November 10 2005

Yves Brun
Systems Biology/Microbiology Faculty Search
Department of Biology
Indiana University
Jordan Hall 142
1001 E 3rd St
Bloomington IN 47405-7005

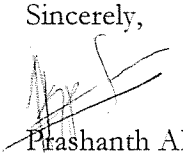
Dear Dr. Brun,

Attached please find my CV and statement of research plans and teaching interests as application for your assistant professor faculty position in computational Systems Biology. I have requested reference letters from Profs. Craig Benham (UC Davis), Carol Newlon (UMDNJ) and Bruce Stillman (CSHL), which should arrive shortly.

My work takes a unique approach to Systems Biology. My research involves quantitative elucidation of emergent complex behaviors of DNA. I study this property at multiple scales in a wide range of functions and organisms, spanning from replication origins in yeast to social/sexual behavior in voles, and often use a combination of computational and experimental approaches. The uniqueness of the work I do and the levels it spans, ranging from molecular to organismal multi-scale biology, will benefit the research efforts at the Biocomplexity Institute, especially given my proclivity for collaborative research. I feel that my multidisciplinary background, broad-ranging interests (my Ph. D. research included development of computational methods for analyzing large datasets) and proven ability to meld diverse approaches make me exceptionally suited for interdisciplinary research in a setting such as the Biocomplexity Institute. My expertise in both experimental and theoretical approaches to research problems will stand in good stead to augment the IU-Bloomington's research efforts in the life sciences. I look forward to the possibility of being a part of your research community.

Thank you very much for considering my application.

Sincerely,



Prashanth Ak

SUMMARY OF PROPOSED RESEARCH

Prashanth Ak

My research program focuses on elucidating fundamental design principles governing genomic function and its organization. DNA strand separation and re-formation in a precisely ordered fashion is fundamental to key genomic events such as transcription and replication, as well as a variety of other normal and pathological activities of DNA. Therefore, strand separation must be stringently and dynamically controlled in living systems. I seek to understand how this ubiquitous property of DNA is organized genome-wide, and the general principles governing the way strand separation regulates genome function and interacts with other multi-component networks that together constitute the genetic regulatory system.

Assessing the propensity of DNA for strand separation entails not only an understanding of the molecular machinery, but also, more fundamentally, of the energetics involved in such processes. Common methods of assessing strand separation using local DNA sequence information (A+T content/thermodynamic stability) are not relevant, for two reasons: First, strand separation *in vivo* happens under isothermal conditions, so there are no changes of temperature to drive a thermal denaturation ('melting'). Second, DNA is held under regulated negative superhelicity *in vivo*, which couples the strand opening behaviors of all the base pairs that experience them. This coupling can occur over long distances, extending over many kilobases. Therefore, to accurately assess strand separation, the interplay between local properties (pair bonding, stacking energies and so on) and the long range coupling induced by superhelical stress energies needs to be considered. Hence, not only the local sequence, but also the genomic context in which a given sequence of interest is located needs to be taken into account.

The calculation methods that I employ use genomic sequence data with **empirically determined** parameters (there are no free variables) to accurately evaluate energetic requirements of any given region of the DNA duplex to strand separate. Sites that are calculated to have lowered energetic requirements have been shown experimentally to separate, both *in vivo* and *in vitro*. Many investigators have shown specific instances of supercoiling-induced unwinding and its importance *in vivo*. (A few examples are listed in refs. 1-8.)

Such calculations, based on an original use of sequence information, represent a novel, structure-based approach that goes beyond existing pattern matching techniques to extract knowledge from genomic sequences. This approach enables insight into novel genome-wide organizational and regulatory features that cannot be identified by other current methods. My recent genome-wide

analysis in yeast shows that regulatory regions such as promoter regions, transcription terminator regions and replication origins have specific patterns of duplex destabilization (9,10).

My main interests, as well as the strengths of these methods, go well beyond pattern extraction and identification of biological features. One of the most important properties of superhelically mediated duplex destabilization is that it can be regulated dynamically by protein-DNA interactions. Protein binding to a duplex destabilized region can stabilize that region, which will transfer the destabilization to another region because of superhelical coupling of strand opening behaviors. By using a combination of experimental and theoretical methods, in collaboration with Juergen Bode's group, I have shown that such a mechanism could operate in the induction of the human and mouse interferon-ß gene (11). Though the current level of knowledge of regulation of genomic function *in vivo* cannot yet easily take into account such precise calculations of duplex destabilization and its dynamic regulation, they are already contributing to our understanding of genomic function at a mechanistic scale, and their systems-level effects (9-13,15).

The main focus of my future research program is twofold: (i) Understanding how DNA duplex destabilization interacts dynamically with other biological factors (such as DNA binding proteins) to regulate genome function at a systems level, and (ii) Elucidating the design principles underlying the genomic organization of such structural properties of DNA. I am interested in an integrated theoretical and experimental approach to these questions. Understanding fundamental design principles across multiple levels requires a three-pronged approach consisting of experimentation, computation, and theory. I have prepared myself for this integrated approach by combining a rigorous training in experimental sciences with a solid background in theoretical methods. My preparation, by going beyond acquisition of facility in both areas to an in-depth understanding, enables me to take unique approaches to answer questions that are not tractable otherwise. Duplex destabilization is a ubiquitous property of all DNA, and therefore, instances of biological phenomena where duplex destabilization plays a functional role occur widely. This has necessitated collaborations across disciplines. Attempts to understand how duplex destabilization plays out at the whole genome, systems level have also required interdisciplinary collaborations. Initiating and maintaining these collaborations has been facilitated by my strong multidisciplinary background, which includes a Ph.D. in experimental neuroscience and a Master's degree in Life Sciences (a comprehensive curriculum in biological research at one of the top research institutions in India, the Jawaharlal Nehru University in New Delhi).

My immediate future plans target three main research areas: regulation of genome function, evolution and the neurogenetic basis of behavior. Specific instances I plan to study in order to uncover general principles of genome function are outlined below.

Genome Function

Duplex destabilization is a universal and intrinsic property of DNA in all organisms. Therefore there are a myriad of phenomena that can be classified under 'genome function' where strand separation might play a role. Here I describe my future research plans to investigate, at genome wide scales, two key genome functions, transcription and replication. I also describe my plans for research regarding the general mechanisms of insertion of foreign DNA into genomes (retrotransposons, retroviruses and vectors in gene therapy).

a) Gene Regulation: My collaborative research with Juergen Bode's group has uncovered a novel mechanism of genetic control mediated by duplex destabilization. Using a combination of theoretical and experimental techniques we have demonstrated how remote control of transcription of the human and mouse interferon-ß genes could be mediated by superhelical duplex destabilization (11). The evidence thus far indicates that many types of duplex destabilization-mediated gene regulating mechanisms occur throughout genomes. The significant duplex destabilization at the terminal regions of yeast genes (10) has given rise to the hypothesis that duplex destabilization might play a role in transcription termination. Intriguing experimental evidence suggests that this may be the case, since mutation in the SEN1 protein (a helicase) prevents proper transcription termination (14). Absence of duplex destabilization in the promoter regions of human genes but its presence in the transcription factor binding regions gives rise to an empirically testable hypothesis that a double strand binding transcription factor binds to the destabilized duplex, stabilizes it, and thereby transfers the destabilization to another location. My future work in this area aims to thresh out the general principles involved in genomic regulation at a whole genome level, in both disease and normal states.

b) Replication: DNA strand separation is of obvious importance in the initiation of replication. Supercoiling-induced strand separation has been experimentally shown to be essential in replication origin function in specific instances. With the methods I use, comprehensive, genome-wide analyses are possible. We recently showed that 38 of 39 (97%) well-characterized replication origins in yeast are highly duplex destabilized (9). Since duplex destabilization also occurs in *E.coli*, EBV and (by preliminary indications) human replication origins, it is likely to be a feature of replication origins of many species. I plan to examine the known replication origins in other species soon. This work has obvious implications for carcinogenesis.

Genome-wide views of duplex destabilization allow comparative analysis in unique ways. Comparisons of destabilization patterns within and between organisms could help reveal general principles of genome function (in this case regarding replication origins) that might exist. Genome-wide views will also help in generating hypotheses regarding mechanisms of regulation. Such

calculations aid in formulating experimental approaches that would not be possible otherwise, for example designing sequences that can alter destabilization at a region (say a replication origin) without altering the local sequence at the site of interest but rather at a location a distance away. I aim to delineate the exact role of duplex destabilization in the functioning of replication origins by a close coupling of theoretical/computational approaches with experimental approaches. This work has already started, using yeast as a model system. At the moment I am forging close collaborative links with experimentalists using a variety of other systems including viruses, plants and humans.

Currently I am completing the analysis of duplex destabilization of the ENCODE regions (the **ENC**yclopedia **O**f **D**NA **E**lements, an NHGRI consortium to study a selected 1% (~30 Mb) of the human genome sequence). This intense focus offers an unprecedented opportunity to investigate genome function by leveraging the wealth of experimental results emerging for these regions. Duplex destabilization analysis is an original approach that complements these efforts and therefore will be a novel contribution that could offer unique insights into genome function and human diseases. More than merely annotating these 30 Mb with the duplex destabilization profile, I am collaborating with experimentalists (Peggy Farnham at UC Davis, David Levens at NIH and Anindya Dutta at University of Virginia) working on RNA polymerase II, the transcription factor binding protein E2F1, the oncogenesis-related FBP (FUSE binding protein) and replication origins in the ENCODE region to arrive at an understanding of the role of duplex destabilization in transcriptional regulation and DNA replication. I also soon plan to include results from mapping of many transcription factor binding sites, measurements of mRNA levels in a variety of cell types, DNAse hypersensitive sites, functional enhancer/promoter assays, and careful annotation of sequence conservation of the ENCODE regions using sequences from an evolutionary diverse set of species.

c) Retrotransposons, retroviral diseases and gene therapy: Genome-wide assessments of yeast retrotransposons and FIV have shown that insertion into host genomes occurs preferentially at duplex destabilized sites (12,13). This observation may illuminate a general property of insertion mechanisms for some types of retrotransposons and retroviruses. By removing sequences of long terminal repeats (a sign of transposition in yeast) as well as existing retrotransposons and re-analyzing the reconstructed 'pre-integration' genome for duplex destabilization, I have shown that duplex destabilization exists *ab initio* of transposition. Novel *in silico* experiments can be used both for hypothesis generation as well as validation. I plan to examine other retroviral phenomena in other organisms as well as retroviral insertions as in the case of HIV and MLV. Such studies could uncover generalities in mechanisms of insertion of foreign DNA into genomes. Even consideration of the role of duplex destabilization in insertion/targeting mechanisms (in specific cases where there is a

significant correlation with duplex destabilization) could have important implications for the elucidation of retroviral integration as well as for gene therapy.

I am also initiating research to explore the DNA structural correlates of genetic diseases involving repetitive expansion of DNA. This is seen in several ataxias, diseases such as Huntington's disease and so on.

Evolution

Understanding the origins of genetic diversity and how it affects the behaviors that play a role in evolution are outstanding questions. Answering these questions entails a clear mechanistic understanding of how changes in sequence can lead to evolutionary changes. A novel, hitherto unexplored mechanism is that change in DNA sequence-dependant duplex destabilization in regulatory regions of genes causes specific alterations in gene expression (e.g. by facilitating / impeding polymerase binding or transcription factor binding). Such changes in expression of key genes underlie evolutionary/adaptive changes. A particular instance of this potentially general mechanism is outlined below.

Sexual and social behaviors are key to evolutionary changes leading to speciation. Variation in the levels of expression of the vasopressin hormone receptor has been shown to be important in social and sexual behaviors such as monogamy, social bonding and parental care in many species, including voles (*Microtus* sp.). Preliminary results from my collaborative work with Larry Young's group show that minor sequence variation in the upstream regulatory microsatellite regions of the vasopressin receptor gene causes significant changes in duplex destabilization. This duplex destabilization is remarkably correlated with gene expression, which in turn has been recently shown to underlie behavioral variation. This could explain the role of microsatellite regions in rapid evolution, in this case by regulating expression of genes involved in behaviors that play a significant role in evolution. My future work in this area will take two directions:

1. Experimental determination of the precise role of duplex destabilization. I will design and analyze sequences that will alter the levels of duplex destabilization at the regulatory microsatellite regions by altering the sequence a distance away from the microsatellite region. (This alteration a distance away can cause changes in duplex destabilization at the site of interest due to superhelical energetic coupling, which can be calculated with precision). These altered sequences will be used in 'knock in' experiments in mice.

2. Comparative analysis of the regulatory microsatellite regions of the vasopressin receptor gene across species. Results thus far reveal a high degree of similarity between humans and bonobos in the duplex destabilization patterns in the microsatellite regions upstream of the

vasopressin receptor gene. However, human and bonobo patterns both differ significantly from the chimpanzee duplex destabilization pattern at the microsatellite region. Efforts are currently underway to sequence the vasopressin receptor microsatellite regions of the next closest species, gorilla and orangutan. Analyses of these regions (in these and other primate species, initially) will enable precise identification of regions of the microsatellite regions involved and their possible roles.

By understanding the molecular basis of specific behaviors and by combining this with comparative analyses of the sort outlined above, one can search for conserved or converged mechanisms and processes of evolution. The high levels of polymorphism in microsatellites not only render them useful as markers, but since they also have a regulatory role, such mechanisms are likely to be specific instances of more general phenomena. I aim to search for and investigate other examples of such processes in both evolution and behavior.

In a second approach, my ongoing collaborative research with the New Zealand Plant Species Radiation Group indicates that mutational hotspots in chloroplasts and mitochondria seem to be localized at regions with specific patterns of duplex destabilization. Duplex destabilization can therefore be used as part of a rational approach to choose regions of a genome for further characterization in phylogenetic/phylogeographic (intraspecific) studies. This represents a significant improvement over the current practice, in which the decision about what regions to characterize for such studies is arbitrary.

I seek to explore this issue further by a combination of theoretical and experimental approaches. Such approaches include:

1. Computationally designing sequences with specific patterns of destabilization and experimentally measuring mutation rates at those locations.
2. Simulating substitution rates and measuring duplex destabilization values for those substitutions (to test the hypothesis that substitutions in particular regions induce further substitutions in those regions because they alter the structural stability of the region).
3. Making maximum likelihood estimates of branch lengths in evolutionary tree-building using duplex destabilization as a marker.

A third approach that I will take to answer questions of evolution is more long-term. I will explore taxon-specific signature patterns of destabilization. In work done thus far, I find that in yeast, the coding regions are very stable, and the promoter regions are destabilized. However, unexpectedly, the terminator regions are highly destabilized. This is not the case in *E.coli*. Preliminary work

indicates that in humans the terminator regions are destabilized, but the promoter regions are highly stable. However, the transcription factor binding sites tend to be destabilized. I plan to do genome-wide analysis in other species to examine and compare such patterns.

While these findings have obvious implications for drawing the *Tree of Life*, the more interesting question to me is how the specific patterns of destabilization might play a role in the functioning of these genomes that establishes a distinct taxon identity.

Finally, in the longer term, I also plan to take another direction distinct from the above. This approach will be highly theoretical initially, but will become increasingly empirical. All the calculations of destabilization thus far are under equilibrium conditions. While that holds considerable promise in its explanatory power and has already shown success, I aim to incorporate dynamics. I also plan to incorporate alteration of energetics by *in vivo* conditions, such as by protein binding. These additions will present a more accurate picture of strand separation in living systems.

In summary, the main thrust of my future research program is to understand the general principles underlying genomic function and organization. Currently I focus on one key organizational and functional feature of genomes, DNA duplex destabilization, its control and its precise role in regulation. To understand general principles that might exist, I plan to study the precise role of DNA duplex destabilization in specific situations, which are of necessity diverse and widespread.

References:

1. *Miller CA, Beaucage SL, Cohen SN.* Role of DNA superhelicity in partitioning of the pSC101 plasmid. *Cell. 1990 Jul 13;62(1):127-33.*

2. *Muskhelishvili G, Travers A.* Transcription factor as a topological homeostat. *Front Biosci. 2003 Jan 1;8:d279-85.*

3. *Travers A, Muskhelishvili G.* DNA supercoiling - a global transcriptional regulator for enterobacterial growth? *Nat Rev Microbiol. 2005 Feb;3(2):157-69.*

4. *Wang Z, Droge P.* Differential control of transcription-induced and overall DNA supercoiling by eukaryotic topoisomerases *in vitro. EMBO J. 1996 Feb 1;15(3):581-9.*

5. *Negri R, Costanzo G, Buttinelli M, Venditti S, Di Mauro E.* Effects of DNA topology in the interaction with histone octamers and DNA topoisomerase I. *Biophys Chem. 1994 May;50(1-2):169-81.*

6. *Parvin JD, Sharp PA.* DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell. 1993 May 7;73(3):533-40.*

7. *Gartenberg MR, Wang JC.* Positive supercoiling of DNA greatly diminishes mRNA synthesis in yeast. *Proc Natl Acad Sci U S A. 1992 Dec 1;89(23):11461-5.*

8. *Sheridan SD, Benham CJ, Hatfield GW.* Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activating sequence *J Biol Chem. 1998 Aug 14;273(33):21298-308.*

9. *Prashanth Ak and Craig Benham.* Susceptibility to superhelically Driven DNA Duplex Destabilization: A Highly Conserved Property of Yeast Replication Origins. *PLoS Comput Biol 1(1): 41-46.*

10. *Prashanth Ak and Craig Benham.* DNA Duplex Destabilization at Yeast Regulatory Regions: Genome-Wide Analysis Reveals A Highly Conserved Property. *Manuscript under review.*

11. *Martin Klar, Erik Stellamanns, Prashanth AK, Angela Gluch and Juergen Bode.* Dominant genomic structures.Detection and potential signal functions in the interferon beta domain *To appear in Gene 2005 doi:10.1016/j.gene.2005.07.023.*

12. *Prashanth Ak Suzanne Sandmeyer and Craig Benham.* Retrotransposon insertion in yeast occurs preferentially at duplex destabilized sites. *Manuscript submitted.*

13. *Yubin Kang, Todd E. Scheetz, Christopher J. Moressi, LitaoXie, Diane Thi Tran, Thomas L. Casavant, Prashanth Ak, Craig Benham Beverly L. Davidson, and Paul B. McCray Jr.* Integration Site Choice of a Feline Immunodeficiency Virus Vector in Somatic Cells *in vitro* and *in vivo. Manuscript under revision.*

14. Personal communication from Jack Greenblatt.

15. *P. Albrecht, J. Bode, K. Buiting, Prashanth AK, D. R. Lohmann.* Recurrent deletion of a region containing exon 24 of the RB1-gene caused by nonhomologous recombination between a LINE-1HS and MER21B element. *J. Med. Genet. 2004 Dec;41(12):e122.*

STATEMENT OF TEACHING INTERESTS

Prashanth Ak

The urge to communicate is fundamental. A desire to teach naturally follows from the urge to communicate - communicate not just the pleasures of science, but also the joys of research. I have always intended to teach and made teaching an integral part of my activities. With this in mind I have sharpened my teaching abilities and skill over the years, using every opportunity to teach, and also taking courses and workshops in teaching skills at UC Davis and NYU. In addition, I regularly strive to improve my teaching by seeking student feedback (both explicit and implicit), discussing with peers and experimenting with new methods. Through these evaluative processes I am continually refining my teaching practices. Teaching, to me, is a privileged position that demands humility. I find it to be a reciprocally educative endeavor, informative and uplifting for teachers and students alike.

While it may not be possible to teach a student how to get good ideas, I feel that students can certainly be helped to tune themselves for inspiration. I consider this to be the first of my tasks as a teacher: to help the students learn how to look and think critically. My omnivorous intellectual curiosity stands me in good stead at this task. However, even more importantly, I would like to aid the students (whether future scientist of not) to move beyond acquisition of knowledge and technical expertise. I would like to encourage ideas and critical thinking in whatever forum is appropriate. I make it a priority to help students learn articulation of ideas, collaborating with technical experts and others with complementary proficiencies as needed. I encourage students to move fluidly from one discipline to another and to seamlessly integrate the transitions.

Specifically in teaching or training young scientists, I feel that a critical skill the aspiring scientist needs to acquire is to learn to transform their ideas or inspiration into distinct tangible/tractable research questions. The most 'teachable' of my tasks as a teacher is probably the enabling of the tools to allow the students to do so. This is my most important role as a teacher of future scientists. In this process, technical prowess is of undoubted importance to the aspiring scientist, and I would try to inculcate in the students a disciplined approach to acquiring technical proficiency.

I do not see a rigid dividing line between my research work and teaching, especially at the graduate level. Scientific research, of course, is not a collection of facts, but rather an endeavor alive with puzzles, contradictions, creative compromises and new directions. I try to "demystify" the process for students, by encouraging them to discover the excitement that can be found in exploring. To me, teaching is not about lecturing to students; it is about empowering them with tools to enable them to understand and integrate this information into a larger context. I try to accomplish this not only in my presentations and lectures, but also in other dialogues and interactions with students. I take Plutarch's quote to heart: "The mind is not a vessel to be filled, but a fire to be kindled." In my teaching, I encourage students to come up with ideas and questions. When they do, instead of answering their questions, I gently counter with further questions for them to answer - the process of which leads them to answer their own questions. These have been some of the best moments of my teaching experience.