Evolution of complexity in signaling pathways

Orkun S. Soyer^{†‡} and Sebastian Bonhoeffer

Theoretical Biology Group, Institute for Integrative Biology, Swiss Federal Institute of Technology (ETH), Universitätsstrasse 16, ETH Zentrum, CHN K12.2, CH-8092 Zürich, Switzerland

Edited by Steven Strogatz, Cornell University, Ithaca, NY, and accepted by the Editorial Board September 14, 2006 (received for review May 30, 2006)

It is not clear how biological pathways evolve to mediate a certain physiological response and why they show a level of complexity that is generally above the minimum required to achieve such a response. One possibility is that pathway complexity increases due to the nature of evolutionary mechanisms. Here, we analyze this possibility by using mathematical models of biological pathways and evolutionary simulations. Starting with a population of small pathways of three proteins, we let the population evolve with mutations that affect pathway structure through duplication or deletion of existing proteins, deletion or creation of interactions among them, or addition of new proteins. Our simulations show that such mutational events, coupled with a selective pressure, leads to growth of pathways. These results indicate that pathways could be driven toward complexity via simple evolutionary mechanisms and that complexity can arise without any specific selective pressure for it. Furthermore, we find that the level of complexity that pathways evolve toward depends on the selection criteria. In general, we find that final pathway size tends to be lower when pathways evolve under stringent selection criteria. This leads to the counterintuitive conclusion that simple response requirements on a pathway would facilitate its evolution toward higher complexity.

robustness | evolutionary simulations | mathematical models | protein pathways | signal transduction

A central question in evolutionary biology is the evolution of complex features. Research in this field has associated complexity at the organism level with complexity at the sequence level (1, 2), with transcriptional control (3, 4), or with cellular differentiation (5). Although these studies provided important insights into how complexity at organism level could have arisen, we still lack an explicit analysis of complexity in a specific system at molecular level. Here, we attempt such an analysis for biological signaling pathways. These pathways, i.e., systems of proteins that act in an orchestrated fashion, mediate the response of a cell toward internal and external signals. It is usually observed that pathways show high complexity, which manifests itself in various forms including number of components in a given pathway, interaction among different pathways, and compartmentalization (6). Although a sequential evolution as proposed by Darwin (7) could explain the existence of such complex pathways, it is not clear why pathways should be driven toward a level of complexity that is higher than required by their function (i.e., producing a specific response). Both experimental and theoretical studies indicate that certain biological responses such as oscillations and switches or phenomena such as pattern formation and chemotaxis could be achieved via pathways that are simpler than found in nature (5, 8–12). One possible explanation for pathway complexity (or biological complexity in general) is that selection for increased fitness necessitates an increase in complexity. Although there is support for such an explanation from studies in digital organisms (13), it is not clear why an increase in complexity should always be accompanied with an increase in fitness and vice versa (14, 15).

There is a much simpler explanation for the emergence of complexity in biological systems that has been first formulated by Saunders and Ho (15). In their words: “a system which optimizes its structure with respect to some criterion (which one may call fitness) will tend to permit the addition of components

more readily than their removal.” This hypothesis does not assume that complexity should be related to increases in fitness or any other property that is under selection. However, complexity arising neutrally because of the aforementioned tendency in time could result in functional advantages and a fitness benefit.

Here, we formulate a similar hypothesis to explain the emergence of complexity in biological pathways: Pathways have an intrinsic tendency to become more complex that results from imbalanced effects of mutational events on pathway response. In other words, mutations resulting in the addition of new interactions or proteins to a given pathway have less deleterious effects on its response than mutations that result in loss of interactions or proteins. If such an imbalance exists, it would drive pathways to become larger and more complex in time. To test this hypothesis, we analyze evolution of pathways by using mathematical models and computer simulations. Starting with a population of small pathways, we let pathways replicate with various mutations affecting their structure. We find that such mutations, coupled with a selective pressure for function, lead to growth of pathways under various model assumptions. These results indicate that pathways have an intrinsic tendency to grow larger than required for proper functioning, arising simply from the nature of mutational events affecting pathway structure.

Results and Discussion

To test the possibility that pathways have an intrinsic tendency to become more complex because of mutational mechanisms affecting their structure, we perform evolutionary simulations and monitor pathway size as a proxy for complexity. These simulations (see *Methods*) start with a population of random pathways composed of only three proteins. Fig. 1 illustrates a sample pathway used in such simulations, its mathematical description, and response to a signal. At each successive generation of the simulation, pathways replicate based on their fitness value and are exposed to mutations with a certain probability during this replication.

To decouple fitness and complexity, we use a step-like fitness function that is adjusted only by the cost of having multiple proteins. In other words, pathways are either viable [fitness = $1 - (\text{number of proteins}) \times (\text{cost per protein})$] or not (fitness = 0) and the only determinants of fitness are pathway size and the ability to fulfill a chosen selection criteria (see *Methods*). Furthermore, in the presented simulations, we assume that mutations leading to an increase in pathway complexity (i.e., mutations adding a protein or an interaction) and mutations leading to a decrease (i.e., mutations deleting a protein or an interaction) have equal chances to occur, ensuring that we do not artificially

Author contributions: O.S.S. and S.B. designed research; O.S.S. performed research; O.S.S. analyzed data; and O.S.S. and S.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission. S.S. is a guest editor invited by the Editorial Board.

Freely available online through the PNAS open access option.

[†]To whom correspondence should be addressed. E-mail: orkun.soyer@msr-unitn.unitn.it.

[‡]Present address: The Microsoft Research-University of Trento, Centre for Computational and Systems Biology, Piazza Mancini 17, 38050 Povo (TN), Italy.

© 2006 by The National Academy of Sciences of the USA

Fig. 1. Cartoon and mathematical representation of a sample three-protein pathway and its response (solid line) to a specific signal (dashed line). The mathematical pathway representation gives coefficients for interactions among proteins in a table format, listing actions of other proteins on a given protein at each row. The absolute values of negative and positive numbers stand for activating and deactivating interactions, respectively. The shown signal is used throughout all evolutionary simulations to assess pathway fitness.

introduce complexity (see *Methods*). Despite such an assumption, this null-model biases toward decreasing pathway size because interaction-deleting mutations can occasionally result in loss of proteins, whereas interaction-adding mutations cannot have a balancing effect. Hence, we find that in absence of any selection (i.e., all pathways are viable) pathways shrink to a size of zero (see Fig. 2).

This situation reverses in presence of selection. Fig. 2 shows that when a population of randomly created three-protein pathways are under selection for existence (i.e., having at least two proteins) or for giving any type of response to a signal (see

Fig. 2. The average pathway size for populations evolving under different selection schemes: black (no selection), red (selection for any type of response to a given signal), blue (selection for existence), and yellow (selection for derivative-type of response to a given signal). The latter simulation starts with a homogenous population of three-protein pathways (shown in Fig. 1) that are capable of a derivative-like response, whereas all other simulations start with a random population. See *Methods* section for details.

Fig. 3. The average pathway size for populations evolving under different selection schemes: black (selection for transient base), yellow (selection for transient magnitude), blue (selection for transient base and magnitude), and red (selection for switch). All simulations start with a random population of three-protein pathways. See *Methods* section for simulation details.

Methods), the average pathway size increases until it settles to stochastic fluctuations around a certain size. Given the nature of the used fitness function and the fact that even small pathways can achieve high fitness under these selection criteria, it is unlikely that the observed increase in pathway size is due to increasing fitness. To make this point more clear, we run a simulation starting with a homogenous population of three-protein pathways that are capable of producing a response that loosely tracks the derivative of the signal (see Fig. 1). Hence, under selection for this type of response, all members of the population have the same fitness value and there is no possible fitness gain from increased complexity but even a small cost associated with having more proteins in the pathway. Still, we find that the average pathway size increases in this population as evolution progresses (see Fig. 2). It is important to note that the observed increase in average pathway size is not due to few pathways dominating the population but rather due to an increase in the frequency of larger pathways (see Fig. 5, which is published as supporting information on the PNAS web site).

To see whether these results are extendable to various biologically relevant responses, we have run additional simulations with the following selection schemes: (i) “transient/base,” a transient response with equal pre- and postsignal levels of active effector protein concentration; (ii) “transient/magnitude,” a transient response with a magnitude that is larger than the presignal level of active effector protein concentration; (iii) “transient/base and magnitude,” a transient response that fulfills both criteria (i) and (ii); and (iv) “switch,” a “switch-like” response where the postsignal level of the active effector protein concentration is higher than the presignal level by at least 0.01. Starting with a random population of three-protein pathways, we find that under such selection criteria pathways grow and reach a certain size level that is different for different selection schemes but always above the minimum possible (see Fig. 3). In Fig. 6, which is published as supporting information on the PNAS web site, we show sample pathways from the last generation of some of these evolutionary simulations.

These results clearly indicate that the effects of the modeled mutational events on pathway response are distributed in an asymmetric manner. Deletion of interactions and proteins render pathways nonfunctional (e.g., not able to produce the required response) more often than additions do, pushing pathways to become larger and more complex. However, the fact that populations reach “equilibrium” with respect to size indicates that the functional effects of these two types of mutations depend on the

Conclusion

The intuitive view for the emergence of complexity is that it is due to increasing fitness, a view that has found support from studies on digital organisms (13). However, this study suggests that an imbalance in the effects of size-decreasing and -increasing mutations on function could lead to increase in complexity, supporting a mechanistic or neutral explanation (15). Hence, as long as there is selection acting on a system, even neutral processes that do not cause any immediate fitness benefit would force the system toward higher complexity.

We find that the amount of the aforementioned imbalance is related to complexity itself. It is highest for the simplest possible pathways that could achieve the task for which there is selection operating and decreases with increasing complexity. Hence, as evolution progresses, pathways achieve a certain size or level of complexity that is always well above the minimum required for their function. Further, we find that although simulations with different selection criteria start with pathways of the same size, for pathways that are under selection for functions that are attainable by many different topologies (i.e., that are simpler to achieve), the final level of complexity achieved tends to be higher.

These findings have implications for our understanding of the evolution of new features. For example, in simulating the evolution of a random population of three-protein pathways under selection for responding to a given signal, we reach a structurally diverse population with an average pathway size of ≈ 14 proteins. Given the exponential relation between pathway size and available pathway topologies, it is plausible to assume that such growth would facilitate the emergence of pathways with various response dynamics that would not have been possible with three proteins only. If such pathways achieve new functions or high-fitness solutions under the current selection criterion, they would be strongly selected. Hence, even though complexity can emerge neutrally, once it results in evolutionary favorable pathways, it may be maintained subsequently by selection.

Finally, we note that the presented results have an interesting connection to robustness. In general, robustness in biological pathways is associated with the ability of such pathways to withstand knock-out mutations or disturbances in interaction parameters. Based on this description, many natural pathways have been found to have high robustness (19–24). This observation becomes obvious in light of the presented results. Our simulations suggest that in general, simplest-solution pathways are nonrobust toward deletions. Thus, populations predominantly consisting of such pathways drift toward larger, more complex solutions, and, in this sense, selection for function leads to emergence of robustness. A similar scenario is given for the evolution of robustness in general (25); through evolutionary mechanisms populations move toward wide plateaus on an imaginary fitness landscape, where they would be protected against small perturbations (i.e., they have high robustness). The idea of an imbalance between the functional effects of various mutations having different effects could be seen as the driving force behind such move, leading both to high robustness and high complexity. This statement overlaps with the theories that see complexity and robustness tying closely together (26).

The presented discussion approximates the behavior of biological pathways with a simple mathematical model and makes several assumptions regarding the nature and rate of mutations affecting their structure. Although the true nature of these pathways clearly is more complicated, we believe that the qualitative conclusions of this theoretical study would remain under biologically relevant parameter values (see *Methods*). These conclusions could also be extendable to any system that is under selection and is subject to mutations affecting its structure.

Under such conditions, the idea of complexity arising neutrally could be more general.

Methods

Pathway Model. We use a generic mathematical model to capture the basic properties of biological pathways. The model has been explained in detail previously (11, 12). In brief, it considers a pathway as a collection of interacting proteins, each of which can exist either in an active (P_i^*) or inactive (P_i) state. Proteins in the active state can interact with other proteins and influence their activation. There are many different biochemical mechanisms that make such influence possible including phosphorylation, methylation, and physical interaction. Here, we do not distinguish among these different interaction types and model the influence of an active protein on others simply through an interaction coefficient. We assume that the energy required for these activation and deactivation reactions are provided from outside sources such as high-energy molecules. Each active protein is free to interact with any other protein in the pathway, resulting in the following chemical equilibrium

where $[P_j^*]$ represents the concentration of active form of protein j and k_{ij} and l_{ij} represent the strength of the interaction between protein i and j . We assume that proteins do not interact with themselves (i.e., $k_{ii} = l_{ii} = 0$) and their influence on other proteins can be only activating or deactivating (i.e., $k_{ij}l_{ji} = 0$), resulting in $n^2 - n$ parameters for a pathway of n proteins. We do not consider processes such as autophosphorylation or intrinsic phosphatase activity in this model.

To allow the pathway to respond to a signal, we arbitrarily define two proteins in the system as the receptor and effector. We couple the equilibrium state of the receptor to the ligand concentration (i.e., signal) and take the concentration of the effector active form as the response produced by the pathway. In a natural pathway, this response can have various forms, depending on the effector, and can range from transcriptional control to enzymatic activity. To summarize, this model defines a biological pathway by the number of proteins and a set of interaction coefficients. Pathway response to a changing ligand concentration thus can be obtained by solving the set of differential equations resulting from the collection of reactions as shown in Eq. 1:

$$\frac{d[P_i]}{dt} = \left[[P_i^*] \cdot \sum_j l_{ij}[P_j^*] \right] - \left[[P_i] \cdot \left(\delta_{i1} \cdot [L] + \sum_j k_{ij}[P_j^*] \right) \right], \quad [2]$$

where $\delta_{i1} = 1$ for $i = 1$ and $\delta_{i1} = 0$ for $i \neq 1$. Note, that the total concentration of each protein $[P_i^{\text{tot}}]$ is constant and set to 1 (i.e., $[P_i^*] = 1 - [P_i]$). Also, we set the maximum value that an interaction coefficient can attain to 1 for computational ease.

To assess the response of a pathway, we initiate the model with equal amounts of active and inactive proteins (i.e., $[P_i^*] = [P_i] = 0.5$). After equilibrating the system by integrating the set of differential equations for 1,000 time steps or until steady state is reached, we introduce a signal in the form of a changing ligand concentration (a Gaussian shape with a variance of 6, as shown in Fig. 1). The integration then is continued in presence of this signal for 1,000 time steps or until the system reaches the steady state. At the end of the integration, we validate that we reached a steady state with an eigenvalue analysis. The pathway response then is deduced from the active effector concentration recorded during the second integration period, in the presence of the

18. May RM (1972) *Nature* 238:413–414.
19. Alon U, Surette MG, Barkai N, Leibler S (1999) *Nature* 397:168–171.
20. Cheng P, Yang Y, Liu Y (2001) *Proc Natl Acad Sci USA* 98:7408–7413.
21. Edwards JS, Palsson BO (2000) *Biotechnol Prog* 16:927–939.
22. Gonze D, Halloy J, Goldbeter A (2002) *Proc Natl Acad Sci USA* 99:673–678.
23. von Dassow G, Meir E, Munro EM, Odell GM (2000) *Nature* 406:188–192.
24. Wagner A (2000) *Nat Genet* 24:355–361.
25. de Visser JA, Hermisson J, Wagner GP, Ancel Meyers L, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, *et al.* (2003) *Evolution Int J Org Evolution* 57:1959–1972.
26. Carlson JM, Doyle J (2002) *Proc Natl Acad Sci USA* 99(Suppl 1):2538–2545.