How DNA Replication Originates

Researchers are at long last getting a look at the cellular machinery that ensures that DNA copies itself once—and only once—per cell division cycle

As the repositories of the genetic material, the chromosomes have long fascinated biologists—but rarely as much as they do today. Researchers are making progress on a series of broad fronts, leading to a better understanding of problems ranging from how chromosomes move during meiosis to how their structure influences gene activity. Many of these developments are discussed in the News reports beginning on this page and in the five Articles that follow, as well as in selected Reports.

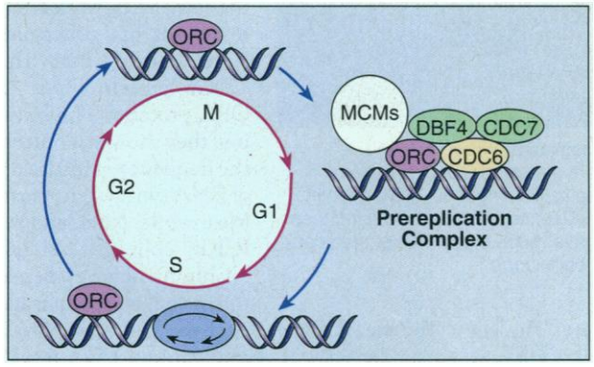
It's an essential prelude to cell division, and it sounds easy in the textbooks: Before a cell divides it has to replicate—or copy—its DNA. But concealed within that simple statement is a logistical nightmare, especially for the more complicated nucleated cells of higher organisms. Not only do they have to copy millions, or even billions, of base pairs of DNA just once, with exquisite accuracy and at just the right time in the cycle of cell division, but the DNA to be copied is distributed over numerous chromosomes—46 of them in the case of humans. And making matters even worse, the copying starts at hundreds to thousands of sites, some of which are triggered early in the synthesis (S) phase of the cell cycle, while others are triggered late.

Until recently, researchers had no idea how cells could complete this task without leaving some segments uncopied or making extra copies of others. But a flurry of recent work has finally begun to crack the puzzle. "It's exciting times in the field at the moment. [The work] is going to give us great insights into how replication is controlled," says molecular biologist Bruce Stillman of Cold Spring Harbor Laboratory, whose lab is among those focusing on DNA replication.

The answer, developed so far largely from studies in yeasts, lies in the interaction between short, specific stretches of DNA scattered throughout the chromosomes and a complex of proteins that binds to these "origins of replication," turning DNA replication on and off at the right moments. Researchers have also started to trace how these proteins are themselves controlled by the proteins that form the machinery of the cell cycle, the highly choreographed series of events that drives the cell to divide.

Understanding of DNA replication in more complex, multicelled organisms is less far advanced. But despite some problems identifying origins of replication comparable to those of yeast in these species, early evidence—including that in three papers in this special issue on the chromosome (pp. 1667, 1671, and 1674)—suggests that these organisms have replication control proteins similar to the yeast proteins. The research "is going very interesting places at the moment," says Julian Blow of the Imperial Cancer Research Fund (ICRF) laboratory in South Mimms, U.K., who studies DNA replication in the frog *Xenopus laevis*. "It looks like the basic processes in all eukaryotes will be the same."

Figuring out how DNA replication is controlled is an important challenge because accurate control is needed to maintain the integrity of the genetic material. Too little DNA synthesis, and essential genes might be lost, but too much could be equally bad. If the DNA is replicated more than once, cells may end up with extra copies of whole or partial chromosomes. Such abnormalities are thought to contribute to cancer development, possibly because they generate extra copies of cancer-causing oncogenes.



Getting ready. An essential step in initiating DNA replication seems to be the building on ORC of a "prereplication complex" containing several proteins, possibly including CDC6, the MCM proteins, the kinase CDC7, and its partner DBF4.

Yeast leads the way

Researchers have known for decades that DNA replication in bacteria and animal viruses is initiated when specific proteins bind to short sequences scattered through their genomes—that is, to the origins of replication. By leading to unwinding of the double helical DNA, a process aided by a helicase enzyme, this interaction eventually exposes the DNA to the polymerase enzyme that copies it. But researchers didn't get their first clue about whether a similar mechanism operates in higher organisms until about 15 years ago. That's when Ronald Davis's team at Stanford University and that of John Carbon at the University of California (UC), Santa Barbara, discovered "autonomously replicating sequences" (ARSs) in the genome of the budding yeast *Saccharomyces cerevisiae*. These are sequences that, when inserted into small circular pieces of DNA called plasmids, allow those plasmids to replicate in synchrony with yeast cell chromosomes.

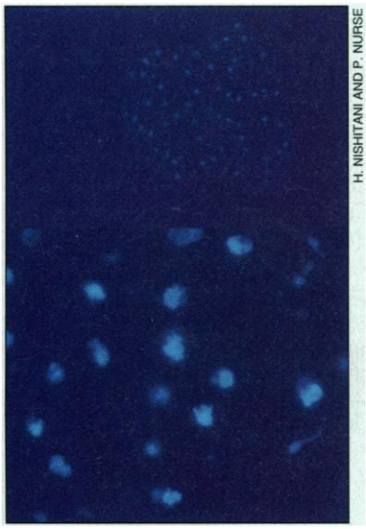
In 1987, teams led by Walt Fangman and Bonnie Brewer of the University of Washington, Seattle, and by Joel Huberman of Roswell Park Cancer Institute in Buffalo, New York, confirmed that the ARSs are the counterparts of the bacterial and viral origin sequences. Since then work by several teams, including Stillman's and that of Carol Newlon at New Jersey Medical School in Newark, has further defined the nature of yeast origins. They consist of an 11 base-pair sequence found in all ARSs, plus two to three additional short sequences that are also needed for efficient initiation of DNA synthesis, all grouped within a 100 to 200 base-pair-long DNA sequence.

Once researchers had found that yeast has well-defined origin sequences, they were able to move quickly to find the proteins that interact with those sequences and presumably tell the DNA-synthesizing enzymes where and when to begin working. In 1992, Stillman and his then-postdoc, Stephen Bell, identified a core group of six proteins, collectively called the origin recognition complex (ORC), that seemed to fit the bill. Subsequently, the Stillman team, and those of Jasper Rine at UC Berkeley, Kim Nasmyth at the I. M. P. Research Institute of Molecular Pathology in Vienna, Austria, and Ira Hershkowitz at UC San Francisco, cloned the genes encoding the proteins, thereby opening the way to genetic dissection of their functions. The researchers found, for example, that mutations in ARS sequences that abolish their ability to initiate DNA replication also abolish ORC binding. And conversely, mutations in ORC

protein genes prevent initiation at ARS sites. Indeed, says Rine, "the ORC complex fulfills all the criteria for a replication initiator."

It soon became apparent, however, that ORC couldn't be acting on its own. In all normal cells, DNA replication occurs only during a specific phase of the cell cycle, called "S" for synthesis. But John Diffley and his colleagues at the ICRF South Mimms laboratory found that ORC remains bound to yeast origins all through the cell cycle. That means something else has to be acting with ORC to restrict initiation of DNA replication to S phase. "It's clear that ORC has an essential role in initiating replication ... but it's not sufficient," says Diffley.

Further evidence that ORC doesn't act alone came from experiments by the ICRF team indicating that one or more additional proteins bind to yeast origins at or near ORC late in the mitosis (cell division) stage of the cell cycle. They remain there until S phase begins and then are apparently lost until mitosis occurs again. Based on these findings, Diffley proposed that the first step toward DNA synthesis is the formation of what he calls the "prereplication complex," consisting of ORC plus the additional protein or proteins.



Overgrown. Yeast cells overexpressing CDC18 *(bottom)* keep replicating their DNA, even in the absence of mitosis, causing them to have much bigger nuclei than control cells *(top)*.

The exact identities of these protein partners aren't known yet, but what researchers have already learned about them is helping them figure out how replication initiation ties into the cell cycle machinery. The key cogs in this machinery are a series of enzymes known as the cyclin dependent kinases (Cdks) because they require additional proteins, the cyclins, to do their job. Now researchers are finding that cells may achieve such tight control of DNA synthesis by having some of those same enzymes also regulate DNA replication.

In the fission yeast *Schizosaccharomyces pombe*, for example, the kinase produced by the *cdc2* gene and its cyclin partner is needed to push cells into mitosis, after which the cyclin is destroyed, inactivating the Cdk. But Paul Nurse of ICRF's London lab and his colleagues have discovered that the kinase has another role: While it's active it prevents replication from being initiated a second time. "That explains why there is only one S phase per cell cycle," says Nurse. "Only when [the mitotic Cdk] is destroyed do you lose the inhibiting signal."

And the Cdks apparently direct their inhibitory signals at components of the prereplicative complex. In a collaborative effort, the Nasymth and Diffley teams have found that the comparable Cdks of budding yeasts block formation of the complex. What's more, these researchers have added an intriguing twist to the story. They've found that these very same kinases are needed to trigger DNA replication during the S phase.

Having the same enzyme first activate DNA replication and then inhibit re-formation of the prereplication complex once it has triggered one round of DNA copying is an extremely efficient arrangement, Nasmyth says: "It's built into the logic of the cell cycle that you can't trigger the prereplication complex until you have gone through mitosis."

The targets of the kinases have not yet been identified, but possible candidates include ORC itself and another protein that combines with ORC in prereplication complexes. This is the product of the *cdc6* gene of the budding yeast *S. cerevisiae*. Diffley's group has genetic evidence that Cdc6 is needed for formation of the complexes, while Stillman and his colleagues have shown that the CDC6 protein binds to ORC proteins. The binding, they showed, controls the frequency of initiation of DNA synthesis in yeast. Meanwhile, Nurse and his ICRF colleague Hideo Nishitani have evidence that a similar protein made by the *cdc18* gene of the fission yeast plays a similar role in that organism. (The results appear in the 3 November issue of *Cell*.) "CDC18 not only initiates, but its high level is sufficient to drive cells into S without other protein synthesis," Nurse says. CDC18 (or CDC6) could be a target of the mitotic kinases, he says, but for now "that's pure speculation."

What about mammals?

The identities of the targets of the cell cycle kinases is only one of a great many questions that still need to be answered about how initiation of DNA replication is controlled. Among other things, for example, researchers want to know how the helicase and DNA polymerase enzymes get plugged into the system. But perhaps the biggest question of all is whether the results researchers are obtaining in yeast will also apply to mammals and other multicelled creatures. The main reason for doubt is that researchers have so far failed to pin down specific DNA sequences as origins of replication comparable to the ones in bacteria, viruses, and the yeast.

Some studies suggest, for example, that DNA synthesis is initiated anywhere over long sequence stretches. Others have come up with much shorter specific sequences, more like those in yeast. "Everyone was confused because all the measurements seemed valid. Yet they were leading to contradictory conclusions," says Roswell Park's Huberman.

Still, researchers are beginning to make progress in resolving the contradictions. One reason for the confusion may be that initiation "depends on the context," not just on specific sequences, as Marcel Méchali at the Institut Jacques Monod in Paris puts it. In the 10 November issue of *Science*, Olivier Hyrien of the École Normale Supérieure in Paris, Méchali, and Chrystelle Maric, also of the Institut Jacques Monod, report that in early *Xenopus* embryos, a time when embryonic genes are not transcribed, DNA replication initiates randomly at sites spread throughout the ribosomal rRNA genes. But later in development, when those genes are being transcribed, initiation sites are restricted to the spaces between the genes. This change from broad to narrower initiation zones may be due to alterations in the structure of the chromatin, the complex of proteins and DNA that form the chromosomes.

What's more, the apparent split between those who find broad initiation zones and those who find short ones may not be as severe as it initially seemed. Recent research suggests that broad initiation zones actually contain a small number of "hot spots," where initiation is more likely to occur than in other parts of the region. The differentiated cells of higher eukaryotes may have specific origins, although they appear to be more spread out than those of yeast.

But there's an even stronger indication that the yeast work may apply to the higher eukaryotes: It's beginning to look as if the rest of their replication machinery is similar to that of yeast, even if their origins aren't. The similarity starts with ORC itself.

In one of the three papers in this issue, the Stillman team describes the cloning of genes from the human and from the plant *Arabidopsis thaliana*, the nematode worm *Caenorhabditis elegans*, and fission yeast whose sequences suggest that they are those species' versions of the *ORC1* and *ORC2* genes of budding yeast. In the second, Mike Botchan's team at UC Berkeley reports the cloning of the *ORC2* and *ORC5* genes of the fruit fly *Drosophila melanogaster*.

Not only do the structures of these genes suggest that ORC proteins have been conserved in evolution, but in the third paper, Botchan's team, in collaboration with Rine's,

reports that the fruit fly's ORC2 gene, when put into mutant yeast, can replace some of the functions of yeast ORC2. "Since the complex is so conserved, we think this means that the basic mechanism of [origin site] recognition is conserved," Botchan says.

And the same may be said for the proteins that combine with ORC. One clue comes from a phenomenon called "licensing," discovered in the late 1980s by ICRF's Blow and Ronald Laskey of the University of Cambridge, U.K. Their studies of DNA replication in extracts of *Xenopus* eggs indicated that initiation of DNA replication requires some factor that can only gain access to the chromosomes during mitosis, a time when the nuclear membrane has broken down. Then, once replication has taken place, this licensing factor, as they called it, is somehow lost or inactivated. Once the nuclear membrane has regenerated, it can't reach the chromosomes again until another round of mitosis begins. This behavior, Blow and Laskey proposed, would ensure that DNA could not replicate more than once in any given cycle.

Now some of the molecules needed for licensing are turning out to belong to a family of proteins, known as MCM proteins, that was originally discovered in yeast but is now known to be present in all eukaryotes. This spring, the Blow and Laskey teams, and also that of Haruhiko Takisawa of Osaka University in Japan, showed, among other things, that antibodies to the *Xenopus* MCM3 protein remove licensing-factor activity from *Xenopus* egg extracts, while adding back the MCM proteins restores licensing activity.

The behavior of the MCM proteins in *Xenopus* resembles that of the proteins that combine with yeast ORC to form the prereplication complex. And because MCM proteins are found in yeast, the supposition is that they may in fact be part of that complex. Indeed, there may be another resemblance as well. Blow's group has found an as-yet-unidentified material in *Xenopus* that's needed for replication initiation, in addition to the MCM proteins. "We speculate that it's going to be a frog CDC6," he says, referring to the protein needed for prereplication complex assembly in yeast.

Taken together, the work in yeast and the other eukaryotes is building a picture of a complex initiation machinery centered on ORC. As Stillman describes it, "The concept has emerged that ORC is a landing pad for a whole bunch of other proteins" that come and go at precise times during the cell cycle to regulate DNA replication. And even though researchers still have a way to go to work out the precise identities and functions of all those proteins, they have also come a long way. "Two years ago we had no idea that all these proteins were involved," Stillman says. "Now at least we know what to focus on."

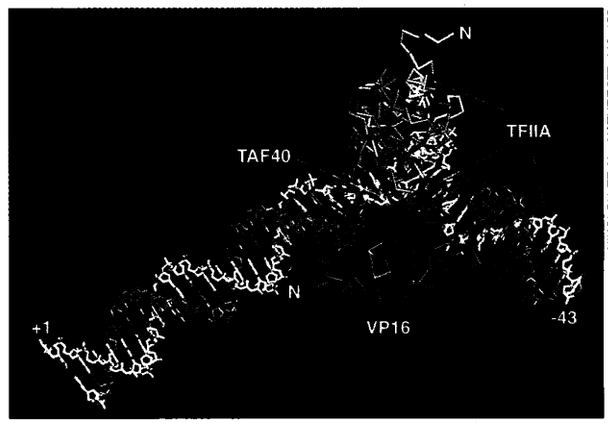
–Jean Marx

ARCHITECTURAL PROTEINS

Protein Sculptors That Help Turn On Genes

Inside a living cell lies perhaps the only place on Earth where a serious accounting problem is solved by something resembling art. In mammals, more than 100,000 genes must blink on or off along the chromosomes at precisely the right times, at just the right intensity, for the body to develop and function normally. But the protein switches that bind to DNA and turn genes on are in such short supply that if each of these proteins, known as transcription factors, had to act by itself, the cell would have far too few of them to give precise control of so many genes. And that's where the cell's artistic talents come into play.

Molecular biologists have known for about a decade that, rather than deploying the gene-regulating proteins one at a time, the cell dispatches groups of them in various combinations to turn on particular genes, thereby greatly expanding the number of distinct switches that can be formed. More recent work now suggests that cells go one step further, using "architectural proteins" to sculpt many of these protein clusters into precise three-dimensional shapes. The architectural proteins, which are themselves part of the transcription factor complexes, usually do this by binding to the DNA and bending it, often sharply, thereby bringing together other members of the complex, which are bound to separate segments of a gene's control region.



Around the bend. DNA bound to TBP *(blue)* and TFIIB *(red, magenta)* is L-shaped. The arrows indicate where other transcription factors in the complex bind.

Without this sculptural talent, the transcription factor clusters would not function normally. "The three-dimensional organization of these complexes is critical for controlling the expression, and level of expression, of various genes," notes Stephen Burley of Rockefeller University. As a result, architectural proteins are turning out to play a key role in physiological processes ranging from inflammation to sex development.

The first indications of the existence of architectural proteins came in the early 1980s when Donald Crothers of Yale University discovered that certain proteins from the bacterium *Escherichia coli* could bend DNA, in one case wrenching 90-degree kinks in the molecule. The significance of that bending wasn't understood at the time, but in the course of his research, Crothers pioneered a method of inferring the existence and location of bends in DNA. The method sifts out bent DNA molecules from straight ones by exploiting the fact that in gel electrophoresis, bent DNA moves more slowly than does straight DNA of the same size.

Crothers's method was soon put to use by Howard Nash and his colleagues at the National Institute of Mental Health in Bethesda, Maryland, who were the first to identify a function for an architectural protein. When certain viruses infect *E. coli*, they insert their DNA into the bacterial genome. At the time, in the mid-1980s, a bacterial protein called integration host factor (IHF) was known to be needed for the insertion, although no one knew exactly what it did. Nash was trying to find out.

After ruling out the possibility that IHF was an enzyme, Nash recalls, he began investigating a structural role for the molecule. He applied Crothers's method and found that IHF bends DNA. He also noted that IHF binds to DNA's minor groove, the small space between its helical twists, an unusual trait for a DNA-binding protein. Nash's subsequent work, together with that of Arthur Landy's group at Brown University, showed that as a result of the bending, two other proteins, bound to separate segments of the bacterial DNA, were able to meet. This interaction, the researchers found, was necessary for viral DNA insertion.

But the DNA bending induced by IHF soon proved to have a much broader role in normal bacterial functioning. In 1990, Sydney Kustu's team at the University of California, Berkeley, noted that the protein