechanism gives us a way of constructing phenotypic surfaces but does not tell us how individuals are distributed on the surface. To locate individuals and populations on the surface, we must know the individual's parameter values. Parameter estimation is typically the most problematic part of mathematical modeling. Even if parameter values have been established experimentally, they are often determined under artificial controlled conditions, and this leads to uncertainty as to how closely they resemble the values that obtain in a normally operating system. In the vast majority of developmental mechanisms, we have little or no information at all about parameter values. Experimental developmental genetics has been concerned primarily with determining the connectivities of regulatory networks rather than the kinetics of their operation. The behavior of these networks under genetic variation is typically tested using mutations of extreme effect. Little or no information is presently available on the actual parameter values or on the range of parameter values in natural populations. This dearth of information greatly limits the predictive power of mechanistic models of evolvability like the one we presented here.

An explicit mathematical model of a developmental mechanism does, however, allow us to examine how genetic variation affects the function of the system. We have explored two ways in which the effects of, and constraints on, this variation can be examined: by means of a steady-state equation and by means of nondimensional equations. The steadystate formulation allows us to examine regions of parameter space where the system is most or least sensitive to genetic variation. Even though in its normal operation the system may never be at steady state, the steady-state equation nevertheless allows one to examine how sensitive the behavior of the system is to variation of parameter values over a broad range of operating points. The nondimensional forms of the equations, by contrast, give us information on trade-offs among parameter values. This information is essential to determine whether some mutations effectively cancel the effects of others and helps us understand how tolerance to variation in one gene can be constrained by variation in another.

Our analysis of the MAPK cascade shows that the response to stimulation saturates under high stimulation over a broad range of parameter values. Phenotypic surfaces level off and are flat at high genetic values, and this illustrates that there are firm limits to which the system can respond to upward selection on the phenotype, even in the presence of genetic variation in the underlying mechanism. Where the phenotypic surface is horizontal and flat there is no phenotypic variation, and thus nothing for selection to act on.

If in its normal operation the cascade merely switches between being maximally activated and completely inactivated, then it is possible that the wild-type ligand levels and kinase rate constants are such that the system saturates and the active phenotypes lie somewhere on the flat portion of the phenotypic surfaces. It is unlikely, however, that phenotypes lie very far from the "edge" of a surface because there is no obvious mechanism of selection that can move a genotype deeply onto a flat phenotypic surface, at least not by selection on this phenotype. If, however, the cascade works by dynamically modulating the level of MAPK-PP rather than as an all-or-none switch, then the wild-type genotypes may be such that regulatory variation in the system occurs mostly in the sloping portion of the phenotypic surfaces. This implies that the system should be intolerant of genetic variation in the region where the slopes are steep, because genetic variation in those regions would mask regulatory variation. Indeed, the findings of Riley et al. (2003) on genetic variation in the MAPK cascade are most easily explained if genetic variation within the steeply sloping regions is efficiently eliminated.

The above observations thus suggest two constraints on parameter values. First, it is unlikely that parameter values occur deeply into a flat region of the phenotypic surface, and second, it is unlikely that parameter values occur where the phenotypic surface is steeply sloped. It is therefore most likely that normal parameter values will lie near the edges of the phenotypic surfaces. Mutational variation in the direction of the slope should be eliminated by selection, because it would interfere with normal regulation, whereas mutational variation into the flat region of the phenotypic surface would be neutral to mutation and could therefore accumulate. One might therefore expect drift to gradually take genotypes deeper and deeper onto the flat regions of a phenotypic surface, although this would require mutations that increase the activity of the kinase, which would be expected to be much rarer than mutations that diminish the activity.

### Acknowledgments

We thank Lou D'Amico, Julia Bowsher, Jeff Marcus, Tomalei Vess, Adrienne Wells, and Andy Yang for critical comments. This work was supported in part by grants from the National Science Foundation. A. Berg and W. T. Gibson were supported by fellow-ships from a grant to Duke University from the Howard Hughes Medical Institute.

### REFERENCES

- Asthagiri, A. R., and Lauffenburger, D. A. 2001. A computational study of feedback effects on signal dynamics in a mitogen-activated protein kinase (MAPK) pathway model. *Biotechnol. Progr.* 17: 227–239.
- Brondello, J.-M., Brunet, A., Pouysségur, J., and McKenzie, F. R. 1997. The dual specificity mitogen-activated protein cascade phosphatase-1 and -2 are induced by the p42/p44<sup>MAPK</sup> cascade. *J. Biol. Chem.* 272: 1368–1376.

- Edelstein-Keshet, L. 1988. *Mathematical Models in Biology*. Random House, New York.
- Falconer, D. S., and Mackay, T. F. C. 1996. An Introduction to Quantitative Genetics. Addison Weseley Longman, Essex.
- Ferrel, J. E. 1996. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem. Sci.* 21: 460–466.
- Ferrell, J. E. 1997. How responses get more switch-like as you move down a protein kinase cascade. *Trends Biochem. Sci.* 22: 288–289.
- Ferrell, J. E. 1999. *Xenopus* oocyte maturation: new lessons from a good egg. *BioEssays* 21: 833–842.
- Ferrell, J. E., and Xiong, W. 2001. Bistability in cell signaling: how to make continuous processes discontinuous, and reversible processes irreversible. *Chaos* 11: 227–236.
- Garrington, T. P., and Johnson, G. L. 1999. Organization and regulation of mitogen-activated protein kinase signaling pathways. *Curr. Opin. Cell Biol.* 11: 211–218.
- Gerhart, J., and Kirschner, M. 1997. Cells, Embryos, and Evolution. Blackwell Science, Malden, MA.
- Haugh, J. M., and Lauffenburger, D. A. 1998. Analysis of receptor internalization as a mechanism for modulating signal transduction. J. Theoret. Biol. 195: 187–218.
- Herskowitz, I. 1995. Map kinase pathways in yeast: for mating and more. *Cell* 80: 187–197.
- Hoek, J. B., and Kholodenko, B. N. 1997. Why do protein kinase cascades have more than one level? *Trends Biochem. Sci.* 22: 288.
- Huang, C.-Y. F., and Ferrell, J. E. 1996. Ultrasensitivity in the mitogen-

activated protein kinase cascade. Proc. Natl. Acad. Sci. USA 93: 10078-10083.

- Keyse, S. M. 2000. Protein phosphatases and the regulation of mitogenactivated protein kinase signalling. *Curr. Opin. Cell Biol.* 12: 186–192.
- Lynch, M., and Walsh, B. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, MA.
- Marshall, C. J. 1995. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* 80: 179–185.
- Nijhout, H. F. 2002. The nature of robustness in development. *BioEssays* 24: 553–563.
- Nijhout, H. F., and Paulsen, S. M. 1997. Developmental models and polygenic characters. Am. Nat. 149: 394–405.
- Rice, S. H. 1998. The evolution of canalization and the breaking of Von Baer's laws: modeling the evolution of development with epistasis. *Evolution* 52: 647–656.
- Riley, R. M., Jin, W., and Gibson G. Contrasting selection pressures in components of the *Ras*-mediated signal transduction pathway in *Dro-sophila*. *Mol. Ecol.* (in press).
- Robinson, M. J., and Cobb, M. H. 1997. Mitogen-activated protein kinase cascades. *Curr. Opin. Cell Biol.* 9: 180–186.
- Roff, D. A. 1997. Evolutionary Quantitative Genetics. Chapman and Hall, New York.
- Schwartz, M. A., and Baron, V. 1999. Interactions between mitogenic stimuli, or, a thousand and one connections. *Curr. Opin. Cell Biol.* 11: 197–202.
- Wagner, G. P., and Altenberg, L. 1996. Complex adaptations and the evolution of evolvability. *Evolution* 50: 967–976.