Why Yeast?

DAVID BOTSTEIN *Stanford University*

Yeast is an apparently strange "model" for the human genome, but it works because of the high degree of conservation in evolution between the primitive eukaryote and mammals. Moreover, the genic concentration in yeast—with almost no noncoding introns—packs the entire yeast genome into 16 chromosomes with only 10% of the DNA of one human chromosome.

Human biology permits limited insight into itself because humans are not experimental animals. Cell biologists have long relied on simpler systems to reveal basic properties of the human cell. Likewise, molecular geneticists are gaining insight into the human genome through studies of model organisms. Several species that seem only remotely related to humans have more in common with *Homo sapiens* on the molecular level than meets the eye. The yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* are examples. Evolutionarily unrelated, these eukaryotic microorganisms share a number of fundamental cellular and molecular properties with humans. This homology, combined with the experimental tractability of yeast, offers researchers a useful model for linking gene structure with protein function and applying the findings to human biology. Without gene-protein-function correlations, the mapping of all 50,000 to 100,000 human genes—a central aim of the Human Genome Project—would have limited value.

Acquiring such data generally requires experimental manipulations. Protein function can be determined by demonstrating the dysfunction that results when the gene that codes for the protein is absent or mutated in an organism. Human experimentation of this sort is clearly out of the question. It is sometimes possible to link a human phenotypic abnormality to a gene, protein, and protein function through a combination of epidemiologic studies, clinical observations, and laboratory research using samples of patient DNA.

Dr. Botstein is Professor and Chairman, Department of Genetics, Stanford University School of Medicine.

Such linkage studies have been carried out for a number of genetic diseases, including Duchenne type muscular dystrophy and, most recently, Marfan syndrome. But in most cases, other experimental modalities are required to link gene structure with protein function. This is where yeast is making a key contribution.

The question might arise, Why not bacteria? After all, bacterial systems (particularly *Escherichia coli*) provided the foundation for classic studies of molecular genetics and for the advent of recombinant DNA technology. Certainly, bacteria have done much to further our understanding of the human genome. The problem is that bacteria are too far away evolutionarily from higher eukaryotes. Bacteria organize, duplicate, and separate their chromosomes by means different from those in eukaryotes; bacteria and eukaryotes also differ in the type and function of their subcellular organelles; and they differ in metabolism and homeostatic regulation. Molecular genetic studies of eukaryotes require experimental systems closer in evolution than bacteria can provide.

Eukaryotic Models

Such experimental systems are provided by eukaryotic microorganisms. They offer the experimental tractability of bacteria (simplicity, rapid growth, amenability to recombinant DNA manipulations) along with many basic biologic properties of complex eukaryotic cells: cytoskeletal organization, subcellular organelles, secretion, receptor and second messenger systems, metabolic regulation, and chromosome mechanics. The eukaryotic microorganism that has been most highly de-

veloped as an experimental system is *S. cerevisiae*, a free-living yeast with excellent classic genetics. It is also the species that is used to produce bread, beer, and wine. *S. pombe*, a less well-developed model system, has many of *S. cerevisiae*'s advantages and some biologic differences that make it more or less useful in different experimental contexts. *S. pombe* has larger chromosomes; *S. cerevisiae* has a more tractable mitochondrial genome. Some laboratories now use both systems in order to make useful comparisons between the two.

The yeast genome contains the basic blueprint of all eukaryotic cells in miniature. *S. cerevisiae*'s genome is one 200th the size of the human genome. (By comparison, the fly's is one 20th the size of the human genome). In other words, there is roughly 10 times less DNA in the entire yeast genome localized on 16 chromosomes than in a single human chromosome.

How is it possible that such a small instruction set yields a eukaryotic cell that shares so many of the properties of cells in much more complex organisms? The answer lies in the packaging of genetic information. Human genes are interrupted by much longer stretches of uninformative DNA. In yeast, coding sequences are compressed into smaller, denser packages, interrupted by fewer introns; a given gene will be 10 or more times larger in the human genome than in the yeast genome. Genes for the proteins tubulin, YPT, and actin lie, for example, within four kilobases of each other in the yeast genome; in the human genome, they would generally be megabases apart. Similarly, yeast sequences coding for HMG-CoA reductase are concentrated in 3.2 kb, whereas human sequences coding for the enzyme are distributed in 25.3 kb (Figure 1).

Gene Conservation

The linking of gene structure with protein function in yeast would have little application to human molecular genetics if it were not for the evolutionary principle of gene conservation. Complex organisms are, to a large extent, elaborations on the design of primitive species; the genes that carry these basic design instructions—and the proteins they encode—have been passed on from generation to generation in much the same form. Many proteins involved in basic biologic processes have such a high degree of amino acid sequence identity in different eukaryotic organisms that they are functionally interchangeable. Yeast and human ubiquitin are 96% identical; for actin and tubulin the rates are 89% and 75%, respectively.

One could say that we are to yeast, worms, or flies what a Cadillac is to a Model T: built on the same principles out of essentially the same parts. If a Cadillac steering wheel were to break, it could be replaced (with some difficulty) with a steering wheel from a Model T. The Cadillac and Model T steering wheels are functionally interchangeable to the extent that both are round, can be grasped, and turn wheels. Some features of the Cadillac, such as the air conditioning, have no counterpart in the Model T, of course; but equivalent parts—wheels, brakes, carburetor—will function essentially the same.

By analogy to the steering wheel, tubulin (a key protein in cell division) is functionally interchangeable in the Model T (yeast) and Cadillac of eukaryotes (humans). Yeasts grow first by budding; buds of different sexes then join to produce a zygote that reproduces much like a mammalian cell. In the division process, yeasts form mitotic spindles, which look nothing like the more elaborate spindles in rat or human cells but function in essentially the same way: Chromosomes are lined up and pulled apart by a mechanism in which tubulin plays a prominent role. A monoclonal antibody directed against rat tubulin will tag yeast tubulin because of the extensive structural homology between the two proteins. (This can be visualized by immunofluorescence microscopy.) Despite billions of years of evolution, the structure of tubulin in yeast, rats, chickens, and humans is more alike than different (Figure 2).

Homologous Proteins

The number of homologous proteins in yeasts and humans is quite large and corresponds to a broad range of functions. Yeasts and humans share homologous secretory proteins, heat-shock proteins, transcription factors, cytochromes, ucyclins, G proteins, and oncogenes. Yeasts and humans have several functionally interchangeable forms of the *ras* oncogene. The yeast STE2 protein that recognizes a mating factor is similar in structure to the human β-adrenergic receptor, which plays a role in the regulation of blood pressure in humans. HMG-CoA reductase—66% identical in amino acid sequence in yeasts and humans—acts at the beginning of the steroid pathway in both species (although yeasts have no cholesterol) and has been found

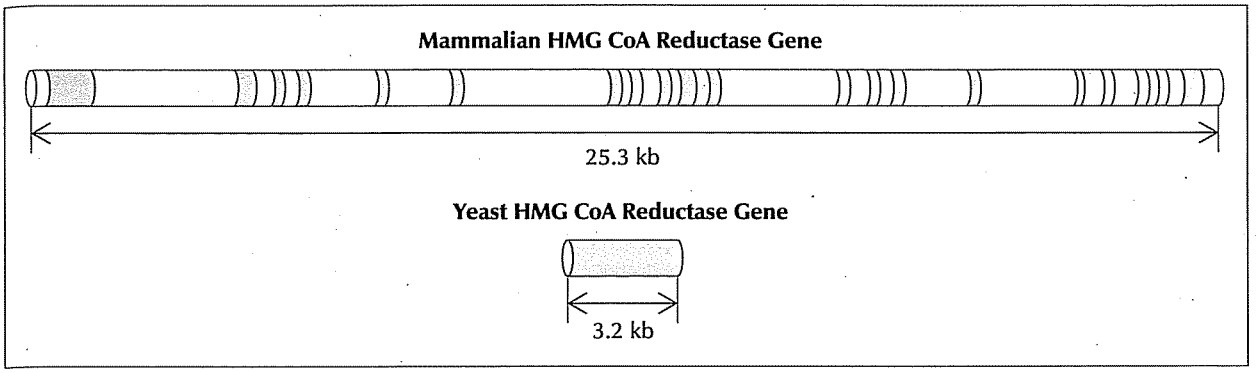


Figure 1. **Because it is more densely packed with coding information, the yeast genome represents a miniature version of genomes in higher eukaryotes. An example is provided by the mammalian and yeast genes that encode an essentially homologous protein, HMG-CoA reductase. The mammalian version extends to 25.3 kb, most of it taken up by noncoding introns (yellow), compared with a mere 3.2 kb in yeast, all of it consisting of exons.**

to be functionally interchangeable in yeasts and humans.

Functional interchangeability of a protein in two different species can be demonstrated by the following experimental method. First, the yeast gene that codes for the protein under investigation is cloned and mutated by chemically altering the sequence of nucleotides in the gene. The simplest way to mutate a gene is to cause it to be deleted. The deleted gene is then transferred into normal cells, where it can be caused to replace the normal gene, disrupting gene function. In most cases, this will prevent the yeast cell from reproducing—in effect, killing the cell.

In the next step, a mammalian protein is expressed in the yeast cells containing the deleted gene. This is done by inserting copies of the gene that codes for the mammalian protein into the cells by way of a yeast expression vector system. (Several different kinds of yeast vector systems have been developed, allowing for different forms of DNA propagation.) Once inside the yeast cells, the mammalian gene initiates synthesis of normal mammalian protein. If the protein brings the cell back to life, this demonstrates that the mammalian protein is homologous in function to the protein produced by the normal version of the mutant yeast gene. The yeast and mammalian proteins are, in other words, similar enough in amino acid identity to be functionally interchangeable. In this way it can be shown, for example, that when the human N-*ras* protein (which is involved in signal transduction and is a major actor in many cancers) is expressed in yeasts, it suppresses a yeast double *ras* mutant protein.

Protein Function Studies

It is possible to learn a great deal about protein function in human cells by studying the function of homologous proteins in yeast. Different experimental approaches to determining yeast protein function can be taken, depending in part on the raw material at hand.

From protein to gene. Let us say a researcher has a purified yeast protein of unknown function encoded by an unknown gene. (Purifying the protein will have involved separating it from other proteins in the cell on the basis of various characteristics, including size, charge, shape, or ability to bind to other proteins.) The researcher's first task will be to sequence the protein, or at least some of its initial amino acids. This is done in a protein sequenator, which chops amino acid residues one by one from the end of a protein and "reads" them. It is generally impossible to sequence an entire protein (which may be up to 3,000 amino acids long). But sequencing the first few residues is usually enough to determine the gene that codes for the protein: The odds that even seven amino acids will appear in a specific order at the beginning of a protein by chance alone are 1 in 64 million.

The researcher's next task is to reverse-translate the amino acid sequence that has been obtained from the protein sequenator into its corresponding DNA sequence. The genetic code can be used either to translate DNA codons into amino acid residues or to reverse-translate residues into codons. This in itself does not reveal the function of the protein, of course. Using the genetic code to translate nucleotides into amino acids, or vice versa, is like translating an Uzbeki text into Urdu when you speak neither. There is no Rosetta Stone in molecular biology on the

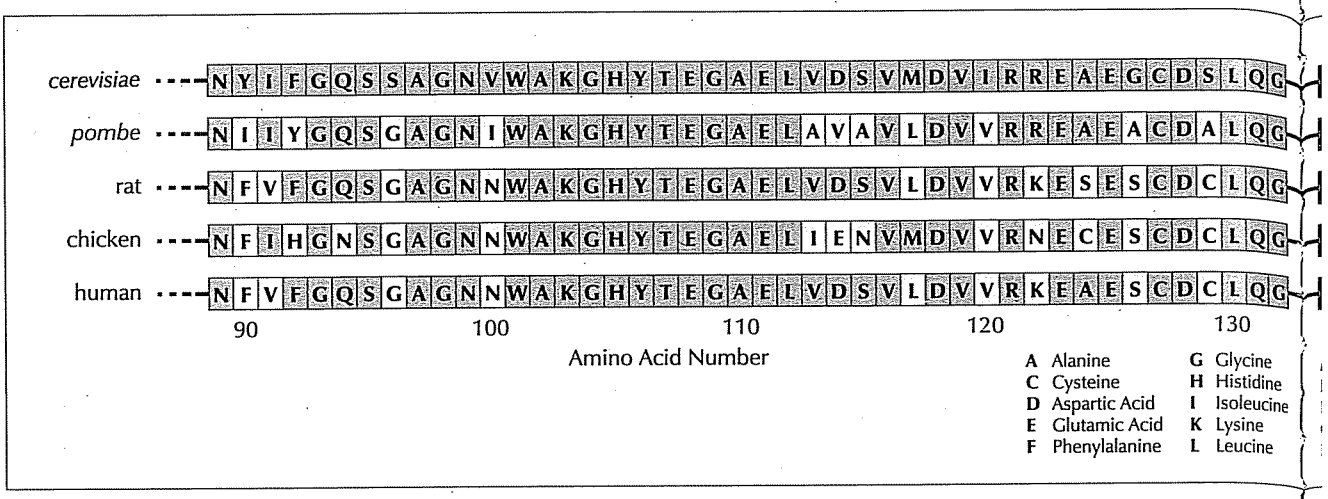


Figure 2. **Gene conservation is a major reason why yeasts serve as an important model organism for studies of the human genome. This is illustrated by the genes coding for the protein tubulin in yeasts, chickens, rats, and humans. A rep-**

basis of which a protein's function can be read from its nucleotide and amino acid sequence.

Because some amino acids are specified by more than one codon, reverse-translation of an amino acid sequence into DNA codons usually yields not one but several possible nucleotide sequences that could code for the protein under investigation. (A stretch of seven amino acids could yield as many as 100 nucleotide sequences, for example.) To determine which of these nucleotide sequences belongs to the gene that codes for the protein, all candidate sequences are synthesized, then radioactively tagged and used to probe a library of clones—yeast cells (in this case) that each contain DNA corresponding to one gene. Only the nucleotide sequence essentially identical to that of the gene that codes for the protein will hybridize with a clone in the library. This can be visualized radiographically, revealing the location of the hybridized clone in the culture.

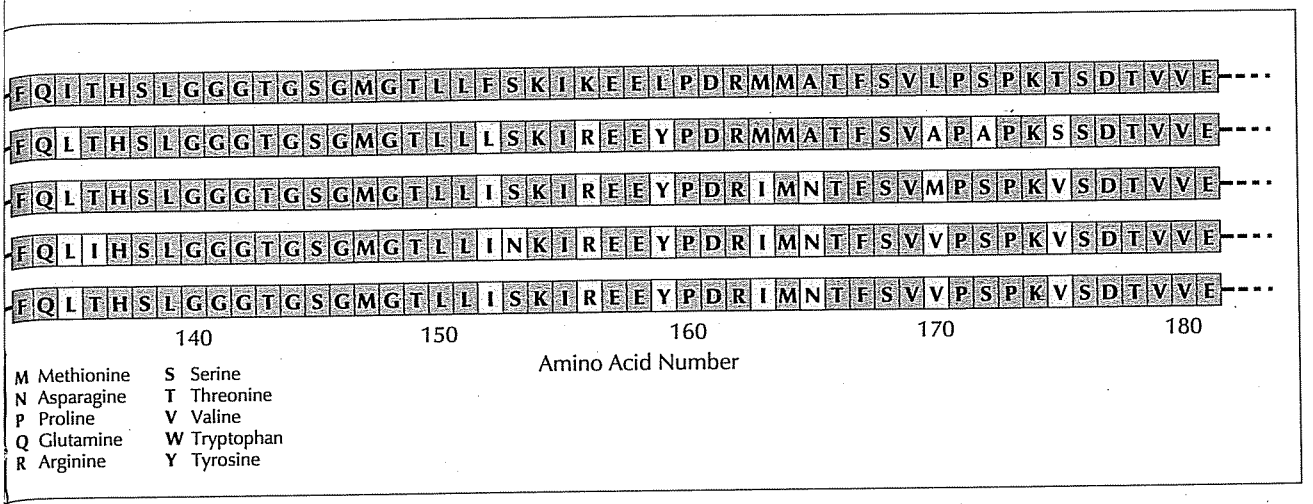
Next, the researcher must isolate the yeast cell containing the hybridized clone from the other clones in the culture. Once isolated, the gene can be sequenced, and its sequence translated into the complete amino acid sequence for the protein. This can then be checked against the amino acid sequence initially generated by the protein sequenator.

Through this process, the researcher obtains the complete sequence of both the protein and the gene that codes for it. The function of the protein can then be analyzed by mutating the gene and observing what effect absence of the normal protein has in the cell.

From gene to protein. In some cases, the point of entry into the gene-protein-function triad will not be the protein but the gene. A molecular geneticist may have as raw material a strain of mutant yeast that does not replicate. The researcher will want to determine which gene in the yeast is mutated, and which protein the normal version of the gene encodes. This can be done by adding a library of normal yeast clones to a plate of the mutant yeast. Each mutant cell will pick up a normal clone. One of these clones will contain the good copy of the mutant gene, which will begin producing normal protein in the cell into which it has been introduced. The cell that has picked up the good copy of the gene will be the only cell on the plate that begins replicating. The colony of replicating cells can then be isolated, and the nucleotide sequence of the mutant gene determined.

Next, the protein coding portion of the gene's nucleotide sequence must be determined by teasing out the exons from the noncoding introns. This is done by scanning a gene's nucleotide sequence for the codon ATG (which codes for methionine, the first residue in every protein) followed by a long open reading frame. Both strands of DNA in the double helix are checked, as it is not known initially which is the coding strand. Once the gene's coding sequence has been located, the amino acid sequence of the protein can be translated from the DNA.

Determining a sequenced protein's function may begin with a computer search of databases containing previously identified protein sequences. That will reveal whether any sta-



resentative sequence from the yeast *Saccharomyces cerevisiae* is depicted in red at top; variances in that sequence in another yeast (*Schizosaccharomyces pombe*) and the chicken, rat, and human are depicted in blue.

tistically significant homologies exist between the protein under investigation and one or more in the database. If such is the case, the proteins may be functionally related. Ideally, this hypothesis would then be tested in an assay. Such an assay, however, might not exist. Or no homologous proteins might turn up in the first place.

ADP Ribosylation Factor

A positive example in this regard is provided by yeast mutant studies of adenosine diphosphate ribosylation factor (ARF). A protein that was obtained from bovine and human cells and purified, ARF had been available in pure form for 10 years without anyone knowing its function. It is, we now know, a guanine triphosphate binding protein common to humans, cows, sheep, frogs, flies, worms, plants, and yeast. Together with cholera toxin, ARF causes G protein ADP-ribosylation—but what does ARF do by itself? We determined this through yeast mutant studies by creating a series of mutations in the genes that code for ARF and another GTP-binding protein, YPT. The mutations resulted in aberrant glycosylation of other proteins, indicating that both ARF and YPT are involved in the pathway of protein secretion. Since glycosylation of proteins occurs during secretion, aberrations in glycosylation are thus symptoms of a defect in secretion.

Through mutant studies, the yeast research community has been been able to determine the entire yeast secretory pathway—which, it turns out, is fundamentally identical to the human secretory pathway. Humans have ARF and YPT proteins that are functionally interchangeable with yeast ARF and YPT. Put a human YPT gene in place of a yeast YPT mutant and the yeast cell replicates. Even more sophisticated yeast mutant studies in our laboratory have allowed us to determine the functional specificity of parts of proteins in the yeast secretory pathway. Such findings can accordingly be applied to human biology.

Yeast molecular genetic research is aided by the extremely advanced stage of genomic mapping in yeast compared with other organisms. The yeast genetic map is second only to that of *E. coli* in detail, and a physical map of overlapping yeast clones now exists, thanks to the pioneering work of Maynard Olsen at Washington University in St. Louis. The next step is sequencing the entire yeast genome. A feasibility study for the sequencing of *S. cerevisiae* is currently under way at Stanford. Ronald Davis and I supervise the project, which is carried out by a group of scientists and technicians who have banded together temporarily to get the job done. Once the project is up to speed, and with a team of about 30, we believe we can sequence the entire yeast genome in a relatively short period, on the order of five years.

The rapidity of the progress in yeast molecular genetic research is due not only to the tractability of yeast, but to the cooperation among the yeast research community, which has tended toward friendly competition and collaboration. Strains of yeast are shared with other laboratories in a straightforward manner once findings are published (and often the

(continues)