Technical Report CoSBi 01/2007

Emergence and Maintanence of Functional Modules in Signaling Pathways

Orkun S. Soyer

The Microsoft Research - University of Trento
Centre for Computational and Systems Biology

soyer@cosbi.eu

This is the preliminary version of a paper that will appear in
BMC Evolutionary Biology
available at http://www.biomedcentral.com/bmcevolbiol/

ABSTRACT

While detection and analysis of functional modules in biological systems have received great attention in recent years, we still lack a complete understanding of how such modules emerge. One theory is that systems must encounter a varying selection (i.e. environment) in order for modularity to emerge. Here, we provide an alternative and simpler explanation using a realistic model of biological signaling pathways and simulating their evolution. These evolutionary simulations start with a homogenous population of a minimal pathway containing two effectors coupled to two signals via a single receptor. This population is allowed to evolve under a constant selection pressure for mediating two separate responses. Results of these evolutionary simulations show that under such a selective pressure, mutational processes easily lead to the emergence of pathways with two separate sub-pathways (i.e. modules) each mediating a distinct response only to one of the signals. Such functional modules are maintained as long as mutations leading to creation of new interactions among existing proteins in the pathway are rare. While supporting a neutralistic view for the emergence of modularity in biological systems, these findings highlight the relevant rate of different mutational processes and the distribution of functional pathways in the topology space as key factors for its maintenance.

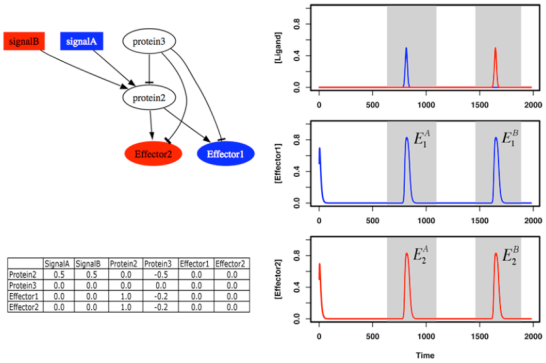
INTRODUCTION

Functional modules are observed at various levels in biology, ranging from sub cellular to the ecosystem. A general definition that holds across these different levels is that a functional module is a discrete entity whose function is separable from those of other modules (1). One straightforward example of such a module in the cell would be a distinct pathway mediating a certain physiological response. Besides the classical biochemical characterization of such pathways, recent analyses have identified many possible modules using multiple high-throughput data sources (2, 3). Analyses of various biological connectivity data have found therein patterns that are overrepresented and might correspond to small modules (4-6) (so-called motifs). Discovered mostly from connectivity and co-expression data, it is not clear whether these "structural" modules correspond to real functional modules that are possibly conserved over evolution (7, 8). So far, it has been only possible to test the functional role of such "discovered" modules in case of few motifs (9).

While such efforts to discover and characterize distinct pathways constituting functional modules continue, we still lack a clear understanding of how modularity emerges in biological systems of multiple interacting proteins. One possibility is that such systems are driven towards having modular structures due to selection for stability. This idea is supported by studies on linear systems of interacting units, which show that stability is enhanced by modular arrangement of interactions (10, 11). An evolutionary analysis on similar systems showed that selection under varying environments (i.e. varying fitness requirements) leads to increase in modularity (12). It is not clear how these findings relate to real biological systems, which are known to have highly nonlinear dynamics. Further, the abstract models used in these studies fall short of capturing real biological systems and might sometimes introduce biases in the analysis (13). Extending on the idea of selection under varying environments, a recent study found that modularity could evolve in electronic circuit and neural network models under an alternating selection scheme (14). In this scheme, selection had to be alternated between two different and modular tasks in order for modularity to evolve. While relevant for engineering modular structures in electronic circuits and neural networks, it is not clear how these findings would relate to real biological systems and their evolution. In particular, it is unlikely that the requirement of modularly alternating selection schemes would be fulfilled in natural evolution.

A more biologically relevant hypothesis suggests that modularity might emerge as a result of simple evolutionary processes. This idea have been put forward in a "thought experiment", to explain modularity in regulatory pathways and bacterial diversification (15). A more detailed

population genetic treatment of similar ideas using toy models of regulatory networks have found that in small populations, separated regulation of genes can emerge neutrally (16). This work suggests that once such separation has emerged, simple selective pressures can then lead to modularity at regulatory pathway and phenotype level. While providing an intuitive evolutionary argument for the emergence of modularity, neither of these studies provides a complete explanation. In particular, they lack a detailed treatment of how different mutational processes affect the emergence and maintenance of functional modules during evolution.



	SignalA	SignalB	Protein2	Protein3	Effector1	Effector2
Protein2	0.5	0.5	0.0	-0.5	0.0	0.0
Protein3	0.0	0.0	0.0	0.0	0.0	0.0
Effector1	0.0	0.0	1.0	-0.2	0.0	0.0
Effector2	0.0	0.0	1.0	-0.2	0.0	0.0

Figure 1. Cartoon and mathematical representation of the ancestral pathway, used in the evolutionary simulations, and its dynamical response to two ligand molecules. The latter is obtained by solving the set of differential equations describing the concentration of each protein in the pathway and is used in the calculation of pathway fitness. Gray areas indicate the time brackets when pathway response through effector 1 (E_1^A and E_1^B) and effector 2 (E_2^A and E_2^B) are evaluated (see *Methods*). Proteins labeled as two and three correspond to a receptor and "global deactivator" (i.e. non-specific phosphotase) respectively. Interaction coefficients are shown as a matrix, listing the actions of other proteins on a given protein row-by-row.

Here, we analyze the emergence and maintenance of functional modules using a realistic model of signaling pathways and their evolution. We assume that signaling pathways have evolved from a simple ancestral pathway containing few proteins, some of which acted as effectors and receptors. The fitness benefit for an organism to mediate separate (and possibly dynamically different) responses to different signals would exert a constant selective pressure on such a pathway for achieving specific signal-response relations. We propose that such a selective pressure would then drive pathways to evolve modular structures. To test this hypothesis, we use mathematical models of signaling pathways and evolutionary simulations. Results of these simulations show that pathways evolve readily distinct sub-pathways or modules that mediate specific signal-response relations. Further analyses highlight duplications and protein recruitment as key mutational processes facilitating modularity. On the other hand, creation of new interactions among existing proteins in a pathway destroy functional modules and lead to crosstalk and complex pathways. The relevant rates of these different mutational processes that shape pathway topology, and the distribution of such topologies in the topology space emerge as the key determinants for the evolution of modularity.

RESULTS AND DISCUSSION

To test the hypothesis that modularity in signaling pathways emerge as a result of evolution towards mediating distinct responses to different signals, we use mathematical models of such pathways and simulate their evolution. These simulations start with a homogenous population of an "ancestral" pathway that contains two effectors (effector one and two), one receptor and one intermediary protein. Both of these proteins are assumed to be non-specific; the receptor has equal affinity towards all ligand molecules present in the medium, and equally activates the two effectors, while the intermediary protein acts as a "global" deactivator inhibiting both the receptor and the two effectors with equal strength. Figure 1 shows this ancestral pathway and its response (the time course of active effectors) to two distinct ligand molecules (signal A and B hereafter). During the course of evolution, each generation is created from the previous one by selecting pathways randomly with replacement and allowing them to replicate with a probability proportional to fitness. Here, we use a fitness function that represents a constant selective pressure on pathways to mediate distinct

responses to the different signals presented. It rewards pathways ability to respond through effector one (two) in presence of signal A (B), and not in presence of signal B (A).

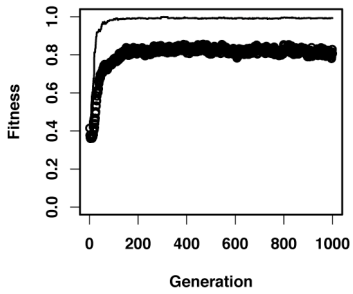


Figure 2. Fitness during an evolutionary simulation starting with a homogenous population containing only the ancestral pathway. Circles and the line represent the average fitness of the population and the highest fitness at each generation respectively.

Figure 2 shows the average fitness during the course of a typical evolutionary simulation. As the ancestral pathway responds in identical fashion to both signals through both effectors, the average fitness is initially low. However, evolution results quickly in high fitness values and pathways in the final population are able to respond specifically to each signal through the corresponding effectors. Figure 3 shows a sample pathway from the final population and its response. As clearly seen in the cartoon representation of this pathway, signals A and B are propagated through the pathway via receptors and over intermediary proteins to the two effectors, following two separate paths. The ancestral pathway has evolved into two separate sub-pathways or modules for processing each of the signals. In fact, such modularity is found in all pathways present in the final population. For each of these pathways there exist a path, connecting signal A (B) with effector one (two), while there is no such path to effector B (A). Additional simulations result in similar fitness curves (see Supporting Figure 7) and final populations that contain only modular pathways. Furthermore, we find that in all these simulations modular, high fitness pathways first emerge in the population after only few generations (19 generations for the simulation shown in Figure 2). These results indicate that evolution under a constant and biologically plausible selective pressure leads readily to the emergence of functional modules in signaling pathways.

To better understand how such modularity emerges in these simulations, we analyze the evolutionary processes that shape pathway structure. Here, we consider duplication and loss of proteins, loss and creation of interactions, and adjustment of kinetic rates as such processes. Creation of new interactions can result when point mutations (or accumulation thereof) on a protein lead to a new binding surface for recognizing another protein or signal, as observed *in vitro* (17, 18). Considering that there are many proteins in an organism that are not participating in a given pathway, it is much more likely that such mutations would lead to formation of a new interaction between a protein that is already participating in this pathway and one that is not (i.e. protein recruitment). This intuition leads to the assumption that creation of new interactions among existing proteins in a pathway are negligibly rare compared to new protein recruitment. Results shown in Figure 2 are obtained under such an assumption (i.e. all interaction creation events were modeled as protein recruitment).

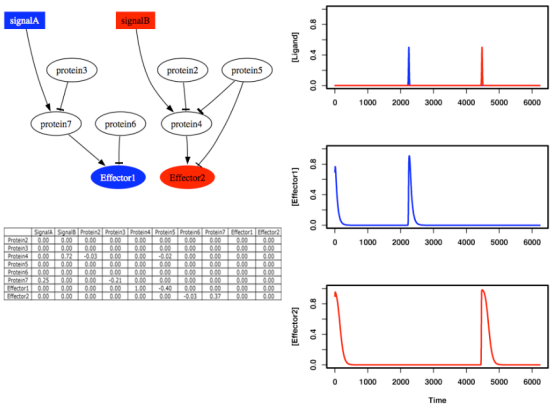


Figure 3. A sample pathway from the final generation of the evolutionary simulation shown in Figure 2, and its response to two ligand molecules. Note the separation of signal-response relations both at dynamic and structural levels. The mathematical description of the pathway shows only the non-zero interaction coefficients, listing the actions of other proteins on a given protein row-by-row.

Relaxing this assumption, we run additional simulations with decreasing probability for

protein recruitment in expense of new interactions forming among existing proteins. Figure 4 shows the frequency of different pathway types in the final populations obtained from these simulations. We find that allowing interaction creation among participating proteins in a pathway diminish the chances of modularity emerging and lead to complex pathways or crosstalk (i.e. from one of the signals there exist two paths leading to both effectors, see sample pathways shown in Supporting Figure 8). Analysis of the distribution of pathway types over the entire evolutionary simulation, we get a clearer picture (see Figure 5); regardless of the relative rate of protein recruitment and interaction creation, modular pathways emerge relatively quickly in the population. However, in presence of the latter process modular pathways get quickly replaced by pathways with crosstalk or complex pathways. The relevant rate of this process determines the fraction of modular pathways that can be maintained in the population. Analyzing the effects of different mutational processes on pathway structure, we find that transitions from modular pathways to pathways with crosstalk are extensively caused by interaction addition (data not shown). The reverse transitions, resulting in modular pathways, are solely driven by protein and interaction loss. Hence, the emergence and maintenance of functional modules is mostly determined by the relevant rate of these different mutational processes.

Another key mutational process is duplication of proteins already participating in the pathway. Without duplication there is no possibility of functional modules emerging. For example, new receptors can only be created through duplication in the model. Furthermore, duplications push pathways to grow in size and make it possible for the pathway structure to get rearranged towards modularity via other mutational events. Pathway growth (see Supporting Figure 9) occurs despite the higher frequency of protein loss because duplications, and also to some extent protein recruitments, are less costly in terms of fitness as shown in Figure 6. This finding is inline with previous studies analyzing pathway growth in similar models (19). Similarly, imposing a high fitness cost for additional proteins in the model prohibits pathway growth and emergence of modularity.

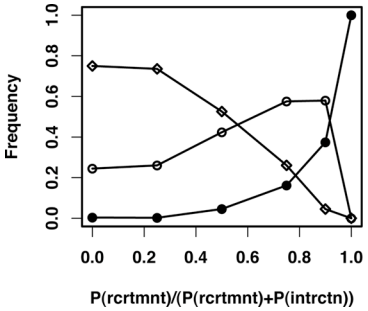


Figure 4. Frequency of different pathway structures in the final generation of the evolutionary simulations with increasing ratio of protein recruitment over the sum of interaction creation and protein recruitment probabilities (see *Methods*). For each probability ratio the frequencies are obtained as an average over seven different runs. We distinguish among three different structural types for pathways. Pathways where there is a path from each signal to only one effector and the other (modular, solid circles), pathways where there is a path from one of the signals to both effectors (crosstalk, open circles), pathways where there is a path from each signal to each effector (complex, open diamonds).

To summarize, results from these evolutionary simulations suggest the following scenario for the evolution of functional modules. Simple, non-specific pathways that arose early in evolution would grow in size due to low fitness costs associated with protein duplication and recruitment events. As pathways grow, mutations leading to loss of proteins or their interaction would lead to rearrangement of the pathway structure resulting in the emergence of functional modules. This process happens surprisingly easily and does not require a complex selective pressure. In fact, we find that functional modules emerge even with an alternative fitness function that is simply based on the ability of the pathway to respond to two signals (i.e. no additional reward for response separation). This supports a neutralistic view for the emergence of modularity, as envisioned in regulatory pathways (16). Once emerged, functional modules would then be maintained depending on the frequency of mutations leading to formation of new interactions

among proteins participating in the pathway. This process causes modular pathways to drift towards complex pathways and crosstalk, which provide equally fit solutions as their modular neighbors.

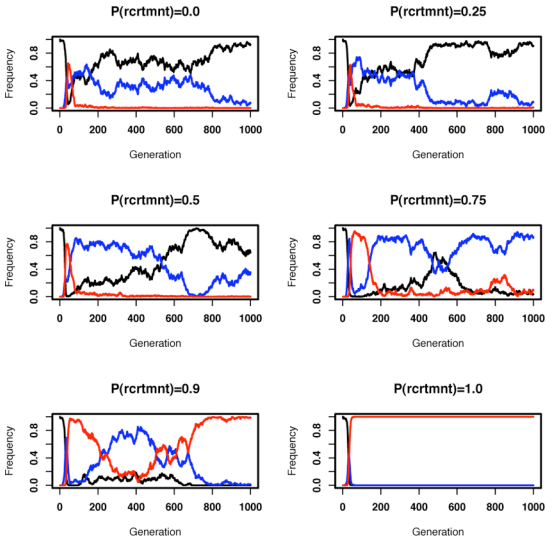


Figure 5. Frequency of different pathway structures during the course of evolution. Different panels show results from sample simulations with increasing probability for protein recruitment in expense of interaction creation (i.e. results from one of the runs used to create Figure 4). Red, blue and black lines show the frequency of modular, crosstalk, and complex pathways (see the legend of Figure 4 for pathway types). Note, that measurements are taken after mutations but before selection, hence there is a small fraction of unconnected pathways at each generation (not shown on the graph).

Such neighbor relations, in other words the distribution of functional pathways (i.e. pathways capable to produce separate responses to separate signals) in the topology space, are the other key determinant for the emergence of modularity. Repeating the evolutionary simulations with an initial population composed of random pathways that contain three or five intermediary proteins, we find high variance in the frequency of modular pathways in the final population even when there are no interaction creation events allowed in the model (results not shown). This suggests that for some pathways it is highly improbable for mutational events to restructure them towards modularity, while for others the opposite is true. Even though the presented study assumes a biologically plausible ancestral structure, it would be highly desirable (but equally difficult) to analyze the available topology space and the distribution of biological topologies in it to exactly understand the role of topology space in the evolution of system level properties like modularity. Finally, we note that horizontal gene transfer, another evolutionary process that we did not consider here, can facilitate modularity in signaling pathways, as it is found that most of such events involve transfer of entire receptor-effector pairs rather than individual proteins in bacteria (20).

CONCLUSION

This study provides a simple and biologically plausible explanation for the widely observed modularity in biological pathways. According to this explanation, functional modules specializing in processing one of the multiple signals an organism could encounter emerges readily under a constant selective pressure. The driving processes behind such emergence are protein duplication and recruitment events leading to pathway growth, and loss of proteins and their interactions leading to rearrangement of pathway topology. Once arisen, the probability that such functional modules will be maintained will depend on the frequency of mutations leading to formation of interactions among proteins already participating in the pathway. Hence, the extent of modularity in a specific pathway will mainly depend on the relevant rates of these different mutational mechanisms and how functional pathway topologies are distributed over the entire topology space for a given function.

These findings are highly relevant for our understanding of modularity in biological systems, and for applying such understanding to mimic biology in engineering applications. Firstly, they validate the previous arguments that modularity can emerge readily in biological pathways (15, 16) without any need for complex selective pressure. We find that the more difficult part is in the maintenance of such modularity, as non-modular pathways can be equally capable of achieving functionality. Here we provide a mechanistic view of how modularity could be maintained. However, it is

equally plausible that functional modules are maintained due to their secondary fitness benefits such as increases in evolvability (21) or robustness. Secondly, we find that the different mutational processes affecting the structure of biological pathways are highly important for the emergence and maintanence of modularity. This indicates that any practical application aiming to mimic the properties of biological systems (such as in (14)) have to pay careful consideration to such biologically plausible mutations that can shape the structure of a given system.

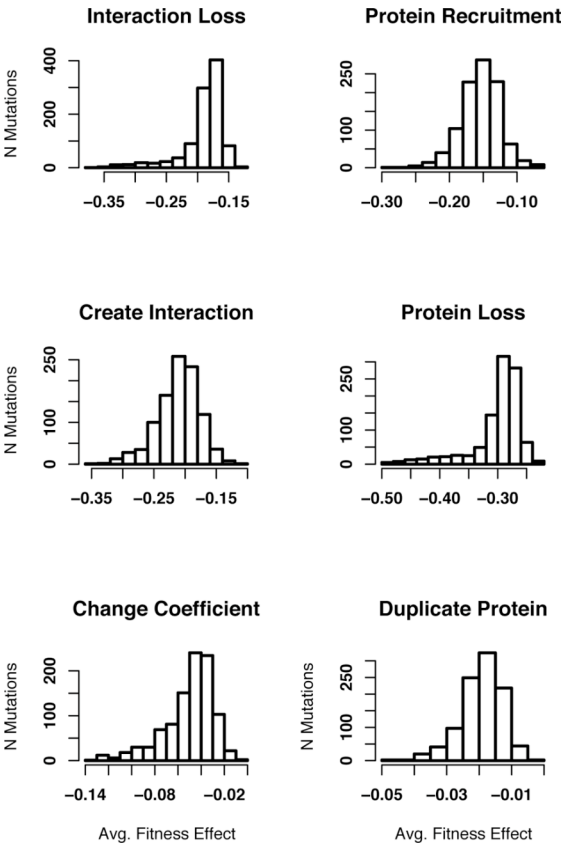


Figure 6. Distribution of different mutations based on their average fitness effects. Results are the average over seven different runs of a simulation where the ratio of protein recruitment probability over the sum of interaction creation and protein recruitment probabilities was 0.5 (one of the simulations used to create Figure 4). Each panel shows the distribution for a different mutational mechanism. Note that the x- and y-axes are scaled differently on each panel.

Finally, we note that the presented scenario for the evolution of signaling pathways is in its essence similar to the one put forward for the evolution of metabolic pathways. According to that theory, current day metabolic pathways with specialized enzymes have evolved, from an ancestral pathway containing non-specific ones, under constant selective pressure for high metabolic yield (22). Evolutionary simulations testing this scenario did not only find it plausible, but has further showed that it could explain the existence of hub molecules in these pathways (23). Other, similar studies have shown that mutational processes and evolutionary mechanisms can be the driving force behind many of the system level properties in biology (16, 19, 24-27). Taken together all these results demonstrate the importance of evolutionary approaches for achieving a complete system level understanding in biology.

ACKNOWLEDGEMENTS

We thank Matteo Cavaliere for insightful discussions, Sebastian Bonhoeffer for technical support and Richard Goldstein for valuable comments on an earlier version of this manuscript.

REFERENCES

1. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. (1999) *Nature* **402,** C47-C52.
2. Spirin, V. & Mirny, L. A. (2003) *Proceedings of the National Academy of Sciences of the United States of America* **100,** 12123-12128.
3. Tanay, A., Sharan, R., Kupiec, M. & Shamir, R. (2004) *Proceedings of the National Academy of Sciences of the United States of America* **101,** 2981-2986.
4. Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M. & Alon, U. (2004) *Science* **303,** 1538-42.
5. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. (2002) *Science* **298,** 824-7.
6. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. (2002) *Nat Genet* **31,** 64-8.
7. Snel, B. & Huynen, M. A. (2004) *Genome Res* **14,** 391-7.
8. Spirin, V., Gelfand, M. S., Mironov, A. A. & Mirny, L. A. (2006)

Proceedings of the National Academy of Sciences of the United States of America **103,** 8774-8779.
9. Mangan, S. & Alon, U. (2003) *Proc Natl Acad Sci U S A* **100,** 11980-5.
10. May, R. M. (1972) *Nature* **238,** 413-4.
11. Variano, E. A., McCoy, J. H. & Lipson, H. (2004) *Phys Rev Lett* **92,** 188701.
12. Lipson, H., Pollack, J. B. & Suh, N. P. (2002) *Evolution Int J Org Evolution* **56,** 1549-56.
13. Gardner, A. & Zuidema, W. (2003) *Evolution Int J Org Evolution* **57,** 1448-50.
14. Kashtan, N. & Alon, U. (2005) *Proc Natl Acad Sci U S A* **102,** 13773-8.
15. Rainey, P. B. & Cooper, T. F. (2004) *Res Microbiol* **155,** 370-5.
16. Force, A., Cresko, W. A., Pickett, F. B., Proulx, S. R., Amemiya, C. & Lynch, M. (2005) *Genetics* **170,** 433-46.
17. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. (2003) *Nature* **423,** 185-90.
18. Ohashi, K., Yamashino, T. & Mizuno, T. (2005) *J Biochem (Tokyo)* **137,** 51-9.
19. Soyer, O. S. & Bonhoeffer, S. (2006) *Proc Natl Acad Sci U S A* **103,** 16337-42.
20. Alm, E., Huang, K. & Arkin, A. (2006) *PLoS Comput Biol* **2,** e143.
21. Kirschner, M. & Gerhart, J. (1998) *Proc Natl Acad Sci U S A* **95,** 8420-7.
22. Kacser, H. & Beeby, R. (1984) *J Mol Evol* **20,** 38-51.
23. Pfeiffer, T., Soyer, O. S. & Bonhoeffer, S. (2005) *PLoS Biol* **3,** e228.
24. Eisenberg, E. & Levanon, E. Y. (2003) *Phys Rev Lett* **91,** 138701.
25. Salazar-Ciudad, I., Newman, S. A. & Sole, R. V. (2001) *Evol Dev* **3,** 84-94.
26. Wagner, A. (2003) *Proc R Soc Lond B Biol Sci* **270,** 457-66.
27. Wagner, A. (2005) *Proc Natl Acad Sci U S A* **102,** 11775-80.
28. Soyer, O. S., Pfeiffer, T. & Bonhoeffer, S. (2006) *J Theor Biol*.
29. Soyer, O. S., Salathe, M. & Bonhoeffer, S. (2006) *J Theor Biol* **238,** 416-25.
30. Stock, A. M., Robinson, V. L. & Goudreau, P. N. (2000) *Annu Rev Biochem* **69,** 183-215.