In silico evolution of allosteric protein modules

Julien F. Ollivier[1] and Peter S. Swain[1]

***Short Abstract* — Signaling proteins typically co-localize in allosterically regulated, multi-state macro-molecular complexes when processing signals. Combining a rule-based modeling approach with a genetic algorithm, we can evolve (or design) protein modules that accomplish specified functions. With the analysis of these modules, we intend to recover known signaling design principles and suggest novel ones.**

***Keywords* — allosteric regulation, cellular signaling, macro-molecular complexes, genetic algorithms, rule-based modeling.**

I. INTRODUCTION

SEVERAL recent theoretical studies have demonstrated the value of an *in silico* evolutionary approach to increasing our understanding of gene networks [1] and protein signaling networks [2, 3]. These studies shed light on the design principles of functional modules and suggest how evolution might have given rise to a module in small incremental steps [1-3].

In contrast to these studies, our approach emphasizes the importance of recruitment and co-localization, allosteric transitions [4], and the modular structure of signaling proteins [5]. Our methodology includes a rule-based modeling tool [6], which takes as input a high-level textual description of an evolved functional module and generates a detailed network of interactions. This high-level description facilitates subsequent analysis: we can delineate functional modules within proteins, ascribe them adaptor/scaffolding roles, and gain an understanding of the interplay between co-localization and allosteric regulation in signaling [4, 5].

II. METHOD

We use an *in silico* evolutionary method to discover signaling networks. A genetic algorithm simulates evolution on a population of virtual genomes. Each genome is translated into a high-level model consisting of a set of proteins and their interactions. This model is "unrolled" by an allosteric network compiler to produce a detailed set of biochemical equations. We simulate the response of the network to an appropriate stimulus. A scoring function ascribes a fitness value to the network quantifying how closely the network output matches the desired response.

A. Genetic algorithm

Our genetic algorithm simulates genome evolution on a population of virtual genomes. From generation to generation, the genomes replicate, mutate and die according to their fitness value, driving the population towards incrementally better designs. Mutational operators include gene duplication, gene deletion, domain shuffling, and point mutations.

B. Genotype-phenotype mapping

A set of mapping rules delineates proteins and protein domains within the virtual genomes and generates a high-level description of the protein interaction network.

C. Allosteric Network Compiler (ANC)

ANC is a rule-based modeling tool [6]. Its input is a high-level, textual description of each signaling protein's structure, allosteric transitions if any, and active site ligands or substrates. ANC infers and outputs the set of biochemical equations implied by the textual description. The equations output by ANC are translated into ODE form for simulation in MATLAB [7].

III. RESULTS AND CONCLUSION

We first used a suitable scoring function to evolve an ultrasensitive switch. We succeeded and the networks thus obtained all operated using 0^{th}-order ultrasensitivity [8]. Recovering this well-known design principle *de novo* is encouraging and validates our methodology.

REFERENCES

[1] Francois P, Hakim V, Siggia ED (2007) Deriving structure from evolution: metazoan segmentation. *Molecular Systems Biology* **3:154**.

[2] Paladugu et al. (2006) *In silico* evolution of functional modules in biochemical networks, *IEE Proc.-Syst. Biol.,* **153:4**, 223-235.

[3] Soyer OS, Pfeiffer T, Bonhoeffer S (2005) Simulating the evolution of signal transduction pathways, *J. Th. Biol.,* **241**, 223-232.

[4] Kuriyan J, Eisenberg D (2007) The origin of protein interactions and allostery in colocalization, Nature, **450:13**, 983-990. Review.

[5] Bhattacharyya RP, Remenyi A, Yeh BJ, Lim WA (2006) Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits, *Annu. Rev. Biochem.,* **75**, 655-80. Review.

[6] Hlavacek et al., (2006) Rules for Modeling Signal-Transduction Systems, *Sci. STKE..*, **344:re6**. Review.

[7] Siso-Nadal F, Ollivier JF, Swain PS (2007) Facile: a command-line network compiler for systems biology. *SMBC Syst. Biol.,* **1:36**.

[8] Goldbeter A, Koshland DE Jr. (1981) An amplified sensitivity arising from covalent modification in biological systems. *PNAS* **78:11**, 6840-4.

[1] Department of Physiology and Centre for Non-Linear Dynamics in Physiology and Medicine, McGill University, 3655 Promenade Sir William Osler, Montreal Québec, H3G 1Y6, Canada.

E-mail: julien.ollivier@mail.mcgill.ca and swain@cnd.mcgill.ca.