



THE ROCKEFELLER UNIVERSITY

1230 YORK AVENUE • NEW YORK, NEW YORK 10021-6399

November 29, 2004

Biocomplexity Faculty Search Committee
c / o Prof. Rob de Ruyter van Steveninck
Department of Physics
Indiana University
Swain Hall West 117
Bloomington, IN 47405-7105

Re: Letter of Recommendation for **Saurabh Sinha**

Dear members of the Search Committee,

It is with great pleasure that I recommend Saurabh Sinha for a faculty position in your department. I have had scientific interactions with Saurabh over a period of two years and have worked more closely with him for a year, in the context of a long-standing collaboration between Eric Siggia's and my own group.

Saurabh studied Computer Science and Engineering at the Indian Institute of Technology in Kanpur. During his PhD at the University of Washington, Seattle, he developed algorithms, now successfully being used in bacteria and yeast, that search for transcription factor binding sites on the basis of statistical overrepresentation of sequence motifs ('YMF'). In Siggia's group, he moved on to transcriptional regulation in higher eukaryotes. Using *Drosophila* segmentation as a paradigm, Siggia's group had developed an algorithm, 'Ahab', which uses a thermodynamic model to detect *cis*-regulatory elements or 'modules'. Ahab uses binding site information for multiple transcription factors in the form of position weight matrices and seeks an optimal binding of the factors to a given sequence window. Subsequent experimental work carried out in my lab showed that Ahab predicts not only the genomic position but also the binding site composition of segmentation modules with excellent success.

An important next step was to make use of the growing number of sequenced genomes and incorporate evolutionary comparison into the analysis of non-coding sequence, both to improve the detection of regulatory elements and, equally important, to study their evolution. This is the problem Saurabh chose to tackle. Over the past two years, he developed and implemented algorithms that integrate multi-species comparison in the search for binding site clusters (an extension of Ahab called 'Stubb') or individual binding sites ('PhyME'). Throughout this work, Saurabh has demonstrated an exceptional command of the computational aspects of the problem and a strong ability to acquaint himself with the underlying biological questions and methods. This is all the more impressive since he had relatively little knowledge of the biology and evolution of higher eukaryotes when he arrived at Rockefeller University.

Over the past year, Saurabh has closely collaborated with my lab on the biological application of Stubb. In a first paper (Sinha et al. BMC: Bioinformatics 2004, 5:129), we examined how cross-species comparison between *D. melanogaster* and *D. pseudoobscura* using Stubb improves the genome-wide prediction of *cis*-regulatory elements, and determined the optimal parameter settings for the algorithm. In a separate study, Saurabh aligned genomic sequences from three *Drosophila* species (*melanogaster*, *yakuba*, *pseudoobscura*) and compared rates of substitutions, insertions and deletions between functional and neutral sequence. He discovered a high sequence turnover rate in functional regions as a result of frequent insertions, and found that short local repetitions (tandem repeats) may account for this turnover. Together, we then examined how these molecular changes affect the predicted binding site composition of orthologous segmentation gene modules in *D. melanogaster* and *D. pseudoobscura*. We found an impressive array of changes (loss or gain of entire modules, large indels carrying multiple sites, sites moving within modules, severe mutations in specific sites) and experimentally tested the effects on module expression patterns for multiple examples in each category. We could show that duplicate modules with overlapping function tend to exchange some of their functionality and that even non-redundant modules show noticeable changes in expression, as a result of fluidity in the distribution of binding sites along the DNA. This project, which will be concluded over the next few months, represents the first comprehensive evolutionary comparison of a regulatory network in higher eukaryotes. It provides novel insights into the molecular mechanisms driving evolution and the mechanisms by which regulatory networks remain robust against change, and has major implications for the bioinformatic treatment of regulatory evolution. The work required the integration of several different heavily researched areas (bioinformatics, developmental biology, evolution) and presented a number of technical challenges (multi-species sequence alignment, systematic evaluation of expression patterns, correlating different axes of information).

With his work, Saurabh has made a very substantial contribution to the burgeoning fields of systems biology and molecular evolution, both areas of great scientific import and potential for growth. He has shown himself to be technically versatile and innovative, but also organized and meticulous, with excellent attention to detail. Saurabh's aptitude for interdisciplinary research is exceptional and lies in a combination of talent and personality: He thinks and writes clearly and genuinely strives for transparency, a rare but crucial skill. His openness to biological concepts and methods combined with his genuine friendliness and cooperative spirit make him a pleasure to collaborate with. I am confident that he will be not only a very successful and innovative researcher, but also a wonderful colleague and teacher.

Please don't hesitate to contact me if you need additional information.

With best regards,



Ulrike Gaul, PhD
Associate Professor, Head
Laboratory of Developmental Neurogenetics

Tel: 212 327 7621
Fax: 212 327 7923
email: gaul@mail.rockefeller.edu