Preface

Why is this handout so long?

The major problem with trying to teach a course like this is that there is no textbook that fits exactly what I want to cover over the semester. Over the last 5 years, I have used a mixture of primary research papers and short reviews as reading assignments for this course, and we will continue to do that this year. However, all of the authors of the reading assignments have their own points to make, and often these include excursions into areas of more detail than is appropriate for us to cover. Students then spend more time getting stressed out over things they don't need to know than is necessary.

Over the years, students have often asked me for my lecture notes. I have not provided them because they are usually just short phrases or words that would only make sense to me. However, I have been sympathetic to the need for additional written materials to review, since I know that I go over things in class very quickly, and a mistake in note-taking can't be checked against the textbook since there isn't one.

So, last summer I planned to whip my notes into a more usable form. Summer turned into Fall, and there was still no draft. Finally, I started to work on what follows as the start of the semester approached.

What this means is that this is very much a first draft, and it probably needs a lot of additional work to be what I would like it to be. There will be places where there are obviously words, sentences or paragraphs that are completely missing. Your feedback on these handouts would be appreciated very much.

Will they all be this long?

Unfortunately, the same things that kept me from sitting down to write this earlier haven't gotten any better. I expect that my written supplements to the lectures will appear sporadically at best. I hope that whatever additional material I can provide will be useful, and that you will still be able to muddle through the lectures that don't get these supplements, as previous classes have.

Introduction - the nature of molecular biology

What is molecular biology anyway?

"I myself was forced to call myself a molecular biologist because when inquiring clergymen asked me what I did, I got tired of explaining that I was a mixture of crystallographer, biophysicist, biochemist and geneticist, an explanation which in any case they found too hard to grasp" - Francis Crick (1965) *Brit. Med Bull* 21:183
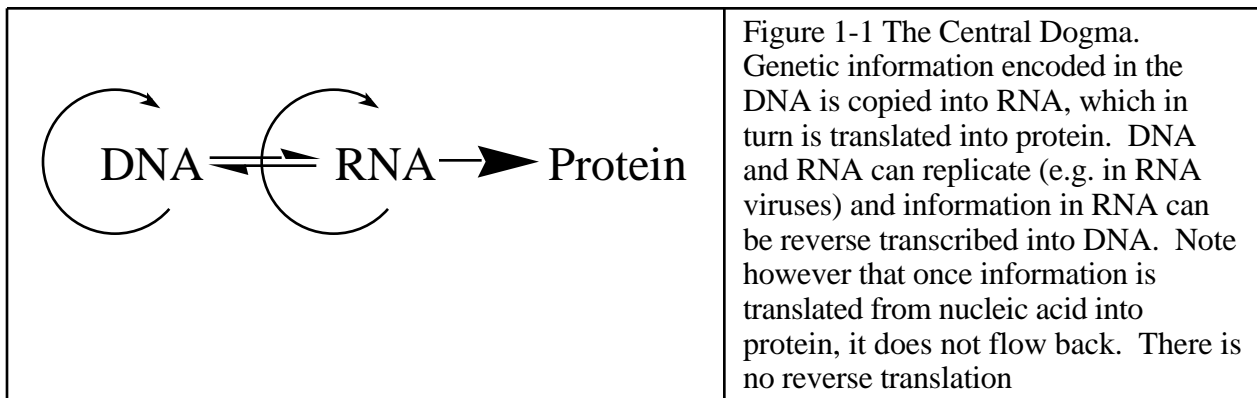
Although the title of this course in the catalog is Biochemical Genetics, it is generally thought of as a course in advanced molecular biology. Like biochemical genetics, the term "molecular biology" involves intellectual approaches that cross traditional disciplinary lines. Biochemical genetics grew out of the studies in the 1940s trying to understand the chemical basis for observable phenotypes such as the different eye pigmentation of lines of mutant flies. Beadle and Tatum realized that approaching the relationship between specific biochemical reactions and specific mutations could also be used to understand the nature of genes and the mechanisms of inheritance. Their studies on mutants unable to synthesize specific metabolites led to the one gene-one enzyme hypothesis and were one of the starting points for the development of what most of us now think of as molecular biology.

As reflected in the quotation above from Francis Crick, molecular biology includes biochemical genetics but includes an even more complex mixture of disciplines (curiously, Crick neglected to include biologist in his list!). A perfectly reasonable definition of molecular biology based on the literal meaning of the words could be the study of the mechanisms underlying biological phenomena at the molecular level. This definition is virtually indistinguishable from a reasonable definition of biochemistry, and probably accounts for "traditional" biochemists describing molecular biology as "biochemistry without a license" (J.C. Kendrew (1967) *Sci Am* 216:141-4 , quoting an unnamed "distinguished biochemist"). This is not altogether unfair, as one will sometimes hear self-described molecular biologists talk about "doing the biochemistry"

as if it is merely filling in the boring details; something they might do in the future if they can't find a more interesting field to switch to.

To my mind, the defining characteristics of molecular biology are ways of thinking that are derived from a mixture of physics, genetics and biochemistry, despite the fact that the work is almost never recognizable as physics, and regardless even of whether the work involves anything that is recognizable of genetics or biochemistry. From genetics, molecular biology borrows the language of abstraction. Insofar as genetics deals with the ways genes are transmitted, recombined or expressed, the materials and mechanisms that account for why those rules are followed are peripheral or irrelevant. The abstract nature of genetics probably accounted in some part for the appeal of genetics to the physicists like Max Delbrück who were involved in the origins of molecular biology (at least according to one school of the history of the field). Physicists entering the field brought with them the urge for unification similar to that which formed the basis for the intellectual triumphs of quantum physics. This had the fortunate consequence of leading the physicists toward the simplest possible model organisms, especially microorganisms and viruses. After all, if there were universal truths to be found in biology, the best chance to find them would be in the simplest systems.

A difference between classical genetics and molecular biology is that the latter uses the language and cartoon representation of abstraction to describe molecular entities including proteins and nucleic acids. Like biochemists, molecular biologists are concerned with the

Figure 1-1 The Central Dogma. Genetic information encoded in the DNA is copied into RNA, which in turn is translated into protein. DNA and RNA can replicate (e.g. in RNA viruses) and information in RNA can be reverse transcribed into DNA. Note however that once information is translated from nucleic acid into protein, it does not flow back. There is no reverse translation

interactions among these entities; however, the interactions include abstract regulatory transactions as well as the physical binding and chemistry. Pathways are not only about the flow of material, but may relate instead to the flow of information, as in the "central dogma of molecular biology" (Figure 1-1).

Consider transcription, the first step in gene expression as represented in the Central Dogma (Figure 1-2). The left side of the figure shows the transcription step as a component of the overall pathway shown in Figure 1-1. Note that the arrow connecting DNA and RNA is not equivalent to a biochemical reaction. A representation of transcription from the point of view of the chemistry involved is shown on the right side of Figure 1-2. Chemically, DNA is not converted into RNA during transcription by the removal of 2' OH groups and demethylation of T residues. In a biochemical reaction, DNA could be thought of as a cofactor along with the proteins involved in transcription that catalyze the synthesis of RNA from small molecule precursors. However, the simpler representation on the left embodies a profound idea that is not as clear from the more chemically rigorous representation on the right: that the *information content* of genes passes from its manifestation as DNA to being physically encoded in RNA.

$$PP_i$$

$$DNA \longrightarrow RNA \qquad\qquad rNTPs \longrightarrow RNA$$
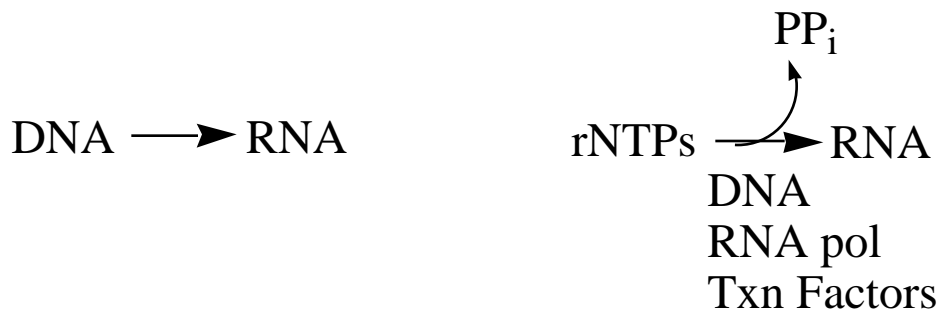DNA
RNA pol
Txn Factors

Figure 1-2. Two ways to represent transcription

As important as the central dogma is to molecular biology, the point of this course is not to pass on a list of dogmas, central or not, of molecular biology. What the central dogma illustrates here is how an extremely complex process can be represented by a simple abstraction. A

"rigorous" diagram showing all of the reactants, products, catalysts and cofactors that are involved in the expression of genetic information would not only be too complex to fit on the page, it would also obscure the elegance of the ideas illustrated by the central dogma.

What do we want to know?

Does the central dogma embody the level of abstraction appropriate to all questions in molecular biology? Of course not. The details do matter; the central dogma does not even explain why *E. coli* makes -galactosidase when grown on lactose but not when grown on glucose, much less explain how to make an immune system or a brain or why plants can be regenerated from cuttings and people can't.

For any biological process or system, we can ask what information is needed to understand what is going on at a detailed level. Bruce Alberts, citing an overhead shown by XX at a conference on molecular motors, has described a "path to enlightenment" toward molecular understanding. As recounted by Alberts - understanding a process requires:

• Complete inventory of components

• Description of all reaction intermediates

• Rates of all reactions

• Structures at atomic resolution

Another way of stating this is that we want an atomic level movie where we know who all the players are, we can tell what happens and we understanding how it all works without special effects that violate the laws of physics.

Levels of understanding

Complete understanding needs both the general structures of things and the details of how they work. However, as is illustrated by some of the important syntheses in the history of molecular biology, such as the central dogma or the operon model, important progress toward

enlightenment can be made by looking for the common principles that are shared by processes that may differ in their molecular details. Also, it should be clear that the experimental paths to enlightenment can be taken from a variety of different approaches, drawing from the different disciplines that intersect in molecular biology.

Each kind of approach operates at a different level of detail and of abstraction, and the level of molecular detail appropriate to study to describe a process depends on the kind of question about the process that needs to be answered. My bias is that the historical strength of molecular biology is in large part derived from trying to address any individual process at the lowest level of detail first in order to frame the general issues before moving on to the specific details. This is not specific to molecular biology - in most areas of science and, one could argue, in everyday life, it is often productive to try to "see the forest" instead of concentrating on the individual trees.

In computer programming, this is described as taking a top down approach. Instead of starting out writing a program by writing instructions for loops and branching, one starts by considering what one actually wants the overall program to do. Then the purposes of the subroutines are determined, and only after the general flow of the program is worked out are the details of the final code added.

In an idealized world (which is what we will often pretend we work in for exams and problem sets), a top-down approach would involve asking the most general questions first, and only later filling in the mechanistic details. Using the central dogma as an example, the idea that information flows from DNA to RNA to protein provided very useful insights into how inheritance and gene expression must work long before more detailed mechanisms for replication, transcription and translation were worked out.

The general picture from the central dogma provides a context for other questions about gene expression. For example, knowing that genetic information is transferred from DNA to RNA and only then to protein leads to different models for how gene expression could be controlled than if it had turned out that proteins had been synthesized directly on DNA, a possibility that

was seriously considered before the evidence for mRNA accumulated in the mid 1950s. The processes of replication, transcription and transcription can then be examined in progressively greater detail. In fact, there are still plenty of important "details" that we still do not have about these processes.

Reality, of course, is complex. The experimental approaches that are taken are very much influenced by the available technology. Projects take unexpected turns, and unexpected results can shed light into different areas than were expected based on the working hypotheses used to design the experiments. Nevertheless, the value of things found by accident is not an argument against trying to take an organized and systematic approach to problem solving as long as one keeps an open mind. It can be argued that carefully thinking through the possible alternative results ahead of time is the best way to prepare for the unexpected.

The material that follows is an attempt to give you examples of how problems have been attacked by molecular biologists in the past. The examples will ask questions at different levels of detail, but it is my hope that you will come away from this with a sense of the kinds of experimental questions that can be asked to address issues related to the Path to Enlightenment described by Alberts.

Will you achieve complete enlightenment by taking this course? I doubt it; if you do, see me about taking over my teaching duties next year! I hope that by sightseeing along the paths taken by others as they attacked a variety of problems, you will learn to appreciate the beauty of the journey and how to make your own path toward understanding the biological problems

Model systems

"The meanest living cell becomes a magic puzzle box full of elaborate and changing molecules, and far outstrips all chemical laboratories of man in the skill of organic synthesis performed with ease, expedition, and good judgement of balance" - Max Delbrück (1949) A physicist looks at biology. *Trans. Connecticut Acad. Arts and Sci.* 38:173-190, reprinted in *Phage and the Origins of Molecular Biology*, Cold Spring Harbor Laboratories, NY.

Biology is ultimately an experimental science; despite my desire to emphasize the importance of universal aspects of molecular biology and the power of abstraction, and although there is a *Journal of Theoretical Biology*, we need experimental systems in order to study life. Since most basic research in molecular biology is funded by government or private institutions associated with human diseases, one might think that for funding reasons if nothing else, scientists would follow the axiom that the best place to study what affects humans is humans. Similar arguments can be made for focusing research on medically or agriculturally important pathogens, crops, or animals. However, molecular biologists have generally chosen model organisms for basic research based on other properties and fortunately, funding agencies have by and large accepted the arguments for the use of model organisms.

The use of model organisms is based on the assumption that mechanisms for processes of interest are conserved through evolution. If this is true, then it is clearly advantageous to study the process in an experimental system that is easier to use than humans or commercial crops. What constitutes ease of use depends on both the process under study and the kind of approach the experimenter wants to take. The ideal systems for genetic approaches are often not the ideal systems for biochemical approaches to the same problems. The organisms that are most amenable to examination by microscopy or electrophysiology are often yet another group. As we go into experiments later in the class, I will try to point out what properties of the specific model system made it a good choice. As you read further, try to think about whether the experimental

approach described was technically feasible in other organisms, and how the choice of system affects how one would attack the problem.

Although a variety of model systems will appear in this class, the majority of the papers will focus on two: *E. coli* and budding yeast, *Saccharomyces cerevisiae*. In the next sections, I will review some of the essential features of these specific systems, and mention some of the related systems that we will see over the course of the semester.

E. coli and its bacteriophages

"...although not everyone is mindful of it, all cell biologists have two cells of interest: the one they are studying and *Escherichia coli*." - Fred Neidhart (1996) in *Escherichia coli and Salmonella: cellular and molecular biology.* ASM Press.

E. coli is a gram-negative, rod-shaped enteric eubacterium that can grow rapidly in defined media and in the presence or absence of oxygen. When grown in glucose minimal medium at 37˚C, a typical cell has a volume on the order of 10^{-15} liters. The cell contains about 150 fg of protein, 60 fg of RNA (of which perhaps 5% is mRNA) and 9 fg of DNA. It has a circular haploid genome of about 4.9 Mbp, which encodes about 4290 open reading frames. As of 1995, it was estimated that 80% of its metabolic pathways were known, and many of its systems for regulating gene expression have been characterized in detail.

E. coli is undoubtedly the most thoroughly studied cellular organism around. Jacques Monod's famous statement that what is true of *E. coli* is true of elephants applies to many of the fundamental processes of life ranging from the genetic code to the rules for protein folding to most of the reactions of central metabolism. As a free-living organism, *E. coli* has to solve the same problems as other free-living organisms, and many of the solutions must have evolved very early in the history of life. In the history of molecular biology, *E. coli* and its viruses (bacteriophage) have been very important for studying the basic problems of molecular biology.

Why was *E. coli* so important? As a relatively nonpathogenic prototrophic bacterium, *E. coli* can be grown to high densities quickly, safely and conveniently. This allows both genetic and biochemical manipulations.   The choice of *E. coli* over other bacteria with similar properties is partly historical. One part of the historical answer related to the choice of *E. coli* as the host for the study of bacteriophages by Alfred D. Hershey, Salvador Luria and Max Delbrück in the late 1930s. Delbrück, who was trained as a theoretical physicist, is generally celebrated as being the leader of what became known as the Phage Group, which was involved in many of the early landmarks in molecular biology. For Delbrück, the most important question in biology was not related to the wonderful complexity of life derived from evolution, but rather "the really marvelous accomplishment: ordinary uni-parental reproduction" (Delbrück, 1949).

 For Delbrück and his colleagues, the ability of a phage particle to replicate in the black box of its bacterial host was the simplest system in which to study this "marvelous accomplishment". Phage experiments provided many important clues about the nature of genes and their expression (see inset). Although many of these achievements relied heavily on the powerful genetics available for phage, biochemical experiments using phage and phage-infected cells complemented the information obtained from genetics.

The genetics of the *E.coli* host were being developed in parallel with the work on bacteriophage. *E. coli* was one of the early choices of bacteriologists looking to study bacterial physiology; studies on variation among strains led to early microbial genetics. The Luria-Delbrück experiment, which showed that mutant bacteria that were resistant to phage infection were preexisting variants in bacterial populations, rather than cells that acquired resistance by a Lamarckian mechanism, led to the realization that bacteria actually had genes and were thus an appropriate system in which to do genetics. The synergy between work on *E. coli* and its phages is evident in the development of the operon model in the late 1950s and early 1960s. Francois Jacob, Jacques Monod and their coworkers were studying the regulation of lysogeny by phage   and regulation of the *lac* operon in parallel and in labs at the opposite ends of the attic of the Institut Pasteur, when they realized that the two phenomena reflected similar

fundamental regulatory mechanisms. The experiments were made possible by technical advances made in many labs working on the same organism. In particular, the work on genetic exchange in *E. coli* from Joshua and Esther Lederberg, and by William Hayes was important in enabling the experiments leading to the operon model.

Some landmark experiments from Bacteriophage Research

1952 -   Hershey and Chase confirm Avery's demonstration that DNA is the genetic material

1955 -   Benzer shows that genes have substructure and that mutations can be mapped into a linear arrangement within genes. Benzer, S. (1955) Fine structure of a genetic region in bacteriophage.Proc. Natl. Acad. Sci. 41, 344-354.

1956 -   Evidence for the existence of mRNA. Volkin, E. and L. Astrachan (1956) Phosphorus incorporation in Escherichia coli ribonucleic acid after infection with bacteriophage T2. Virology 2, 149-161.

1961 -   Crick provides genetic evidence for a triplet genetic code. Crick, F.H.C., L. Barnett, S. Brenner, and R.J. Watts-Tobin (1961) General nature of the genetic code for proteins. Nature 192, 1227-1232.

1965 -   Genetic evidence for the nature of nonsense codons. Brenner, S., A.O.W. Stretton, and S. Kaplan (1965) Genetic code: the "nonsense" triplets for chain termination and their suppression. Nature 206, 994-998.

1967 -   Purification of a repressor protein. Ptashne, M. (1967) isolation of the phage lambda repressor. Proc. Natl. Acad Sci USA 57:306

1969 -   Identification of transcription termination factor rho. Roberts, J. (1969) Termination factor for RNA polymerase. Nature 224, 1168

1972 -   First DNA sequences of regulatory elements. Blattner, F. R., and Dahlberg, J. E. (1972). RNA synthesis startpoints in bacteriophage lambda: are the promoter and operator transcribed? Nature New Biol 237, 227-32.

  -        First sequences of complete viral genome -   X174

-         First crystal structures of transcription factors

Regardless of how *E. coli* came to be the most prominent eubacterial model system, the concentration of work on *E. coli* as a specific model system means that we know more of the detail about how things work in *E.coli* than any other organism. Learning about an organism is cooperative: each advance makes the next advance easier to achieve.

As a genetic system, *E. coli* is very powerful for a variety of reasons. A detailed survey of E. coli genetics is beyond the scope of this course, but some of the critical features will be

summarized here.

One of the first characteristics that makes *E. coli* good for genetics is simply the fact that it grows fast and is small. Genetics involves the isolation of rare mutants and recombinants from complex populations. The size of the populations is perhaps the most important determinant in whether or not one can observe such events. It helps a lot if the starting material is derived from population with a single genotype; rapid generation times allow this condition to be met.

The fact that *E. coli* is haploid and propagates by binary fission means that mutant phenotypes can be examined by applying selections and screens to isolated bacterial colonies, which are the clonal descendants of individual cells. No additional interbreeding is necessary to observe a recessive phenotype. A haploid genotype has disadvantages, however. In particular, complementation studies are complicated by this fact. There are several ways in which E.coli strains can be constructed that have regions of the chromosome carried on low copy plasmids or duplicated in different chromosomal locations. These cells are said to be **merodiploid**; most of the genome remains haploid.

Genetic exchange is important for any genetic system. A variety of tools are available to move genetic markers between different strains of *E. coli*. These can be exploited to generate genetic maps as well as to examine the interactions among the effects of different alleles of various genes (the epistasis relationships).

The DNA phages such as T4 and propagate even faster than their bacterial hosts. Although the length of a generation is comparable, the number of progeny from an infection is on the order of 100, rather than the 2 generated by binary fission. In addition, one can mimic diploid genetics by infecting host cells with a mixture of phages at a concentration that assures more than one phage particle per infected cell. From these kinds of mixed infections, complementation and recombination studies are possible.

Yeast

Universality obviously does not apply to everything, if it did, everything would be literally just like *E. coli*. Although there are many common components and they are presumably derived from common ancestors, there are features that are distinct among the three domains of life, the eubacteria, the archaea and the eukaryotes.

As a model eukaryote, the budding yeast *Saccharomyces cerevisiae*, also known as bakers or brewers yeast, has advantages similar to the advantages of *E. coli* for studying processes universal to life (I will often refer to *S. cerevisiae* simply as yeast). Like *E. coli*, yeast is a unicellular organism that grows rapidly on defined medium. The 12MBp yeast genome (only 2.5X *E. coli*) *S. cerevisiae* is distributed over 16 chromosomes. The whole genome encodes only about 6000 genes (about 1.5X *E. coli*).
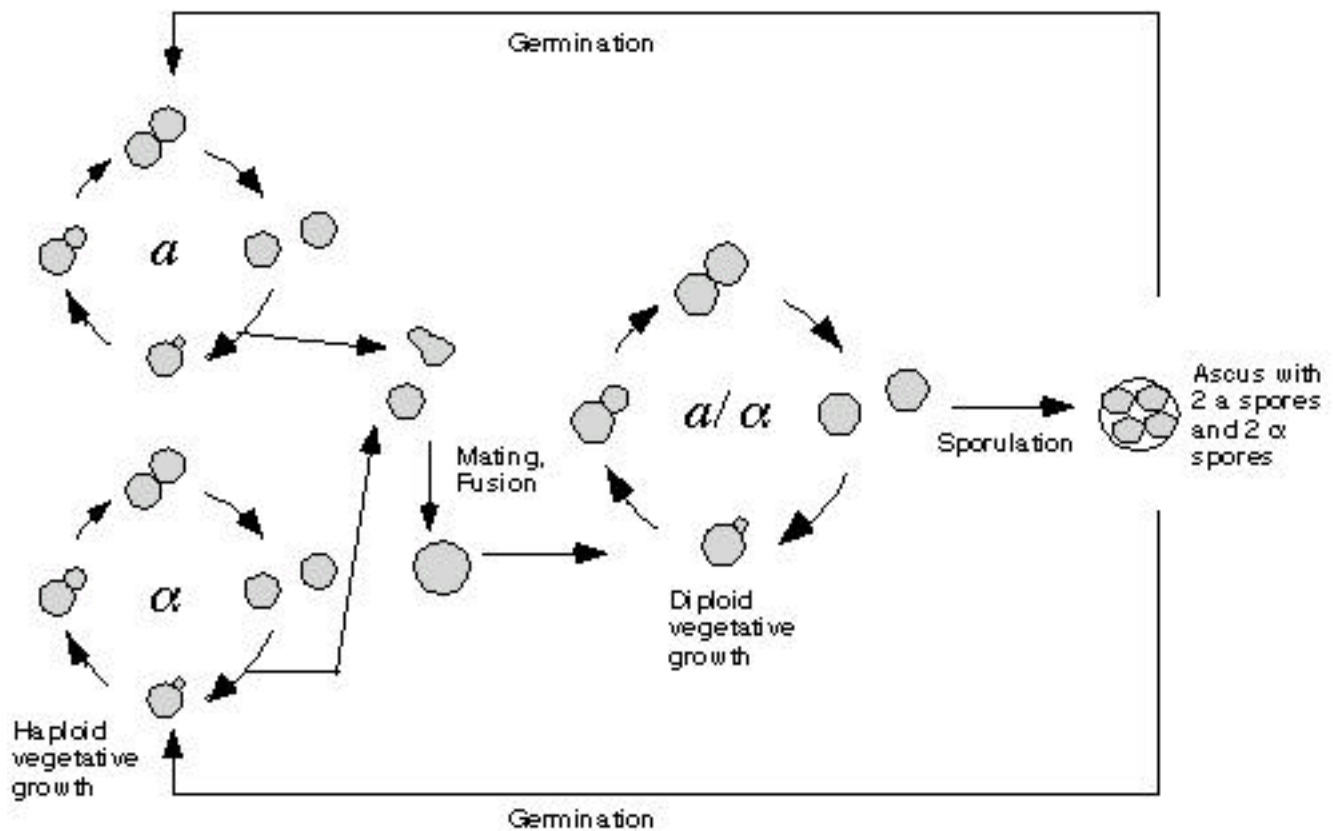

Figure 1-3. Life cycles for *S. cerevisiae*

The yeast life cycle is actually two life cycles that can interconvert (Figure 1-3). Haploid yeast cells can grow vegetatively, with daughter cells budding off in each generation. Although the details of the cell cycles are very different, from the point of view of doing mutant hunts and other genetic manipulations, the growth of haploid yeast is similar to bacteria. Unlike *E. coli*, however, haploid yeast strains can form true diploids by mating, as long as they are of the opposite mating type.

The diploid cells can grow vegetatively by mitosis to give a culture of diploid cells. However, when exposed to the right environmental conditions, the diploid cells go through meiosis to generate four haploid spores enclosed in a sac called an ascus. The group of spores is known as a **tetrad**. The spores from a culture of yeast can be isolated and germinated to give haploids again. The proportions of offspring that have a specific genotype will follow the rules of Mendelian inheritance (reviewed below). In additon, the asci are large enough to see in a dissecting microscope, where micromanipulators can be used to pull the spores apart. Each spore can be placed on a different spot on a plate, and the resulting colony can be analyzed to determine its genotype and phenotype. This is often called a tetrad analysis. Thus unlike most other model organisms, this allows yeast geneticists to examine the products of individual meiotic divisions (this works even better in *Neurospora*, where the arrangement of the spores reflects their lineage from the two meiotic divisions).

As with *E. coli*, there are important historical reasons for why yeast became a widely used model system. Yeast is certainly nonpathogenic (most of us eat and drink it without complaint), and its industrial importance in brewing was a factor in the funding of early research on yeast genetics. Yeasts have an honored place in the history of biochemistry - the word "enzyme" literally means "from yeast". However, there are problems with using yeast for studying the biochemistry of complex systems; often is easier to get materials for study from plants, animal tissues or cultured insect or animal cells.

Other model systems

Just as *E. coli* is not an adequate model for understanding those phenomena that are specific to eukaryotes, the same properties that make yeasts useful also limit their applicability as model systems. Yeasts are unicellular, and although processes in yeast share mechanisms with developmental programs in metazoans, yeasts are not thought of as organisms with significant developmental biology. The intrinsic biology of yeasts doesn't provide anything that is a version of an immune system, neurobiology or circadian timing, just to name a few examples.

As molecular biologists moved from studying processes that were viewed as universal to processes that were more specialized problems, the number of model systems being studied increased dramatically. Many systems were either pioneered or popularized by scientists who moved their research interests from phage and microorganisms into more complex systems (see inset; many of the older generation of yeast molecular biologists were also active in phage research earlier in their careers, e.g. Ira Herskowitz and David Botstein).

Some model systems with roots in phage, bacterial or yeast molecular biology	
System	Comments
Caenorhabditis elegans (nematode)	Model organism for development; powerful genetics and cell biology; virtually nonexistent biochemistry. Genome sequence published (90% finished). Genome of 97MBp, encoding about 19,000 ORFs Popularized by Sydney Brenner, who used to work on phage T4
Arabadopsis thalliana	Model for higher plants. Powerful genetics. Popularized by Jerry Fink who is well known for studies on a variety of aspects of yeast.
Danio rerio (sp?) (zebrafish)	Model system for vertebrate development. Good genetics recently augmented by a massive mutant hunt by Nusslein-Vollhard and others (see reading assignment #2). Embryonic development takes place outside the mother (unlike mice). System chosen as suitable genetic model system by George Streisinger, who used to work on phage T4

Over the course of the semester, we may read several papers that address how to examine problems in other systems, including those listed above, as well as mice. As genetic systems, higher organisms have one major disadvantage by comparison with microorganisms. In general, it is just not possible to grow enough organisms to observe the rare events that are easily found with microbial systems.

Model systems as surrogate organisms

Although *E. coli* and yeast don't have many of the complex biological features of other organisms, both can be very useful for the study of more complex processes. Recombinant DNA technology allows us to make chimeric organisms with genes from evolutionarily distant sources. This allows *E. coli* or yeast to act as a living test tube in which to reconstitute various complexes and reactions. The specific uses of these kind of chimeras will be discussed in detail as we reach specific topics over the semester. However, the general point should be mentioned now to encourage you to pay attention to the basic biology of microorganisms; their usefulness and relevance to your own research may be greater than you think.