The structure and evolution of sequence niches in cellular regulation and signaling

Christopher R. Myers[1]

***Short Abstract* — In order to function appropriately, living cells must convey reliable biomolecular signals despite the stochastic nature of molecular activity. One potential source of signaling noise is the promiscuous interactions that characterize many important regulatory and signaling processes. Sequence niches driven by negative selection have been proposed as a mechanism for avoiding crosstalk. The structure and evolution of such niches, as well as their connection to phase transitions in constraint satisfaction problems and message coding, is discussed.**

The functioning of complex biomolecular pathways hinges on conveying molecular signals reliably in the stochastic and evolving milieu of living cells. These signals are mediated by molecular interactions that distinguish physiological binding partners from myriad other cellular constituents. But molecular recognition is subtle: many of the interactions involved in cellular regulatory and signaling pathways do not involve highly specific lock-and-key binding, but instead are characterized by somewhat fuzzy and promiscuous recognition of families of sequences and configurations [1,2]. This motivates some key questions: to what extent is molecular distinguishability achieved, and if it is, by what mechanisms? How is specificity aggregated from intrinsic sequence-dependent signals and extrinsic context-dependent ones (e.g., localization to scaffolds)? The distribution of interaction specificity is an important element in the organization of complex information processing networks, both those engineered [3] and evolved [4,5]. Theories of message coding – which strive to convey signals reliably by organizing the geometry of high-dimensional message spaces – are also applicable to the study of biomolecular regulation and signaling [6].

Recent experiments aimed at understanding the specificity of SH3-mediated signaling in yeast [7] have suggested one possible mechanism for achieving signaling specificity in crowded sequence spaces: negative evolutionary pressure against crosstalk that drives the coevolution of protein sequences such that substrate motifs hide in niches in sequence space where they are recognized only by one target protein and not by any other competing proteins. Earlier theoretical work explored aspects of distinguishability among sets of competing proteins [8,9].

Acknowledgements: This work was funded by USDA/ARS grant 1907-21000-017-05.

[1]Cornell Theory Center, Cornell University, Ithaca, NY. E-mail: myers@tc.cornell.edu

I have distilled the scenario suggested by this experimental work to formulate the Sequence Niche Problem, which asks under what conditions such sequence niches are possible, i.e., for which crosstalk can in principle be avoided [10]. Simulations of random instances of the Sequence Niche Problem illustrate a transition from a satisfiable regime (where crosstalk can be avoided) for small numbers of competing proteins to an unsatisfiable regime (where no niche exists) for large numbers of proteins. This transition is reminiscent of those seen in other NP-complete constraint satisfaction problems, such as k-SAT [11]. Fragmentation of the neutral network of allowable solutions is also evident near the transition to unsatisfiability; fragmentation of this sort has been observed in k-SAT and leads to increasingly solution difficulty in the "hard SAT" phase [12].

Ongoing work is focused on probing the evolution of such sequence niches under selection pressure. This includes extending simple models of competitive protein binding to study evolution under selection, as well as examining the structure of evolved sequence niches in nature as deciphered from experimental studies of protein-protein interaction.

REFERENCES

[1] M. Ptashne and A. Gann, *Genes & Signals*, Cold Spring Harbor Lab Press, 2002.

[2] L. Castagnoli, *et al.,* "Selectivity and promiscuity in the interaction network mediated by protein recognition modules", *FEBS Letters* 567, 74-79 (2004).

[3] C.R. Myers, "Software systems as complex networks: structure, function, and evolvability of software collaboration graphs", *Phys. Rev. E* 68, 046116 (2003).

[4] M. Kirschner and J. Gerhart, "Evolvability", *Proc. Natl. Acad. Sci.* 95(15), 8420-8427 (1998).

[5] A.M. Sengupta, M. Djordjevic and B.I. Shraiman, "Specificity and robustness in transcription control networks", *Proc. Natl. Acad. Sci.* 99(4), 2072-2077 (2002).

[6] S. Itzkovitz, T. Tlusty and U. Alon, "Coding limits on the number of transcription factors", *BMC Genomics* 7, 239 (2006).

[7] A. Zarrinpar, S.-H. Park and W.A. Lim, "Optimization of specificity in a cellular protein interaction network by negative selection", *Nature* 426, 676-680 (2003).

[8] R.P. Sear, "Specific protein-protein binding in many-component mixtures of proteins" *Phys. Bio.* 1(2), 53-60 (2004).

[9] R.P. Sear, "Highly specific protein-protein interactions, evolution and neagtive design", *Phys. Bio*. 1(3), 166-172 (2004).

[10] C.R. Myers, "Satisfiability, sequence niches, and molecular codes in cellular signaling", www.arxiv.org/q-bio/0702042 (2007).

[11] S. Kirkpatrick and B. Selman, "Critical behavior in the satisfiability of random Boolean expression", *Science* 264(5163), 1297-1301 (1994).

[12] M. Mézard and R. Zecchina, "Random K-satisfiability problem: from an analytic solution to an efficient algorithm*", Phys. Rev. E* 66, 056126 (2002).