THE
SCRIPPS
RESEARCH
INSTITUTE

10550 North Torrey Pines Road
La Jolla, California 92037

1 November 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 Professor Brun,

I am writing to apply for an Assistant Professor position in Systems Biology and Microbiology, as advertised in Nature. I am currently a Post-doctoral Fellow at the Scripps Research Institute in La Jolla, California. I was briefly at IU when I began my tenure in Andy Ellington's lab as a graduate student. During that time, I met a number of the faculty, and was thoroughly impressed with the Biology Department.

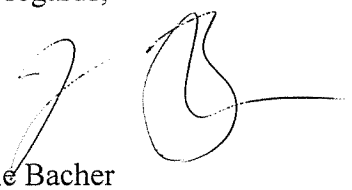
My background includes evolutionary work on the origin and early evolution of the genetic code. This is an exciting, new field of study, in which the genetic code is targeted for directed evolution. By studying the evolution of the genetic code, I hope to better our understanding of the origins of life. As you'll see in my research proposal, I intend to further explore the effects of genetic code ambiguity by better understanding the biochemical and evolutionary responses to perturbations of the translation system, the heart of genetic decoding. Applications of this work will target systems of small groups of highly interdependent proteins, ultimately focusing on viruses and disease states. Models for these areas of investigation have already been worked out, and initial work has begun. In particular, this work will support your stated interests as being applicable to mechanisms of bacterial cell function and biomolecular networks.

As a researcher with a broad background ranging from evolutionary biology through microbial genetics, I could bring to the department an enthusiasm for evolution that readily encompasses other areas. My teaching perspective would strongly place molecular biology and biochemistry within the framework of evolution. Along with my experience at top-ranked academic institutes, I have industrial experience as well. This will enable me to train researchers as highly qualified personnel, as well as the drive to expand my own research in molecular evolution to practical application. Finally, I am very collaboration-oriented, including being a member of the Microbial Comparative Genome Analysis Consortium (headed by Peg Riley at U-Mass Amherst). This group

will analyze and mine the genome sequence of *Burkholderia phymatum* STM815, part of a biochemically and ecologically diverse group of organisms.

Please find enclosed my CV, a research proposal, a teaching statement and three reprints. My collaborators have written letters attesting to our collaborations, and I have included those as well. My references will send letters under separate cover. Indiana University has an exciting, dynamic faculty, with a breadth of experience that I would be honored to join. I look forward to the opportunity, and to your kind consideration.

Best regards,



Jamie Bacher

The Scripps Research Institute
10550 N. Torrey Pines Rd.
La Jolla, CA 92037

858-784-8974 (phone)
858-784-8990 (fax)
jbacher@scripps.edu

GENETIC CODE AMBIGUITY: DISRUPTION AND ADAPTATION OF NATURAL SYSTEMS

Overview

Cellular systems exist in a delicate balance, with small changes often resulting in unpredictably large effects. One form that this equilibrium takes is in the use and distribution of amino acids in protein structure. Although conservative mutations often have predictably small results, unpredictable effects occur as well. A more subtle version of a genetic mutation would involve the broad incorporation of amino acid analogues. This proposal will detail investigating the chemical genetic and evolutionary implications of such an "ambiguous genetic code." Evidence exists that cellular subsystems can be specifically targeted by amino acid misincorporation. I propose to better understand how such a genetic code ambiguity can have specific effects by

1) Disruption of specific systems,
2) Characterization of strains with ambiguous genetic codes, and
3) Evolving new genetic codes.

Overall the goal will be to understand how small, broadly applied chemical perturbations result in critical disruptions of specific molecular systems (including viruses and cancerous states), as well as how cellular systems can adapt genetically to such perturbations.

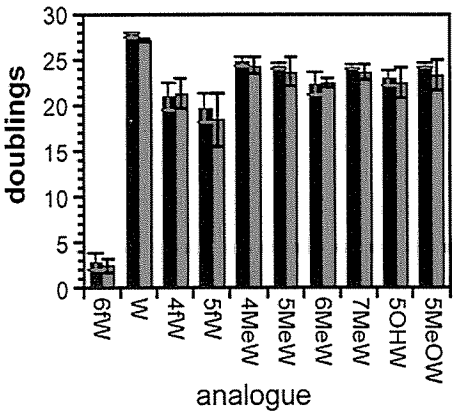


Figure 1: Effect of tryptophan analogues on fitness of bacteriophage Qβ. Tryptophan auxotrophic bacterial hosts of Qβ were grown on media in which 95% of the available tryptophan was the analogue indicated. These conditions result in ~50% incorporation of 4-fluorotryptophan (4fW), and ~65% incorporation of 6-fluorotryptophan (6fW). The number of doublings of two clones of Qβ was determined after 20 hours. Only 6fW was found to significantly diminish the fitness of the phage. Shown are mean and standard deviations of triplicate measures of fitnesses of two clones (red and black).

Disruption of biological systems with amino acid analogues

Subtle perturbations of biological systems can have catastrophic results. Incorporation of amino acid analogues into proteins generally has small effects, and yet it has been shown that subsystems within a cell can be completely disrupted at sublethal levels of misincorporation. For example, it has been shown that *Bacillus subtilis* has been adapted to incorporation of the tryptophan analogue 4-fluorotryptophan (4fW) in rich, defined media[1]. However, if the adapted strains are denied access to methionine or tyrosine, they are incapable of growth (Wong, pers. comm.) This indicates that the subsystems required for biosynthesis of these amino acids is fatally disrupted when 4fW replaces W. Furthermore, in a panel of nine tryptophan analogues, I determined that only 6-fluorotryptophan (6fW) had a significant effect on the fitness of the bacteriophage Qβ under conditions of ~65% incorporation of the analogue[2] (**Fig. 1**). At the same level of incorporation, 6fW had minimal effects on the *E. coli* host[3]. This suggests that a broad

search for amino acid analogues may reveal specific inhibitors of cellular function and viral activity. A broad range of systems is available for targeting. Initial experiments will test an array of amino acid analogues and bacteriophage in order to better understand the frequency with which amino acid analogues cause specific and targeted viral inhibition. I propose to determine whether amino acid analogues are feasible antiviral agents, as has been suggested[4]. A number of additional biological systems can also be targeted, including biosynthetic machinery and synthetic circuits. The ultimate goal will be to target specific disease states in humans; however, a broad knowledge base is required first.

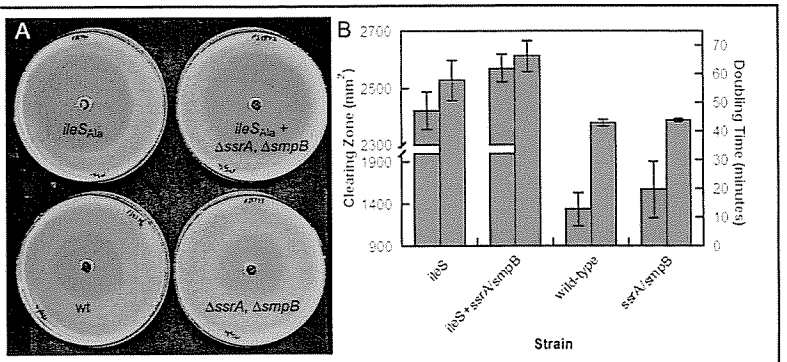


Figure 2: Effect of combining editing-deficiency- and trans-translation-mutations. The $ileS_{Ala}$ editing deficient mutation was combined with Δ*ssrA-smpB* mutations. Sensitivity to norvaline (an isoleucine analogue) was determined by halo assay (**A** and **B**, blue) in which the area of the clearing zone correlates to the sensitivity to the analogue. Sensitivity was also determined by measuring the doubling times of the strains in the presence of 5 μM norvaline (**B**, red). Both assays reveal a significant difference between $ileS_{Ala}$ and $ileS_{Ala}$+Δ*ssrA-smpB* ($P < 0.05$, two-tailed t-test), but no significant differences between wild-type and Δ*ssrA-smpB*. This reveals a synergistic effect of combining the two mutations. Shown are means and standard deviations of measures of clearing zones for three clones and triplicate measures of doubling times for each of three clones.

Chemical genetics and genetic code ambiguity

While it has been shown that genetic code ambiguity can be beneficial to the growth of organisms under a small set of specific conditions (ref [5] and see below), it is more generally detrimental[6]. This has often been assumed to be broadly due to protein misfolding. However, it is also likely that cellular subsystems have been disrupted, to the detriment of the cell. We have introduced translational ambiguity into *E. coli* with the ileSAla allele, in which a number of residues, critical for the editing function of the isoleucyl-tRNA synthetase, have been mutated to alanine in the chromosomal copy of *ileS*[5]. For example, it has been shown that EF-Tu preferentially binds to tRNA charged with cognate amino acid as compared with misacylated tRNA[7]. This may lead to slowed translation under conditions of mischarging. Trans-translation with tmRNAs is a mechanism for freeing stalled ribosomes[8]. Bacteria lacking trans-translation are more sensitive to specific targeting of translational ambiguity for isoleucine with the isoleucine analogue norvaline (**Fig. 2**), indicating that trans-translation may ameliorate the effects of genetic code ambiguity.

A number of additional hypotheses regarding the specific effects of translational ambiguity can be proposed. For example, heat shock proteins are known to assist in the folding of proteins. Under conditions of genetic code ambiguity, it is possible that folding pathways are disrupted. In collaboration with Jean-Hervé Alix at the CNRS in Paris, France, I am currently examining strains that combine the $ileS_{Ala}$ mutation with a

temperature sensitive variant of *dnaK*, which is normally active in post-translational quality control[9]. Similarly, long-term evolution experiments adapting *E. coli* to minimal media conditions resulted in mutations in *spoT*[10] (among other mutations). This gene is associated with the stringent response, which signals starvation[11]. CgtA associates with SpoT at the ribosome[12], while the mutations found in *spoT* diminished these interactions. It is hypothesized that this may in turn result in an increased rate of translation at a cost of accuracy. If true, then combining the *spoT* mutations with the $ileS_{Ala}$ mutation may result in an increased level of translational ambiguity, to the detriment of the cells. This work is being undertaken in collaboration with Dominique Schneider at Université Joseph Fourier at Grenoble, France.

Additional systems for targeting will be identified by screens in which strains carrying the $ileS_{Ala}$ allele recover insensitivity to temperature. The chemical genetics of these mutant strains will allow a deeper understanding of the effects of genetic code ambiguity on the cellular subsystems.

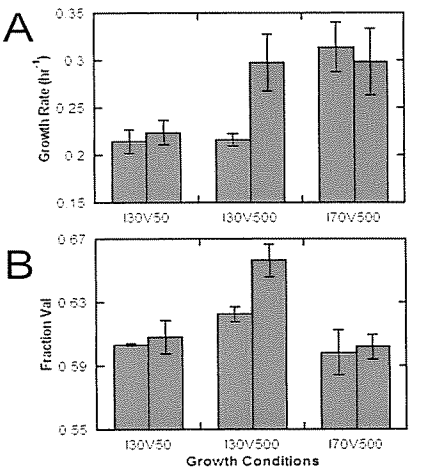


Figure 3: Relative concentrations of Ile and Val affect the growth of *A. baylyi* strains. Error-prone (blue) and wild-type (red) strains were grown under conditions of limiting Ile and Val (I30V50), excess Val (I30V500) and excess Ile and Val (I70V500). **A)** The error-prone strain was found to have a growth rate advantage in the presence of excess Val. **B)** Amino acid analysis of whole-cell protein extracts showed that excess Val resulted in an increased amount of Val relative to total Val + Ile in the error-prone strain. Shown are mean and standard deviations of four strains.

Evolving new genetic codes

The coevolutionary theory (CET) of the genetic code states that the genetic code expanded in concert with biochemical pathways to supply amino acids for protein synthesis[13,14]. A specific prediction of the CET is that the genetic code remains evolvable despite >3 billion years of stasis[1,15,16]. In particular, isoleucine is a practical target for genetic code evolution: it is considered primitive but also "new"[16]. An *A. baylyi* strain carrying the $ileS_{Ala}$ allele has been constructed. This strain achieves a growth rate advantage in the presence of excess valine, along with an increase in the relative incorporation levels of valine and isoleucine (**Fig. 3**). These conditions provide avenues for evolution to replace isoleucine (separately) with valine and norvaline in this strain. The major goal of this work will be to generate organisms with substantially altered genetic codes. This is an ambitious goal; however, my experience uniquely positions me for success[2,3,6,17]. Intermediate successes are likely, and may be considered evolutionary intermediates of the genetic code. A number of additional approaches are possible, including the use of *Deinococcus radiodurans* TrpRS II, which is especially tolerant of substitutions at the 4 position of tryptophan[18]. Generating a pathway for the directed evolution of the genetic code will afford a deeper understanding of what might have occurred in evolutionary history. Finally, as sequencing bacterial genomes cheapens (fast approaching $1,000[19]), studies of evolved strains will invoke hypotheses concerning the evolution of the genetic code.

References:

1. Wong, J.T. Membership mutation of the genetic code: loss of fitness by tryptophan. *Proc Natl Acad Sci U S A* **80**, 6303-6 (1983).
2. Bacher, J.M., Bull, J.J. & Ellington, A.D. Evolution of phage with chemically ambiguous proteomes. *BMC Evolutionary Biol* **3**, 24 (2003).
3. Bacher, J.M. & Ellington, A.D. Selection and Characterization of Escherichia coli Variants Capable of Growth on an Otherwise Toxic Tryptophan Analogue. *J Bacteriol* **183**, 5414-25 (2001).
4. Tai, J., Cheung, S., Chan, E. & Hasman, D. In vitro culture studies of Sutherlandia frutescens on human tumor cell lines. *J Ethnopharmacol* **93**, 9-19 (2004).
5. Pezo, V. et al. Artificially ambiguous genetic code confers growth yield advantage. *Proc Natl Acad Sci U S A* **101**, 8593-7 (2004).
6. Bacher, J.M., de Crecy-Lagard, V. & Schimmel, P.R. Inhibited cell growth and protein functional changes from an editing-defective tRNA synthetase. *Proc Natl Acad Sci U S A* **102**, 1697-701 (2005).
7. Dale, T., Sanderson, L.E. & Uhlenbeck, O.C. The affinity of elongation factor Tu for an aminoacyl-tRNA is modulated by the esterified amino acid. *Biochemistry* **43**, 6159-66 (2004).
8. Haebel, P.W., Gutmann, S. & Ban, N. Dial tm for rescue: tmRNA engages ribosomes stalled on defective mRNAs. *Curr Opin Struct Biol* **14**, 58-65 (2004).
9. Alix, J.-H. The Work of Chaperones. in *Protein synthesis and ribosome structure* (eds. Nierhaus, K.H. & Wilson, D.N.) 529-554 (Wiley-VCH, 2004).
10. Cooper, T.F., Rozen, D.E. & Lenski, R.E. Parallel changes in gene expression after 20,000 generations of evolution in Escherichiacoli. *Proc Natl Acad Sci U S A* **100**, 1072-7 (2003).
11. Chatterji, D. & Ojha, A.K. Revisiting the stringent response, ppGpp and starvation signaling. *Curr Opin Microbiol* **4**, 160-5 (2001).
12. Wout, P. et al. The Escherichia coli GTPase CgtAE cofractionates with the 50S ribosomal subunit and interacts with SpoT, a ppGpp synthetase/hydrolase. *J Bacteriol* **186**, 5249-57 (2004).
13. Wong, J.T. A co-evolution theory of the genetic code. *Proc Natl Acad Sci U S A* **72**, 1909-12 (1975).
14. Wong, J.T. The evolution of a universal genetic code. *Proc Natl Acad Sci U S A* **73**, 2336-40 (1976).
15. Wong, J.T. Evolution of the genetic code. *Microbiol Sci* **5**, 174-81 (1988).
16. Wong, J.T. Coevolution theory of the genetic code at age thirty. *Bioessays* **27**, 416-25 (2005).
17. Bacher, J.M., Hughes, R.A., Tze-Fei Wong, J. & Ellington, A.D. Evolving new genetic codes. *Trends in Ecology & Evolution* **19**, 69-75 (2004).
18. Buddha, M.R., Keery, K.M. & Crane, B.R. An unusual tryptophanyl tRNA synthetase interacts with nitric oxide synthase in Deinococcus radiodurans. *Proc Natl Acad Sci U S A* **101**, 15881-6 (2004).
19. Marusina, K. Whole genome sequencing in 24 hours. *Gen Eng News* **25**, 26-27 (2005).

Teaching Statement

In order to teach students to be scientists, it is crucial to convey the importance of critical and independent thinking. I believe that the most effective method for teaching is when the students are engaged in discovery. My strategy for teaching can be summed up as a teaching by doing. In the classroom, this will involve not only lecturing, but also a more extensive attempt to involve the students in a dialogue. I will strongly encourage questions from the students. After presenting ideas and information, follow-up will include questions about these ideas that should hopefully force the students to follow the logic of the ideas and arguments (ie, using the Socratic method of teaching). For example, after presenting an evolutionary principle, questions will oblige the students to discuss how the principle might apply in particular cases. Similarly, after presenting a biochemical pathway, follow-up might involve questions regarding mechanisms for the regulation and evolution of the pathway. The use of an interactive format will engage the students in a way that traditional lecturing does not. Furthermore, I intend to teach from the perspective of the centrality of evolution. An attempt will be made to explain not only the facts of biology in the classroom, but also to support those facts with the concept that those characteristics arose through evolution. This will provide an important continuity to ideas presented in the classroom. Finally, I believe that delivery is as important as content. One of the most vital aspects of higher education is that students are in the process of deciding on courses of study. I am very effective at conveying my own enthusiasm for science and for research in general. I believe that this will be critical in capturing and maintaining the interest of the students.

Owing to my broad interests, ranging from evolutionary biology, microbiology, genetics, molecular biology and biochemistry, I expect to be capable of teaching a diverse array of introductory courses, such as Molecular Biology and Microbiology. Furthermore, in laboratories, I expect to involve a number of techniques that are my strengths, including techniques that I assisted in developing. Specifically, I published as part of a group that described using *Acinetobacter baylyi* strain ADP1 for genetic manipulation[1]. This strain was used in a related study to confirm the identities of genes[2]. *A. baylyi* simplifies genetic manipulation, and could easily be used as the basis of a laboratory course that would include bioinformatics, microbiology, molecular biology, protein expression and characterization. In addition, I would be interested in teaching upper-level courses that would cover advanced topics in Evolution, Honors and Genetics. Finally, I would be interested in starting a graduate-level course discussing the intersection of evolution and practical molecular biology and biochemistry. This would introduce students to applied evolution, as it exists in state of the art biotechnology companies. I am uniquely suited to teach such a course, having worked directly with the founder of Maxygen, Inc., the biotechnology company that holds much of the intellectual property regarding DNA shuffling and molecular breeding.

My past teaching duties have included two teaching assistantships for biochemistry courses. The first was the second part of an undergraduate course introducing basic biochemistry. In addition to grading and normal office hours (during which I was available for discussion) I also instituted pre-testing discussion sessions.

These sessions were designed to bring students together to go over material in class, to answer questions more in-depth as well as to foster discussion between students. Furthermore, I summarized class material in handouts that I gave to students prior to exams, in order to help to organize and highlight important points. The second course that I TA'd was an undergraduate and graduate level course in biochemical methods, which was a broad overview of the techniques used in biochemical studies. My responsibilities for this course included office hours, discussion sessions and grading exams and papers. In particular, the term papers for this course were especially interesting: the students were required in discussion with me, to use techniques described in class to address a particular problem. In effect, I helped a class of 40 to write their first grant proposal.

My laboratory will be a furtherance of the concept of teaching by doing. I plan to involve undergraduates in my laboratory. The best way of learning about research is by doing it in an organized setting, with supervision of experienced members of the lab. As the techniques in my laboratory will mainly involve microbiology and molecular biology, certain techniques are straightforward, in particular the genetic manipulation of *A. baylyi*. Projects involving such genetic manipulation will ground students in the basic skills required for students in my lab, and also will be relatively achievable. These sorts of projects are therefore skill and confidence builders for beginning researchers in my lab. Undergraduate researchers in my lab will begin with one project under the supervision of graduate students.

References:

1. Metzgar, D. et al. *Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering. *Nucleic Acids Res* **32**, 5780-90 (2004).
2. Reader, J.S., Metzgar, D., Schimmel, P. & de Crecy-Lagard, V. Identification of four genes necessary for biosynthesis of the modified nucleoside queuosine. *J Biol Chem* **279**, 6280-5 (2004).