Genetic-Algorithm Selection of a Regulatory Structure that Directs Flux in a Simple Metabolic Model

Alex Gilman and John Ross
Department of Chemistry, Stanford University, Stanford, California, 94305 USA

ABSTRACT A genetic algorithm (GA) is used to optimize parameters for allosteric regulation of enzymes in a model of a metabolic futile cycle, in which two metabolites are interconverted by a pair of irreversible enzymatic reactions. The cycle is regulated by end products of the surrounding pathway. The optimization criterion for the GA is the proper direction of chemical flux in the regulated cycle toward one or the other end product in response to a simple, time-dependent model of biochemical "need" based on externally imposed variation of the end product concentrations. An energetic cost, to be held to a minimum, is also imposed on the operation of the cycle. The best-performing individuals selected by the GA are found to switch rapidly the direction of net flux according to need. In different "environments" (specific time courses of end product concentrations), the GA produces better- or poorer-performing individuals. In some cases "generalists" and "specialists" are produced. The present approach provides, purely as a consequence of formally specifying the task of flux direction, the new result of numerical confirmation, in a simple model, of the intuition that negative feedback and reciprocal regulation are important for good flux direction in arbitrary environments, and gives rise to a diversity of structures, suggestive of the results of biological evolution.

INTRODUCTION

A fruitful hypothesis for the effect of evolution on biochemical pathways of metabolism is that these pathways in their present form have undergone optimization with respect to some discernible criteria over the course of their natural history. In line with this hypothesis, we examine a simple biochemical reaction model in which certain functional parameters are deliberately left unspecified and are made the object of an optimization procedure. Specifically, we study an idealized futile cycle embedded in a larger metabolic pathway. Futile cycles are of interest because they occur widely throughout metabolism and are in some cases implicated as sensitive points of metabolic control. Our model cycle consists of two "irreversible" enzymes that are regulated by signals from external species (that is, species in the pathway but downstream or upstream from the cycle). We fix the reaction structure of the pathway and certain intrinsic kinetic parameters for the enzymes (the Michaelis-Menten parameters K_m and V_{max}). We also specify that the regulation of the enzymes occurs only by modulation of their V_{max} through noncompetitive binding of the external species. The parameters governing this modulation, however, are not specified but are left to be optimized. The criterion for optimization is that the regulated futile cycle be able to carry out a metabolic function that we specify, and that it do so under time-varying constraints on the end point species of the pathway. It is thus a functional and time-dependent criterion. For the present study, the cycle's metabolic task is chosen to be flux direction, in which net flux through the cycle is required to be sent in the direction given according to a simple model of biochemical "need." The cycle carrying out this task idealizes an animal cell that metabolizes blood glucose for energy as long as the glucose concentration in the blood is adequate but synthesizes glucose for export if the glucose concentration in the blood drops too low. The input or output of a single cell can hardly affect the overall glucose concentration in the blood, so for the cell the blood is a reservoir of glucose. Likewise, the cell's store of ATP may be crudely seen as a reservoir, because various pathways may be active simultaneously in producing and degrading it. The idealized cell must funnel material in the correct direction from one reservoir to another in response to concentration levels in those reservoirs, converting its surplus energy to glucose when the glucose level is low, but supplying itself with energy from glucose when its ATP level is low. This task can therefore be seen as a mechanism in the maintenance of homeostasis in both the cell itself and its parent organism.

We optimize the regulatory parameters of the model numerically with respect to performance of the task. For added realism, an energetic cost constraint is also included. The optimization problem is, not surprisingly, quite elaborate, so a genetic algorithm (GA) is used as the optimization procedure (Davis, 1991) (this choice is motivated in greater detail under Model and Methods).

Several authors have previously studied the optimization of metabolic models. Models of pathways have been optimized for such criteria as maximization of pathway fluxes under fixed "evolutionary effort" (Heinrich and Holzhütter, 1985; Heinrich et al., 1987), minimization of the total osmolarity of pathway intermediates (Schuster and Heinrich, 1987), and the maximization of the largest control coefficient in the pathway (Schuster and Heinrich, 1987). These approaches were limited by the simple topology of the

Received for publication 13 March 1995 and in final form 20 July 1995.
Address reprint requests to Dr. John Ross, Department of Chemistry, Stanford University, Stanford CA 94305. Tel.: 415-723-9203, Fax: 415-723-4817; E-mail: ross@chemistry.stanford.edu.

0006-3495/95/10/1321/13 $2.00

pathways examined (typically linear or branched, although an analysis for a more complex pathway modeling glycolysis was undertaken; Heinrich et al., 1987) and by the fact that only the intrinsic kinetic parameters of the enzymes were considered. The reaction structure of a pathway has also been the subject of optimization, under the constraints of a specific overall chemical transformation and maximum simplicity of the pathway (Meléndez-Hevia and Isidoro, 1985). In a recent work, Bray and Lay used a limited GA to optimize reaction rates in a small but topologically nontrivial model of signal transduction (Bray and Lay, 1994). Their optimization criterion was that the model transform given input waveforms of extracellular signal into specific output waveforms of an intracellular signaling molecule, an example of a functional, time-dependent criterion. To our knowledge, only Savageau has treated the optimization of regulation for metabolic networks, presenting analyses for feedback (Savageau, 1974, 1975) and feedforward (Savageau, 1979; Savageau and Jacknow, 1979) inhibition of linear and branched pathways. He examined the strength and site of action of the inhibition in light of a number of optimality criteria, including fast responsiveness to changes in inhibitor concentrations, insensitivity to perturbations in the structure of the pathway, decreased production of intermediates in response to exogenous supply of end product, and stability of the steady state of the pathway with respect to perturbations in the initial substrate concentration.

Several interesting findings arise from our optimization. First, the most optimal individuals observed do respond rapidly to changes in both food supply and energy charge, fully reversing the direction of net flux in accordance to the need state. The regulatory pattern evident in these systems shows negative feedback and reciprocal effect on the opposing branches of the cycle. Although these regulatory motifs are fully consistent with intuitive expectations, the finding is significant in that *it arises purely as a consequence of specifying the task to be performed.* It thus serves as a numerical confirmation of the intuition.

The other findings are more surprising and appear in connection with the specific form of the time-varying external constraints. Optimization runs were performed on five different time courses of end point species concentrations, and strikingly, no global winner was found. That is, individuals found to be optimal on one or several of the courses proved not to be optimal on others. Indeed, we observe the appearance of "generalists," which perform well, if not optimally, on all of the course, and "specialists," which perform well on a single course but poorly on the others. The performance of generalists is generally accompanied by a higher expenditure of energy cost, although high cost does not appear to be sufficient for good performance. On some courses, the regulatory structures selected as optimal did not conform to intuitive expectations and in fact did not perform well relative to those selected on other courses. In this regard, some courses proved to be less stringent than others. This finding corresponds to the notion that under less stringent conditions, even a suboptimal way of doing things is adequate.

It is interesting to find results reminiscent of biodiversity and ecological evolutionary effects in a system as simple as ours. The analogy arises because of the fact that while the GA is not intended to model the process of biological evolution, the two processes share certain limitations on their effectiveness at optimization.

These findings are presented under Results and discussed in the subsequent two sections. Model and Methods details the model and the computation involved in the study. The concluding section summarizes the results and briefly discusses the significance of the method used. The implementation of the GA is given in the Appendix.

MODEL AND METHODS

The model, shown in Fig. 1, consists of three parts: an enzymatic cycle in which a pair of metabolic intermediates, A and B, is interconverted by a pair of enzymes, α and β; and two chemical "reservoirs," containing metabolic species F and T, respectively. These reservoirs represent concentrations of the respective species that at any instant are specified externally. Each of the reservoir species interconverts with one of the cycle intermediates by means of reciprocal first-order reactions. F is converted into A by a first-order reaction having rate constant k_1, and A is converted back into F by a reaction having rate constant k_{-1}. However, because F is a reservoir species, its concentration is unaffected by these reactions, whereas the concentration of A does change. A similar relationship holds between B and T, the reaction $B \rightarrow T$ having rate constant k_2, and $T \rightarrow B$ having rate constant k_{-2}. (The kinetic parameters used in the model are given in the figure caption.) The reaction converting A to B is catalyzed by enzyme α and proceeds with rate v_α; the conversion of B to A is catalyzed by

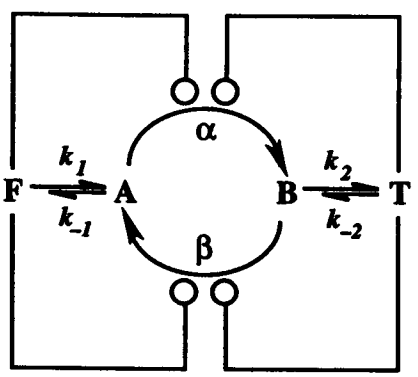


FIGURE 1 Diagram of the model. F and T are the reservoir species. A and B are the cycle intermediates, interconverted by enzymes α and β. Arrows indicate reactions, knobs indicate regulation. The kinetic parameters are: $k_1 = 10^{-2}\,s^{-1}$, $k_{-1} = 8 \cdot 10^{-3}\,s^{-1}$, $k_2 = 10^{-2}\,s^{-1}$, $k_{-2} = 4 \cdot 10^{-3}\,s^{-1}$. For enzyme α, $V_{max} = 1.6 \cdot 10^{-3}$ M s^{-1}, $K_m = 1.5 \cdot 10^{-6}$ M. For enzyme β, $V_{max} = 3.5 \cdot 10^{-3}$ M s^{-1}, $K_m = 2.0 \cdot 10^{-6}$ M.

enzyme β and proceeds with rate v_β. The kinetic equations describing the temporal variation of A and B are thus

$$\dot{A} = k_1F + v_\beta - k_{-1}A - v_\alpha$$
$$\dot{B} = k_{-2}T + v_\alpha - k_2B - v_\beta. \quad (1)$$

The reservoirs affect the concentration of the cycle species in two ways. The first is through the direct influx represented by the first term in each of Eqs. 1. The second and more interesting way is through control of the enzyme activities, where the reservoir species F and T are allowed to become effectors of the enzymes. The type of control modeled in this study is noncompetitive (allosteric) binding of the effectors, where each effector binds to the enzyme independently, according to the scheme (Fersht, 1984) shown in Fig. 2. In this scheme, the enzyme with effector bound is assumed to have altered catalytic activity toward its substrate compared to that of the enzyme without effector bound. The scheme as shown also relies on the simplifying assumptions that 1) the association and dissociation between enzyme and substrate are unaffected by the binding of the effector, and 2) the binding of substrate to enzyme is much faster than the conversion of bound substrate to product. Under these assumptions, the Michaelis constant K_M represents the equilibrium constant for the dissociation of the enzyme-substrate complex. It is easy to show that the expression for the rate of an enzyme in the presence of effectors is of the form (here shown for enzyme α):

$$v_\alpha = \frac{V_{\max,\alpha}A}{K_{M,\alpha} + A} \cdot R_{\alpha,F} \cdot R_{\alpha,T}, \quad (2)$$

where the form for the factors modifying the intrinsic Michaelis-Menten rate expression is

$$R_{\alpha,\epsilon} = \frac{K_{\alpha,\epsilon} + r_{\alpha,\epsilon}\epsilon}{K_{\alpha,\epsilon} + \epsilon} \quad (3)$$

The parameter $K_{\alpha,\epsilon}$ is the dissociation constant for the complex of enzyme α and the effector T or F, labeled ϵ, and $r_{\alpha,\epsilon}$ is the ratio of the catalytic rate constants for the enzyme with and without effector bound, respectively. This general form models both activators and inhibitors of the enzyme, depending on whether r is greater than or less than 1. The response described is hyperbolic, with the half-maximum effect exerted when $\epsilon = K$, and saturating at the maximum effect r for very high effector concentrations. The cumulative multiplication of modifying factors reflects the assumption that all effectors function at different sites on the enzyme and that the sites do not communicate.

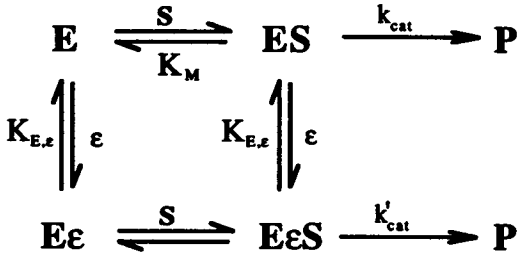


FIGURE 2 Mechanism of non-competitive binding used to model the interaction between an enzyme (E) and an effector (ϵ). In this approximation, K_M is the dissociation constant of the enzyme-substrate complex. The binding interaction between substrate and enzyme is unperturbed by the presence of the effector. Formation of product from the enzyme-substrate complex with effector bound proceeds with the altered rate constant k'_{cat}.

Because the model contains two enzymes, each affected by two reservoir species, there is a total of eight parameters (four K's and four r's) needed to describe the behavior of the system for given reservoir concentrations. This set of parameters specifies a regulatory pattern that determines the response of the system under differing conditions. The system constitutes a network in which the elements (chemical species) are connected by the described reactions and by regulatory links specified by the regulatory pattern.

The task of flux direction to be carried out by a network is quantified as follows. For each reservoir, a particular concentration value is chosen as optimal. As externally imposed variations decrease the reservoir concentration below this target, a positive state of need is induced. As the reservoir concentration rises above the target, again because of externally imposed variations, a negative need state is induced. The need state is labeled ξ_F for the reservoir F and ξ_T for the reservoir T. A functional form for ξ may be chosen so that there is an acceptable window of concentration around the target, within which the numerical value of the need is close to zero, but at the edges of which the value rapidly changes to a positive or negative one. To resolve conflicts, ξ_F is taken as a function of both the concentration of F and of the need state of T. In this way, T may override F, an analogy for a cell giving precedence to its own internal state over that of the organism. The particular functional forms for ξ_F and ξ_T used in this study are shown in Fig. 3. The target window for the concentration of T was chosen to be 3 mM to either side of 30 mM, and the target window for F was chosen to be 5 mM to either side of 60 mM.

The need states are used to compute the following function:

$$f \equiv \xi_T(k_2B - k_{-2}T) + \xi_F(k_{-1}A - k_1F). \quad (4)$$

The term multiplying ξ_F, the need state for the reservoir F, $k_{-1}A - k_1F$, is the net flux of concentration into the reservoir. Likewise, the term $k_2B - k_{-2}T$ that multiplies ξ_T is the net flux of concentration into the reservoir T. If the net flux into a reservoir has the same algebraic sign as the reservoir's need state, there will be a positive contribution to f, and opposite signs result in a negative contribution. A positive overall value of f thus indicates that the net flux through the cycle is directed in proper accord with the need states, and a negative value indicates that the net flux is directed backward. A zero value results when both reservoirs are within their target concentration ranges. In this case, any value of net flux may be considered proper.

The quantity f is useful in evaluating the performance of a network under conditions independent of time, such as

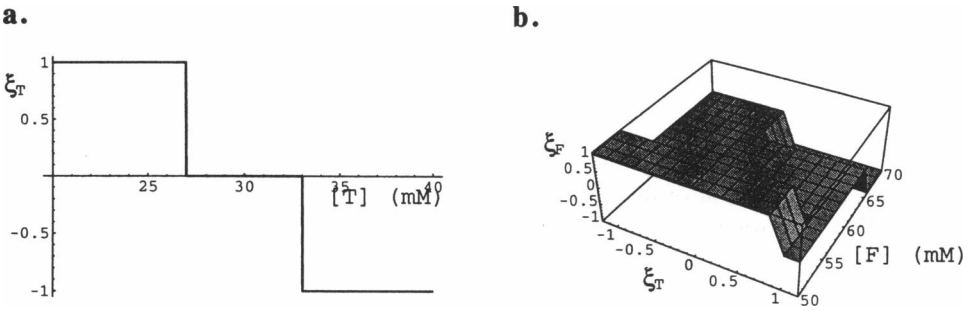


FIGURE 3 Plots of the need functions. (*a*) ξ_T versus the concentration of T, mM. (*b*) ξ_F versus ξ_T and the concentration of F, mM. The functions used are:

$$\xi_T = -\frac{1}{\pi}[\arctan a(T - T_1) + \arctan a(T - T_0)]. \qquad \xi_F = \frac{1}{\pi}\left[\frac{A(\xi_T) - B(\xi_T)}{2} - A(\xi_T)\arctan a(F - F_0) + B(\xi_T)\arctan a(F - F_1)\right].$$

$$A(\xi_T) = \frac{1}{2} - \frac{1}{\pi}\arctan a(\xi_T - 0.9), \quad B(\xi_T) = \frac{1}{2} + \frac{1}{\pi}\arctan a(\xi_T + 0.1). \quad a = 10^5, \quad F_0 = 55\text{ mM}, \quad F_1 = 65\text{ mM}, \quad T_0 = 27\text{ mM}, \quad T_1 = 33\text{ mM}.$$

instantaneous or steady-state conditions. However, time-dependent responses are of physiological interest as well. The integral of f over a period of time τ,

$$\Im \equiv \int_0^\tau f\,dt, \tag{5}$$

gives some indication of the fraction of the period during which the flux was directed properly. The reservoir concentrations may be changed by external means during the course of such a period, and consequently the need states may change. To achieve a high value of $\Im$, a network must be able to respond to these changes correctly.

The measure $\Im$ may be adjusted to include the effect of an energy "cost" imposed on the operation of the network. Such cost is incurred when, for instance, ATP is hydrolyzed in an enzymatic reaction. In this study, a cost is imposed on the operation of enzyme α. The cost is assumed to be expended at the same rate as that of the reaction catalyzed by α, namely v_α. This models a reaction that is driven by the hydrolysis of one stoichiometric equivalent of ATP. The reverse reaction requires no direct metabolic input. The cost in this model is taken to be independent of the concentration of species T, even though T serves as a rough analog of an energy carrier. The overall cost expended over a period of time τ is thus the overall flux through α,

$$C \equiv \int_0^\tau v_\alpha\,dt. \tag{6}$$

The cost is weighted and subtracted from $\Im$ to give a modified (cost-adjusted) measure,

$$\Im' \equiv \Im - mC. \tag{7}$$

The weighting factor m used in this study was arbitrarily chosen as 0.001. To achieve a high value of $\Im'$, a network must not only be able to respond quickly and correctly to changes in need state, but also simultaneously to minimize the energy expended for these responses.

The evaluation of a network's performance of the task proceeds as follows. First, a course of time variation for the concentrations of the reservoirs is constructed. In this study, the courses (see Fig. 4 for illustration) were 3000 s long. Two concentration regimes, "high" and "low," are chosen. For the reservoir F, the high regime in this study centered on 60 mM, and the low regime centered on 20 mM. For T, the high regime centered on 30 mM, and the low regime centered again on 20 mM. A sequence of concentration values, typically spaced at 60-s intervals, is then generated so that at each step there is some probability of switching from the current regime to the other, and if no switch occurs, the concentration fluctuates from the central point of the current regime by some percentage drawn uniformly from ±20%. The probability of switching between regimes was chosen so that there is on average a switch every 1200 s, i.e., a 1 in 20 chance of switching per 60-s step. The concentration values thus generated for each reservoir are then interpolated with a cubic spline.

Second, the Eqs. 1 are numerically integrated over the course for the network being evaluated, and the results are used to compute $\Im'$ (Eq. 7). Initial conditions for the integration are determined by setting the reservoir concentrations to their optimal values and letting the network relax to steady state. The higher the value of $\Im'$ achieved by a network the better is that network's performance.

To optimize a network for the performance of our task, a set of parameters (K's and r's) must be found that maximizes $\Im'$. The nature of $\Im'$ makes this a difficult problem for traditional optimization methods, particularly those requiring the computation of derivatives (Press et al., 1988). The combination of this fact, the high dimensionality of the parameter space, and the possibility

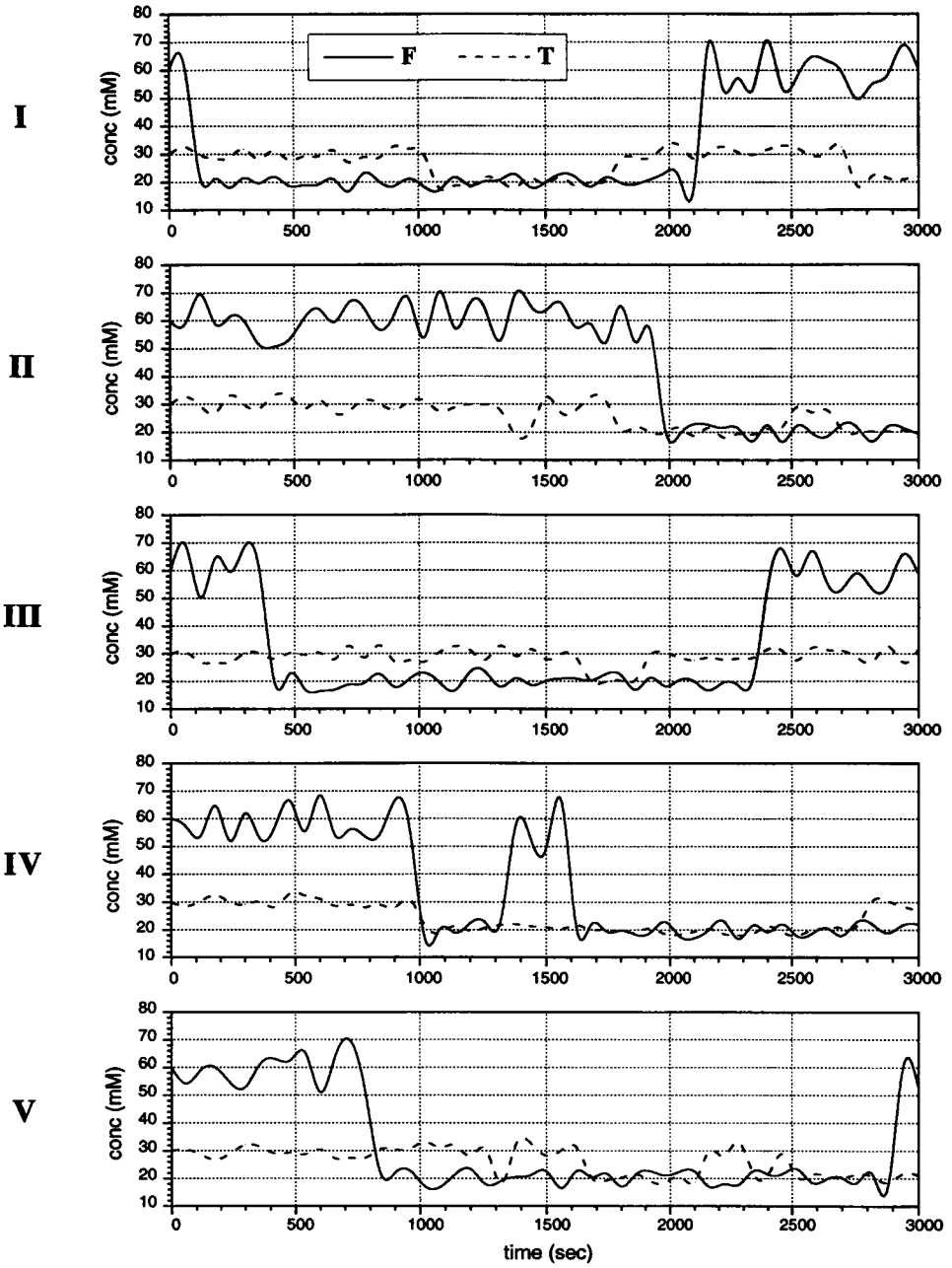


FIGURE 4 Plots of the five courses, designated by roman numerals. The concentrations of the reservoir species F and T are plotted versus time. Each varies within two regimes. For F, these center at 60 mM and 30 mM; for T, 30 mM and 20 mM.

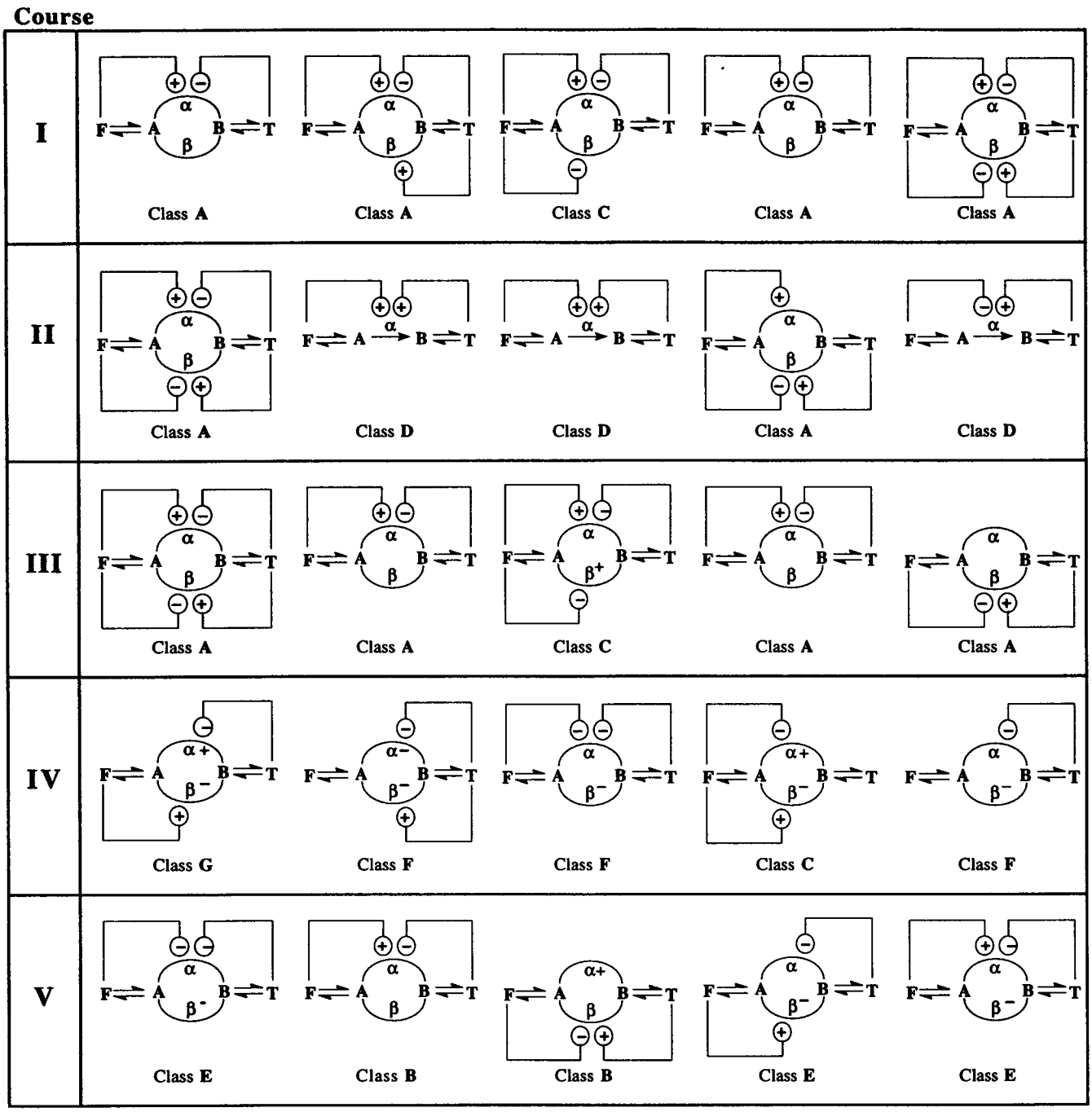
of a great many local maxima for $\mathfrak{I}'$, positively suggests using a genetic algorithm, a procedure that relies on a parallel search of the parameter space for the optimization method (Davis, 1991). The implementation of the particular GA used in this study is described in the Appendix. No attempt was made to optimize the performance of the GA itself.

The optimization study of the model system was conducted in two phases. In the first phase, the GA was run five times on each of five different courses, shown in Fig. 4. The second phase was a cross-course comparison, in which each network produced in the first phase was run on the four courses other than the one that produced it.

RESULTS

In this section we state the results obtained; these are then discussed in the following two sections. The findings of the first phase are given in Table 1. The five rows of the table correspond to the five courses shown in Fig. 4. The GA was run to select the network giving the highest value of $\mathfrak{I}'$, Eq. 7, determined as described in the previous section. This was repeated five times for each course, and the five resulting networks for each course are shown in the rows of the table in the order they were generated. In the remainder of this paper, networks are designated by a roman numeral indicating the course on which they were generated, followed

TABLE 1. Limiting network diagrams of the 25 networks selected in phase I, grouped by the course on which they were selected



The class designation refers to the classes shown in Fig. 5.

by an arabic numeral indicating their position in the row. For example, the third network generated on the second course is named II.3. The networks were generated in the form of parameter lists, which are not shown because they do not readily convey the important information. (As an illustration, one of the networks resulting from the optimization, network I.1, is specified as $K_{F,\alpha} = 1.965$, $r_{F,\alpha} = 1.291 \times 10^4$, $K_{F,\beta} = 0.9949$, $r_{F,\beta} = 1.000$, $K_{T\alpha} = 3.913 \times 10^4$, $r_{T,\alpha} = 3.982 \times 10^{-37}$, $K_{T,\beta} = 1.410 \times 10^{226}$, $r_{F,\alpha} =$ 0.9947.) Instead, Table 1 illustrates regulatory structures of the networks with limiting network diagrams. Such diagrams exaggerate the trends in responsiveness appearing in a network and thus indicate the gross regulatory structure. For example, an effector has no effect on an enzyme if the r value is 1. In the limiting diagram, if an r value is close to 1, the connection is not shown. An extremely large value of K will also cause the connection to be missing in the limiting diagram, because it indicates that the concentration of the effector must become

very large for an effect to be exerted. When the K value is extremely small, the connection is again missing in the limiting diagram, but the enzyme is marked as modified. Because effector concentrations never come very close to zero in our model, an extremely small K value means that the effector's influence on the enzyme is saturated for all occurring effector concentrations. The result is the same as if the intrinsic V_{max} of the enzyme were itself altered. A plus indicates that the new enzyme has a larger maximum activity than before, and a minus indicates the opposite. The table shows a wide variety of network structures, even on the same course. The source of this variety and the similarities of the networks are discussed in the next section.

The temporal behavior of a selected network (I.5) on each of the five courses is shown in Fig. 5, and should be compared to the courses themselves in Fig. 4. Although networks may not always be at steady state while running on a course, network behavior may be summarized approximately in a plot of steady-state A and B concentration surfaces versus F and T. Although each network produced in the optimization phase gave a different steady-state surface, the surfaces tended to fall into seven classes of similar shape. Fig. 6 shows representative plots from each of the observed classes. The caption for Fig. 6 also gives the number of networks from the 25 total that fall into each class.

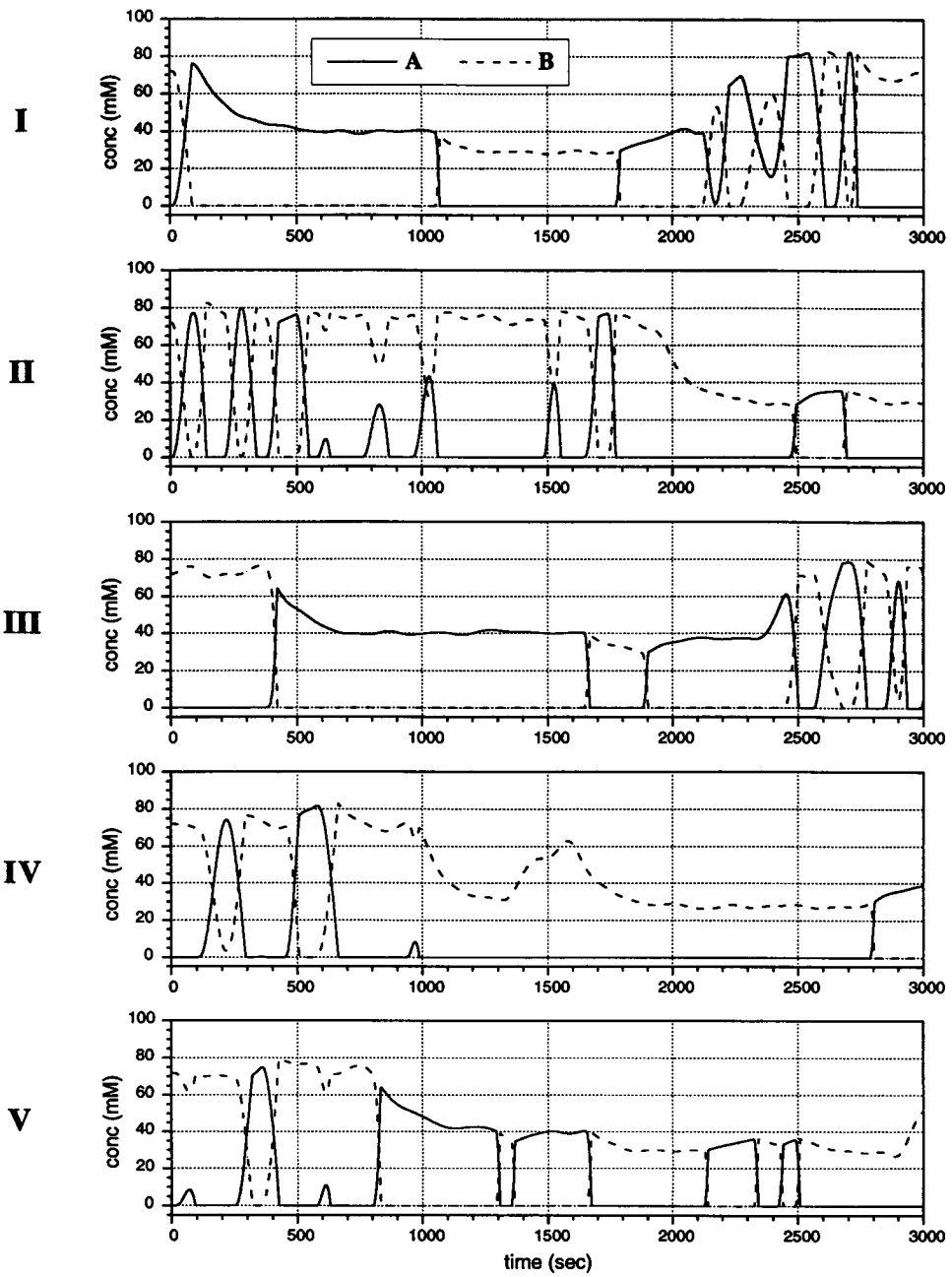


FIGURE 5 Behavior of network I.5 on each of the five courses, designated by roman numerals. The concentrations of cycle intermediates A and B are plotted versus time.

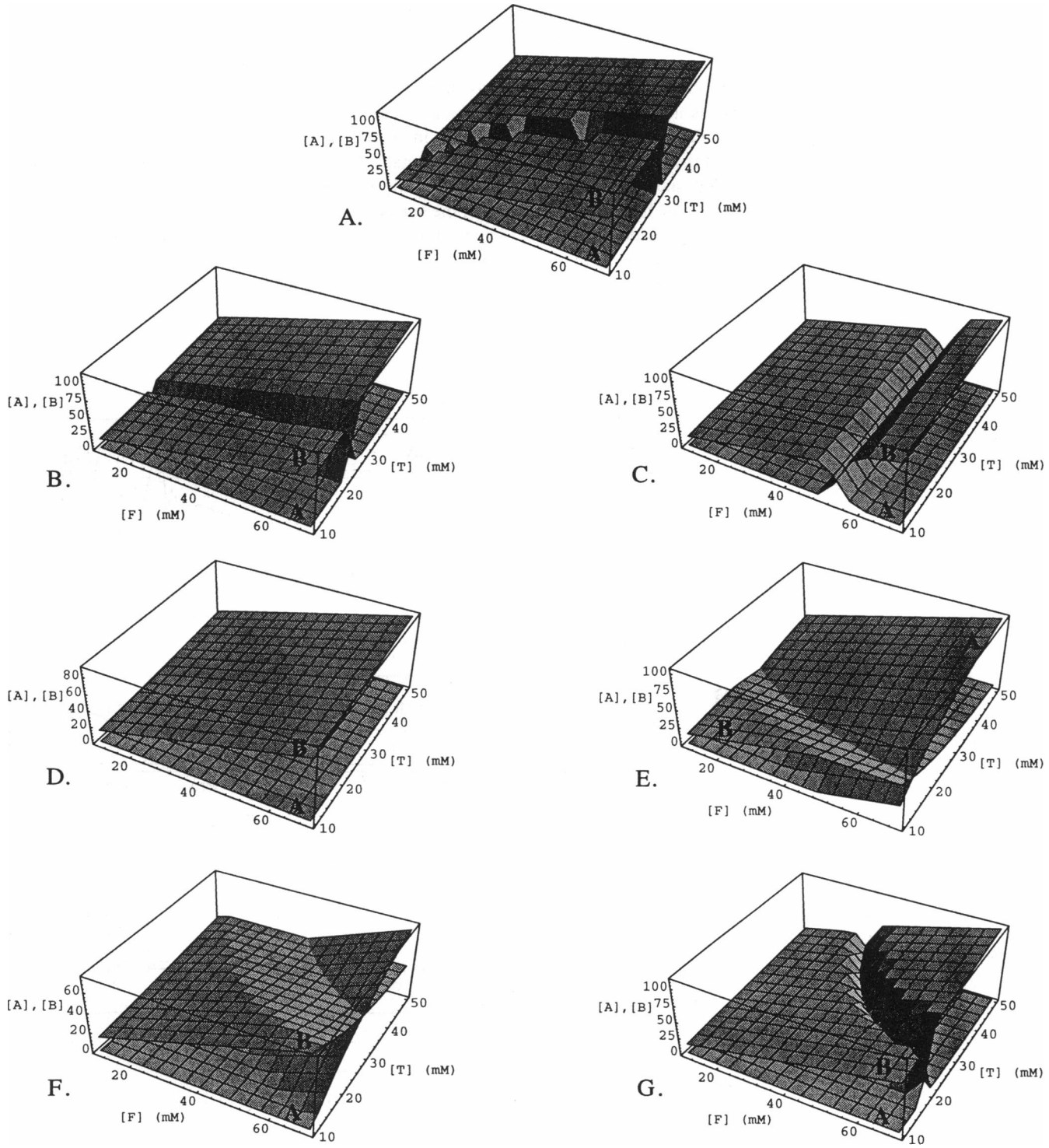
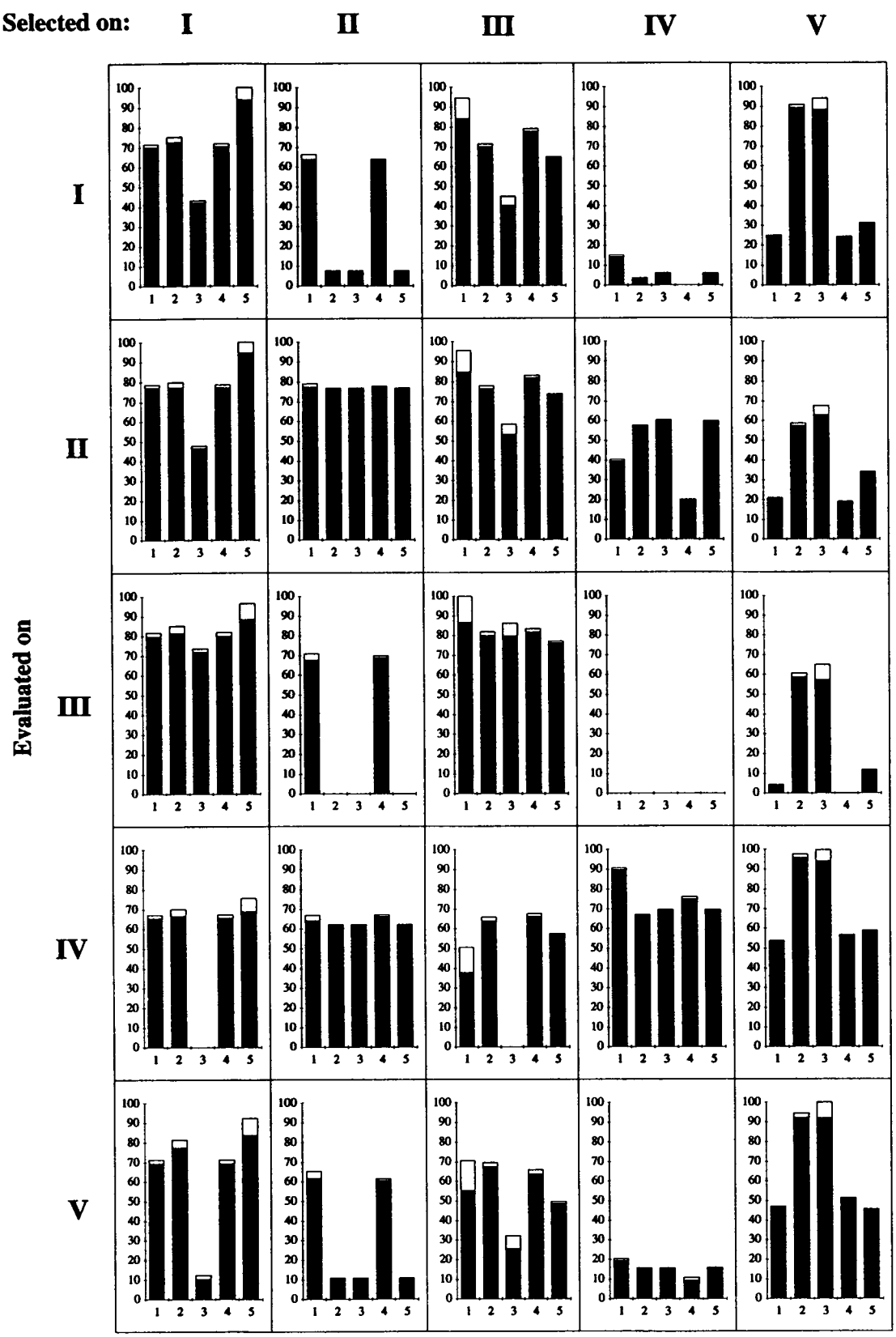


FIGURE 6 Steady-state surfaces representative of each of the seven observed classes. (*A*) Network I.5, 10 members in the class; (*B*) network V.2, 2 members; (*C*) network IV.4, 3 members; (*D*) network II.3, 3 members; (*E*) network V.5, 3 members; (*F*) network IV.5, 3 members; (*G*) network IV.1, 1 member.

The findings of the cross-course comparison are shown in Table 2. Here, the performance of each network has been normalized to that of the best performance observed for the course. The overall height of each bar (black and white portions) shows the normalized raw fitness, $\Im$. The height of the white portion of the bar shows the cost incurred, mC (Eq. 7). The black portion thus shows the normalized cost-adjusted score, $\Im'$. Some networks give negative fitness values, which have simply been cut off the chart. The cost scores appear rather small compared to the fitness, but this

TABLE 2. Cross-course comparison



The rows of the table compare all 25 networks on the course labeled by roman numerals on the left. The columns compare the five networks generated on the course indicated by the roman numeral on top across all five courses. The diagonal entries show the performance of the networks on their "home" courses. The scores, shown as percentages, are normalized to the single best raw fitness for the given course. Bars: ■, adjusted fitness; □, cost. Negative fitness scores have been cut off.

is mainly a result of the relative weighting used to adjust the raw fitness.

DISCUSSION

One of the most important findings of the optimization phase of the study is the trend in the regulatory pattern as seen in Table 1. In all but a few cases, whenever F has an effect on α, it is an activation. Similarly, in most cases where T affects α, the effect is inhibitory. For β, the situation is reversed (i.e., F inhibits β and T potentiates it), but the trend is again general. The exceptions are networks II.2, II.3, II.5, IV.1, IV.3, IV.4, V.1, and V.4, but in none of these networks do *all* of the connections contradict the trend. Also, as seen from Table 2, most of these exceptions perform poorly relative to their siblings (networks generated on the same course), but some nonexceptions do also. The finding of the trend is important because it is the regulatory pattern expected from a simple, qualitative analysis of homeostasis, which is the task that the system is roughly expected to represent. In a physiological system, the concentration of a reservoir is affected by the fluxes into and out of the reservoir. A homeostatic mechanism should then seek to control these fluxes so that the concentration does not change, at least not significantly. For example, a mechanism to buffer the concentration of F should inhibit influx into F as the concentration of F rises (or potentiate efflux, or both), but do the opposite as the concentration of F drops. If the signal for these changes were F itself, and if the reaction structure of the pathway were the one we have modeled, one would expect F to inhibit β, the enzyme that produces its precursor, and to potentiate α, which consumes its direct product. A similar action for T is expected. Although our model stops short of describing a homeostatic system, in that the reservoir concentrations are not affected by the behavior of the futile cycle, we may imagine reservoir concentrations being changed by the simultaneous action of a large number of identically regulated cycles. In that case, the regulatory pattern we find matches the one expected for homeostasis.

Another important finding is seen in Fig. 6: the existence of a switching region in four of the seven representative steady-state surfaces (and so the 16 of the 25 networks that fall into those four classes), where the concentrations of A and B rapidly exchange places. In classes E and F, the concentration surfaces cross but do so gradually, not in a switch. In class D, the switching region is absent entirely. The temporal manifestation of a switching region is seen clearly in the behavior of network I.5 shown in Fig. 5. For example, in the upper plot of the figure, the network switches shortly after the course begins. A faster switch appears at around 1080 s and subsequent ones are seen later. In plot II of the figure, "aborted" switches are seen near 600, 800, 1000, and 1500 s. The switching appears to occur when certain thresholds in reservoir concentrations are crossed. These thresholds are not at the same concentration levels as the thresholds in the need functions but are independently determined by the network parameters, and in the particular case of network I.5, are fairly close to the need thresholds. Switches seem to be "aborted" when a threshold is crossed, then recrossed a short time later. Switching behavior is due to a phenomenon akin to zeroth-order ultrasensitivity, described by Goldbeter and Koshland (1981). It occurs because one or the other of the enzymes is almost always saturated at the concentrations of substrate used in the model and thus acts in the zeroth-order regime, that is, the rate of conversion of the substrate is constant. If both enzymes operate at zeroth order, the net flux through the cycle is likewise constant, and its direction is determined by the greater V_{max}. The flux continues until the concentration of the substrate of the faster enzyme drops low enough for the enzyme no longer to be saturated. For the kinetic parameters used in the model, this occurs at a very low (micromolar) concentration of substrate. The switching region appears where the countervailing effects on the V_{max}'s of the enzymes shift from favoring one enzyme to favoring the other. In the case of class D (networks II.2, II.3, and II.5), this cannot occur, as these networks have lost enzyme β altogether.

Networks of class D are easily identified from their limiting network diagrams because of the absence of an enzyme. But such gross features do not mark members of the other classes. In fact, for the other classes, the limiting network diagram is not sufficient to determine the class and thus the response behavior. This means that the specific regulatory parameters must be known in addition to the gross regulatory structure. It also means that a class of behavior may be realized by more than one regulatory structure. Furthermore, members of a given class may show rather different performances. For example, class C is represented by networks I.3, III.3, and IV.4. Table 2 shows how different their performances are on all the courses. Again, detailed knowledge of the regulation is required even for qualitative relative predictions.

A final observation on Table 1 is that certain classes of behavior are more likely to appear on certain courses than others. For instance, course I gave more members of class A than any other; in fact, only class C also appeared on that course. On the other hand, course V gave no members of classes A or C and gave more members of class E than B. This is an important point and will be discussed further in the next section.

Cross-course comparisons

An important finding that comes from the cross-course comparisons is that no single network does best on all five courses. The absence of a global winner may be due to an insufficient sample size but is nevertheless important with a view to contexts where sample size is perforce limited. If the "environments" represented by the five courses may be allowed to coexist in the same world, the absence of a global

winner implies that no single "species" of network will come to dominate the world and thus a diversity will be maintained. On the other hand, Table 2 shows that in some cases the performance of networks raised on a particular course is surpassed on that "home" course by networks raised on a different one. In other words, the GA procedure used in this study does not always find the global optimum for a given course. This is clearly seen in, for example, the second row of Table 2, where both networks I.5 and III.1 outperform the best "native" network II.1. The other example is row IV, where two networks from course V outperform the native winner. Thus, if these environments were to be brought into contact, a "takeover" by non-native networks might be expected.

In discussing the question of how this comes about, it is useful to consider why a given behavior class or network structure, which we may call a strategy, might fail to be selected in a series of GA runs on a given course. An obvious reason is that the strategy is truly suboptimal and is rejected on that basis by the procedure. But a second reason is that it is just not likely to be encountered in the search, or is relatively less likely with respect to the strategies that do appear. A good example is course II, where the strategy represented by class D (no switching, missing β enzyme) appears in three of five cases. There is a vast region of K-r space where K and r are sufficiently small to eliminate an enzyme, in effect. The region of K-r space that fulfills the criteria for a viable regulatory connection is comparatively much smaller. Thus, even though the course II networks of class A consistently outperform the class D networks, they are less likely to appear. The even better networks of class A that do appear on courses I and III seem even less likely to appear on course II. We may regard the different courses as giving rise to different "fitness landscapes" (Wright, 1932) that are searched in the GA procedure. We envision that on the fitness landscape arising on course II, there is a large fitness plateau in the region of class D realizations, with relatively narrow but taller peaks in the regions of well-performing class A realizations. The peaks yield greater fitness but are more likely to be missed by the search. These effects contribute to the "tolerance" of a course for different strategies that may be suboptimal. Such a tolerance may explain why certain courses appear to produce a wider variety of strategies in the networks selected on them.

In the context of biochemical regulation, the interpretation of the lack of global winner depends strongly on how well our courses represent actual biochemical environments. For example, courses I and IV show very different patterns of performance in Table 2. We presume that the courses themselves differ in certain fundamental features that have an impact on the GA selection procedure, although we have not identified these features. If two biochemical environments that were consistently characterized by the same features as courses I and IV were to persist over an evolutionary time scale, then assuming some verisimilitude for our model, we might expect to find differently regulated metabolisms evolved in the two environments, one of them being capable of functioning "better" than the other in either environment. On the other hand, if over evolutionary time scales a biochemical environment exhibits features of all the courses we have used, and perhaps others, then the interpretation becomes simpler: we then need only examine networks that perform well on all the courses.

Table 2 shows that networks I.5, III.1, and V.3 exhibit good generalized performance. In fact, four of the five networks from both course I and course III show decent generalized performance, as do two of the five from course V. In contrast, course II produced three networks that do very poorly on all but courses II and IV, and course IV shows a similar pattern. These two courses, especially course IV, are thus more likely to produce specialists. But as mentioned above, specialization on these courses is not enough to produce an overall winner. Generalist I.5 wins on course II, and generalist V.3 wins on course IV. The probable explanation is that the strategies produced on these courses, while not optimal, are sufficient, whereas the optimal strategies are unlikely to be found.

The similarity of the performance patterns of the course IV networks and the class D course II networks is suggestive of a similarity in the courses themselves. We conjecture that both courses tolerate a nonswitching strategy (as embodied by classes D, E, and F), even though a switching strategy might do better.

The success of the generalists may at least in part be understood by examining the strategies represented by their behavior classes. Networks I.5 and III.1 are members of class A. The important feature of this class is that (the projection of) the switching region curves when viewed in the F-T plane (see Fig. 6). This curving indicates that the network's response to one effector is modulated by the other effector: taking I.5 as an example, for T concentrations close to 30 mM, the network will switch when the F concentration drops, whereas when T concentrations are in the low regime around 20 mM, the network will not switch on a drop in F. This event of not switching maintains the concentration of B at a high level and thus keeps the net flux directed toward the T reservoir, as required by the need function under such conditions. This is obviously a good strategy because the function being maximized by the network also depends on both effectors. (Other members of class A do not do quite as well because the curve of their switching regions is not as pronounced and the response favors one effector more than the other.) The other generalist, network V.3, is a member of class B, the members of which switch only in response to T. This is also a good strategy because of the precedence given to T in resolving need conflicts. T is thus more likely to be the more important effector. The strategy of responding to both effectors wins three out of five courses, whereas responding only to T wins two out of five. Thus, neither strategy is unambiguously better than the other. Overall, it appears that switching strategies perform better than nonswitching strategies. Three courses, II, IV, and V, produced examples of both switching and nonswitching networks. On courses II and V, the switching networks consistently outperform the nonswitching ones. Course IV gives somewhat

ambiguous results; however, the one switching network produced there is of class F, an anomalous class in that the curve of its switching region is opposite to expectation.

Another interesting feature of the networks using good generalized strategies is that they all expend significant cost in their operation. We believe a high cost to be a consequence of the specific type of regulation employed in the model. With the hyperbolic form we have used for the regulatory factors in Eq. 2, sensitivity (the change in V_{max} due to a change in effector concentration) increases with r. This can be seen by taking derivatives:

$$\frac{\partial R}{\partial \epsilon} = \frac{K(r-1)}{(K+\epsilon)^2} \tag{8}$$

$$\frac{\partial^2 R}{\partial \epsilon \partial r} = \frac{K}{(K+\epsilon)^2}. \tag{9}$$

The right-hand side in Eq. 9 is always positive. Thus, to gain a large difference in activity at one end of an effector's concentration range versus the other, larger values of r are favored.

A high cost is not, however, sufficient for good generalized performance, as network III.3 clearly shows. High cost may be indicative of high sensitivity, but the sensitivity may be for the wrong effector. Network III.3 responds almost exclusively to F (with high sensitivity, indicated by a narrow switching region), but as discussed above, this does not seem to be a good generalist strategy.

CONCLUSION

We have presented a method employing GA optimization for obtaining patterns of regulation that allow a metabolic network to carry out a specified task. For a simple task (flux direction) and a simple network structure (a single futile cycle), our approach develops patterns of regulation that confirm the intuitive expectations of negative feedback and reciprocal effects. The patterns arise purely as a result of specifying the task. That there was no built-in bias toward such results is confirmed by the selection, in certain cases, of non- or counter-intuitive regulatory patterns that proved to be suboptimal in comparison with other patterns. We found that certain "environments" (time courses of variation of external constraints, under which the GA optimization runs were carried out) are more likely to lead to the selection of suboptimal individuals. Indeed, in comparing individuals selected in different environments, we observed the appearance of "generalists" and "specialists," that is, individuals performing relatively well in all environments and individuals performing well only in a certain environment and poorly in others. We did not attempt to improve the degree of optimization provided by the specific GA that was used, and we believe this contributes to the observed diversity among selected individuals. More generally, while the GA is not intended as a model for the process of biological evolution, the two processes share certain limitations on their effectiveness at finding global optima. Two such limitations are a finite (possibly even small) population size and a finite number of generations. The efficacy of biological evolution as optimization is an open problem, so a procedure that is not extremely effective at optimization but shares certain features with biological evolution may provide better insight into the results of biological evolution than an excellent optimization procedure that provides no analogy whatsoever. Thus, we believe it is no coincidence that effects reminiscent of environmental influences on evolution are observed in our procedure.

We represent this work to be a new and useful approach for the study of network-level metabolic regulation. A method for deriving metabolic network structures that are potentially optimal with respect to a functional criterion is important for several reasons. First, it allows a guess of a known network's function based on its structure to be assessed for plausibility, by comparing the known structure to those that are selected by the method to carry out the putative function. Moreover, the formalization of intuitive notions of function that this method requires may lead to refinements and quantification of such notions. Also, if the function of a known structure is known with confidence, the method may be helpful in determining how close to optimal the structure is and thus may point to unconsidered criteria. Furthermore, it provides a tool for exploring the biological realizability of given functions, that is, questions such as how wide a variety of structures is capable of carrying out a given function, and what the effects of specific structural or kinetic constraints might be. The method provides an aid for the investigation of the limits on, and possibilities of, network structures in light of required metabolic functions.

APPENDIX

Networks are encoded as strings ("chromosomes") of eight real numbers ("genes") that are Briggsian logarithms of the K and r regulatory parameters. The genetic operators are a two-point cross and a mutation operator. Population size is a constant 100 individuals. The top 5 and the worst single individual in each generation are reproduced exactly in the next generation. The other 94 slots are filled by recombinant offspring. The parents for each pair of offspring are chosen by roulette-wheel selection, with probability proportional to the given parent's share of the combined fitness for the generation (unscaled). The two-point cross is then carried out on the parents by randomly selecting two points along the chromosome. One offspring receives a chromosome that consists of the first parent's genes up to the first cross-point, then the second parent's genes up to the second cross-point, then the first parent's again until the end of the chromosome. The other offspring receives the complementary set of genes. Each gene of an offspring has a 20% mutation rate, where the mutation operator changes the gene by a random percentage, drawn uniformly from -100 to 100%.

A run begins with a population for which the log$K$ has a Gaussian distribution with a mean of -3.0 and a variance of 3.0, and the log$r$ had a Gaussian distribution with a mean of 0.0 and a variance of 4.0. During the run, some individuals fail to be integrated properly (usually owing to the very high stiffness of the resulting kinetic equations) and are replaced by individuals generated randomly according to the above distributions. Most integrator failures occur within the first few generations. The highest fitness for each generation is recorded, and when no improvement in the highest fitness is seen for 10 generations, the run is ended.

We wish to thank Prof. Lubert Stryer and Dr. Adam Arkin for many helpful discussions. This work was supported, in part, by the National Science Foundation.

REFERENCES

Bray, D., and S. Lay. 1994. Computer simulated evolution of a network of cell-signaling molecules. *Biophys. J.* 66:972–977.

Davis, L., editor. 1991. Handbook of Genetic Algorithms. Van Nostrand Reinhold, New York. 385 pp.

Fersht, A. 1984. Enzyme Structure and Mechanism, 2nd ed. W. H. Freeman and Company, New York. 475 pp.

Goldbeter, A., and D. E. Koshland, Jr. 1981. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci. USA.* 78:6840–6844.

Heinrich, R., and H.-G. Holzhütter. 1985. Efficiency and design of simple metabolic systems. *Biomed. Biochim. Acta.* 44:959–969.

Heinrich, R., H.-G. Holzhütter, and S. Schuster. 1987. A theoretical approach to the evolution and structural design of enzymatic networks; linear enzymatic chains, branched pathways and glycolysis of erythrocytes. *Bull. Math. Biol.* 49:539–595.

Meléndez-Hevia, E., and A. Isidoro. 1985. The game of the pentose phosphate cycle. *J. Theor. Biol.* 117:251–263.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1988. Numerical Recipes in C, The Art of Scientific Computing. Cambridge University Press, Cambridge. 735 pp.

Savageau, M. A. 1974. Optimal design of feedback control by inhibition: steady state considerations. *J. Mol. Evol.* 4:139–156.

Savageau, M. A. 1975. Optimal design of feedback control by inhibition: dynamic considerations. *J. Mol. Evol.* 5:199–222.

Savageau, M. A. 1979. Feedforward inhibition in biosynthetic pathways: inhibition of the aminoacyl-tRNA synthetase by the penultimate product. *J. Theor. Biol.* 77:385–404.

Savageau, M. A., and G. Jacknow. 1979. Feedforward inhibition in biosynthetic pathways: inhibition of the aminoacyl-tRNA synthetase by intermediates of the pathway. *J. Theor. Biol.* 77:405–425.

Schuster, S., and R. Heinrich. 1987. Time hierarchy in enzymatic reaction chains resulting from optimality principles. *J. Theor. Biol.* 129:189–209.

Wright, S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *In* Proceedings of the Sixth International Congress on Genetics, Vol. 1. D. F. Jones, editor. Brooklyn Botanic Garden, Ithaca, NY. 356–366.