10-15-05

Leonard Duncan
Cumbre, Inc.
1502 Viceroy Dr.
Dallas, TX 75235

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

Dear Dr. Brun,

Please consider my application for the position of Assistant Professor in the Department of Biology and the Biocomplexity Institute. I have included a copy of my *curriculum vitae*, a summary of my current and future research activities, a description of my teaching philosophy/interests and copies of relevant papers on which I am an author.

I received my Ph.D. from Harvard University (where I uncovered the biochemical mechanisms that control the cell-type specific activation of σ^F in *B. subtilis*) and, as a Jane Coffin Childs Post-doctoral fellow at Washington University in St. Louis, I studied the development and molecular genetics of the green alga, *Volvox*. I then accepted a position as a Research Scientist at Cumbre, Inc., which has enabled me to expand my repertoire of experimental techniques, manage other researchers, and learn to write a successful research grant. Throughout this time, I have continued my *Volvox* research, as demonstrated by my recently submitted manuscript that reports on my discovery of a previously unknown family of putative transcription factors (the GARL domain proteins) that are present within the volvocalean green algae. The volvocaleans form an unexploited and unparalleled group of organisms for studying fundamental mechanisms of cell-fate determination and the evolution of multicellularity and terminal differentiation. These algae range in complexity from unicellular forms like *Chlamydomonas* to large, multicellular organisms like *Volvox* that exhibit a complex embryogenesis that leads to the generation of fully differentiated cell types. My research program will employ biochemical, molecular genetic and phylogenetic approaches toward two primary goals: 1) understanding the biochemical function of the GARL-domain proteins and how they control cellular differentiation in *Volvox*; and 2) determining how the evolution of these GARL proteins contributed to the independent origins of terminally differentiated cell types in several species of volvocine algae. Importantly, the genome sequences of *V. carteri* and its unicellular relative *C. reinhardtii* are now available for comparative analysis.

I am eager to uncover the biochemical, developmental and evolutionary secrets that this novel family of proteins is ready to reveal.

Thank you for your consideration of my application.

Best wishes,

Leonard Duncan

RESEARCH INTERESTS

The volvocales are a group of green algae that represent an unexploited and unparalleled opportunity to investigate fundamental mechanisms of cell-fate determination and the evolution of multicellularity and terminal differentiation. These algae range in complexity from unicellular forms like *Chlamydomonas*, to colonial forms with only one cell type, to large, multicellular organisms like *Volvox* that exhibit a complex embryogenesis that leads to the generation of fully differentiated cell types. My research program will employ biochemical, molecular genetic, and phylogenetic approaches toward two primary goals: 1) understanding how a novel family of transcription factors (the GARL-domain proteins, which I recently discovered) control cellular differentiation in *Volvox*, and 2) determining how the evolution of these GARL proteins contributed to the origins of cellular differentiation in the volvocine algae.

BACKGROUND

A *Volvox carteri* spheroid (≈individual) is composed of just two differentiated cell types: several thousand small, biflagellate, "*Chlamydomonas*-like" somatic cells embedded in the surface of a sphere of extracellular matrix, and ~16 large, asexual reproductive cells called gonidia that lie just beneath the somatic cell layer (Fig. 1A). Whereas the somatic cells are specialized for providing motility, grow very little and ultimately undergo programmed death, the gonidia are non-motile, specialized for growth and reproduction, and potentially immortal. When mature, each gonidium initiates a stereotyped set of embryonic cell divisions that produce all of the cells of both types that will be present in an adult of the next generation.

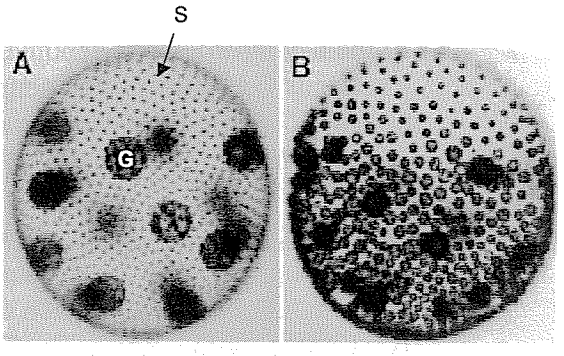

Fig. 1. Wild-type (A) and Reg (B) spheroids. G=gonidium. S=somatic cell. Note regenerating somatic cells in (B). From Kirk *et al.* (1999) Development 126; 639-47.

***regA*, master regulator of somatic cell development:** The *regA* locus, which will be the main focus of my initial research, is of central importance in controlling the normal development of *V. carteri*. In *regA*$^-$ spheroids, the somatic cells initially differentiate normally, but then de-differentiate, grow, and redifferentiate as gonidia that divide a day later to produce offspring with the same phenotype (Fig. 1B). *Thus, regA is required to maintain the terminally differentiated state of V. carteri somatic cells.*

Analysis of the *regA* gene revealed that it is only expressed in somatic cells and that its product, RegA, is a large, nuclear protein. Analyses of several genes that are inappropriately expressed in *regA*$^-$ somatic cells revealed that they all encode important chloroplast proteins. Together these findings led to the hypotheses that: (1) RegA is a transcriptional repressor that specifically inhibits expression of the photosynthetic regulon in somatic cells, and (2) because *V. carteri* is an obligate photoautotroph, repression of chloroplast biogenesis would prevent growth of the somatic cells.

Although attractive, this hypothesis could not be tested for some time, because RegA did not bear significant resemblance to any other proteins in the databases. However, I reasoned that the analysis of *regA* orthologs from multicellular algae closely related to *V. carteri* f. *nagariensis* (the "subspecies" of *V. carteri* that I study) would permit identification of conserved protein regions that are important for RegA function

(and that could then be subjected to further investigation). I therefore cloned and characterized the *regA* gene from the most closely related known taxon, *V. carteri* f. *kawasakiensis*. Interestingly, despite the fact that f. *kawasakiensis* and f. *nagariensis* are indistinguishable in the asexual phase, I found that their corresponding *regA* genes and proteins are surprisingly different and that *regA* may be a "fast-evolving" gene.

The most galvanizing finding of these studies was the discovery of a paralog of *regA* that I call *rlsA* (for *regA*-like sequence), and that is highly conserved and expressed in f. *kawasakiensis* and f. *nagariensis*. This conservation is limited to an ~100 amino acid region that I have termed a GARL domain (green algal *regA*-like). Several findings suggest that the GARL domain is of unusual functional significance. First, the GARL domain is the most conserved region of the RegA proteins. Second, protein motif searches revealed that the GARL domain might be distantly related to a DNA-binding domain called a SAND domain. Third, as described in the next section, BLAST searches of the *Chlamydomonas* genome and the recently released *V. carteri* genome revealed that numerous GARL-encoding genes are present in both genomes. *Thus, I hypothesize that the GARL domain represents a novel DNA binding motif and that I have identified a family of transcription factors that is likely to be widespread within the volvocine algae.*

The GARL-protein family: The *V. carteri* f. *nagariensis* genome contains four predicted genes (*rlsA-D*) that encode GARL domains closely related to the RegA GARL domain. *C. reinhardtii*, which diverged from *V. carteri* ~50 MYA, also encodes a protein (C_170011) with a RegA-like GARL domain.[1] Most strikingly, *C_170011*, *regA* and the four *rls* genes also contain a homologously positioned intron within the GARL-encoding region. These six proteins are dissimilar in primary amino acid sequence outside of their respective GARL domains; thus, with the exception of their GARL domains, these proteins are "fast-evolving."

PHYLOGENETIC ORIGINS OF RegA FUNCTION: The developmental behavior of somatic cells in a *regA*$^-$ mutant parallels the development of cells in several closely related volvocine algae (such as *Chlamydomonas*) that have only one cell type. In these species, each cell alternates between a biflagellate, motile phase and a non-motile, reproductive phase. My working hypothesis is that the *regA* function is one of a handful of key ontogenic innovations that were added during volvocine evolution to convert the ancestral "first biflagellate, and then reproductive" pathway into a dichotomous germ–soma pathway of development.[2]

Interestingly, comparison of the *V. carteri* and *C. reinhardtii* genomes reveals that the *C_170011* (Cr) and *rlsD* (Vc) genes are each next to, and divergently transcribed from, the acetate kinase gene (Fig. 2A). Moreover, the *rlsA-C* and *regA* genes are found in a tandem array elsewhere in the *V. carteri* genome and no such tandem array exists within the *C. reinhardtii* genome (Fig. 2A). Phylogenetic analysis[3] of the GARL domains (Fig. 2B and data not shown) suggests that C_170011 is most closely related to RlsD. Based on these findings, it is attractive to hypothesize that: (1) *C_170011* and *rlsD* are orthologs and (2) a duplication of the ancestral *rlsD* gene, transposition of one copy

[1] Several more distantly related GARL proteins are also encoded by both genomes.
[2] Other genes that I do not discuss here account for the development of gonidia that bypass the "first biflagellate" half of the ancestral pathway, and these genes are attractive prospects for later-stage projects.
[3] L. Duncan, S. Miller & I. Nishii, unpublished

and subsequent further duplications led to the generation of the *rlsA-C/regA* gene array.[4] Thus, *regA* may ultimately have been derived from the ancestor of *C_170011* and *rlsD*. Testing this hypothesis will require analysis of *regA/rlsA-D/C_170011* gene function and the identification of more distantly related GARL homologs for phylogenetic analysis. These experiments are described below.

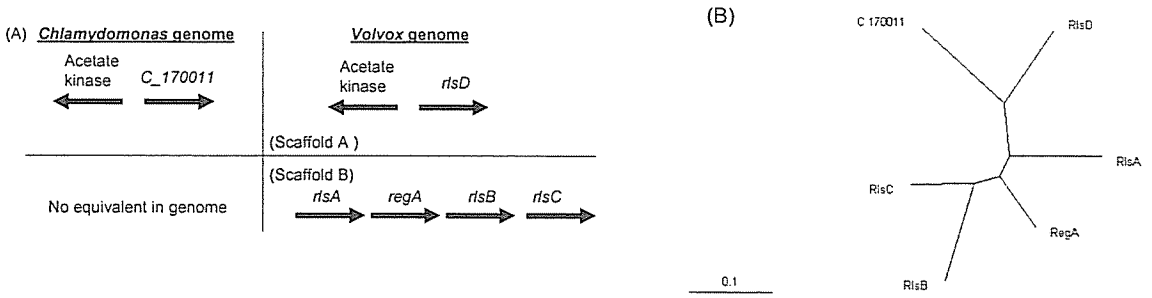

Fig. 2. (A) Genomic organization of several GARL-encoding genes in *V. carteri* and *C. reinhardtii*. (B) Neighbor-Joining tree showing the relationships between the GARL domains in six *Volvox* and *Chlamydomonas* proteins.

RESEARCH AREAS

A wide variety of tools are available for the study of *Volvox* development and cell biology. Most importantly, the genomic sequences of *V. carteri* f. *nagariensis* and *C. reinhardtii* are now available for comparative studies. An impressive array of molecular genetic tools has been developed for *Volvox* research, including: strain repositories, BAC and cDNA libraries, transposon tagging tools, generation of stable transformants by biolistic methods, reporter genes and positive and negative selectable markers. Finally, although *Volvox* normally grows asexually as a haploid organism, a sexual phase can be induced at will. Thus, it is possible to utilize Mendelian analysis under some circumstances.

1. RegA FUNCTION AND THE GARL DOMAIN: *My lab's first primary goal will be to test the hypothesis that RegA and other GARL proteins are transcriptional factors and that the GARL domain is a novel DNA binding motif.*

a. Investigating GARL-domain function. We will test the hypothesis that the RegA GARL domain constitutes a DNA-binding motif by using standard methods, such as nitrocellulose filter assays, EMSA (using promoter regions from photosynthetic genes as putative binding targets) and chromatin immunoprecipitation (ChIP). If our hypothesis is correct, we will next determine the specific DNA-binding site for RegA. Ultimately, we will undertake similar studies for the remaining GARL proteins.

b. **Molecular genetic analysis of *rlsA-D* and *C_170011*.** To gain important clues about gene function, we will determine the temporal and spatial expression patterns of *rlsA-D (Volvox)* and *C_170011 (C. reinhardtii)*. For instance, a correlation of *C_170011* expression with photosynthetic activity in *C. reinhardtii* (which can grow heterotrophically) would suggest that this gene is involved in controlling the

[4] Other hypotheses—e.g., involving loss of a homolog from *Chlamydomonas*—are also possible and will be tested.

photosynthetic regulon. It will also be of great interest to determine if any of the *rlsA-D* genes are expressed in a developmentally and cell-type specific manner analogous to *regA*.

b. *In vivo* analysis of GARL domain function. First, we will use site-directed mutagenesis to determine if the GARL domain is *required* for RegA function. In a second experiment, we will determine whether over-expressing the RegA GARL-domain is sufficient to confer a dominant negative phenotype. Similar experiments will evaluate the effects of over-expressing both full-length and truncated versions of *rlsA-D* (in *Volvox*) and *C_170011* and *regA (*in *Chlamydomonas);* the phenotypes of such strains are likely to provide clues about the functions of these genes. Third, we will use gene-silencing techniques to evaluate the "knock-down" phenotype of *C_170011* in *C. reinhardtii* and *rlsA-D* in *V. carteri.*

c. GARL domain structure. As an important side-project, we will collaborate with a protein NMR or crystallography laboratory to determine the structure of the GARL domain.

2. THE PHYLOGENETIC ORIGINS OF THE GARL DOMAIN AND TERMINAL DIFFERENTIATION: Comparative genomics suggests that a subfamily of the GARL-encoding genes has expanded in *Volvox* since it last shared a common ancestor with *Chlamydomonas*, and that one of these newly-formed genes—*regA*—has assumed a unique developmental role. *My second primary goal is to determine the phylogenetic origins of regA function and how GARL domain proteins contributed to the evolution of terminally differentiated cells types in the volvocales.*

a. Functional analysis of *rlsA-D* with respect to the Reg phenotype. Despite the fact that RegA and RlsA-D all contain a similar GARL domain, genetic evidence suggests that *regA* is the only *V. carteri* gene that can mutate to produce a Reg phenotype. Is this because *regA* developmental expression or RegA protein function is somehow unique when compared to *rlsA-D*? To begin to answer this question, we will carry out rescue experiments and promoter and/or domain swapping experiments (guided by the expression and functional analyses of *rlsA-D* described above) to investigate potential functional similarities between *regA* and *rlsA-D*. Based on these results, we will be able to make inferences about the evolution of *regA* and its role in cellular differentiation.

b. Reconstructing the evolution of the GARL protein family. The discovery of the GARL protein family immediately raises several questions. For example, how ancient is the GARL domain? When was the *C_170011/rlsD* progenitor gene duplicated? Was a RegA-like GARL protein utilized in each of the several independent instances of the evolution of terminally differentiated cells in the volvocaleans? To answer these and related questions, my lab will: (1) clone the *regA* orthologs from two other closely-related *V. carteri* subspecies to calculate K_a/K_s values; (2) analyze all of the GARL-domain protein sequences encoded in the *V. carteri* and *C. reinhardtii* genomes; (3) analyze the GARL-encoding genes in other, select volvocalean taxa to determine when the *rlsA/regA/rlsB/rlsC* gene cluster was created; (4) analyze GARL-encoding genes in green algal taxa outside of the volvocales; and (5) use these data to construct robust trees that describe the phylogenetic history of the GARL proteins and how this history relates to the evolution of terminal differentiation.

LONGER TERM RESEARCH PROJECTS

1. THE CELL-SIZE REGULATED CONTROL OF *regA* EXPRESSION: The somatic cell-specific expression of *regA* is controlled, at least in part, by two activating enhancer elements in introns 3 and 5, plus a silencer element in intron 7 of *regA*. Importantly, cell fate is determined by cell size at the end of embryogenesis in *Volvox*: large cells always become gonidia, and small cells always differentiate as somatic cells. How the cells measure their size is unknown, but it seems likely that the ultimate output of the "large cell" signal is activation of the *regA* intron 7 silencer element. We will use a variety of methods (e.g., yeast one hybrid screens, cDNA library screening with labeled DNA probes and EMSAs with fractionated gonidial cell extracts) to identify trans-acting factors that bind to introns 3, 5 and 7. This will help us understand how *regA* expression is regulated, and more importantly, will provide an experimental pathway for my laboratory to follow back to the cell size measuring and signaling machinery.

2. **A NEWLY RECOGNIZED THIRD CLASS OF RETROTRANSPOSONS.** I discovered a novel, active retrotransposon called *kangaroo*[5] that is present in numerous copies in the *Volvox* genome. *Kangaroo* apparently integrates into the genome as a closed-circle, double-stranded DNA form in a reaction that is catalyzed by a recombinase related to λ integrase. Several other unorthodox transposable elements (including several vertebrate elements) that were previously difficult to classify also share these features with *kangaroo*. Thus, this group constitutes a third class of retrotransposons (the DIRS1 class) that is distinct from the traditional LTR and non-LTR groups of transposable elements. However, very little is known about the replication of the DIRS1 retrotransposons, and an entire grant could be written around experiments to further study the biochemistry and molecular biology of *kangaroo* (and related retrotransposons in *C. reinhardtii*).

[5] Duncan et al. (2002) Genetics 162:1617-1630

Teaching Philosophy and Interests

My teaching philosophy has been influenced by three major factors:

(1) *My father*, who has been a professor of mathematics for 40 years. I have learned from him that excellence in teaching requires a passion and that it is a tremendous amount of work, even for someone as gifted as he.

(2) *My experiences as an undergraduate at the University of Cambridge.* Much of my coursework involved researching particular topics, writing weekly essays (e.g., "Discuss the mechanisms used by enzymes to accelerate chemical reactions.") and asking and answering questions about this topic in small-group tutorials. This style of teaching leads to a much deeper level of understanding than a standard lecture course that involves little discussion between the instructor and students. Whenever possible, I will emulate this model of teaching. However, I fully realize that large courses, where this method is less practical, are also a necessity and require different teaching methods to fully engage the students. One simple tool that I will use in the large class setting is a very short (and easily graded) quiz before each lecture. It encourages attendance and continuously focuses the students' minds on the material being covered.

(3) *My own teaching experiences.* As a graduate student, I taught two semesters of Introduction to Biology laboratory. This involved preparing and delivering lectures, holding office hours, writing/grading quizzes and ensuring that the experiments (which ranged from yeast genetic crosses to Southern blots) went smoothly. As a post-doc, I guided the senior-year research of two undergraduates and the work of our laboratory technician. I have also trained three research assistants during my time as a Scientist at Cumbre, Inc. In each of these cases, my goal has been to teach my students to think critically about the particular problem at hand, using Socratic methods whenever possible. The most important piece of information that I learned from these experiences is that I *do* have a passion for teaching. I enjoy lecturing, and I love the expression on a student's face when my explanations lead them to comprehend a difficult topic.

My Teaching Interests:

My unique background will allow me to teach courses in molecular biology, biochemistry, genetics, developmental biology, cell biology and microbiology. I would be pleased to accept teaching assignments that range from large undergraduate-level courses to small graduate-level classes.

I will make all relevant course material available on the campus intranet, and I will supplement the primary textbook(s) with appropriate web-based materials that are designed to enhance understanding of difficult topics (e.g., with animation). Whenever it would significantly aid learning, I will design "practical" assignments for my students. As one example, if I were teaching a unit on protein structure in an introductory biochemistry class, my hypothetical assignment might consist of: (1) locating a particular protein sequence and structure using Genbank; (2) analyzing the protein's 2° and 3° structure and (3) locating amino acids with critical functions within the 1° and 3° structure.

Finally, I would like to develop an upper level undergraduate course that explores the history, methodology and consequences of some of the most important experiments in modern molecular genetics (e.g., the discovery of the triplet code by Crick and colleagues). I would also develop an intensive, literature-based seminar called "Experimental Design" where graduate students and upper-level undergraduates would discuss the merits and flaws of one or two assigned papers per session. Through discussions in which each student would be asked to speak, we would explore such questions as: Why were the experiments done? What questions were being asked? Were the methods and controls appropriate? What other methods and controls could/should have been used? What are the conclusions? Are they supported by the data? And are they important?