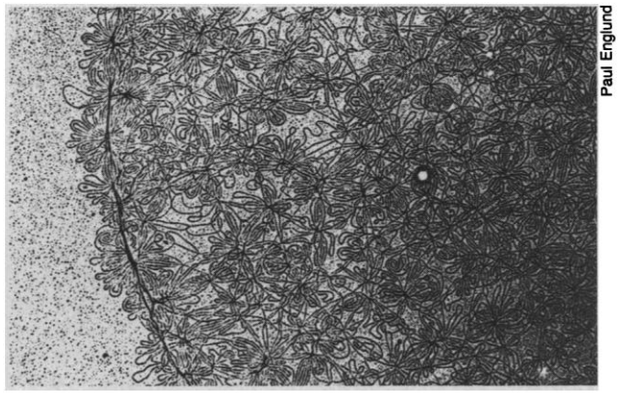
Mathematics Untwists the Double Helix

Once a refuge for mathephobes, molecular biology has recently seen an influx of equations that are helping to explain why macromolecules such as DNA behave as they do

"MATHEMATICIANS ARE LIKE THE FRENCH," cautioned Johann Wolfgang von Goethe (himself a German). "They take whatever you tell them and translate it into their own language—and from then on it is something entirely different."

Goethe's warning notwithstanding, a growing number of intrepid molecular biologists are talking to mathematicians about problems such as gene sequencing and the structure of DNA, and listening attentively to the mathematical language of complexity theory, differential geometry, and even quantum field theory. At a recent conference* on mathematical approaches to DNA, biologists and mathematicians got together to talk their two versions of shop: gels and nucleosomes on the one hand, invariants and Monte Carlo methods on the other.



Paul Englund

All in a tangle. *This tracery is DNA from an organelle called the kinetoplast, which is found in the single-celled trypanosome.*

The spirit of cross-pollination was summed up at the meeting—from the mathematical side, at least—by De Witt Sumners, a mathematician at Florida State University: "The biologists are more and more becoming convinced that these types of mathematical analysis can and will be useful to them."

To be sure, there has always been contact between biologists and mathematicians. Population ecology has strong ties with mathematics because of its inherently quantitative nature. Physiologists make use of mathematical models to describe such things as drug metabolism and the ion flows across the membrane of a nerve cell that underly the transmission of the nervous impulse. Biomechanics relies on equations to analyze how fish swim, birds fly, and people climb stairs.

Molecular biology, however, has long been a haven for mathephobes, in part because the huge molecules that are the subject's stock in trade have until now seemed too complicated for any mathematical model to deal with. But no more. The biologists' desire to understand the structure and dynamics of macromolecules and their need to grapple with the huge amounts of data they have collected about those molecules have brought them face to face with sophisticated mathematical techniques and problems in theoretical computer science. This process has taken even the biologists involved by surprise.

"I was the last person in the world anybody would ever think would get into mathematics, including myself. It was completely unintended," says Nicholas Cozzarelli, a molecular biologist at the University of California at Berkeley. Intentionally or not, Cozzarelli has wound up spearheading the movement to bring mathematicians and molecular biologists together: he and biophysicist Sylvia Spengler, also at Berkeley, are co-directors of the NSF-funded Program in Mathematics and Molecular Biology, which began in late 1988 as a consortium of ten mathematicians and biologists. The program has now grown to include eleven mathematicians and six biologists.

Why the surge of interest in the mathematics of macromolecules? For one thing, modern computers are verging on the power needed to calculate their dynamics on an atom-by-atom basis. For example, Michael Levitt and his colleagues at Stanford University have run computer simulations of molecular systems with upwards of 10,000 atoms. They do so by writing down an enormous equation describing a ball and spring of the atomic interactions—no fancy quantum mechanics, please—and then solving the equation numerically.

In earlier simulations the macromolecules tended to break apart; the new models do much better, Levitt says. Indeed, the latest simulations of short strands of DNA surrounded by water—the computer looks at a dozen base pairs buffeted by several thousand water molecules—are "beginning to approximate reality," according to Levitt. The tiny time steps required by the numerical solution—on the order of 10^{-15} seconds—make it impractical to follow the molecule for more than about one ten-billionth of a second, but that's enough to see what the DNA is up to. "We'd love to look at it longer, but even on the short time scales there's plenty of interesting things happening," Levitt says.

But that hardly exhausts what is becoming a very active field. Other researchers are using mathematically simpler models to study the structure and motion of DNA. Stephen Levene, a biochemist at the University of California at Berkeley, models DNA as a chain of linked rods similar to a segmented toy snake. Levene's snake wriggles its way through a simulated thicket of obstacles representing the gel used in electrophoresis—a fundamental technique in molecular biology. The simulated motion agrees with an experimental observation: DNA travels more slowly when it is bent (a condition brought on most notably by stretches of consecutive adenines in the nucleotide sequence). In Levene's model, a single bend as small as 30 degrees causes a "substantial reduction in mobility."

Another reason for the influx of mathematics into the kingdom of the double helix is that the structure of DNA lends itself to the simplifications required for mathematical analysis. "It has the advantage of being essentially a linear molecule," says Sumners. This means that mathematicians can disregard a lot of the biochemical details and still get useful results by thinking of DNA as little more than a curve winding through three-dimensional space—grist for the mill of differential geometers and topologists.

But DNA doesn't only come in a linear form. Much of the experimental work in molecular biology has been done with "cir-

*Mathematical Approaches to DNA, 24 to 28 January, 1990, Sante Fe, NM.

All Wound Up in DNA

There are few formulas in science that are instantly recognizable. There's $E = mc^2$, of course, but not many others.

Now molecular biology has a candidate: $Lk = Tw + Wr$ or, in plainer English, the linking number of a pair of curves is equal to their twist plus their writhe.

What? you ask. Since this formula hasn't reached pop culture yet, perhaps a word or two of explanation is in order.

To begin with, although molecular biologists may claim it, the equation doesn't really come from biology at all but from the pure mathematics of differential geometry. James White, now at UCLA, proved it in his 1968 Ph.D. thesis. White's formula is a theorem about the geometry of mathematical ribbons: pairs of curves traveling side by side through three-dimensional space. For the formula to hold, the ribbon must be "closed," meaning that you can travel around it and wind up back where you started, like a belt that's been buckled.

The exact definitions of the variables *Lk* (*Link*), *Tw* (*Twist*), and *Wr* (*Writhe*) are rather technical, involving integrals from vector calculus. But the names are suggestive. *Link*, or the linking number, for example, is a measure of how intertwined the two curves are. It can be positive or negative (depending on the ribbon's orientation), but it is always an integer.

The linking number is a topological invariant of the two curves: No matter how you stretch or deform them, it does not change; the number can only be changed by cutting the ribbon. In particular, if the linking number is not zero, you cannot separate the curves without cutting one of them.

Twist and *Writhe*, on the other hand, are "metric" qualities: They change value as the ribbon is deformed, and they need not be integers. Geometrically, the twist is a measure of the ribbon's rotation around its axis, or center line. This may sound like the same thing as the linking number, but it's not—because the axis itself may be wandering in space. And that's what writhe measures: the winding around in space of the ribbon's axis.

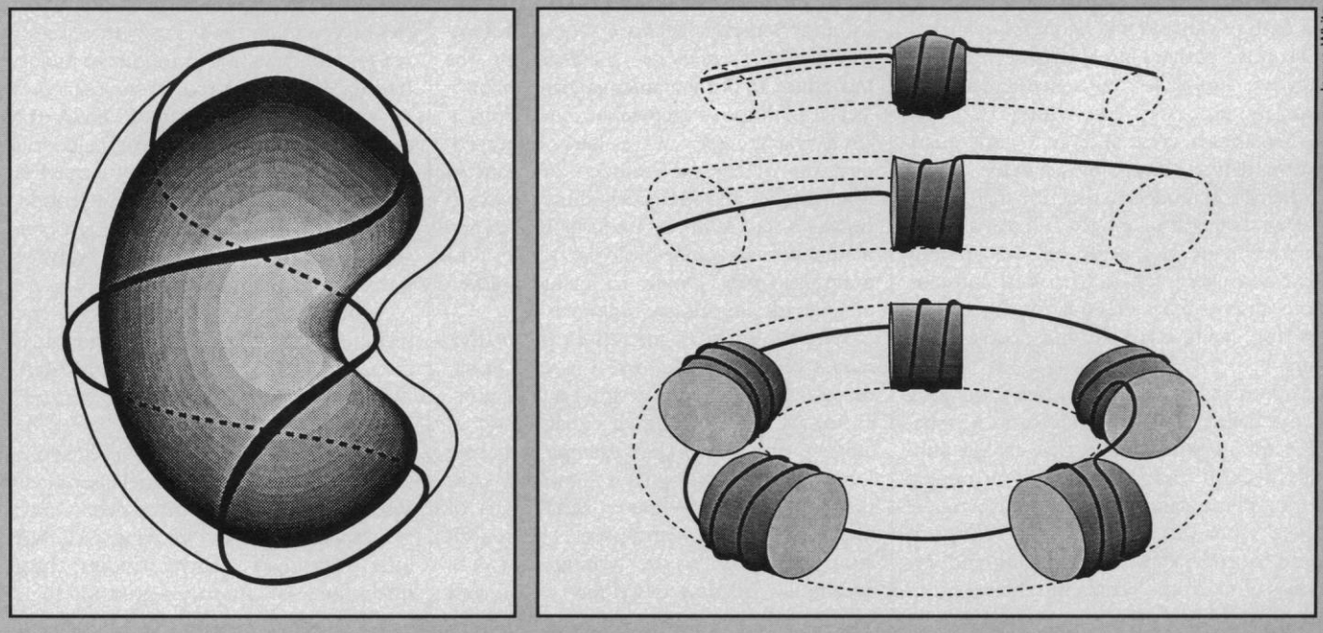
So what's this got to do with molecular biology? Just this: DNA is not quite the simple double helix shown in the textbooks, with the major axis running straight as an arrow. If it were, all those genes wouldn't fit into the nucleus, and you'd have DNA sticking out of you like quills on a porcupine. Instead, the molecule is highly wound.

A mathematician can look at at this wound-up DNA and see a ribbon whose edges are the molecule's phosphate backbones. Moreover, circular DNA, which experimentalists are partial to, corresponds perfectly to the mathematician's closed ribbon. White's formula is a key theoretical tool for analyzing the structure of loops of DNA and how various enzymes act on it.

Recently White and William Bauer, a microbiologist at the State University of New York at Stony Brook, have developed a new formula for the linking number that takes into account the fact that DNA is often wrapped on the surface of a protein. Instead of writing it as *Twist* + *Writhe*, which are not measurable in biological experiments, they write *Lk* as the sum of two other integers: *SLk*, the "surface linking number," and a "winding number," designated Φ.

The virtue of *SLk* and Φ is that they are both experimentally measurable in addition to being topological invariants. *SLk* depends only on how the DNA axis is wrapped about the protein's surface; it can be determined by x-ray crystallography. Φ is a measure of helical periodicity and can be determined by nuclease digestion of the protein or by chemical probes.

Because it is experimentally measurable, the new formula has great advantages. Indeed, in some respects it provides a measure of how mathematics is beginning to penetrate into the actual lab work of molecular biology. But it probably won't ever usurp the original. Somehow it just isn't as punchy as "*Link* equals *Twist* plus *Writhe*." ■ **B.C.**



James White

Twist and shout. *Diagrams show work by James White on DNA topology. At left, axis of DNA wraps around a protein. At right, minichromosome wraps around nucleosomes. Above: nucleosomes with slightly different shapes, which change the DNA's "surface linking number."*

cular" DNA, in which each nucleotide chain forms a closed loop. This form of DNA—frequently found in bacteria, among other creatures—is tailor-made for mathematical analysis. Differential geometry provides a precise mathematical context for the "linking," "twisting," and "writhing" that molecular biologists see in a phenomenon known as supercoiling: the tendency of circular DNA to wrap around itself like a twisted-up rubber band. (Supercoiling is familiar in everyday terms as the tangled mess telephone cords always seem to wind up in.)

James White, a differential geometer at the University of California, Los Angeles, has studied the geometry of circular DNA for more than a decade (see article on p. 914). In his 1968 Ph.D. dissertation, he proved that the "linking number" of two mathematical curves (a measure of how intertwined they are) is equal to the sum of one curve's writhing and the other curve's twisting about the first. White's dissertation was pure math, but his formula was just what the biologists needed. In essence it says that supercoiling is the result of an imbalance between the twisting of the double helix and the intertwining of its twin phosphate backbones.

More recently White has sought to refine the notions of linking and twisting to correspond more closely to quantities that are experimentally measurable. He and several other researchers are also investigating mathematical models to explain exactly how DNA divvies up the linking number into twist and writhe. White is trying a technique called the finite element method, which mechanical engineers have used for decades in analyses of stress in elastic rods.

White's work is complemented by that of researchers like Maxim Frank-Kamenetskii and co-workers at the Institute for Molecular Genetics in Moscow. Using a statistical technique known as a Monte Carlo method, the Soviets jiggle an initially random configuration of a "wormlike" model of DNA into equilibrium. And Wilma Olson and coworkers in the chemistry department at Rutgers University are using both Monte Carlo methods and techniques borrowed from computational geometry to simulate the three-dimensional structure of DNA.

Molecular biologists are also boning up on knot theory and topology to understand reactions that take place in DNA replication and recombination. It's been known for a while that enzymes called topoisomerases can tie and untie knots in DNA; it has also been known that site-specific recombination—in which two stretches of DNA are brought together, cut, and rejoined—often produces knots or links. But it took some help from mathematicians to make sense of what biologists were seeing in the DNA.

Jane Scherr

Translator Spengler. *Mathematicians and biologists don't speak "the same language."*

One of the breakthroughs occurred in 1984, and by a happy coincidence, it occurred in Cozzarelli's own backyard. Vaughn Jones, a mathematician at Berkeley, discovered a new way of classifying knots by means of a polynomial invariant, an easily computed algebraic expression that can distinguish one knot from another. Jones's starting point was a far cry from molecular biology: he was led to knots from research in the mathematics of quantum field theory. Nevertheless, his discovery was just what Cozzarelli needed to solve the problems that had cropped up in recombination experiments with circular DNA.

Cozzarelli's lab had found that throwing circular DNA in with the enzyme called resolvase resulted in a slew of knots and links. The biologists figured that resolvase was acting methodically and that what they were seeing was a sequence of products. But to make sense of the sequence they needed to know exactly what sort of knots and links they were dealing with and how they were related. Jones's polynomial invariant provided a rigorous way of doing this, thereby helping to clarify the sequence of steps in the reaction.

More recently, Sumners and Claus Ernst, a mathematician at Western Kentucky University, have developed a "tangle" model for analyzing the mechanisms of site-specific recombination. They compute the topology of the pre- and post-recombination complex from knowledge of the knots and links that occur. This is particularly effective when several recombination events occur at the same site. Their model calculates all possible enzyme mechanisms that yield the observed results. Given enough information—usually three rounds of recombination suffice—the model determines a unique mechanism.

Meanwhile, Cozzarelli, Spengler, and White, along with Paul Englund and Carol Rauch at the Johns Hopkins School of Medicine, are teasing out the secrets of a knotty mess of DNA in the trypanosome, a flagellated protozoan. That tangle is found in a membrane-bound organelle called a kinetoplast, which is associated with the basal body of the flagellum. "The structure of this DNA is wild and wonderful," says Spengler, who explains that the kinetoplast "has such a multiplicity of packing for its material that it could keep graduate students busy for centuries."

Kinetoplast DNA comes not as a single, knotted molecule, but as several thousand "mini-circles" intertwined with a few dozen "maxi-circles." The mini-circles are approximately 2,500 base pairs long, while the maxi-circles run to about 37,000 base pairs. These circles are woven together into something like an elaborate fishnet. Yet somehow this fishnet manages to reproduce itself. Figuring out the topology of the kinetoplast network and its role in replication poses substantial problems in both biology and mathematics, Spengler says.

So far the collaboration between mathematicians and molecular biologists has been largely a one-way street: mathematicians charging in to answer questions in biology. Cozzarelli, for one, doesn't see this as changing anytime soon. Still, some of the biological applications have followed close on the heels of recent advances in pure mathematics—Cozzarelli's use of Jones's discovery in knot theory is a case in point. Perhaps the time will come when the street becomes two-way, as biological problems demand new mathematics, rather than simply taking advantage of what's already there.

This is not to say molecular biologists have to start majoring in mathematics or poring over papers in math journals. "I'm not convinced that having a biologist go through the gory details of a proof will elucidate anything for him at all," says Sumners. Cozzarelli agrees. "I have no intention of trying to become a mathematician," he says. "But I need to know enough mathematics to collaborate with mathematicians."

Both sides agree that despite their successes there remains a gap between mathematicians and biologists. "It's not just that they're not speaking the same language, it's that they're not thinking the same way," Spengler says. But this culture gap—stemming from different views of the world—may itself be very fruitful in the long run. As Goethe's mathematician might have added: "Vive la différence." ■ **BARRY CIPRA**

Barry Cipra is a contributing correspondent of Science.