combinatorial synthesis, which may well lead to greater diversity of integrated components. DNA-based computation and algorithmic assembly is another active area of research, and one that is impossible to separate from DNA nanotechnology (see Box 1).

The field of DNA nanotechnology has attracted an influx of researchers over the past few years. All of those involved in this area have benefited from the biotechnology enterprise that produces DNA-modifying enzymes and unusual components for synthetic DNA molecules. It is likely that applications in structural DNA nanotechnology ultimately will use variants on the theme of DNA (for example, peptide nucleic acids, containing an unconventional synthetic peptide backbone and nucleic acid bases for side chains), whose properties may be better suited to particular types of applications.

For the past half-century, DNA has been almost exclusively the province of biologists and biologically oriented physical scientists, who have studied its biological impact and molecular properties. During the next 50 years, it is likely they will be joined by materials scientists, nanotechnologists and computer engineers, who will exploit DNA's chemical properties in a non-biological context. □

doi:10.1038/nature01406

1. Hoffmann, R. DNA as clay. *Am. Sci.* **82,** 308–311 (1994).
2. Cuberes, M. T., Schlittler, R. R. & Gimzewski, J. K. Room-temperature repositioning of individual C-60 molecules at Cu steps: operation of a molecular counting device. *Appl. Phys. Lett.* **69,** 3016–3018 (1996).
3. Caruthers, M. H. Gene synthesis machines: DNA chemistry and its uses. *Science* **230,** 281–285 (1985).
4. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99,** 237–247 (1982).
5. Seeman, N. C. Molecular craftwork with DNA. *Chem. Intell.* **1,** 38–47 (1995).
6. Jaeger, L., Westhof, E. & Leontis, N. B. Tecto-RNA: modular assembly units for the construction of RNA nano-objects. *Nucleic Acids Res.* **29,** 455–463 (2001).
7. Zhang, X., Yan, H., Shen, Z. & Seeman, N. C. Paranemic cohesion of topologically-closed DNA molecules. *J. Am. Chem. Soc.* **124,** 12940–12941 (2002).
8. Chen, J. & Seeman, N. C. The synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350,** 631–633 (1991).
9. Seeman, N. C. Nucleic acid nanostructures and topology. *Angew. Chem. Int. Edn Engl.* **37,** 3220–3238 (1998).
10. Li, X., Yang, X., Qi, J. & Seeman, N. C. Antiparallel DNA double crossover molecules as components for nanoconstruction. *J. Am. Chem. Soc.* **118,** 6131–6140 (1996).
11. Winfree, E., Liu, F., Wenzler, L.A. & Seeman, N.C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394,** 539–544 (1998).
12. Mao, C., Sun, W. & Seeman, N. C. Designed two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy. *J. Am. Chem. Soc.* **121,** 5437–5443 (1999).
13. LaBean, T. *et al.* The construction, analysis, ligation and self-assembly of DNA triple crossover complexes. *J. Am. Chem. Soc.* **122,** 1848–1860 (2000).
14. Mao, C., Sun, W., Shen, Z. & Seeman, N. C. A DNA nanomechanical device based on the B–Z transition. *Nature* **397,** 144–146 (1999).
15. Yurke, B., Turberfield, A. J., Mills, A. P. Jr, Simmel, F. C. & Newmann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406,** 605–608 (2000).
16. Yan, H., Zhang, X., Shen, Z. & Seeman, N. C. A robust DNA mechanical device controlled by hybridization topology. *Nature* **415,** 62–65 (2002).
17. Niemeyer, C. M., Koehler, J. & Wuerdemann, C. DNA-directed assembly of bi-enzymic complexes from *in vivo* biotinylated NADP(H):FMN oxidoreductase and luciferase. *ChemBioChem* **3,** 242–245 (2002).
18. Robinson, B. H. & Seeman, N. C. The design of a biochip: a self-assembling molecular-scale memory device. *Protein Eng.* **1,** 295–300 (1987).
19. Keren, K. *et al.* Sequence-specific molecular lithography on single DNA molecules. *Science* **297,** 72–75 (2002).
20. Alivisatos, A. P. *et al.* Organization of 'nanocrystal molecules' using DNA. *Nature* **382,** 609–611 (1996).
21. Taton, T. A., Mucic, R. C., Mirkin, C. A. & Letsinger, R. L. The DNA-mediated formation of supramolecular mono- and multilayered nanoparticle structures. *J. Am. Chem. Soc.* **122,** 6305–6306 (2000).
22. Pena, S. R. N., Raina, S., Goodrich, G. P., Fedoroff, N. V. & Keating, C. D. Hybridization and enzymatic extension of Au nanoparticle-bound oligonucleotides. *J. Am. Chem. Soc.* **124,** 7314–7323 (2002).
23. Dekker, C. & Ratner, M. A. Electronic properties of DNA. *Phys. World* **14,** 29–33 (2001).
24. Fahlman, R. P. & Sen, D. DNA conformational switches as sensitive electronic sensors of analytes. *J. Am. Chem. Soc.* **124,** 4610–4616 (2002).
25. Seeman, N. C. The construction of 3-D stick figures from branched DNA. *DNA Cell Biol.* **10,** 475–486 (1991).
26. Eckardt, L. H. *et al.* Chemical copying of connectivity. *Nature* **420,** 286 (2002).
27. Adleman, L. Molecular computation of solutions to combinatorial problems. *Science* **266,** 1021–1024 (1994).
28. Winfree, E. in *DNA Based Computers. Proceedings of a DIMACS Workshop, April 4, 1995, Princeton University* (eds Lipton, R. J & Baum, E. B.) 199–219 (American Mathematical Society, Providence, 1996).
29. Winfree, E. Algorithmic self-assembly of DNA: theoretical motivations and 2D assembly experiments. *J. Biol. Mol. Struct. Dynamics Conversat. 11* **2,** 263–270 (2000).
30. Mao, C., LaBean, T., Reif, J. H. & Seeman, N. C. Logical computation using algorithmic self-assembly of DNA triple crossover molecules. *Nature* **407,** 493–496 (2000).

Acknowledgements
This work has been supported by grants from the National Institute of General Medical Sciences, the Office of Naval Research, the National Science Foundation, and the Defense Advanced Research Projects Agency/Air Force Office of Scientific Research.

DNA replication and recombination

Bruce Alberts

National Academy of Sciences, 2101 Constitution Avenue, Washington DC 20418, USA

Knowledge of the structure of DNA enabled scientists to undertake the difficult task of deciphering the detailed molecular mechanisms of two dynamic processes that are central to life: the copying of the genetic information by DNA replication, and its reassortment and repair by DNA recombination. Despite dramatic advances towards this goal over the past five decades, many challenges remain for the next generation of molecular biologists.

"Though facts are inherently less satisfying than the intellectual conclusions drawn from them, their importance should never be questioned." James D. Watson, 2002.

DNA carries all of the genetic information for life. One enormously long DNA molecule forms each of the chromosomes of an organism, 23 of them in a human. The fundamental living unit is the single cell. A cell gives rise to many more cells through serial repetitions of a process known as cell division. Before each division, new copies must be made of each of the many molecules that form the cell, including the duplication of all DNA molecules. DNA replication is the name given to this duplication process, which enables an organism's genetic information — its genes — to be passed to the two daughter cells created when a cell divides. Only slightly less central to life is a process that requires dynamic DNA acrobatics, called homologous DNA recombination, which reshuffles the genes on chromosomes. In reactions closely linked to DNA replication, the recombination machinery also repairs damage that inevitably occurs to the long, fragile DNA molecules inside cells (see article in this issue by Friedberg, page 436).

The model for the DNA double helix[1] proposed by James Watson and Francis Crick is based on two paired DNA strands that are complementary in their nucleotide sequence. The model had striking implications for the processes of DNA replication and DNA recombination. Before 1953, there had been no meaningful way of even speculating about the molecular mechanisms of these two central genetic processes. But the proposal that each nucleotide in one strand of DNA was tightly base-paired with its complementary nucleotide on the opposite strand — either adenine (A) with thymine (T), or guanine (G) with cytosine (C) — meant that any part of the nucleotide sequence could act as a direct template for the corresponding portion of the other strand. As a result, any part of the sequence can be used either to create or to recognize its partner nucleotide sequence — the two functions that are central for DNA replication and DNA recombination, respectively.

In this review, I discuss how the discovery of the structure of DNA half a century ago opened new avenues for understanding the processes of DNA replication and recombination. I shall also emphasize how, as our understanding of complex biological molecules and their interactions increased over the years, there have been profound changes in the way that biologists view the chemistry of life.

Structural features of DNA

The research that immediately followed the discovery of the double helix focused primarily on understanding the structural properties

© 2003 Nature Publishing Group

NEOGENOMICS EX.1027.001

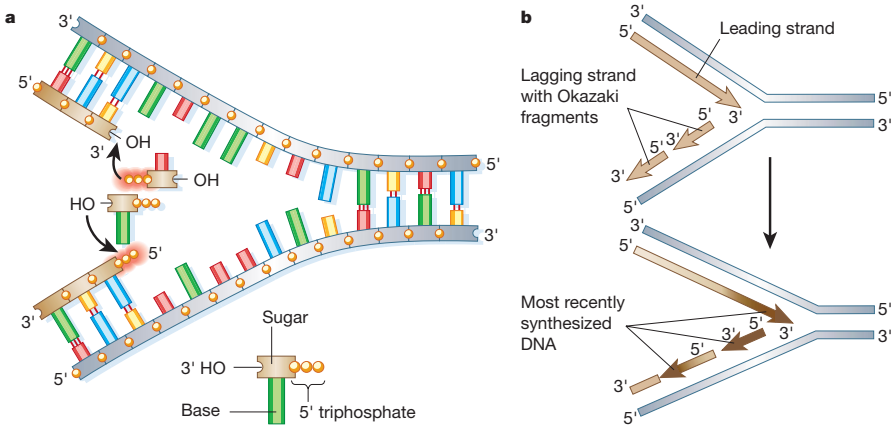


Figure 1 The DNA replication fork. **a**, Nucleoside triphosphates serve as a substrate for DNA polymerase, according to the mechanism shown on the top strand. Each nucleoside triphosphate is made up of three phosphates (represented here by yellow spheres), a deoxyribose sugar (beige rectangle) and one of four bases (differently coloured cylinders). The three phosphates are joined to each other by high-energy bonds, and the cleavage of these bonds during the polymerization reaction releases the free energy needed to drive the incorporation of each nucleotide into the growing DNA chain. The reaction shown on the bottom strand, which would cause DNA chain growth in the 3′ to 5′ chemical direction, does not occur in nature. **b**, DNA polymerases catalyse chain growth only in the 5′ to 3′ chemical direction, but both new daughter strands grow at the fork, so a dilemma of the 1960s was how the bottom strand in this diagram was synthesized. The asymmetric nature of the replication fork was recognized by the early 1970s: the 'leading strand' grows continuously, whereas the 'lagging strand' is synthesized by a DNA polymerase through the backstitching mechanism illustrated. Thus, both strands are produced by DNA synthesis in the 5′ to 3′ direction. (Redrawn from ref. 27, with permission.)

of the molecule. DNA specifies RNA through the process of gene transcription, and the RNA molecules in turn specify all of the proteins of a cell. This is the 'central dogma' of genetic information transfer[2]. Any read-out of genetic information — whether it be during DNA replication or gene transcription — requires access to the sequence of the bases buried in the interior of the double helix. DNA strand separation is therefore critical to DNA function. Thus, the Watson–Crick model drove scientists to a search for conditions that would disrupt the hydrogen bonds joining the complementary base pairs, so as to separate the two strands of the DNA double helix.

Physical chemists found that heating a solution of DNA to temperatures near boiling (100 °C), or subjecting it to extremes of pH, would cause the strands to separate — a change termed 'DNA denaturation'. The so-called 'melting temperature' (or T_m) of a stretch of DNA sequence depends on its nucleotide composition: those DNAs with a larger proportion of G–C base pairs exhibit a higher T_m because of the three hydrogen bonds that Watson and Crick had predicted to hold a G–C base pair together, compared with only two for the A–T base pair. At physiological salt concentrations, the T_m of mammalian DNA is nearly 90 °C, owing to the particular mix of its base pairs (47% G–C and 53% A–T)[3].

Initially it seemed inconceivable that, once separated from its complementary partner, a DNA strand could reform a double helix again. In a complex mixture of DNA molecules, such a feat would require finding the one sequence match amongst millions during random collisions with other sequences, and then rapidly rewinding with a new partner strand. The dramatic discovery of this unexpected phenomenon[4], called 'DNA renaturation', shed light on how sequences could be rearranged by DNA recombination. And it also provided a critical means by which DNA could be manipulated in the laboratory. The annealing of complementary nucleotide sequences, a process called hybridization, forms the basis of several DNA technologies that helped launch the biotechnology industry and modern genomics. These include gene cloning, genomic sequencing, and DNA copying by the polymerase chain reaction (see article by Hood and Galas on page 444).

The arrangement of DNA molecules in chromosomes presented another mystery for scientists: a long, thin molecule would be highly sensitive to shear-induced breakage, and it was hard to imagine that a mammalian chromosome might contain only a single DNA molecule. This would require that a typical chromosome be formed from a continuous DNA helix more than 100 million nucleotide pairs long — a massive molecule weighing more than 100 billion daltons, with an end-to-end distance of more than 3 cm. How could such a giant molecule be protected from accidental fragmentation in a cell only microns in diameter, while keeping it organized for efficient gene readout and other genetic functions?

There was no precedent for such giant molecules outside the world of biology. But in the early 1960s, autoradiographic studies revealed that the chromosome of the bacterium *Escherichia coli* was in fact a single DNA molecule, more than 3 million nucleotide pairs in length[5]. And when — more than a decade later — innovative physical techniques demonstrated that a single huge DNA molecule formed the basis for each mammalian chromosome[6], the result was welcomed by scientists with little surprise.

DNA replication forks

How is the enormously long double-stranded DNA molecule that forms a chromosome accurately copied to produce a second identical chromosome each time a cell divides? The template model for DNA replication, proposed by Watson and Crick in 1953 (ref. 7), gained universal acceptance after two discoveries in the late 1950s. One was an elegant experiment using density-labelled bacterial DNAs that confirmed the predicted template–anti-template scheme[8]. The other was the discovery of an enzyme called DNA polymerase, which uses one strand of DNA as a template to synthesize a new complementary strand[9]. Four deoxyribonucleoside triphosphate nucleotides — dATP, dTTP, dGTP and dCTP — are the precursors to a new daughter DNA strand, each nucleotide selected by pairing with its complementary nucleotide (T, A, C or G, respectively) on the parental template strand. The DNA polymerase was shown to use these triphosphates to add nucleotides one at a time to the 3′ end of the newly synthesized DNA molecule, thereby catalysing DNA chain growth in the 5′ to 3′ chemical direction.

Although the synthesis of short stretches of DNA sequence on a single-stranded template could be demonstrated in a test tube, how an enormous, twisted double-stranded DNA molecule is replicated was a puzzle. Inside the cell, DNA replication was observed to occur at a Y-shaped structure, called a 'replication fork', which moves steadily along a parental DNA helix, spinning out two daughter DNA helices behind it (the two arms of the 'Y')[5]. As predicted by Watson and Crick, the two strands of the double helix run in opposite chemical directions. Therefore, as a replication fork moves, DNA polymerase can move continuously along only one arm of the Y — the arm on

© 2003 Nature Publishing Group

NEOGENOMICS EX.1027.002

Box 1
Core proteins at the DNA replication fork

Proteins at the Y-shaped DNA replication fork are illustrated schematically in panel **a** of the figure below, but in reality, the fork is folded in three dimensions, producing a structure resembling that of the diagram in the inset **b** (cartoons redrawn from ref. 27, with permission).

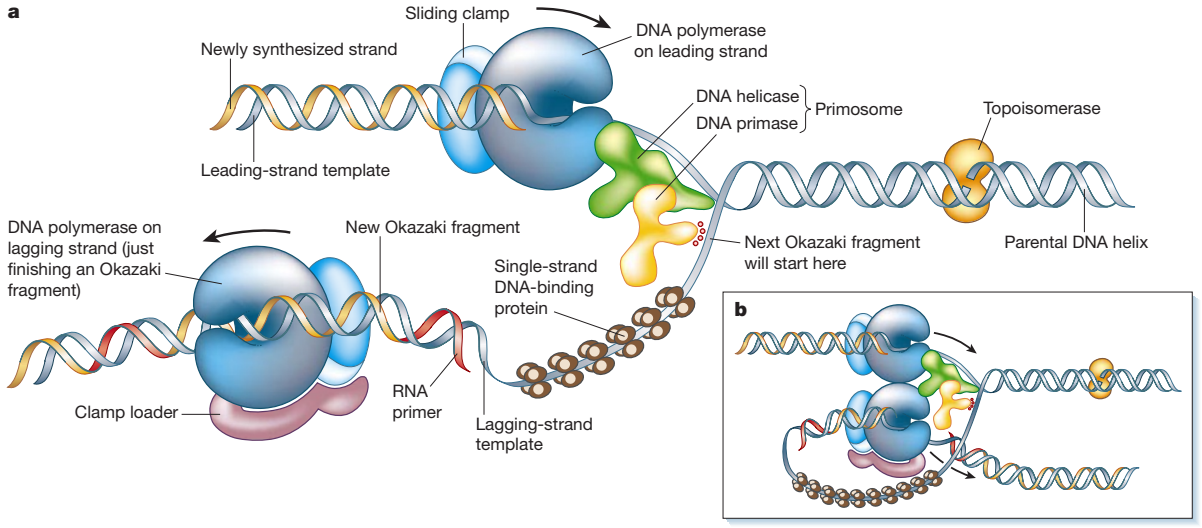
Focusing on the schematic illustration in **a**, two DNA polymerase molecules are active at the fork at any one time. One moves continuously to produce the new daughter DNA molecule on the leading strand, whereas the other produces a long series of short 'Okazaki DNA fragments' on the lagging strand. Both polymerases are anchored to their template by polymerase accessory proteins, in the form of a sliding clamp and a clamp loader.

A DNA helicase, powered by ATP hydrolysis, propels itself rapidly along one of the template DNA strands (here the lagging strand), forcing open the DNA helix ahead of the replication fork. The helicase exposes the bases of the DNA helix for the leading-strand polymerase to copy. DNA topoisomerase enzymes facilitate DNA helix unwinding.

In addition to the template, DNA polymerases need a pre-existing DNA or RNA chain end (a primer) onto which to add each nucleotide. For this reason, the lagging strand polymerase requires the action of a DNA primase enzyme before it can start each Okazaki fragment. The primase produces a very short RNA molecule (an RNA primer) at the 5′ end of each Okazaki fragment onto which the DNA polymerase adds nucleotides. Finally, the single-stranded regions of DNA at the fork are covered by multiple copies of a single-strand DNA-binding protein, which hold the DNA template strands open with their bases exposed.

In the folded fork structure shown in the inset, the lagging-strand DNA polymerase remains tied to the leading-strand DNA polymerase. This allows the lagging-strand polymerase to remain at the fork after it finishes the synthesis of each Okazaki fragment. As a result, this polymerase can be used over and over again to synthesize the large number of Okazaki fragments that are needed to produce a new DNA chain on the lagging strand.

In addition to the above group of core proteins, other proteins (not shown) are needed for DNA replication. These include a set of initiator proteins to begin each new replication fork at a replication origin, an RNAseH enzyme to remove the RNA primers from the Okazaki fragments, and a DNA ligase to seal the adjacent Okazaki fragments together to form a continuous DNA strand.



which the new daughter strand is being elongated in the 5′ to 3′ chemical direction. On the other arm, the new daughter strand would need to be produced in the opposite, 3′ to 5′ chemical direction (Fig. 1a). So, whereas Watson and Crick's central predictions were confirmed at the end of the first decade of research that followed their landmark discovery, the details of the DNA replication process remained a mystery.

Reconstructing replication

The mystery was solved over the course of the next two decades, a period in which the proteins that constitute the central players in the DNA replication process were identified. Scientists used a variety of experimental approaches to identify an ever-growing set of gene products thought to be critical for DNA replication. For example, mutant organisms were identified in which DNA replication was defective, and genetic techniques could then be used to identify specific sets of genes required for the replication process[10–12]. With the aid of the proteins specified by these genes, 'cell-free' systems were established, where the process was re-created *in vitro* using purified components. Initially, proteins were tested in a 'partial replication reaction', where only a subset of the protein machinery required for the full replication process was present, and the DNA template was provided in a single-stranded form[13]. New proteins that were identified were added one at a time or in combination to test their effects on the catalytic activity of DNA polymerase. Further advances in understanding replication then depended on creating more complex *in vitro* systems, in which, through the addition of a larger set of purified proteins, double-stranded DNA could eventually be replicated[14–15].

Today, nearly every process inside cells — from DNA replication and recombination to membrane vesicle transport — is being studied in an *in vitro* system reconstructed from purified components. Although laborious to establish, such systems enable the precise control of both the concentration and the detailed structure of each component. Moreover, the 'noise' in the natural system caused by side reactions — because most molecules in a cell are engaged in more than one type of reaction — is avoided by eliminating the proteins that catalyse these other reactions. In essence, a small fraction of the cell can be re-created as a

© 2003 Nature Publishing Group

NEOGENOMICS EX.1027.003

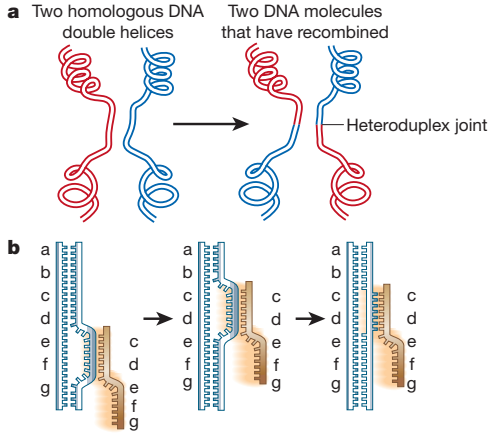
Box 2
DNA recombination

Homologous DNA recombination involves an exchange between two DNA double helices that causes a section of each helix to be exchanged with a section of the other, as illustrated schematically in panel **a** in the figure below (redrawn from ref. 27, with permission). Critical to the reaction is the formation of a heteroduplex joint at the point where the two double helices have been broken and then joined together. To form this joint, which glues two previously separate molecules together, a strand from one helix must form base pairs with a complementary strand from the second helix. This requires that the two DNA helices that recombine have a very similar sequence of nucleotides, that is, they must be homologous.

The DNA double helix poses a major problem for the DNA recombination process, because the bases that need to pair to form a heteroduplex joint are buried in the interior of the helix. How can two DNA helices recognize that they are homologous, in order to begin a recombination event, if their bases are not exposed?

The breakthrough came from the isolation and characterization of the RecA protein[17] from the bacterium *Escherichia coli*, which would turn out to be the prototype for a family of strand-exchange proteins that is present in all organisms, from bacteria to humans. The human equivalent of the RecA protein is the Rad51 protein. These proteins catalyse the central synapsis step of homologous DNA recombination — the process that brings two matching DNA helices together and causes them to exchange parts, resulting in either the reassortment or the repair of genetic information (panel **b** below). Powered by the energy generated from ATP hydrolysis, the RecA protein assembles into long filaments on a single-strand DNA molecule (brown strand). Because the RecA protein has a second DNA-binding site that recognizes a DNA double helix, a RecA-coated strand has the remarkable ability to scan for a complementary strand in any double helix (blue strand) that it encounters. Once found, the complementary strand is pulled from the helix to form a new 'hybrid helix' with the RecA-coated single strand, thereby initiating the formation of the heteroduplex joint needed for recombination, as illustrated schematically in panel **b** (RecA protein not shown).

DNA recombination makes it possible for a damaged chromosome to repair itself by using a second copy of the same genetic information as a guide. It also causes the extensive breakage and reunion of chromosomes that occurs during the development of eggs and sperm, which greatly increases the genetic variation produced by sexual reproduction. Many of the atomic details of the RecA protein reaction are still uncertain, remaining as a future challenge for scientists.



bounded set of chemical reactions, making it fully amenable to precise study using all of the tools of physics and chemistry.

By 1980, multiprotein *in vitro* systems had enabled a detailed characterization of the replication machinery and solved the problem of how DNA is synthesized on both sides of the replication fork (Fig. 1b). One daughter DNA strand is synthesized continuously by a DNA polymerase molecule moving along the 'leading strand', while a second DNA polymerase molecule on the 'lagging strand' produces a long series of fragments (called Okazaki fragments)[16] which are joined together by the enzyme DNA ligase to produce a continuous DNA strand. As might be expected, there is a difference in the proteins required for leading- and lagging-strand DNA synthesis (see Box 1). Remarkably, the replication forks formed in these artificial systems could be shown to move at the same rapid rates as the forks inside cells (500 to 1,000 nucleotides per second), and the DNA template was copied with incredibly high fidelity[15].

As more and more proteins were found to function at the replication fork, comparisons could be made between the replication machinery of different organisms. Studies of the replication machinery in viruses, bacteria and eukaryotes revealed that a common set of protein activities drives the replication forks in each organism (Box 1). Each system consists of: a leading- and a lagging-strand DNA polymerase molecule; a DNA primase to produce the RNA primers that start each Okazaki fragment; single-strand DNA binding proteins that coat the template DNA and hold it in position; a DNA helicase that unwinds the double helix; and additional polymerase accessory proteins the tie the polymerases to each other and to the DNA template. As one progresses from a simple virus to more complex organisms, such as yeasts or mammals, the number of subunits that make up each type of protein activity tends to increase. For example, the total number of polypeptide subunits that form the core of the replication apparatus increases from four and seven in bacteriophages T7 and T4, respectively, to 13 in the bacterium *E. coli*. And it expands to at least 27 in the yeast *Saccharomyces cerevisiae* and in mammals. Thus, as organisms with larger genomes evolved, the replication machinery added new protein subunits, without any change in the basic mechanisms[15,18–20].

While the work I have described on DNA replication was advancing, other groups of researchers were establishing *in vitro* systems in which homologous DNA recombination could be reconstructed. The central player in these reactions was the RecA type of protein[17], named after the bacterial mutant defective in recombination that led to its discovery (Box 2).

Protein machines

As for all other aspects of cell biochemistry, the DNA replication apparatus has evolved over billions of years through 'trial and error'— that is, by random variation followed by natural selection. With time, one protein after another could be added to the mix of proteins active at the replication fork, presumably because the new protein increased the speed, control or accuracy of the overall replication process. In addition, the structure of each protein was fine-tuned by mutations that altered its amino acid sequence so as to increase its effectiveness. The end results of this unusual engineering process are the replication systems that we observe today in different organisms. The mechanism of DNA replication might therefore be expected to be highly dependent on random past events. But did evolution select for whatever works, with no need for elegance?

For the first 30 years after Watson and Crick's discovery, most researchers seemed to hold the view that cell processes could be sloppy. This view was encouraged by knowledge of the tremendous speed of movements at the molecular level (for example, it was known that a typical protein collides with a second molecule present at a concentration of 1 mM about 10^6 times per second). The rapid rates of molecular movement were thought initially to allow a process like DNA replication to occur without any organization of the proteins involved in three-dimensional space.

© 2003 Nature Publishing Group

NEOGENOMICS EX.1027.004

Quite to the contrary, molecular biologists now recognize that evolution has selected for highly ordered systems. Thus, for example, not only are the parts of the replication machinery held together in precise alignments to optimize their mutual interactions, but energy-driven changes in protein conformations are used to generate coordinated movements. This ensures that each of the successive steps in a complex process like DNA replication is closely coordinated with the next one. The result is an assembly that can be viewed as a 'protein machine'. For example, the DNA polymerase molecule on the lagging side of the replication fork remains bound to the leading-strand DNA polymerase molecule to ensure that the same lagging-strand polymerase is used over and over again for efficient synthesis of Okazaki fragments[18,20,21] (Box 1). And DNA replication is by no means unique. We now believe that nearly every biological process is catalysed by a set of ten or more spatially positioned, interacting proteins that undergo highly ordered movements in a machine-like assembly[22].

Protein machines generally form at specific sites in response to particular signals, and this is particularly true for protein machines that act on DNA. The replication, repair and recombination of the DNA double helix are often considered as separate, isolated processes. But inside the cell, the same DNA molecule is able to undergo any one of these reactions. Moreover, specific combinations of the three types of reactions occur. For instance, DNA recombination is often linked directly to either DNA replication or DNA repair[23]. For the integrity of a chromosome to be properly maintained, each specific reaction must be carefully directed and controlled. This requires that sets of proteins be assembled on the DNA and activated only where and when they are needed. Although much remains to be learned about how these choices are made, it seems that different types of DNA structures are recognized explicitly by specialized proteins that serve as 'assembly factors'. Each assembly factor then serves to nucleate a cooperative assembly of the set of proteins that forms a particular protein machine, as needed for catalysing a reaction appropriate to that time and place in the cell.

A view of the future

It has become customary, both in textbooks and in the regular scientific literature, to explain molecular mechanisms through simple two-dimensional drawings or 'cartoons'. Such drawings are useful for consolidating large amounts of data into a simple scheme, as illustrated in this review. But a whole generation of biologists may have become lulled into believing that the essence of a biological mechanism has been captured, and the entire problem therefore solved, once a researcher has deciphered enough of the puzzle to be able to draw a meaningful cartoon of this type.

In the past few years, it has become abundantly clear that much more will be demanded of scientists before we can claim to fully understand a process such as DNA replication or DNA recombination. Recent genome sequencing projects, protein-interaction mapping efforts and studies in cell signalling have revealed many more components and molecular interactions than were previously realized. For example, according to one recent analysis, *S. cerevisiae*, a single-celled 'simple' eukaryotic organism (which has about 6,000 genes compared with 30,000 in humans), uses 88 genes for its DNA replication and 49 genes for its DNA recombination[24].

To focus on DNA replication, fully understanding the mechanism will require returning to where the studies of DNA first began — in the realms of chemistry and physics. Detailed atomic structures of all relevant proteins and nucleic acids will be needed, and spectacular progress is being made by structural biologists, owing to increasingly powerful X-ray crystallography and nuclear magnetic resonance techniques. But the ability to reconstruct biological processes in a test tube with molecules whose precise structures are known is not enough. The replication process is both very rapid and incredibly accurate, achieving a final error rate of about one nucleotide in a billion. Understanding how the reactions between the many different proteins and other molecules are coordinated to create this result will require that experimentalists determine all of the rate constants for the interactions between the various components, something that is rarely done by molecular biologists today. They can then use genetic engineering techniques to alter selected sets of these parameters, carefully monitoring the effect of these changes on the replication process.

Scientists will be able to claim that they truly understand a complex process such as DNA replication only when they can precisely predict the effect of changes in each of the various rate constants on the overall reaction. Because the range of experimental manipulations is enormous, we will need more powerful ways of deciding which such alterations are the most likely to increase our understanding. New approaches from the rapidly developing field of computational biology must therefore be developed — both to guide experimentation and to interpret the results.

The Watson–Crick model of DNA catalysed dramatic advances in our molecular understanding of biology. At the same time, its enormous success gave rise to the misleading view that many other complex aspects of biology might be similarly reduced to elegant simplicity through insightful theoretical analysis and model building. This view has been supplanted over subsequent decades, because most biological subsystems have turned out to be far too complex for their details to be predicted. We now know that nothing can substitute for rigorous experimental analyses. But traditional molecular and cell biology alone cannot bring a problem like DNA replication to closure. New types of approaches will be required, involving not only new computational tools, but also a greater integration of chemistry and physics[20,25]. For this reason, we urgently need to rethink the education that we are providing to the next generation of biological scientists[22,26]. □

doi:10.1038/nature01407

1. Watson, J. D. & Crick, F. H. C. A structure for deoxyribose nucleic acid. *Nature* **171,** 737–738. (1953).
2. Crick, F. H. C. The biological replication of macromolecules. *Symp. Soc. Exp. Biol.* **12,** 138–163 (1958).
3. Doty, P. *Inside Nucleic Acids* (Harvey Lecture, 1960) (Academic, New York, 1961).
4. Marmur, J. & Doty, P. Thermal renaturation of deoxyribonucleic acids. *J. Mol. Biol.* **3,** 585–594 (1961).
5. Cairns, J. The bacterial chromosome and its manner of replication as seen by autoradiography. *J. Mol. Biol.* **6,** 208–213 (1963).
6. Kavenoff, R., Klotz, L. C. & Zimm, B. H. On the nature of chromosome-sized DNA molecules. *Cold Spring Harb. Symp. Quant. Biol.* **38,** 1–8 (1974).
7. Watson, J. D. & Crick, F. H. C. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171,** 964–967 (1953).
8. Meselson, M. & Stahl, F. W. The replication of DNA in *E. coli*. *Proc. Natl Acad. Sci. USA* **44,** 671–682 (1958).
9. Kornberg, A. Biological synthesis of DNA. *Science* **131,** 1503-1508 (1960).
10. Epstein, R. H. *et al.* Physiological studies of conditional lethal mutants of bacteriophage T4D. *Cold Spring Harb. Symp. Quant. Biol.* **28,** 375 (1963).
11. Bonhoeffer, F. & Schaller, H. A method for selective enrichment of mutants based on the high UV sensitivity of DNA containing 5-bromouracil. *Biochem. Biophys. Res. Commun.* **20,** 93 (1965).
12. Kohiyama, M., Cousin, D., Ryter, A. & Jacob, F. Mutants thermosensible d'*Escherichia coli* K/12. I. Isolement et caracterisation rapide. *Ann. Inst. Pasteur* **110,** 465 (1966).
13. Huberman, J. A., Kornberg, A. & Alberts, B. M. Stimulation of T4 bacteriophage DNA polymerase by the protein product of T4 gene 32. *J. Mol. Biol.* **62,** 39–52 (1971).
14. Morris, C. F., Sinha, N. K. & Alberts, B. M. Reconstruction of bacteriophage T4 DNA replication apparatus from purified components: rolling circle replication following *de novo* chain initiation on a single-stranded circular DNA template. *Proc. Natl Acad. Sci. USA* **72,** 4800–4804 (1975).
15. Kornberg, A. & Baker, T. A. *DNA Replication* 2nd edn (Freeman, New York, 1992).
16. Okazaki R. *et al.* Mechanism of DNA chain growth: possible discontinuity and unusual secondary structure of newly synthesized chains. *Proc. Natl Acad. Sci. USA* **59,** 598–605 (1968).
17. Radding, C. M. Recombination activities of *E. coli* RecA protein. *Cell* **25,** 3–4 (1981).
18. Davey, M. J. & O'Donnell, M. Mechanisms of DNA replication. *Curr. Opin. Chem. Biol.* **4,** 581–586 (2000).
19. Waga, S. & Stillman, B. The DNA replication fork in eukaryotic cells. *Annu. Rev. Biochem.* **67,** 721–751 (1998).
20. Benkovic, S. J., Valentine, A. M. & Salinas F. Replisome-mediated DNA replication. *Annu. Rev. Biochem.* **70,** 181–208 (2001).
21. Alberts, B. M. The DNA enzymology of protein machines. *Cold Spring Harb. Symp. Quant. Biol.* **49,** 1–12 (1984).
22. Alberts, B. The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* **92,** 291–294 (1998).
23. Radding, C. Colloquium introduction. Links between recombination and replication: vital roles of recombination. *Proc. Natl Acad. Sci. USA* **98,** 8172 (2001).
24. Dwight, S. S. *et al. Saccharomyces* Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucleic Acids Res.* **30,** 69–72 (2002).
25. Trakselis, M. A. & Benkovic, S. J. Intricacies in ATP-dependent clamp loading: variations across replication systems. *Structure* **9,** 999–1004 (2001).
26. National Research Council. *Bio2010: Undergraduate Education to Prepare Biomedical Research Scientists* (The National Academies Press, Washington DC, 2002).
27. Alberts, B. *et al. Molecular Biology of the Cell* 4th edn (Garland, New York, 2002).

© 2003 Nature Publishing Group

NEOGENOMICS EX.1027.005