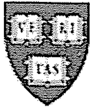

HARVARD UNIVERSITY
BAUER CENTER FOR GENOMICS RESEARCH

BAUER LABORATORY
7 DIVINITY AVENUE, ROOM 206
CAMBRIDGE, MASSACHUSETTS 02138
KURT THORN
617-384-9456
kthorn@cgr.harvard.edu

17 October 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 and Search Committee:

I wish to apply for the assistant professor position in Systems Bilogy, as advertised in the September 2nd issue of *Science.*

I have been a research fellow at the Bauer Center for Genomics Research for four years, where I run my own lab of four people. During this time, I have developed fluorescence resonance energy transfer (FRET) microscopy as a tool for *in vivo* structural biology. I recently used this methodology to construct a model of the yeast septin complex, a key component of the cytokinesis machinery in fungi and animals. This was accomplished by fusing fluorescent proteins to each of the five proteins in the complex and using the FRET signal to determine the pairwise distances between them. Fitting coordinates of the individual proteins against this data resulted in a structural model of the complex.

I am now extending this into a method for determining structures of large protein complexes *in vivo*. Combining pairwise distance constraints between proteins from FRET microscopy with structures of the individual proteins will enable the modeling of protein structures that cannot be purified or reconstituted *in vivo*. I will use this technique to determine a structure of the yeast bud neck cortex, a complex of ~100 proteins that carries out cytokinesis in yeast. This approach, combined with a study of the dynamics of complex assembly and of the regulatory pathways controlling cytokinesis initiation will enable a comprehensive understanding of cytokinesis, as well as a method for determining structures of similar protein complexes.

As requested, I have enclosed a copy of my CV, a joint description of my past research and future plans, and a statement of teaching interestes. I have also attached a preprint of my paper "Cell-cycle regulated structural dynamics of the septin polymer", which is under consideration at Nature Cell Biology, and reprints of two additional papers. In addition, I have arranged for three letters of reference (from the individuals indicated on my CV) to be sent to you. I am very excited about this position and look forward to further contact with you.

Sincerely,

Kurt Thorn

An integrated approach to determining the structure and dynamics of the yeast bud neck cortex and its function in cytokinesis

Kurt Thorn

Summary: Multiprotein complexes are involved in nearly every aspect of biology. Many of these are soluble, and modern biology has made great strides in determining the structures of these. However, many other biologically important complexes function only in the context of the intact cell and cannot be purified for study. These complexes play crucial roles in biological processes such as synaptic transmission (the post-synaptic density) and immunological recognition (the T-cell synapse). Here, I propose to determine a structure of the yeast bud neck cortex, a complex of ~100 proteins that carries out cytokinesis in *S. cerevisiae*. To do so I propose to use fluorescence resonance energy transfer (FRET) microscopy to measure, *in vivo*, the pairwise distances between the proteins that make up the complex. These distance constraints will then combined with structures of the individual proteins and large-scale constraints from techniques like electron tomography to construct structural models of the complex. Furthermore, because the FRET measurements are made *in vivo*, we can observe how the structure is constructed throughout the cell cycle and how it rearranges to carry out cytokinesis. I therefore additionally propose to characterize the assembly pathway and dynamics of this complex and to integrate these structural measurements with chemically sensitive versions of two kinases that allow us to specifically block the onset of cytokinesis at defined points. Together these experiments should define the structure of the yeast bud neck complex and how it evolves throughout the cell cycle. These experiments will also develop a methodology to solve structures of other large supramolecular complexes in the future.

Structure determination of soluble protein complexes is now a mature field. However, many biologically important complexes, such as the post-synaptic density, the immunological synapse, or the yeast bud neck cortex, exist only in the context of the intact cell and cannot yet be purified and studied *ex vivo*. Furthermore, these complexes are not necessarily homogeneous – the composition and structure of the complex can differ between cells or even within the same complex over time. The proteins associated with the complex, as well as their locations within the complex, can change in response to signals both from within the complex and from elsewhere in the cell.

This class of protein complexes (referred to here as large supramolecular complexes, LSCs) has so far not been amenable to standard methods of structural analysis. Because they are difficult or impossible to purify intact from cells, traditional electron microscopy and crystallographic techniques cannot be used to study their structure. While some electron microscopy techniques, in particular electron tomography, can be used to study protein complexes in their native cellular context, the inherent contrast of these techniques is low because the electron scattering of these protein complexes is only slightly higher than that of the surrounding cytoplasm. This limitation has traditionally been circumvented by averaging many images of the complex under study, but this approach cannot be used for the heterogeneous LSCs.

Here, I propose an alternative approach for studying the structure of LSCs. Rather than trying to determine a structure of the whole complex by an imaging method, I

propose to reconstruct the structure of the complex from distance constraints measured between individual proteins in the complex. The distance constraints will be determined by fluorescence resonance energy transfer (FRET) measurements made between fluorescently tagged proteins *in vivo*. Although this technique should be applicable to any such complex, I propose to study the bud neck cortex of budding yeast, which is responsible for carrying out cytokinesis. This is an attractive target, as cytokinesis is currently a poorly understood process, and the ease of genetic manipulation in yeast simplifies the experimental strategy. While this proposal focuses on the yeast bud neck cortex and its role in cytokinesis, I hope to use the same techniques to study the structure and function of other LSCs in the future.

Previous Work:

Bogan, A.A. and **Thorn, K.S.** Anatomy of Hot Spots In Protein Interfaces. *J. Mol. Biol.* **280**: 1 – 9, 1998.

In this work, we compiled a database of alanine-scanning mutations to study the energetic contribution of different amino acids to protein-protein interactions. From this data, we showed that the contribution made by an individual amino acid to the free energy of binding varies greatly over the binding interface of a protein. Contrary to results obtained from averaging over the whole interface, the free energy contribution ($\Delta\Delta G$) is not proportional to the change in solvent accessible surface area on binding. However, significant burial in the complex (low solvent accessible surface area) is a necessary condition for a significant contribution to $\Delta\Delta G$, suggesting that solvent access weakens the interaction between proteins and that solvent attack may be involved in the dissociation pathway. We also found that certain amino acids (in particular, tryptophan, tyrosine, and arginine) are overrepresented in the class of energetically important amino acids, suggesting that it may be beneficial to include these amino acids when trying to design interacting interfaces (or to avoid them when trying to block interactions between proteins).

Thorn, K.S., Ubersax, J.A., and Vale, R.D. Engineering The Processive Run Length Of The Kinesin Motor. *J. Cell Biol.* **151**: 1093-1100, 2000.

Unlike other molecular motors, kinesin is highly processive, traveling ~100 steps along a microtubule before dissociating. In this work we used a single molecule fluorescence assay to measure the run length of individual kinesin molecules. Using this assay, we measured the effect of mutating different portions of kinesin on its run length. This showed that a positively-charged sequence, known as the neck coiled-coil, was required for high processivity. We further showed that this region interacted with the negatively charged C-terminus of tubulin to tether the kinesin near the microtubule. By adding positively charged residues to the neck coiled-coil we were able to further increase the processivity of kinesin. As kinesin processivity can be easily changed by mutation it seems likely that this parameter of has been selected through evolution.

Thorn K.S. and Sheff, M.A. Cell-Cycle Regulated Structural Dynamics Of The Septin Polymer. Submitted to Nat. Cell Biol.

I have spent much of my four years at the Bauer Center developing fluorescence resonance energy transfer (FRET) microscopy as a practical tool for cell biology. Here, we used FRET microscopy to measure the pairwise distances between the five proteins

which make up the septin polymer, one of the core components of the bud neck cortex. To do so, we fused cyan (CFP) and yellow (YFP) fluorescent proteins to the C-terminus of each protein, and mated these strains to generate all 25 possible pairs. By measuring the fluorescence intensities of CFP, YFP and FRET (YFP emission on CFP excitation), we were able to calculate the efficiency of energy transfer (which is proportional to distance) for each protein pair. To do so accurately requires careful accounting for cellular autofluorescence, photobleaching, and crosstalk between the fluorescent proteins and the FRET channel. To properly correct for these effects, we developed new algorithms and extended existing ones. These are embedded in a software suite I developed for the analysis of FRET microscopy data. Equally important, we have optimized the microscope and data collection procedures to minimize error.

This work enabled us to measure all pairwise distances between the five proteins which make up the septin polymer, in both G1 and metaphase cells and in the presence of mutations of three key regulatory proteins. The total data set comprises 125 FRET efficiency measurements. We then fit coordinates of the five proteins to this data, enabling us to construct a structural model of the septin complex in both G1 and metaphase, and to determine the effect of the three regulators on the structure. This work revealed a previously unknown role for the kinase Gin4 early in the cell cycle, and produced the first quantitatively tested structural models of the septin complex (Figure 1).

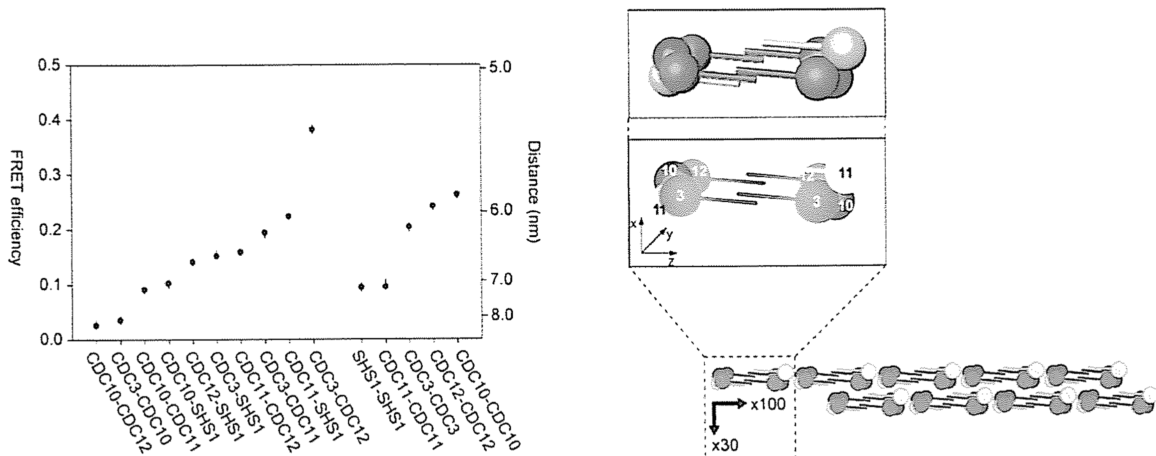

Figure 1: Determination of the structure of the yeast septin complex. On the left are the FRET efficiencies measured from CFP- and YFP-tagged septins imaged at the bud neck in live yeast. The approximate distances corresponding to these FRET efficiencies are shown. On the right is the model derived from fitting protein coordinates against these FRET efficiencies.

Future Research Plans:

In budding yeast, cytokinesis occurs at the bud neck between mother and daughter cells. Approximately 100 proteins are localized to this region, including many proteins essential for cytokinesis (e.g. myosin, actin, chitin synthase, and the septin complex). This complex, the bud neck cortex, is highly dynamic. It is assembled in a stepwise fashion between G1 and metaphase; once chromosome segregation is completed, it undergoes a dramatic structural rearrangement. The septin ring splits in half, forming two rings, one closer to the daughter cell and one closer to the mother cell, while actomyosin and the other components of the contractile complex relocalize between these

two between these two rings and contract to carry out cytokinesis. The complex is also asymmetric, with some proteins localized at the mother-cell side of the ring and others localized at the daughter-cell side .

Despite the importance of cytokinesis, the underlying organization of the protein complex which carries it out is very poorly understood. While a number of specific interactions between individual proteins in the bud neck have previously been determined and the order in which proteins assemble has been partially mapped, we lack an understanding of the structural organization of the complex. To better understand the structure of this complex, I intend to answer three questions: what the structure of the complex is, how the complex is assembled, and how the structure of the cortex is regulated to carry out cytokinesis. Although here I propose to carry out these experiments in *S. cerevisiae*, I am also considering *Sz. pombe* as an experimental system, as it shares the experimental tractability of *S. cerevisiae* and its cytokinesis more closely resembles that of higher eukaryotes.

1. Determination of the *in vivo* structure of the bud neck cortex

Because obtaining an atomic resolution structure of the bud neck cortex is not currently possible, I propose to generate a lower-resolution structure of the complex by combining data from a number of techniques. These data will include structures or structural models of the individual proteins in the complex, measurements of the overall structure of the complex from electron microscopy and tomography, and crucially, a map of pairwise distances between the proteins in the bud neck cortex measured by FRET microscopy. We recently used a very similar approach to generate a model of the septin complex, a core component of the bud neck cortex (Figure 1). In that case, the model was only weakly overdetermined. However, the number of FRET measurements grows as the square of the number of proteins, while the number of coordinates that needs to be determined grows linearly with the number of proteins. Thus the large data set measured here will overdetermine (by ~10-fold) the structure of the bud neck. In practice many of the proteins will likely be too far apart to undergo FRET, and thus contribute little information, but the excess of constraints over parameters to be fit should make the fit of this larger data set more robust than our fit of the septin complex was.

However, although the FRET measurements will overdetermine the positions of proteins within the bud neck cortex, they lack information about both the high resolution (<1 nm) and low resolution (>10 nm) length scales. Because the FRET measurements are made between the C-termini of each protein, they provide no high-resolution information about the location of the rest of the proteins. To overcome this drawback, we will include structures of each protein in the cortex from experimentally determined structures or from homology modeling. Experimental structures of homologous proteins are available for at least individual domains of many of the bud neck proteins. For those proteins or domains of proteins where we lack structures, we will attempt to obtain them from either homology modeling (which has grown increasingly accurate in recent years) or experimentally. I have considerable experience with protein crystallography and will attempt to determine structures of those bud neck cortex proteins which play central roles in the bud neck cortex.

The FRET data measured here will also be subject to limitations on long length scales. Like NMR structure determination, we will be using many short distance

constraints to determine the structure of a large complex. While these measurements will provide accurate constraints over short distances, they only poorly constrain the overall structure of the complex. Constraints on the large scale organization of the bud neck cortex can greatly reduce this source of error. The best source for such a constraint would be electron tomography, which would allow determination of the overall structure of the complex at a resolution of ~10 nm. As I have little experience in electron microscopy or tomography, this would best be carried out in collaboration. A potential additional long distance constraint would be to measure the radial distances of proteins from the cell membrane by immunoelectron microscopy. Although this can be quite time consuming, one advantage we have is that each of the proteins will be fused to GFP, allowing a common procedure based on antibodies against GFP to be used. It may also be possible to measure these radial distances by optical microscopy if the proteins are far enough apart.

To construct a model of the bud neck cortex from this data, we will use a nonlinear optimization procedure to adjust the positions of the proteins within the bud neck, minimizing differences between the modeled structure and our experimental constraints. Minimizing deviations from the FRET data and radial distance constraints can be handled by a simple cost function. Structural data from proteins in the complex will be incorporated primarily by minimizing steric clashes between the individual proteins. Electron tomography data will be incorporated by fitting the calculated electron density for the individual proteins to that determined from the tomogram, as is commonly done to dock protein structures into electron microscopy data.

The final goal of this work is the construction of a structural model of the yeast bud neck cortex. However, even before this point is reached, the data will greatly enhance our understanding of the organization of the yeast bud neck cortex. The initial maps of distances between proteins will allow us to identify the overall organization of the cortex; for instance, if it is hierarchical, or if subcomplexes exist within the bud neck cortex. If we do identify such subcomplexes, we will attempt to further characterize them by purification from yeast or by *in vitro* reconstitution. For small, stable subcomplexes, direct structural determination may be possible. For other subcomplexes, chemical crosslinking combined with mass spectrometry to identify the cross-linked residues can be used to identify additional distance constraints between the proteins. I am currently collaborating with Gavin MacBeath's lab to use this method to test our model of the yeast septin complex.

The FRET microscopy planned here is a straightforward extension of our existing work determining the structure of the yeast septin complex. Our previous study required 125 efficiency measurements; the project proposed here requires ~5000. To enable collection of this large set of measurements, we are working to automate the microscopy system for collecting the data, and to improve the fluorescent protein pair used to tag the proteins. We have already begun testing and implementation of automation on our existing microscope, which we expect to reduce the time required to collect this data set to ~6 months. This, together with established techniques for high-throughput construction of yeast strains, makes this a practical project for a single investigator.

Identification of improved fluorescent proteins for FRET is a more challenging task. The commonly used CFP/YFP FRET pair suffers from both a high degree of crosstalk between the individual proteins and the FRET signal and from rapid

photobleaching. While the research proposed here would be possible using the CFP/YFP pair, identification of an improved fluorescent protein pair would increase both sensitivity and accuracy. Therefore, my lab has embarked on a systematic screen of existing fluorescent proteins to identify improved FRET pairs. In our initial screen we measured the brightness, photobleaching rate, folding time, and effect on fusion partners of each protein alone as well as its crosstalk to potential FRET partners. From this initial screen we have identified seven pairs which show greatly improved properties over CFP/YFP: <20% crosstalk, and 5-10x more photostable. We are currently testing all seven of these FRET pairs both *in vivo* and *in vitro* and expect to identify one or more as a suitable replacement, which will result in a greatly improved signal to noise ratio in our experiments.

2. Determining the assembly pathway of the yeast bud neck cortex

To identify how the bud neck cortex is assembled, I will initially determine which of the other proteins in the bud neck cortex are required for its proper localization. This is easily assessed by testing for proper localization of each cortex component in the presence of a deletion or loss-of-function mutation in each other component. Those proteins which fail to localize in the presence of a given deletion by definition require the deleted protein for proper localization. From this systematic analysis I expect to identify both structural proteins required for localization, as well as regulatory proteins that are localized to the cortex. The resulting network defines a 'genetic assembly pathway' detailing the hierarchical dependencies between proteins required to assemble the bud neck cortex. Furthermore, by fixing tagged cells at different points in the cell cycle, or by identifying the cell cycle stage from the microscopy images, we can also measure the temporal order of assembly of these proteins. This data is straightforward to acquire and will complement the data acquired in the first project by identifying not just which proteins are close together, but which ones are functionally important for the structure of the complex.

In addition to identifying the genetic and temporal order of assembly of this complex, I will also study the dynamics of this complex over shorter timescales, by assessing the timescales on which proteins exchange between the bud neck and the cytoplasm. Experiments of this sort have previously revealed that the septins readily exchange with the cytoplasm early in the cell cycle, but become immobilized after bud emergence. Characterization of dynamics in this way will shed light on the mechanism of assembly. If, during assembly, proteins freely exchange between the bud neck and cytoplasm, this suggests that the assembly process is reversible and occurs as a (quasi-) equilibrium process. If no exchange of assembled subunits occurs with the cytoplasm, this suggests that the process occurs far from equilibrium and that there is a substantial free energy change on incorporation of these proteins into the complex. Proteins which exchange freely early in the assembly process, but become immobilized later are candidates for core components of the complex that have become buried by other proteins preventing their exchange, or have been locked into place by some other mechanism.

Measurement of dynamics on these time scales is typically done by photobleaching of a fluorescently-tagged protein in the complex and following replacement of the bleached protein with unbleached protein by monitoring the recovery

of fluorescence. It can equally well be done by photoactivation or a photochromic shift in the emission of the fluorophore. These techniques are required to achieve high time resolution of the exchange dynamics but are technically challenging and require specialized equipment. Because of the large number of proteins that I wish to study, I propose to do an initial prescreen to rapidly identify those proteins which show either rapid exchange or lack of exchange. This prescreen will be done by arresting cells carrying a GFP-tagged protein at a defined point in the cell cycle, followed by induction of an RFP-tagged copy of the same protein. If we see the RFP-labeled protein incorporated into the bud neck cortex, this suggests that that protein exchanges rapidly between the bud neck cortex and the cytoplasm. Conversely, if it is not incorporated, that suggests that the protein is not exchanging rapidly. Those that show interesting behavior will then be followed up by more traditional means. This data should shed considerable light on the dynamics and energetics of bud-neck cortex assembly.

3. Understanding the regulation of cytokinesis

How cytokinesis is regulated, and in particular, the identity of the signal which triggers the initiation of cytokinesis, remains poorly understood. While we will likely identify some of these regulators through the systematic genetic analysis proposed above, here I aim to take a more conventional molecular genetic approach to identifying regulators of events at the bud neck cortex. In particular, I hope to better characterize the regulatory machinery underlying the commitment to and initiation of cytokinesis. Previous work implicates two kinases involved in mitotic exit, Cdc15 and Cdc5 (the yeast homolog of polo kinase) in controlling initiation of cytokinesis. These results were obtained using temperature sensitive kinase alleles combined with additional mutations that bypass their requirement for mitotic exit. When shifted to the restrictive temperature, cells bearing these mutations completely fail cytokinesis, but continue through the cell cycle, accumulating as long chains of cells with connected cytoplasm. The requirement of a temperature shift, however, makes it difficult to rapidly inactivate and reactivate these kinases, and so little followup work has been done on them.

To address these shortcomings, we have, in collaboration with Kevan Shokat's lab, constructed equivalent strains where the temperature sensitive kinase is replaced with a drug-sensitive kinase. These strains behave identically to the previously described strains, but unlike them, kinase inhibition and release from inhibition can be achieved rapidly (in minutes) by addition or washout of the drug. We are using these strains to identify the specific effects that these kinases have on the bud neck by examining how the structure of the bud neck changes before and after kinase activation. Currently we are examining whether protein localization changes following kinase activation, but we hope to identify more detailed structural changes by using the higher resolution FRET experiments described in project 1. We are also examining how cytokinesis activation is integrated with other cell-cycle processes. Initial results show that bud necks where cytokinesis had been blocked by inhibition of Cdc15 or Cdc5 cannot reinitiate cytokinesis in subsequent cell cycles if the inhibition is removed. This suggests that the timing of kinase activation is critically coordinated with other processes to successfully trigger cytokinesis. I hope to identify these processes and how they are coordinated with cytokinesis initiation.

Conclusion: The proposed work aims to use my expertise in FRET microscopy to develop a new approach for determining *in vivo* structures of large supramolecular protein complexes. Pairwise distances between proteins in the complex, measured by FRET microscopy, combined with structural models of the proteins in the complex, should provide sufficient information to determine a low resolution structure of the protein complex. I will use this methodology to determine the structure of the yeast bud neck cortex, which carries out cytokinesis in budding yeast. Determining the structure of this complex will shed light both on how large protein complexes are organized, and also on how this structure carries out cytokinesis, which is currently poorly understood. I will combine this work with a systematic study of the requirement of each cortex protein for localization of the other cortex proteins to better understand the assembly pathway for this structure. Finally, I plan to integrate the work above work with a study of the regulation of cytokinesis, focusing on the role of the kinases Cdc5 and Cdc15. The proposed work should provide a framework for studying large supramolecular complexes and a better understand understanding of how these structures are organized. These studies will also lead to a better understanding of cytokinesis, a crucial cell-cycle process that remains poorly understood.

Teaching Interests

Kurt Thorn

My experience in teaching comes from both teaching graduate students as a graduate student at UC San Francisco, and from mentoring undergraduates in my lab at Harvard. In particular, at UCSF I taught six lectures in the 1st year graduate course on structural biology and biochemistry. These lectures discussed ongoing research in the focus areas of the course, and focused on areas of groundbreaking or controversial research. These lectures drew extensively on recently published work, and focused on dissecting this work to understand the experimental design and interpretation of results to illustrate how scientific knowledge is arrived at. In this regard, using two papers which address the same problem but arrive at different results showed very clearly the complexities of real scientific work.

While this approach cannot be directly transferred to general undergraduate classes, I think its central idea can be maintained: that at its core, scientific enquiry is a means and not an end in itself. Science classes should therefore focus more on the process by which scientific knowledge is obtained and not solely on the facts already known. Of course, some facts are necessary knowledge, and these must be taught. In my opinion the most important facts are those general principles that provide a lens to see aspects of biology through. For example, from undergraduate biochemistry, I came away with the understanding that there are core biochemical transformations that recur again and again in different pathways using the same mechanism. I came to this realization myself, after memorizing dozens of reactions from intermediate metabolism. This is perhaps not the best way to teach this, but the concept that the general principles should be supported by specific examples is a good one. Finally, I think that biology education should become more scientific: physics education has greatly benefited from careful assessment of the effectiveness of specific lessons and teaching strategies. This iterative assessment and improvement should be useful in biology as well.

Below I discuss several classes I would be interested in teaching:

Biochemistry and Structural Biology: This would be a standard sophomore/junior level course on protein and nucleic acid structure, the energetics of macromolecule folding and interaction, basic biochemistry, and then the biochemistry of biological processes (DNA replication, transcription, translation, and metabolism).

Programming for Biologists: In my experience at Harvard, very few undergraduate biology majors have any experience in programming, which is quickly becoming an essential skill in modern biology. This course would be an introduction to programming for biology undergraduates with no previous programming experience. I would teach one or two programming languages (probably Perl and Matlab) and show how to use them to solve common biological problems. The first half of the course would deal with bioinformatics problems: sequence searching and matching, file parsing, and related areas. The second half of the course would teach more mathematical problems: data fitting and kinetic modeling. The course would aim to introduce some of the biology behind the topics studied, while focusing on developing competency in programming.

Microscopy: This would be a graduate class in microscopy. It would begin with an introduction the physics of image formation and some basic optical principles. This would be followed by a description of different commonly used imaging modalities (wide-field, confocal, spinning-disk confocal, two-photon, total internal reflection), including practical discussions of detectors, light sources, objectives, and other optical elements, and would preferably include a hands on lab. More exotic imaging modalities, such as optical trapping and FRET, would be discussed at the end of the course.