Selections and screens

Papers to read for this section :

Szybalski, W. (1992) Use of the HPRT gene and the HAT selection technique in DNA-mediated transformation of mammalian cells: first steps toward developing hybridoma techniques and gene therapy. *Bioessays* **14**:495-500

Unlike the last two papers, the reading assignment for this lecture is more of a memoir than a description of experiments. This is one of the places where the material in this handout should be especially useful for the review.

As we have seen, genetics relies on having large numbers of organisms, preferably with short generation times. Isolation of mutants depends on having enough individuals and a high enough mutation rate to have a population that contains mutants. We have discussed some of the issues involved in obtaining such a population. A major concern of the Mullins paper was how to treat the fish to get the rate of mutation high enough to find mutant fish. Large numbers are also useful for other aspects of genetic studies. As described in Chapter 2, the resolution of recombinational linkage mapping depends on the size of the population you have available for study.

Even if you have a population that contains the desired mutants or recombinants, you have to find them. Finding the mutants or recombinants requires a phenotypic difference between the parents and the desired mutants or recombinants. There are two basic approaches to finding variants in a population: screening and selection.

The most important thing for you to take away from this chapter is an understanding of the difference between a selection and a screen:

- **Screen** - Experimenter examines a population with a mixture of phenotypes and identifies individuals with the a particular phenotype.
- **Selection** - Experimental design allows recovery of only phenotypic class, the experimenter only has to look at individuals that have a particular phenotype.

 ©1999 J. Hu all rights reserved

The rest of this chapter will be devoted to examples of selections and screens. On the one hand, these should illustrate what the difference means in practice. On the other hand, many of the examples represent genetic tools that are so commonly used that you should remember the basic mechanisms of how they work for practical reasons. For some of you, these methods will apply directly to your research; in any case, understanding them will help with the reading assignments later in the course.

Screens

Both of the mutant hunts we have examined identify the mutants by screening. Carlson had to look at about 44,000 colonies (one colony on the master plate and one on the replica plate for each survivor tested) in order to find the 31 mutants that she studied. Mullins screened for mutations by looking at lots of microscope fields containing zebrafish embryos. In both cases, the vast majority of the individuals examined by the experimenter had the phenotype she didn't want.

Screening is limited by the number of individuals the experimenter is willing to examine; a lot of individuals can be screened by brute force if you have the personnel and the patience. The massive screen for developmental mutants in the zebrafish that was done by the Nüsslein-Vollhard and Driever labs looked at about 10^6 fish embryos. This took hundreds of person-years, and was only possible because they had lots of people all doing it at once.

A screen can be based on anything you can observe. This can range from looking at the different colors of colonies on an indicator plate to grinding up organisms and purifying a specific protein. Genetic screens in higher eukaryotes often involve careful visual inspection. Suma Datta found an interesting *Drosophila* mutant in a screen that involved examining the brains that she dissected out of larvae. It is important to note in these cases, as in the screen for embryonic lethals described by Mullins, the screen detects a mutation in a dead animal. In order to make more mutants, one has to maintain the allele as a heterozygote which can be inbred (or bred with wild-type males in the case of an X-linked mutation) to give animals that express the phenotype.

 ©1999 J. Hu all rights reserved

It is possible to screen individuals for specific changes in their DNA sequences that have no observable physiological effect on the organism. The presence or absence of a DNA marker is a phenotype. It is now not only possible to screen for specific DNA sequences, it will become common practice in everyday life. Examining the BRCA1 gene by PCR to determine which women are at higher risk for breast cancer is an example of a genetic screen.

Improving the efficiency of an assay goes a long way toward increasing the throughput in a screen. As already mentioned, the change from using individual liquid cultures to looking at colonies on solid medium allowed very rapid visual screens. Reactions that lead to a color change are especially easy to use in colony assays; although reactions with purified enzymes can be easily examined by looking at things like the spectral shift in the absorbance of NAD^+ vs NADH, there's just too much other material that absorbs in a living cell to see that kind of reaction.

Indicator plates have dyes incorporated to cause colonies to turn different colors depending on whether or not the cells can perform certain reactions. Some of the simplest indicator plates use pH sensitive dyes to detect the acidification of the medium that occurs when sugars are fermented. One of the most useful chromogenic indicators is 5-Bromo-4-chloro-3-indolyl- -D-galactoside, commonly known as Xgal. Xgal is colorless, but the hydrolysis of Xgal by -galactosidase (the product of *E. coli*'s *lacZ* gene) releases an insoluble dark blue dye. This is the basis for the well-known "blue-white" screen used to find plasmids containing inserts that inactivate a fragment of *lacZ.* A small number of blue colonies are very easy to find in a sea of white colonies; this makes Xgal very useful for finding cells that express *lacZ.* Related compounds known as Xgluc and XP give the same blue color in the presence of -glucuronidases or phosphatases, respectively.

The usefulness of compounds like Xgal goes far beyond the study of the *lac* operon in *E. coli.* Long before molecular cloning as we currently know it, *E.coli* molecular biologists realized that engineering *E. coli* or its phages so that their favorite operons contained *lacZ* allowed them to use Xgal and other assays for -galactosidase to study the other operon. In this situation, *lacZ* is being used as a **reporter gene**, something that is more convenient to detect than the gene is substituting for.

 ©1999 J. Hu all rights reserved

E. coli lacZ is now used as a reporter in organisms ranging from bacteria to mammals (Plant molecular biologists prefer glucuronidase, also known as GUS).

There are a variety of reporter genes that are now commonly used. Each has different advantages and disadvantages, and it is always important to test whether something you learn using a reporter construct reflects the phenomenon you want to study, or if it reflects something related to the reporter or how the reporter construct alters the genes of interest.

Humans (i.e. students and postdocs) can screen large numbers of individuals, and, as noted above, screens that involved up to 10^6 individuals can be done with enough people, money and time. As with virtually every other form of human labor, however, automation can improve the speed with which assays can be done, which in turn improves the numbers of individuals that can be tested. Robots are now available that will pick colonies into microtiter dishes and perform plasmid extractions, enzyme assays or PCR tests. Susan Golden's lab screens cyanobacteria for mutants that affect the circadian expression of genes by using a computer-controlled CCD camera plate reader that records the light intensity generated by luciferase reporters at different times. Fluorescence activated cell sorters (FACS) pass a stream of cells by a laser light source; detectors measure fluorescence of each cell as it passes, and the stream can be diverted into different collection vessels. Using fluorescent stains or Green Fluorescent Protein (GFP) reporters, FACS machines can screen hundreds of cells per minute.

Selection and enrichment

Recall the definition of a selection. In a selection, the experimental design allows recovery of only phenotypic class, the experimenter only has to look at individuals that have a particular phenotype. What do I mean by this and how is it different from a screen? Let's look at an example of a selection and think about what would be needed to accomplish the same thing by screening.

Selection for ampicillin resistance as a means to isolate bacteria with recombinant DNA plasmids is a method that is probably used in almost every molecular biology lab in the world. In a typical experiment, you want to put a particular plasmid or mixture of plasmids into a strain of *E. coli*. To do this, you mix plasmid DNA with competent cells that you either made or bought. You do something to

 ©1999 J. Hu all rights reserved

the mixture to encourage the cells to take up the DNA, add some rich medium such as LB or 2XYT broth, and plate an aliquot on an agar plate containing ampicillin. The next day there are ampicillin resistant colonies on the plate, assuming that everything worked.

Why is this a selection? In the aliquot that you plate from the transformation mixture of competent cells and DNA, there are probably about 10^8 cells. Imagine that you get 1,000 colonies. This means that out of about 100,000,000 cells that you put onto the plate, you don't have to deal with about 99,999,000 of them at all. The experimental design, i.e. drugs in the agar, got rid of them for you.

How might you do this as a screen? Suppose you wanted to introduce a plasmid that expressed *lacZ*, but did not carry any drug resistance markers into *E. coli*. In principle, you could do the transformation exactly in the same way, but instead of ampicillin plates you would use plates containing Xgal. To screen for plasmid-containing cells you would have to look at the plates and find the blue colonies. If you plate the cells so that there are 1,000 colonies per plate, you would about 100 plates to have an even money chance of finding one blue colony. For the ampicillin selection, you only needed one plate, and you get 1,000 colonies. This is an advantage of about 5 orders of magnitude in sensitivity!

This example illustrates the principle advantage of selections: they allow you to to find rare individuals among large populations. Selections lose their practical advantage if either the individuals you want are not rare, or if the population available is so small that you might as well check every individual.

Microbial geneticists love selections. Microbial systems usually grow to very high population densities compared to multicellular eukaryotes. However, it is still often practical to perform selections on higher organisms; the utility of trying a selection will depend on how hard it would be to find the same thing by screening. As described in Szybalski's paper, selections are very useful in genetic manipulation of tissue culture cells. We will see how that is an important factor in our ability to generate mutations in specific genes in mice.

Historically, selections have been used to identify mutations genes involved in important cellular functions and to provided quantitation for rare events such as mutation and recombination. We've

 ©1999 J. Hu all rights reserved

already seen an example of the use of selection in fine-structure recombination mapping, both in the Carlson paper, where the ability to select for sucrose utilization allowed quantitation of rare recombinants between *suc2* alleles, and in the description in Chapter 2 of Benzer's work on the rII mutants of phage T4. We've also seen how Carlson also selection for simultaneous reversion of two amino acid auxotrophies to select for amber suppressors. The cloning example above illustrates the usefulness of selection in constructing strains; Szybalski describes how selections were useful in making hybrid tissue culture cell lines.

Different kinds of selections

Selections usually involve some set of conditions where the majority organisms are killed or do not grow. For example, selecting for bacterial that survive infection by phage T1 tends to yield mutants that have lost the surface receptor for T1 attachment. Cell that have not lost the receptor are killed by the phage; the treatment is **bacteriocidal** (a treatment that kills yeast would be fungicidal). In contrast, selection for sucrose utilization allows Suc$^+$ cells to grow while Suc$^-$ cells starve for carbon. Differential reproduction allows the former to take over the population; however, the Suc$^-$ cells aren't dead, they're just sitting there waiting for a carbon source. Starving yeast cells for a nutrient is an example of a treatment that is **fungistatic** (for *E. coli* it would be bacteriostatic).

Although the most commonly used selections are based on death or growth, there are other kinds of selections. **Physical selections** remove either the desired or the undesired subset from a population. For example, a culture of cells could be immunoprecipitated with an antibody against a particular surface antigen. If you start with a population that has the antigen, then the supernatant will be enriched for mutants that have lost the ability to react with the antibody.

This kind of physical selection can be used with any kind of ligand that can be used to physically separate individuals in the population. In a method called **phage display**, proteins or peptides are expressed as fusion proteins so that they are on the outside of a phage particle. A population of phage containing particles that bind or don't bind a particular ligand can be separated by incubating the phage particles in the plastic plates where the ligand is used to coat the surface. The binders stick to the plate,

 ©1999 J. Hu all rights reserved

and the nonbinders can be washed away. This process is called **biopanning**, by analogy to panning for gold particles in the sediments of a stream. The variety of ligands that can be immobilized is limited only by chemistry and the imagination of the investigator. Lately, a common way to do physical selections involves coupling the ligand of choice to biotin. The cells that bind the ligand will display biotin on their surface; the biotin can be bound by magnetic beads coated with avidin, a protein that binds biotin with very high affinity. A simple magnet can then be used to pull the beads out of the mixture, bringing anything directly or indirectly attached with it.

Selections can also be based on behavior. A famous example comes from the isolation of the *sevenless* mutation in *Drosophila*. The same Seymour Benzer who studied the rII mutations in phage T4, switched to studying behavioral mutants in flies. Benzer and his coworkers have isolated a number of mutants that alter the ability of flies to respond to different stimuli (XX find ref for sevenless isolation - need a diagram of the apparatus) or to learn to associate a specific stimulus with a negative reinforcement.

An example of this kind of selection is shown in figures 4-1 and 4-2. Even animals as simple as Drosophila can be given Pavlovian training (Figure 4-1). Flies are placed in a tube with a wire grid where they land. A volatile odorant is wafted through the chamber, and at the same time an electric current is passed through the grid to teach the flies to associate that odor with pain. This can be repeated until the flies learn to avoid the odorant.

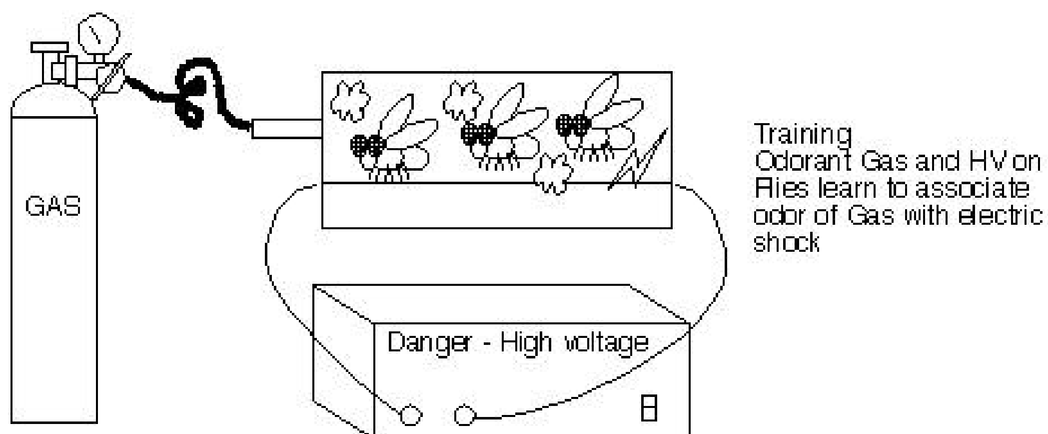


Figure 4-1 Pavlovian training for fruit flies

 ©1999 J. Hu all rights reserved

Mutants that fail to learn can be selected in an apparatus like the one shown in Figure 4-2. F3 flies from a mutagenized population are trained and then placed in the bottom of a tube shaped like a T. At the top, the training odorant is injected at one end. Normal flies will fly up the tube, smell the odorant and fly the other way to avoid the expected electric shock. Mutants that can't learn to associate the odor with the shock will go one way or the other at random; flies that are at the end of the T-tube where the odorant is entering will tend to be mutants.

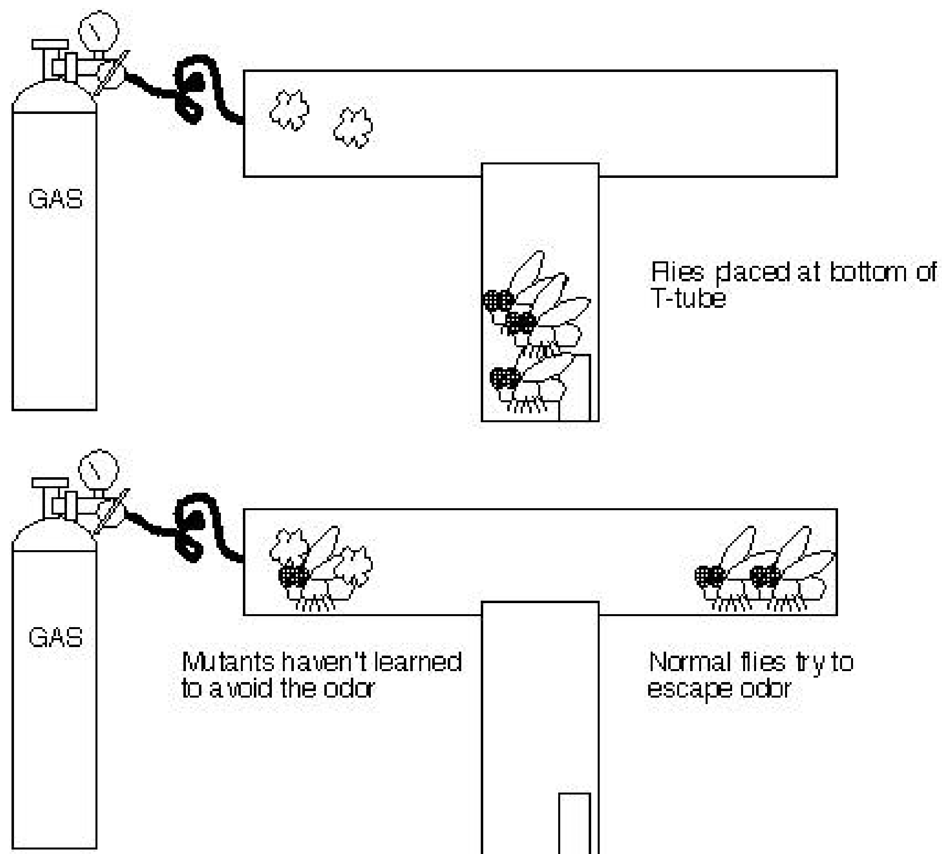


Figure 4-2. Selection for behavioral mutants

This kind of setup can be used to find mutations in genes that affect important processes in the brain, such as the ability of a fly to place a learned behavior into long term memory. On the other hand, you will also get mutants that can't learn because they have defects in their olfactory organs - the genetic equivalent of a stuffy nose. If the flies can't detect the odorant, then they can't associate it with the electric shock. The training will just seem like random shocks to them.

 ©1999 J. Hu all rights reserved

Some selections you should know about

In the rest of this chapter, I will review some of the selections that you should try to learn in more detail. Many of these are used in later reading assignments, or are things you are likely to encounter in the lab or at seminars.

Drug resistances

A wide variety of natural products kill or inhibit the growth of prokaryotes, eukaryotes or both. These include the drugs we call antibiotics, such as ampicillin, tetracycline, streptomycin etc. Antibiotics generally act by binding to a specific target enzyme and either inhibiting it or subverting its normal activity. Resistance to antibiotics can be acquired by several different mechanisms, including

- Modifying the target
- Overexpressing the target
- Bypassing the need for the target
- Modifying the drug
- Preventing the drug from reaching the target

Mutations that modify a drug target to a resistant form have been very useful in identifying genes encoding proteins involved in essential cellular activities. For example, rifampicin is an antibiotic that shuts down RNA synthesis in sensitive bacterial cells. It could be shown that rifampicin blocks RNA synthesis by purified RNA polymerase *in vitro.* Rifampicin resistant mutants of *E. coli* can be isolated by plating a sensitive strain on plates containing the drug. Since purified RNA polymerase from these strains is also resistant to inhibition by rifampicin *in vitro*, we can conclude that the enzyme has been altered; thus mutations to rif resistance identify one or more of the subunits of *E. coli* RNA polymerase. In fact, this is how the first gene for an RNA polymerase subunit was identified. *rif* mutations map to *rpoB*, the gene that encodes the subunit of *E. coli* RNA polymerase.

Table 4-1 lists some antibiotics and the target proteins that can be mutated to give a resistant phenotype. In many of these cases, the resistant phenotype allowed mapping or cloning of genes that encoded the target protein. Note that this is far from a comprehensive list.

 ©1999 J. Hu all rights reserved

Table 4-1 Inhibitors of specific functions and their targets		
Drug		Target
RNA synthesis inhibitors		RNA polymerase subunits
	Rifampicin	
	Streptolydigin	
	-amanatin (euk)	
Protein synthesis inhibitors		
	Streptomycin	ribosomal protein S12
	Kasugamycin	ribosomal protein S2
	Spectinomycin	ribosomal protein S5
DNA synthesis inhibitors		
	Naladixic acid, Coumermycin, Novobiocin	DNA gyrase (*gyrA*, *gyrB*)
	Acridine	DNA polymerase III large subunit (*dnaE*)
	Trimethoprim, Methotrexate, aminopterin	Dihydrofolate reductase (needed to make purines and thymidine)
	Fluorouracil	Thymidylate synthetase
Structural proteins (euk)		
	benomyl, taxol, colchicine, vinca alkyloids	tubulin

The mutations described above alter the target of the drug. Another important route to drug resistance is through the acquisition of drug resistance genes. These are genes that are not found in the wild-type host, but are introduced by horizontal gene transfer. Drug resistance genes were mostly discovered on transmissible elements from clinical isolates of bacteria. Clinically, they their horizontal transfer among pathogens represents a serious threat to human health. However, molecular biologists think of drug resistance genes as convenient genetic markers; they allow selection of cells that contain the gene (the specific genes used in most molecular biology are so widespread already that the possibility of even a large-scale accidental release from a lab would be an insignificant contribution to what's already in the environment). How these kinds of selectable cassettes can be used should become apparent later in the course.

The ampicillin resistance conferred by plasmid-encoded -lactamase is an example of a drug

 ©1999 J. Hu all rights reserved

resistance gene of this kind. -lactamase modifies the drug, ampicillin or one of its relatives in the -lactam family of antibacterials, by cleaving it. Other drug resistance genes encode enzymes that modify their substrates by acetylation (e.g. chloramphenicol acetyl transferase) or phosphorylation (e.g. neomycin phosphotransferase).

Resistance to tetracycline conferred by most laboratory plasmids operates by a different mechanism. The tet genes encode a membrane protein that pumps the drug out of the cell. Active export of antibiotics has recently been recognized as an important mechanisms of drug resistance. Many tumor cells become resistant to chemotherapy because treatment with the chemotherapeutic agents selects for cells that have deregulated the expression of export pumps. These pumps, which are also found in bacteria, are often able to export a wide variety of drugs, and are called multidrug resistance (MDR) pumps.

Selections for and against the same function

In a screen, you can pick individuals that either have or don't have a particular phenotype. While the power of selections is in their ability to discard large fractions of a population, it can be very useful to be able to isolate individuals with the other phenotype. Szybalski's paper describes a pair of protocols that are especially useful because one selects *against* the function of an enzyme, hypoxanthine guanine phosphoribosyl transferase (HGPRT, or HPRT), while the other protocol selects *for* the function of the same enzyme.

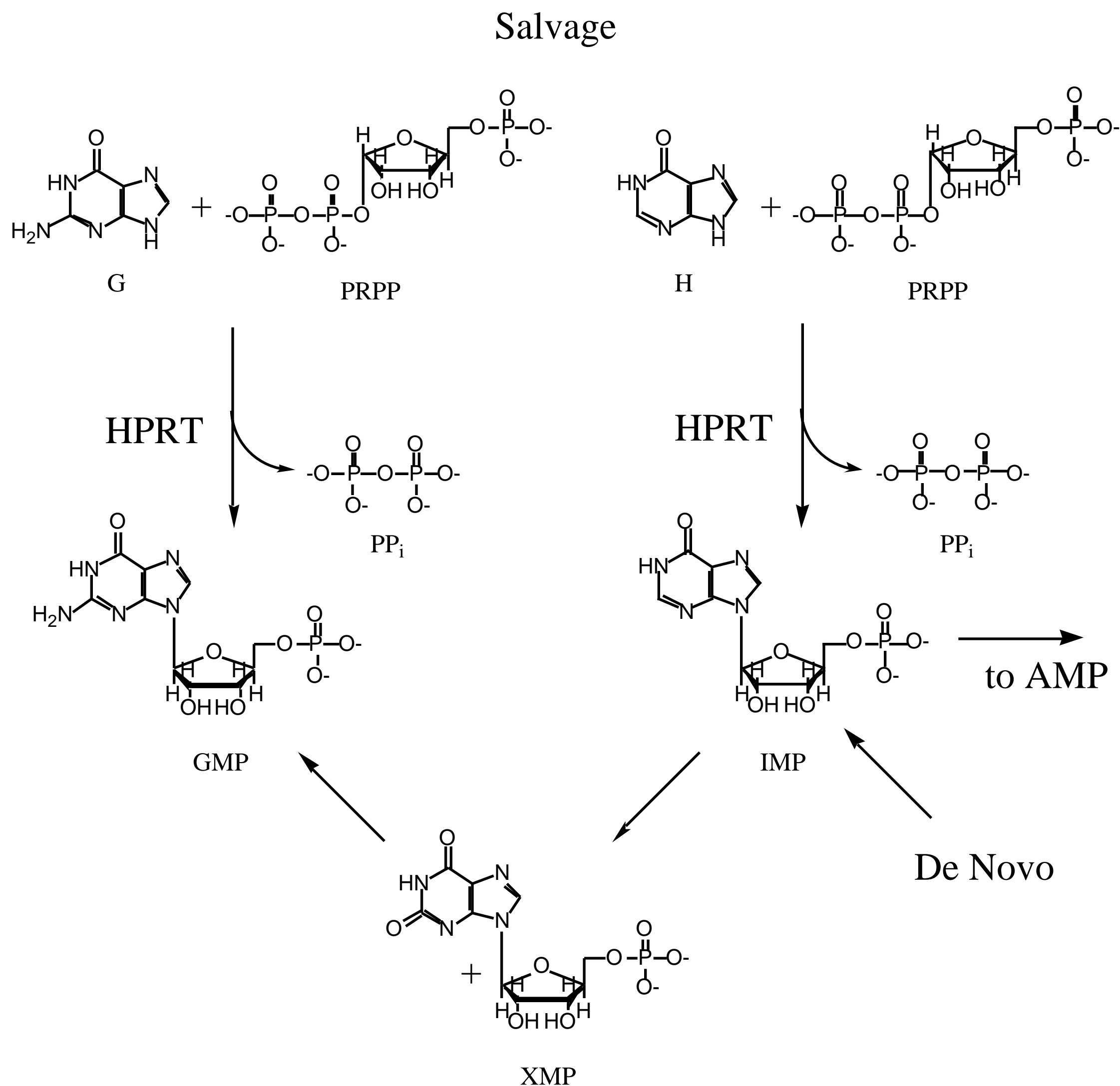
In this section, we will examine two pairs of selection protocols; each pair allows selection for or against the same gene. First we will go through the selections described by Szybalski, which select for and against hprt. Then we will describe selection for and against the activity of orotidine 5'-P decarboxylase, the product of the *URA3* gene in yeast and the *pyrF* gene in *E. coli*. These are probably the two most common selections used to go back and forth between opposing phenotypes, although there are others.

Both pairs of selections we will discuss are based on the biochemistry of nucleotide metabolism. To understand what is going on, it will help to review some general principles as well as some specific

 ©1999 J. Hu all rights reserved

steps in nucleotide metabolism.

For both purine and pyrimidine nucleotides, there are two kinds of pathways that feed into the cellular NTP pool. In the *de novo* pathways, nucleotides are synthesized by a series of steps leading to IMP in the purine pathway, or UMP in the pyrimidine pathway. In the salvage pathway, preformed bases or nucleosides are converted into nucleosides and nucleotides, respectively. In organisms with both pathways, the *de novo* pathway can make all of the needed NTPs and dNTPs without the need for any added bases or nucleosides in the medium. However, if bases or nucleosides are in the medium, the salvage pathway provides an energetically cheaper way to get to nucleotides.



 ©1999 J. Hu all rights reserved

Figure 4-3. Pathways leading into purine nucleotides.

Purines, azaguanine and HAT

Some of the key steps in purine metabolism are shown in Figure 4-3. The key player in this pair of selections is HGPRT, which can couple free guanine or hypoxanthine to 5-phosphoribosyl 1-pyrophosphate (PRPP); the base goes onto the 1' position of the sugar and inorganic pyrophosphate is released as the leaving group. If cells are fed guanine, HGPRT is used to make GMP. If cells are fed hypoxanthine, HGPRT makes inosine monophosphate (IMP), which can be converted to either GMP or AMP. Thus, HGPRT is a key player in the salvage pathway for purines. A different enzyme, APRT, can convert free adenine into AMP.

To select against HGPRT activity in tissue culture cells, Szybalski fed the cells 8-azaguanine (Figure 4-4). In 8-azaguanine, one of the carbons in the 5-membered ring is replaced by a nitrogen atom. If this base analog is converted by HGPRT into 8-azaGMP, it can be phosphorylated into the diphosphate and triphosphate and incorporated into nucleic acids. Although I am not sure of the mechanism of toxicity, 8-azaguanine is nontoxic unless it can be metabolized through the salvage pathway by HGPRT. Thus, the rare cells that carry mutations that inactivate HGPRT will survive the treatment and will grow in the presence of the base analog. Wild-type cells will die. 8-azaguanine selects for mutants defective in HPRT activity.

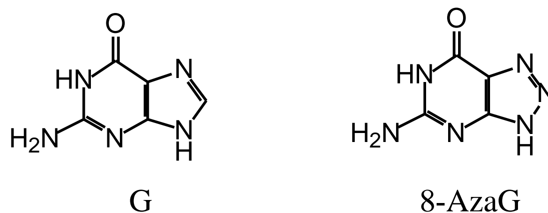


Figure 4-4. Guanine and 8-azaguanine.

To select for HGPRT activity, we need a condition where only the cells that have it can grow. We need to make the salvage pathway essential. Szybalski found that such a condition was met by HAT medium, which contains hypoxanthine, aminopterin and thymidine. Aminopterin is an inhibitor of dihydrofolate reductase, which converts dihydrofolate into tetrahydrofolate. Inhibiting DHFR blocks the *de novo* synthesis of both purines and thymidine, since 1-carbon units carried by tetrahydrofolate

 ©1999 J. Hu all rights reserved

are required in two of the steps in purine synthesis and in the methylation of uracil by thymidylate synthase. In wild-type cells, the hypoxanthine is converted by HPRT into IMP, which can then be converted into both AMP and GMP. The thymidine in the HAT medium is converted by thymidine kinase (TK) into TMP, overcoming the other block due to aminopterin. Note that HAT medium also selects for the presence of active TK.

Selections for or against URA3

One of the most commonly used markers in yeast genetics is the *URA3* gene. Like HGPRT, it is possible to exploit our understanding of the biochemistry of nucleotide metabolism to select for or against the presence of active orotidine 5'-P decarboxylase (Ura3p). The relevant reaction is part of the *de novo* pathway of pyrimidine synthesis (Figure 4-5)

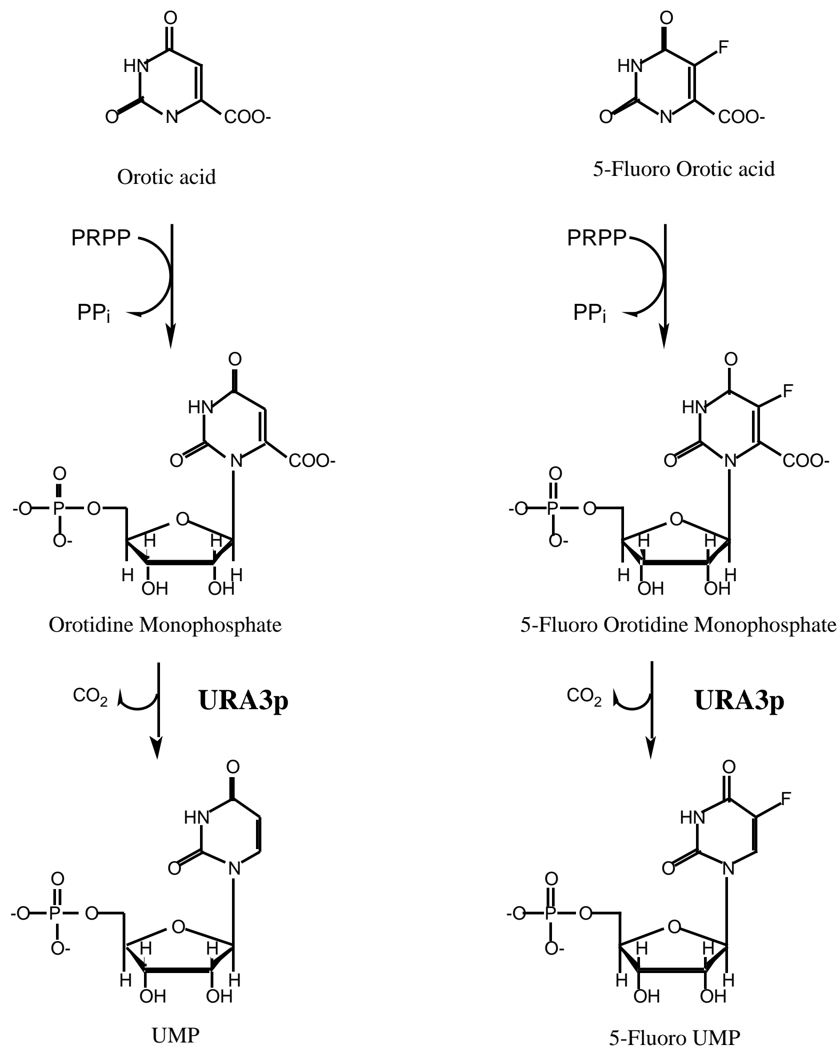


Figure 4-5 Reactions catalyzed by the *URA3* gene product, orotidine 5'-P decarboxylase.

 ©1999 J. Hu all rights reserved

The left side of Figure 4-5 shows the last step in the biosynthesis of UMP. Unlike purines, where the hypoxanthine base of IMP is assembled onto PRPP, pyrimidine biosynthesis builds a free base in the form of orotic acid, which is then assembled onto the sugar-phosphate. Orotidine monophosphate (OMP) is then decarboxylated by the product of the *URA3* gene to yield UMP, which by a series of reactions is converted to UTP, dTTP, CTP and dCTP.

A *ura3* strain is a uracil auxotroph, it requires uracil or uridine in the medium to grow. Thus selection for *URA3* is very easy; you just plate on minimal media lacking uracil or uridine. Since URA3p is part of the *de novo* pathway, it will be needed to provide a source of UMP and everything derived from UMP.

Selection against *URA3* (i.e. for *ura3*) is performed by using media containing 5-fluoroorotic acid (5FOA) and uracil. 5FOA is converted by orotate phosphoribosyl transferase to 5-FOMP, consuming PRPP and releasing pyrophosphate. URA3p converts the 5-FOMP into 5-FUMP.

Fluorouracil and fluorouridine are potent anticancer drugs. Like 5-FOA, they are converted into 5-FUMP; however, they get there through the pyrimidine salvage pathway. 5-FUMP is a suicide inhibitor of thymidylate synthase, forming a covalent complex with the enzyme's active site and permanently inactivating it. Cells treated with fluorouracil can't make dTTP, and are thus unable to replicate their DNA. The same thing will happen to *URA3* cells treated with 5FOA, since URA3p converts the nontoxic 5-FOMP into the toxic F-FUMP. Since ura3 cells can't create the toxic product, they grow in the presence of 5-FOA, as long as you provide the uracil that they can't make *de novo*.

Why doesn't this also select for loss of orotate phosphoribosyl transferase? Think about some of the reasons discussed in Chapter 2 for why one would fail to isolate mutations in a gene. It turns out that there are two genes for orotate phosphoribosyl transferase, *URA5* and *URA10*. Two mutations would thus be required to lose this enzyme activity; as we have seen, this is a low probability event.

The selections against HGPRT and URA3p both involve the conversion of substrate analogs to toxic products. This is a general strategy for selections against enzymatic activities; in practice, its application is limited by the availability of the appropriate substrate analogs.

 ©1999 J. Hu all rights reserved

HAT selection and growth in the absence of uracil both select for the functions of the genes of interest. However, note that HAT selection can be applied to tissue culture cells, while the selection for URA3p activity is limited to organisms that grow in defined media.

 ©1999 J. Hu all rights reserved