

# Sing a Song of Benzene, A Pocket Full of $\pi$

by Douglas L. Smith

*Herein lies a paradox: Organic chemicals and oily gunk are synonymous, as anyone who's ever taken an organic lab knows, yet water is the solvent of life.*

Greasy hair went out in the seventies, and greasy cooking in the eighties, but greasy amino acids are very much in vogue in the nineties, thanks to Caltech research showing that these gucky, oily residues are not just molecular filler, but a vital part of the protein in which they live.

Proteins, of course, are the molecular machines that actually do the work of the cell. A protein is a chain of hundreds of small building blocks, called amino acids, strung together in a specific order that differs for every protein. The amino acids have various functions endowed by their physical and chemical properties. Some functions are structural—making a hairpin turn that folds the protein back on itself, or creating a sheetlike surface that might form the docking site for another molecule. Other amino acids actually *do* things—they have side chains that can participate in chemical reactions. Others form links between the protein strands, and hold everything in proper alignment. But, says Professor of Chemistry Dennis Dougherty, “the aromatic, or benzene-containing, amino acids (phenylalanine, tyrosine, and tryptophan) were primarily considered to just be greasy organic stuff—hydrophobic, repelled by water—and that was it.” Now these molecular underachievers stand revealed as movers and shakers—the middlemen in communications between nerve cells.

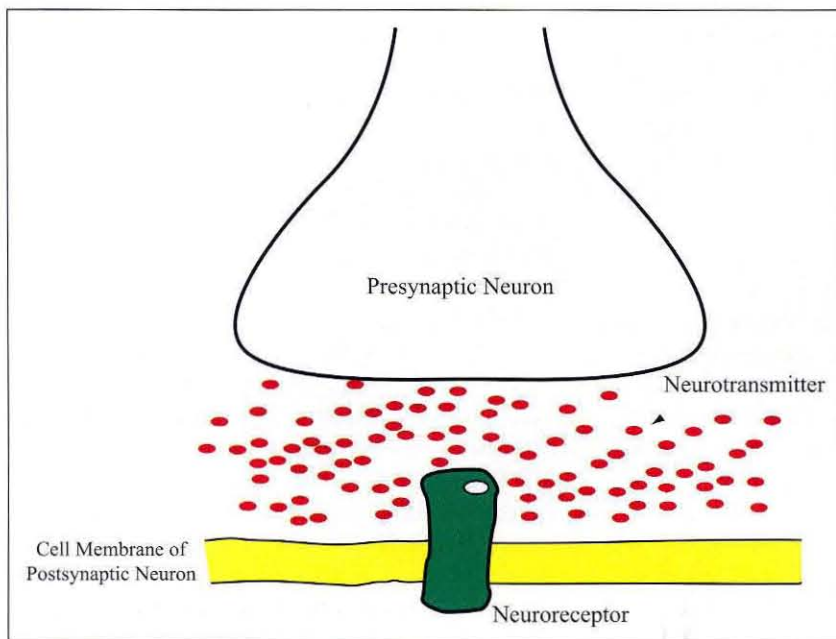
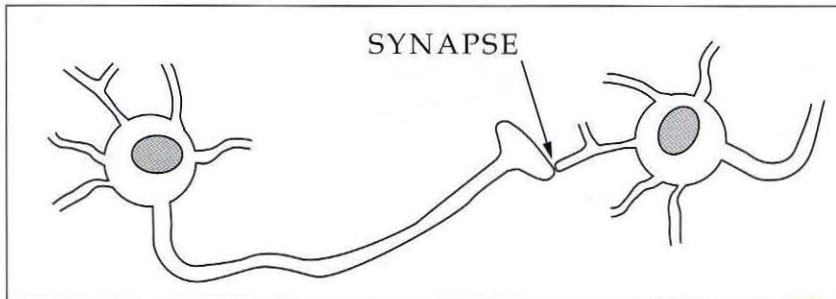
Nerve cells make very few direct electrical connections with each other; instead they rely on a chemical process called molecular recognition. They secrete chemical messengers, called neurotransmitters, that leap across the gaps between cells. A message is delivered when the neuro-

transmitter finds its receptor—a protein embedded in the surface of the receiving cell—and binds to it, causing a change in the receptor that triggers an electrical response within the cellular addressee. It's a wonderfully flexible system—there are about 50 known neurotransmitters, and any one cell can have receptors for several of them, each of which conveys a different message. A signaling system based on molecular recognition needs two things. First, the receptor must be selective enough to pluck out its messenger—it needs a pocket that fits the neurotransmitter just right. Second, the pocket's bond must be the molecular equivalent of Velcro—strong enough to hold the neurotransmitter, yet weak enough to let go quickly once the message is delivered. The greasy amino acids, aptly enough, appear to do part of the dirty work—they're in the part of the protein that actually recognizes the messenger.

The late Linus Pauling (PhD '25) won the Nobel Prize in chemistry in 1954 for figuring out the nature of the chemical bond, but the chemistry of biology today is increasingly the chemistry of these Velcro bonds. They are much weaker than chemical bonds, and form and come undone without affecting the underlying chemical structures. Says Dougherty, “Pauling also recognized that these weak interactions were going to be the key to biology, because biological systems are dynamic—they're not etched in stone, locked in place; things are flexible. And the molecules are large, so biological systems can amass a very large number of weak interactions to produce a strong effect. The amazing chemistry of life, in the end, often involves a lot of very weak interactions

**It is the very model of a major neurotransmitter. Professor of Chemistry Dennis Dougherty demonstrates how a molecule of acetylcholine, in his left hand, slips neatly into a molecule he designed. Acetylcholine is a common neurotransmitter—a chemical messenger that nerve cells use to communicate—and its correct fit into a receptor molecule is crucial to the transmission of its message.**

**Below: Neurons, or nerve cells, don't generally make direct electrical connections with one another, but are separated by a gap called a synapse. Bottom: An outbound nerve impulse triggers the release of neurotransmitter molecules into the gap. The molecules jump across the synapse in a millionth of a second and bind to neuroreceptor molecules that protrude from the surface of the receiving cell.**

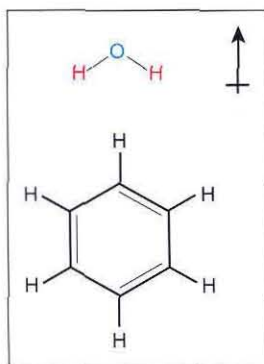


working in concert." (The receptor for a 25-atom neurotransmitter may contain some 30,000 atoms, for example. If the messenger were the size of your hand, the receptor would stand five feet tall, and you wouldn't be able to get your arms around its girth.)

The best known Velcro bond is the hydrogen bond, not to be confused with the ordinary covalent chemical bond that hydrogen usually forms. In fact, a hydrogen atom can't form a hydrogen bond to something unless that hydrogen atom has already entered into a covalent bond with some other atom—usually oxygen—first. Covalent bonds are based on the principle of share and share alike, with each atom contributing one electron to the binding pair. But the oxygen atom likes electrons a lot more than the hydrogen atom does—a phenomenon called electronegativity, which Pauling also elucidated—and greedily draws the hydrogen atom's electron toward itself. The hydrogen atom acquires a slight positive charge; the oxygen atom an equal negative one. This charge distribution is called a dipole. Opposites attract, so the positive end of one dipole will seek out and snuggle up to the negative end of another one. This dipole attraction between a hydrogen atom and a negative charge elsewhere is the hydrogen bond. (Nitrogen, and, to a lesser extent, sulfur, affect hydrogen the same way, although they aren't quite as electronegative as oxygen.)

And herein lies a paradox: Organic chemicals and oily gunk are synonymous, as anyone who's ever taken an organic lab knows, yet water is the solvent of life. Cells are about 70 percent water. Much of the common coinage of the cell—neurotransmitters, metabolic intermediates, regulatory molecules, and even pharmaceutical visitors from the outside world—carry positive charges in their biologically active forms. And water molecules have a huge dipole, with their two hydrogen atoms perched atop their oxygen atom like Mickey Mouse ears. The negative, or oxygen-atom, ends of these dipoles get right in there and nuzzle up to the positive charges on the active molecules as the water molecules cluster around them, dissolving them and making them available to the cell. So proteins make themselves soluble by wadding up in a way that exposes their dipolar, hydrogen-bond forming amino acids—handles for the water molecules to grab onto, so that they can drag the otherwise water-repellent proteins into solution.

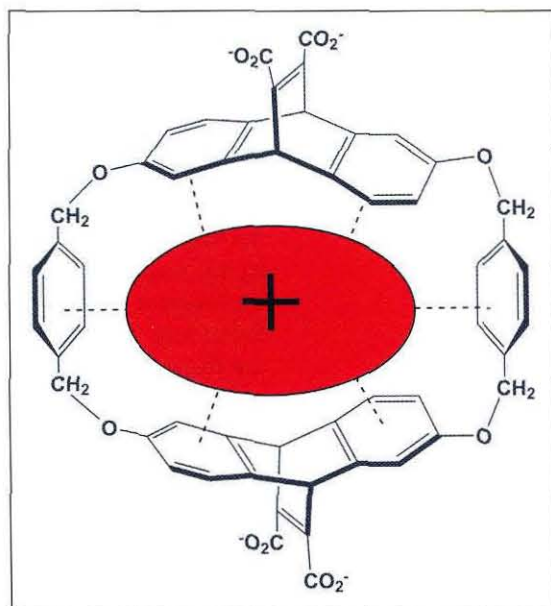
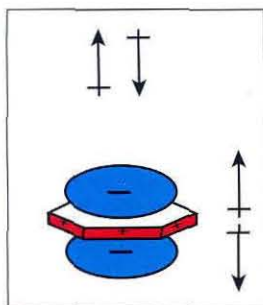
Three of the water-averse amino acids are made slick by benzene rings within their structure. Like dissolves like, and benzene—perhaps best known to the layperson as a nasty carcino-



**Above: A water molecule (top) has an asymmetric charge distribution, as shown by its red (for positive) and blue (for negative) atoms. This creates a dipole, symbolized as an arrow whose head points toward the negative charge and whose tail is a + sign. Benzene (bottom) is perfectly symmetrical and has no dipole.**

**Right: Dougherty's model system has six benzene rings that form Velcro bonds (dotted lines) to positive ions.**

**Below: A quadrupole is two dipoles pointing in opposite directions. They usually line up head to tail, as in the upper drawing, so that there's no net charge palpable to the outside world. But in benzene, they line up back to back and present substantial regions of negative charge, as shown at the bottom.**



gen—is an organic chemist's best friend, because it dissolves all those greasy things that water won't. Benzene is oil to water's water—they don't mix. Unlike water, benzene has no electronegative atoms, and its molecule is perfectly symmetrical. It's a flat, hexagonal thing that looks like a Susan B. Anthony dollar—remember them?—but with fewer edges. "Benzene doesn't have a charge, like a cation [positive ion] does," says Dougherty, "and it doesn't even have a dipole. It avoids water, and so it tends to be buried in the interiors of proteins"—and thus was slighted as filler.

This is not to say that grease doesn't have its place. The cell membrane is a double layer of fatty molecules that separates the water outside the cell from the water within. Thus proteins spanning the cell membrane—and there are a lot of them—obviously need fat-soluble regions in order to reside in that neighborhood. Other suety organic molecules play vital roles in the cell, too; they bind to one another weakly through nonpolar interactions.

Low-level interactions—polar and otherwise—pose problems for theorists, says Dougherty. "The quantum mechanics of bonding that Linus and others worked on aren't easily applicable to these weak interactions. It's *much* more difficult to describe them rigorously. Proteins are gigantic, complicated molecules, so we organic chemists design and build model systems—smaller, more manageable systems that we hope exhibit the same basic physical properties, and that can be studied much more rigorously."

Dougherty's model system is a doughnut-

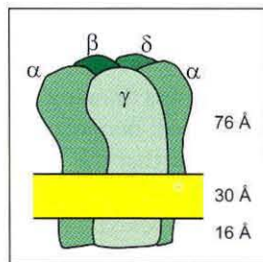
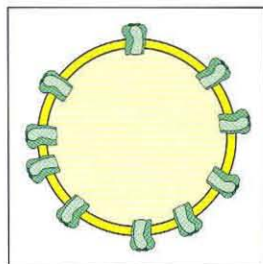
shaped molecule—the "host"—whose interior is lined with the flat faces of six benzene rings. Actually, the molecule looks more like a sandwich-sliced Kaiser roll, with each half of the roll containing two benzenes linked edge-on. Two spacers, each containing another benzene ring, keep the halves of the roll a set distance from each other, thereby defining a slot into which slips the sandwich filling—a small "guest" molecule. Sprinkled like sesame seeds around the Kaiser roll's crust are carboxylate groups—negatively charged ions (anions) that make the entire sandwich, including the hydrophobic benzenes, water-soluble. The researchers use nuclear magnetic resonance, or NMR, a common analytical technique, to see what happens next. When a guest molecule enters the host's cavity, the NMR signal suddenly changes in a manner that allows the strength of the interaction—those Velcro bonds—to be calculated.

"Initially, we—Tim Shepodd, Mike Petti [both PhD '88], and I—emphasized neutral organic molecules as potential guests," Dougherty recalls. "Like others, we saw that the strength of the interaction with our host was directly proportional to how insoluble the guest was in water. That is, guests were going into our host cavity not because they liked the host, but because they were so unhappy in water." But mere solvent repulsion is not the same as molecular recognition, so Dougherty "decided to emphasize guests that were still organic, but had considerable water-solubility—structures that really would make a choice between our host and an environment (the water) in which they were not entirely unhappy. If this kind of guest chose to go into the host, that would signal a true attraction between host and guest (versus repulsion between guest and solvent) and that would be molecular recognition. The way to make an organic molecule water-soluble is to add charge, and there is a greater variety of structures for organic cations than for anions." Although Dougherty was looking for weak interactions between nonpolar molecules, "we ended up seeing, to our surprise, that the benzene rings kept binding cations."

But how could this be? What was drawing these ions to the ultimate uncharged, nonpolar, unwaterlike molecule? It turns out that the organic chemists, and thus the molecular biologists and the biochemists, had overlooked something that the physical chemists had known all along, but which wasn't of great relevance to them—benzene has a quadrupole. A quadrupole is not *one* dipole but *two*, arranged so that they point in opposite directions; as a result, there's no net dipole. But just as a dipole has a more com-

**Top: The acetylcholine receptor (green) pierces the cell membrane (yellow), forming a channel into the cell through which ions can flow.**

**Bottom: The receptor is actually made of five closely associated protein strands, called subunits, each of which has its own gene and is assembled independently before coming together to form the receptor. There are two identical  $\alpha$  subunits, and one each of three others ( $\beta$ ,  $\gamma$ , and  $\delta$ ), per receptor. The measurements at right are in Angstroms, or ten-billionths of a meter. The acetylcholine molecule itself, in comparison, is about 10 Angstroms long.**



plex electric field than does a single charge, a quadrupole has a more complex field than a dipole. In benzene's case, a set of six electrons called  $\pi$ -electrons live above and below the plane of the hexagon, zipping around it like midjet racers on a dirt track. This swirl of electrons creates a substantial negative charge on the hexagon's two flat faces, balanced by an equal-magnitude (but more diffuse) band of positive charge along the milled edge of the coin, as it were. And there you are—opposites attract, so it should be no surprise that the cations came a-flocking in what Dougherty has christened the cation- $\pi$  interaction. Quadrupoles are not terribly exotic—electrical engineers and physicists use them in all sorts of ways, from focusing beams of charged particles to describing Jupiter's magnetic field.

"The novel thing about benzene is that it is simultaneously hydrophobic yet polar. Most organic chemists think of that as a contradiction, but it's a fact. Fundamentally, everybody knew that benzene had a quadrupole; and everybody knew that in principle it could do something—the surprise was that the interaction was much bigger than any of us had anticipated. *So* big, that the quadrupole of benzene is able to compete with the dipole of water to bind cations. They will actually leave water, where they tend to be very happy, and race into an otherwise hydrophobic environment. This is totally backward from the way things normally operate." The cation- $\pi$  interaction is roughly one-fifth the strength of a covalent bond and about five times stronger than a hydrogen bond—a middle ground that, happi-

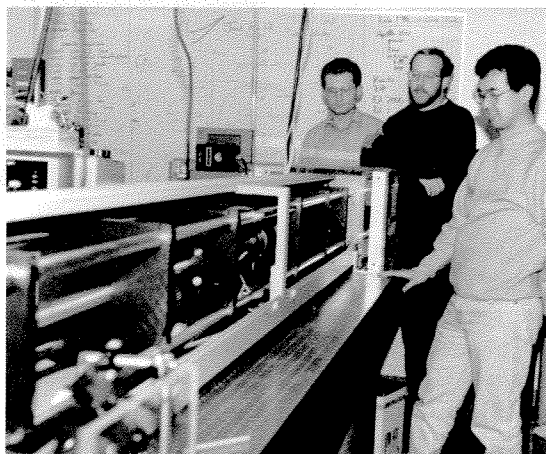
ly, is just the right degree of stickiness to grab hold and let go readily at room temperature. And this is why benzene is ideal for forming Velcro bonds.

Further experiments showed that the cations the model system bound especially well looked a lot like acetylcholine, the first to be discovered of those 50-odd neurotransmitters mentioned earlier. The acetylcholine receptor is a ring of five large protein strands, bunched like a fistful of cigars, that penetrates the cell membrane. When the receptor binds to an acetylcholine molecule, a shiver runs down the length of the proteins, causing them to shift their bulk slightly away from each other and opening a channel into the cell's interior. A torrent of ions—an electric current—courses through the channel, galvanizing the cell into action. "So," recalls Dougherty, "we asked ourselves, 'Does nature bind acetylcholine the same way we do in our model?'" Dougherty posed this question to Henry Lester, professor of biology, and to Norman Davidson, the Chandler Professor of Chemical Biology, Emeritus, who've been jointly studying the acetylcholine receptor for many years.

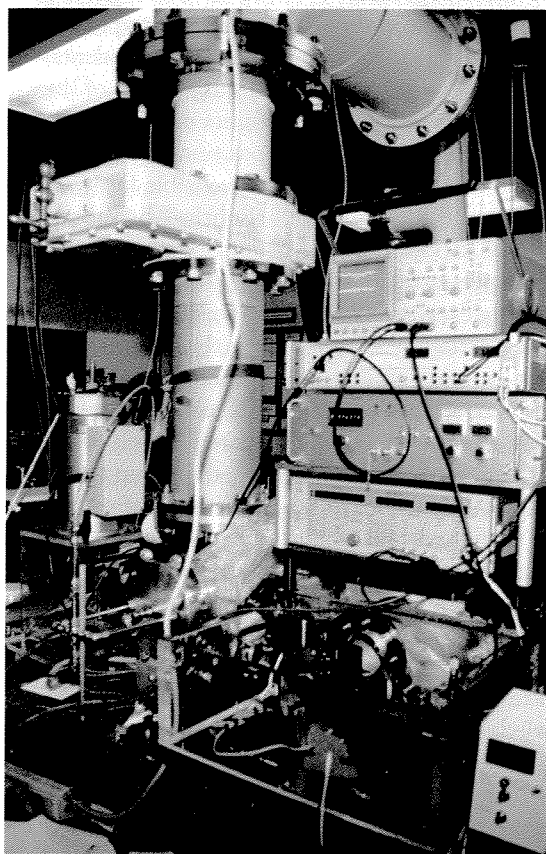
This isn't an easy question to answer. Back in the early 1980s, Davidson's group had helped to find the sequence in which the protein's amino acids are strung together. But the biologically active protein bears as much resemblance to that sequence as a tangle of Christmas lights fished out of the bottom of the decoration box does to the lights when strung along the eaves. The usual way to figure out a protein's structure is to purify a sample of the protein, dissolve it in something from which it will slowly crystallize out, and determine the three-dimensional structure of the crystal by bombarding it with X-rays and analyzing how they're scattered. But most proteins that span the cell membrane have so far defeated attempts at crystallization. Separated from the membrane's embrace, the proteins lose their all-important shape, and the resulting crystal structure is meaningless. And left in the membrane, the doggone proteins just won't crystallize, because the membrane glop prevents the molecules from stacking neatly. "There are thousands of these incredibly important proteins," says Dougherty. "These are the molecules of thought. This *is* the brain, at the molecular level, and we don't have structures of them."

Chemical intuition comes to the fore in such situations, and Dougherty's told him that benzene's oleaginous mien might mask a clean-cut pillar of the molecular community. So David Stauffer (PhD '89) looked up all the known amino acid sequences of acetylcholine receptors,

**Top: (From left) Rodham, Blake, and Suzuki contemplate how much more room they'll have in their lab next year, once the carbon-dioxide laser before them is replaced by the shoe-box-sized model. The CO<sub>2</sub> laser pumps a far-infrared laser (not shown).**



**Bottom: Since far-infrared lasers aren't tunable, the group makes light of the wavelength they need by mixing the laser beam with a microwave signal of adjustable wavelength. To prevent airborne moisture from absorbing the infrared light, the mixing apparatus and all of the infrared optical components live in the Plexiglas box. The bank of electronics above it is the microwave generator. The column to the left that resembles battleship plumbings is the exhaust system for the vacuum chamber, which is the horizontal cylinder to the left of the mixer.**



and discovered few of the anions one would normally expect to stick to cations, but scads of benzene rings. "So we went public with our prediction that acetylcholine binding sites would be rich in benzene rings. That was in 1990. And in 1991, the structure of the first acetylcholine-binding protein was solved. And to make a long story short, that structure validated our prediction. Spectacularly, in fact—14 benzene rings all over the place. And it's absolutely clear that the binding is due to benzene rings." (This molecule was actually an enzyme called acetylcholinesterase, which binds to used acetylcholine molecules and breaks them down into choline and acetic acid. The enzyme's business end drifts in the watery intercellular medium and is anchored to the cell—or, in some cases, the gel that fills the synapse—by a long, fat-soluble tail. The people who solved the enzyme's structure cut the anchor line and recrystallized only the water-soluble portion.)

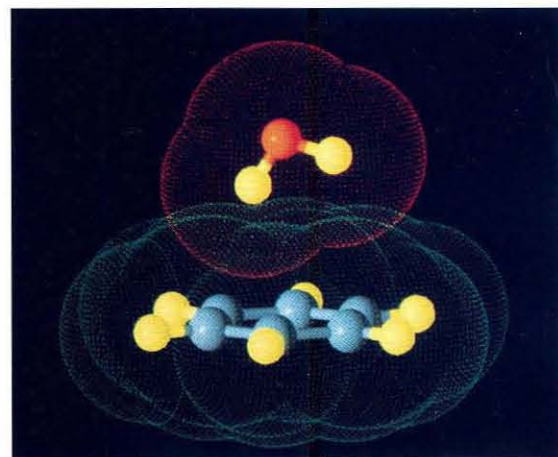
Meanwhile, Linus Pauling turned 90. Since he was only a decade younger than the Institute itself, Caltech seized the occasion—February 1991—to throw him a birthday bash as part of the Centennial celebration. Among the speakers who gave papers on current work in fields Pauling had tilled over his long career was Nobel Laureate Max Perutz, who spoke on the significance of the hydrogen bond in physiology. In the audience that day was Associate Professor of Cosmochemistry Geoffrey Blake (PhD '86), who had been studying how clusters consisting of two or three small molecules form hydrogen bonds with one another. Such clusters are simple models for the water- and methanol-rich ices present in the interstellar medium and outer solar system—cosmic dust bunnies that slowly accrete into stars, comets, planets, and what have you. Recalls Blake, "Perutz gave a talk saying how unusual these benzene interactions were. And we went off and looked in the literature and almost no work on mixed benzene clusters had been done. There had been a lot of work on snowballs of pure benzene, but that was it. We were absolutely shocked! It's hard to think of a more important set of clusters to look at."

His interest piqued, Blake and grad students Sakae Suzuki and David Rodham and postdoc Peter Green (now a senior scientist in Caltech's Bank of America Environmental Analysis Center) started looking at clusters consisting of one benzene molecule and a molecule of either water or ammonia. The experimental method was quite simple—spray a benzene-water mist, carried by an inert gas, into a vacuum chamber and shine a laser through the cloud to look for spectroscopic

*The water molecule spins freely around the hydrogen bond like a figure skater doing an arabesque—spinning on one hydrogen with the other one sticking straight out behind.*

evidence of hydrogen bonding. Since all the molecules in the mist are traveling in the same direction at essentially the same speed, and in the molecular world speed equals temperature, “they think they’re cold,” Blake explains. “Their relative velocities are characterized by temperatures of only a few degrees Kelvin. So the collisions are very soft, and that’s why things stick together.” If the collisions had any more oomph, the molecules would rebound too hard for hydrogen bonds to form.

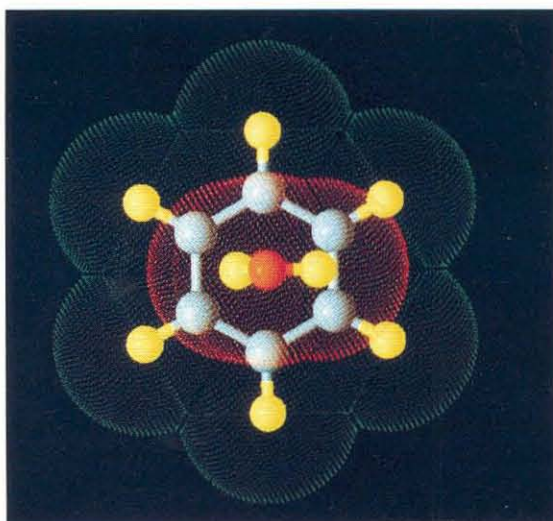
The Blake group had their first data in hand by April and, in collaboration with a theorist whose specialty is computer simulations, William Goddard III (PhD ’65), the Ferkel Professor of Chemistry and Applied Physics, sat down to interpret the results. The art of spectroscopic interpretation consists of assigning every line in your spectra to a specific physical action by atoms in your sample—a certain bond bending or stretching, for example, or one part of the cluster rotating with respect to the rest of it. Blake’s group would tell Siddharth Dasgupta, a member of the Beckman Institute in Goddard’s group, what they thought the cluster’s structure was. Dasgupta would then determine whether that structure was energetically favorable, predict exactly where all the atoms should be, and ascertain how hard it would be to rotate the water or pull it out of alignment with respect to the benzene. Armed with that knowledge, the cosmochemists would figure out where the spectral lines should fall. Then everyone would twiddle with the hypothetical structure—jinking the water molecule about, cocking its spin axis



various ways, twirling it at different speeds—to try to make the lines generated by the proposed structure match the lines in the real spectra. An exact match indicated that the hypothesis accurately depicted the real molecules. “In theory,” says Blake, “there are ways to do it with computers, but in practice the human mind is much better at recognizing incomplete patterns and making extrapolations, so the students and I—mostly the students—have spent long hours staring at lists of lines and plots and trying to figure out what the assignments are.” A line’s position depends on the masses of the atoms responsible for it, so varying one atom’s mass slightly—by substituting deuterium for hydrogen, for example—causes its lines to shift, making them stand out against the fixed background of the other atoms’ unmoving lines. So by repeating the experiment over and over again with minor variations in the masses of the atoms, Blake’s group refined the calculations, predicting more precisely where the peaks should be, allowing the group to take better data, and so on.

Every line contains vital information because a hydrogen bond, unlike a covalent bond, is a dynamic beast even at low temperatures. It doesn’t show up directly as spectral lines, so its presence must be deduced from a detailed analysis of the lines you *do* see. For example, the microwave frequencies tell how the cluster is tumbling, and seeing that the water molecule and the benzene molecule are spinning around a common axis could be a sign that a hydrogen bond between the two runs along that axis. This is the best place to start interpreting the spec-

**In a discovery that gives a whole new meaning to the term “water ballet,” it turns out that a water molecule (red) can pirouette gracefully atop a benzene molecule (green), as shown on the opposite page. The dot patterns represent the atoms’ surfaces, while the ball-and-stick models within show the positions of the atomic nuclei. The water molecule is actually centered over the benzene ring, as shown at right. One hydrogen atom thrusts down into the center of the ring’s  $\pi$ -electron cloud. The other lines up with the thickest part of the cloud, which lies directly over the carbon atoms.**



trum, says Blake, because “that’s a positive, definite number—you can’t have negative rotational frequencies. And you know roughly how heavy the molecule is, so you know within 10 percent or better where something’s going to show up.” Things get a lot hairier in the far infrared, where the vibrational frequencies lurk. The number of lines grows beyond belief. And worse, each line’s precise location can wander greatly—the vibrational motions are strongly coupled to one another, like pendulums tied together by a spring, so that what happens to one vibrational mode affects the spectra of others.

“In the end,” says Blake, “when we really assign a spectrum, all the lines fit to within a part in  $10^6$ . There’s no uncertainty. And that’s the attractive thing about this kind of spectroscopy compared to, for example, protein studies, where you can have a few tenths of an Ångström’s slop in the electron diffraction. But it also means that you’d better know something about what you expect to see going in. What saves these experiments is that no matter where you look, you see something. The challenge is to figure out exactly what it is.”

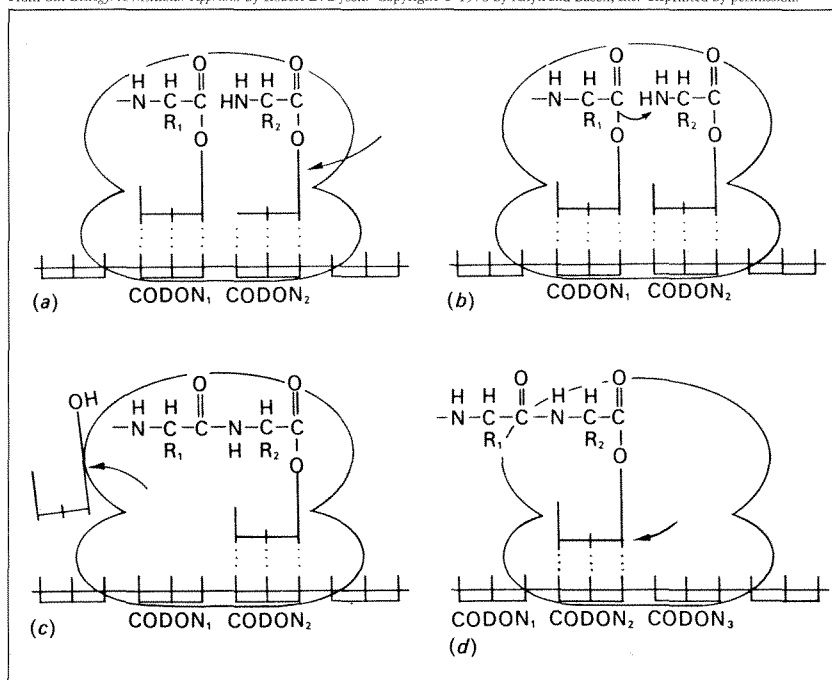
What it was—proof positive that the flat face of a benzene molecule readily makes a strong hydrogen bond with a water molecule hovering over that face—made the cover of *Science* on August 14, 1992. (The water molecule spins freely around the hydrogen bond like a figure skater doing an arabesque—spinning on one hydrogen with the other one sticking straight out behind.) A companion paper with similar results for benzene and ammonia followed in *Nature* in

early 1993. Both papers included a description of the deepest valley in what chemists call a potential-energy surface—a multi-dimensional description of the strength of the interaction between the molecules, depending on their separation and relative orientation. The lowest point on the potential-energy surface—the configuration in which the system has the least potential energy—is equivalent to the strongest interaction.

The group is now climbing out of that valley and exploring the hills around it. Says Blake, “We’d like to find out more details—for example how much energy does it cost to stretch that bond or to twist the water or the ammonia—and that requires moving up in energy.” When a protein kinks up into its active shape, it will do its best to minimize its potential energy by squirming around like a restless traveler in an airplane seat until it’s most comfortable, but there may be no way to bring two amino acids that want to form a hydrogen bond into the orientation corresponding to the deepest valley. Thus it’s important to know what other valleys may be found at higher energies. This information will ultimately be rendered mathematically, in collaboration with the computational chemists, as force fields describing how amino acids attract or repel one another. The idea is that eventually one will be able to type the amino-acid sequence of a protein into a computer, and the computer will use the force fields to pull the protein into its natural shape.

Making the clusters is easy, compared to making laser light of the right wavelength. “That’s the technical area where things are really changing,” says Blake. “The experiment itself now lives on a five-by-twelve-foot optical table. We think there are some new techniques that will make it fit in a shoe box.” And the big lasers have limited operating ranges, so you have to keep changing lasers as you scan across the spectrum. But the shoe box model will contain a single tunable laser. Just punch in a frequency, and—bingo!—there you are. This part of the project has brought Blake into collaboration with a lot of laser and detector gurus—Assistant Professor of Physics Jonas Zmuidzinas (BS ’81), Associate Professor of Applied Physics Kerry Vahala (BS ’80, MS ’81, PhD ’85), Associate Professor of Astrophysics Kenneth Libbrecht (BS ’80), and Professor of Physics Jeff Kimble. “That’s the thing about Caltech,” remarks Blake. “It’s small enough so that you get to meet people like Kerry and Jeff, whereas at a bigger university you might not.” The mixers—which make the desired wavelength of light by combining two photons of other wavelengths—are being built





**The protein-manufacturing process. The mRNA runs across the bottoms of the figures; the ribosome is the blob sitting on it. (a) A tRNA-amino acid complex enters stage right, recognizes codon 2, and forms hydrogen bonds to it. This places the incoming amino acid next to the protein's growing end, which is Velcroed to the mRNA at codon 1. (b) The chemical bond attaching the protein's terminal amino acid to its tRNA transmutes into a new chemical bond that attaches the new amino acid to the protein. (c) The tRNA that formerly bound the growing protein to the mRNA departs in search of another amino acid. (d) The ribosome creeps down the mRNA one codon, setting up codon 3 as the next recognition site.**

at MIT's Lincoln Labs by Elliott Brown (MS '81, PhD '85, and a labmate of Blake's when they were both grad students), who will escape some Massachusetts cold by bringing them out to California in December. Blake hopes to spin this technology off into environmental studies in a few years by flying the shoe box on NASA's ER-2 spy plane or Perseus unmanned aircraft, where it could replace several instruments now used to track nitric acid, ozone, carbon monoxide, and various chlorine compounds. And the shoe box may also ride on a European Space Agency mission called FIRST (Far Infrared and Submillimeter Space Telescope), which will search the cosmos for various gases, plus those water- and methanol-containing ices that got Blake into this line of work in the first place.

Getting back to the brain, or at least to neuroreceptors, we now have two independent lines of evidence showing that benzene can make Velcro bonds in two different ways. But having a crystal structure that puts benzene rings at the scene of the bind, and spectroscopic analyses that show an M.O.—that benzene's negatively charged face will indeed interact with even a partial positive charge—is not the same as an eyewitness account of a neurotransmitter being recognized by benzene rings in a living cell.

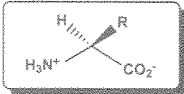
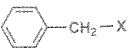
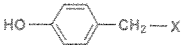
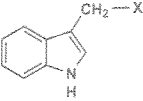
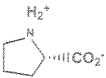
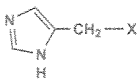
So Dougherty, Lester, Davidson, and John Abelson, the Beadle Professor of Biology, are hoping to become the star witnesses. Lester's group alters the gene that tells the cell how to make the protein, swapping out a benzene-containing amino acid for a different one. Then the researchers inject the modified gene into an

unfertilized frog egg, which obligingly churns out the protein molecules and inserts them into the cell membrane. The group then assays the protein's function by a series of electrical measurements. If the modified protein behaves like the original one, then the change obviously wasn't important. But if the new protein behaves oddly or doesn't work at all, then the missing amino acid does something vital. And by replacing that vital amino acid with ones having a range of different properties, the researchers can sometimes infer what that something is. Dougherty notes, "You need such a broad range of disciplines for this project—organic chemistry, molecular biology, electrophysiology—that it would be very hard to do at larger places. But at Caltech, I talk to Henry, he calls John, and five minutes later we have a collaboration."

The 20 amino acids on nature's palette limit one's freedom to experiment with the structure. But in the late 1980s, Peter Schultz (BS '79, PhD '84), a professor of chemistry at UC Berkeley, figured out how to put an amino acid that nature had never designed into a protein. The trick was to suborn the molecules of tRNA that bustle about the cell looking for the right amino acids to feed to the protein-assembling machinery. The tRNAs get their instructions by "reading" a blueprint molecule called mRNA that encodes each amino acid as a sequence of three "letters" chosen from a four-letter alphabet—A (for adenine), C (cytosine), G (guanine), and U (uracil). Such a sequence is called a codon. And, as Watson and Crick discovered in DNA, G binds only with C, and A only with U, a pairing that—

The 20 naturally occurring amino acids. They share the structure in the box (symbolized by the "X" in the detailed structures), which is how they make links with their fellows—the  $\text{CO}_2^-$  of one amino acid reacts with the  $\text{H}_3\text{N}^+$  of its neighbor. The "R" in the boxed structure represents the rest of the amino acid. Thus glycine (the top entry in the right column) is the boxed structure when "R" is replaced by a hydrogen atom. Note that, in this "traditional" classification, tyrosine is considered to be polar by virtue of its OH group, which can form hydrogen bonds.

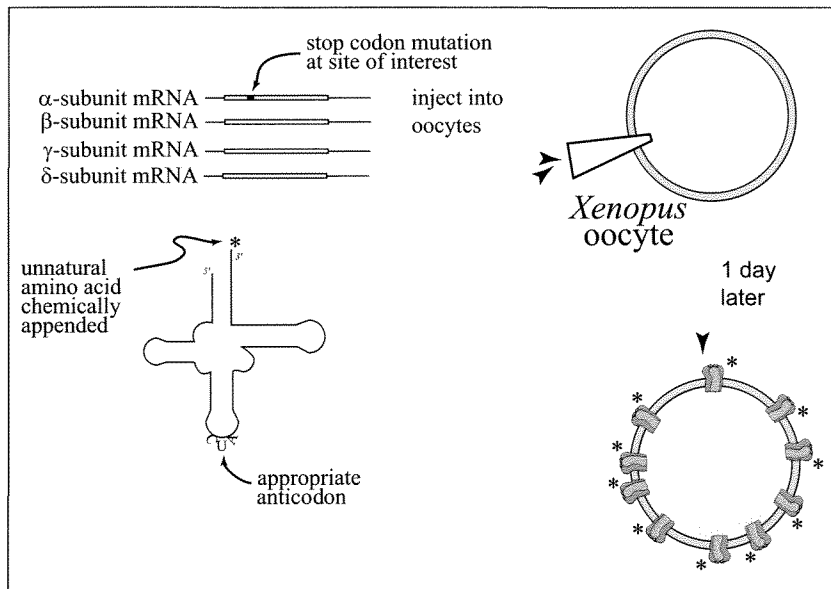
"Amber," in this case, has nothing to do with Jurassic Park. The codon was named for a chap named Bernstein, who discovered it at Caltech some 30 years ago. "Amber," in German, is "Bernstein."

THE COMMON AMINO ACIDS			
<b>NON-POLAR</b>		<b>POLAR, BUT NEUTRAL</b>	
ALANINE Ala - <b>A</b>	$\text{CH}_3\text{---X}$	GLYCINE Gly - <b>G</b>	$\text{H---X}$
VALINE Val - <b>V</b>	$\begin{array}{c} \text{CH}_3 \\   \\ \text{CH---X} \\   \\ \text{CH}_3 \end{array}$	SERINE Ser - <b>S</b>	$\text{HO---CH}_2\text{---X}$
LEUCINE Leu - <b>L</b>	$\begin{array}{c} \text{CH}_3 \\   \\ \text{CH---CH}_2\text{---X} \\   \\ \text{CH}_3 \end{array}$	THREONINE Thr - <b>T</b>	$\begin{array}{c} \text{HO} \\   \\ \text{CH---X} \\   \\ \text{CH}_3 \end{array}$
ISOLEUCINE Ile - <b>I</b>	$\begin{array}{c} \text{CH}_3\text{---CH}_2\text{---CH---X} \\   \\ \text{CH}_3 \end{array}$	CYSTEINE Cys - <b>C</b>	$\text{HS---CH}_2\text{---X}$
PHENYLALANINE Phe - <b>F</b>		TYROSINE Tyr - <b>Y</b>	
TRYPTOPHAN Trp - <b>W</b>		ASPARAGINE Asn - <b>N</b>	$\begin{array}{c} \text{O} \\    \\ \text{H}_2\text{N---C---CH}_2\text{---X} \end{array}$
METHIONINE Met - <b>M</b>	$\text{CH}_3\text{---S---CH}_2\text{---CH}_2\text{---X}$	GLUTAMINE Gln - <b>Q</b>	$\begin{array}{c} \text{O} \\    \\ \text{H}_2\text{N---C---CH}_2\text{---CH}_2\text{---X} \end{array}$
PROLINE Pro - <b>P</b>		<b>BASIC (CATIONIC)</b>	
<b>ACIDIC (ANIONIC)</b>		LYSINE Lys - <b>K</b>	$\text{H}_3\text{N}^+\text{---CH}_2\text{---CH}_2\text{---CH}_2\text{---CH}_2\text{---X}$
ASPARTIC ACID Asp - <b>D</b>	$\begin{array}{c} \text{O} \\    \\ \text{---C---CH}_2\text{---X} \end{array}$	ARGININE Arg - <b>R</b>	$\begin{array}{c} \text{NH}_2 \\   \\ \text{H}_2\text{N}^+\text{---C---NH---CH}_2\text{---CH}_2\text{---CH}_2\text{---X} \end{array}$
GLUTAMIC ACID Glu - <b>E</b>	$\begin{array}{c} \text{O} \\    \\ \text{---C---CH}_2\text{---CH}_2\text{---X} \end{array}$	HISTIDINE His - <b>H</b>	

wouldn't you know it!—involves hydrogen bonds. Each tRNA has two working ends. One end, called the anticodon, contains the three-letter sequence that recognizes the codon and pairs up with it. The other end binds only to a specific amino acid. The amino-acid-laden tRNAs and the mRNA are brought together by a marvelous device called a ribosome—part protein and part RNA—that helps rivet the amino acids together.

However, three codons—UAG, UAA, and UGA—don't have a corresponding tRNA. (Molecular geneticists of a previous generation dubbed these codons "amber," "ochre," and "opal." "Amber," in this case, has nothing to do with *Jurassic Park*. The codon was named for a chap named Bernstein, who discovered it at Caltech some 30 years ago. "Amber," in German, is "Bernstein.") When the ribosome reaches one of these codons, there's no amino acid to process, and the riveter lurches to a halt. These three sequences tell the assembly line that the protein is finished.

In 1986, Abelson and his group had learned how to change the anticodon of an otherwise normal tRNA so that it would pair with the amber codon. Then, using the standard tools of molecular biology, they could insert an amber codon into the gene at a location of their choosing, replacing the codon that normally occurred there. The amber-binding tRNA would then insert its habitual amino acid into the nascent protein at that point, and the ribosome would carry on assembling the altered protein. (But first, one has to make sure that none of the active



**Below: Physics grad student Haiyun Zhang now injects frog eggs in Lester's lab—another discipline heard from!**



**Above: To grow a neuroreceptor, two solutions are injected into the *Xenopus* oocyte, or frog egg, simultaneously. (*Xenopus* is a genus of South African clawed frogs.) One solution contains the four mRNAs needed to make the subunits of the acetylcholine receptor. An amber codon has been introduced into the  $\alpha$  subunit at the point where the unnatural amino acid is to be inserted. The other solution contains amber-binding tRNA molecules to which the unnatural amino acid (the purple asterisk) has been attached. (The egg already has a full complement of normal tRNAs and amino acids.) Then the cell gets to work building the receptors and dispatching them to the membrane.**

mRNAs in the cell have an amber stop codon. Any amber codons must be changed to ochre or opal in order for the protein to be assembled normally.) Schultz extended this idea by building tRNA molecules that had the amber-binding anticodon on one end, and his choice of doodad on the other. As long as the doodad was an amino acid—even if it wasn't found in nature—the ribosome would happily rivet it into the protein.

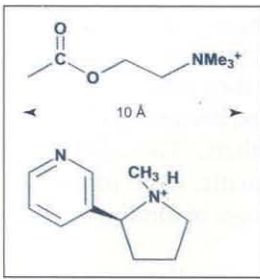
This proved to be a trickier proposition than it undoubtedly first appeared, but Schultz finally got it to work. However, it would only work in a test tube. This isn't a problem for many biologists—it might even save them the bother of having to extract the protein from the cell—but it's fatal for neurobiological work. It takes a living cell to put the protein in a lifelike pose in the membrane. Postdocs Mark Nowak and Patrick Kearney (who got his PhD in '94 from Dougherty) in Lester's group teamed up with senior research fellows Jeffrey Sampson and Margaret Saks in Abelson's group, says Dougherty, and worked "with Pete to modify the Schultz protocol in very clever ways—I can say that, because none of them were my ideas—so that we've gotten it to work in a living cell."

Once the protein has carpeted the cell surface, Lester's lab uses sensitive electrical instruments, including a device called a patch clamp (whose inventors, Erwin Neher and Bert Sakmann, received the Nobel Prize in 1991) to zoom in on a single channel—i.e., the receptor protein—and follow its behavior. The patch clamp is basically a glorified eyedropper—a piece of glass tubing drawn out to a blunt tip about one millionth

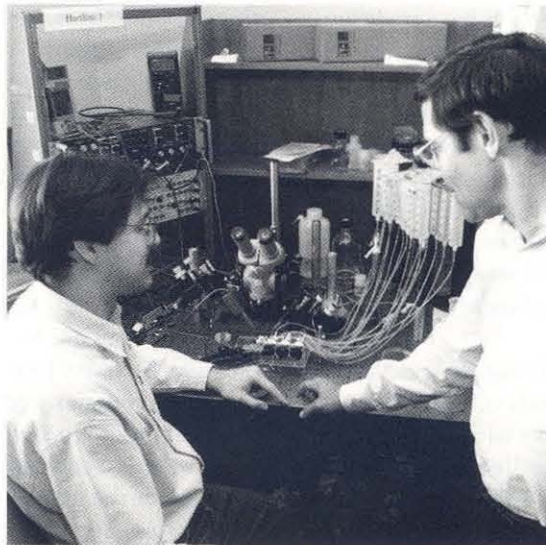
of a meter in diameter. Fill the dropper with conducting liquid, place the point against a frog egg and apply a little gentle suction, and the cell seals against the dropper. Putting a wire in the conducting liquid creates an electrode that measures the current through the patch of cell membrane to which it's sealed. And if conditions are right, only one channel will open at a time in that piece of membranous real estate. "The patch clamp is an amazing tool," Dougherty exclaims. "You're getting a signal from one molecule in real time! Physicists get all excited when they see a signal from a single molecule, and biologists have been doing this for a decade."

The frog eggs live in a Petri dish filled with liquid nutrients and awash with the neurotransmitter in question, which is supplied through a metering apparatus that looks like an IV drip. The receptor is put through its paces by varying the concentration of its neurotransmitter, making it open and close its channel as it binds and releases the messenger, like the machine that tests car doors by slamming them over and over and over again. The eggs are pretty durable, says Lester. "They're good for anywhere from a couple of days to a couple of weeks, depending on how carefully we handle them."

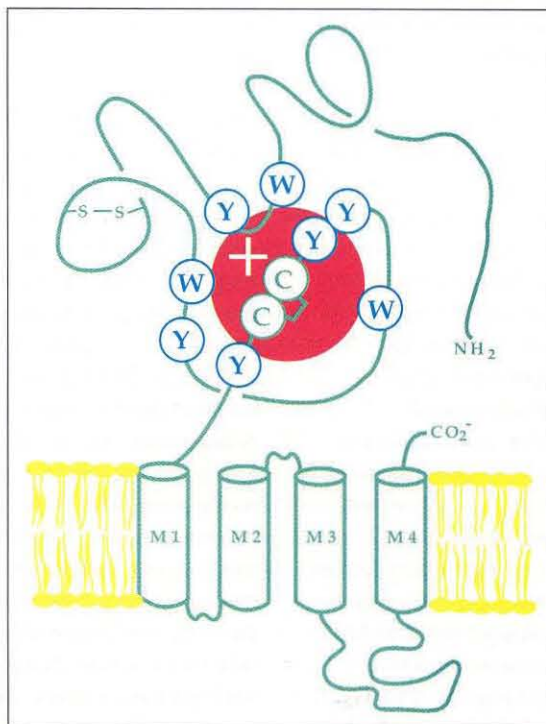
A typical experiment begins with the channel closed; the voltage through the patch clamp electrode is the baseline. Then the messenger binds to the receptor and the channel opens, changing the voltage. The electrode stays at this voltage until the receptor lets go, and the channel closes again. The trace from the electrode resembles a series of mesas on the Arizona desert floor—up



**Above: Acetylcholine (top) and nicotine (bottom) may look different to us, but the receptor can't tell them apart. Both molecules are about the same size, and both have a positively-charged nitrogen atom on one end and an exposed electron pair (in either the C=O or the second nitrogen atom) on the other. Far right: A serving of *Xenopus* caviar, injected and ready to go. Right: Nowak (left) and Lester (right) at a patch-clamp setup. The dish full of oocytes goes under the microscope, where the patch clamp is held in place by adjustable screws. The drip lines at right deliver nutrients and neurotransmitters.**



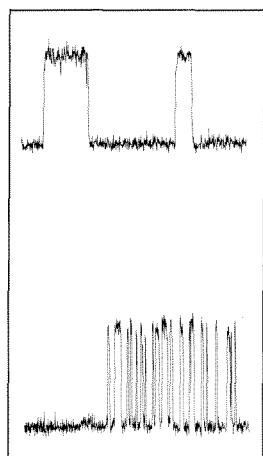
**Right: The acetylcholine binding site flies above the cell membrane as part of the channel's  $\alpha$  sub-unit (green). One might expect to find the anionic amino acids (aspartic acid and glutamic acid, abbreviated D and E, respectively) here. Instead, cysteine (C) is electrically neutral, and tryptophan (W) and tyrosine (Y) contain benzene rings. The cylinders labeled M1-M4 are the membrane-spanning regions that anchor the protein, and the yellow two-tailed polliwogs are the fatty molecules that make up the cell membrane.**



for a while, down for a while, up again, then down, and so on. By measuring how the width and number of the mesas varies with the concentration of the neurotransmitter, one is actually measuring the strength of its attraction to the receptor.

Most of the experiments to date have been done with the nicotinic acetylcholine receptor, so called because nicotine binds to it as strongly as does its intended messenger molecule, acetylcholine. There are about half a dozen different types of nicotinic acetylcholine receptors. One type makes muscles contract, but all the rest are found in the brain. The interaction between nicotine and one or more of these receptors is the first step in nicotine addiction, says Lester, so it's hardly surprising that this molecule is of intense interest to a lot of people. The receptor's binding site includes several tyrosines, one of the benzene-containing amino acids, and it's known that these tyrosines somehow contribute to the binding. The Lester-Davidson-Dougherty-Abelson collaboration has substituted unnatural amino acids for three of those tyrosines. Each one of the three appears to make a different kind of contact with the messenger molecule, Lester explains, because when they are replaced in turn by the same set of unnatural amino acids, a variant that grips acetylcholine tightly when standing in for one tyrosine has only a weak effect, or none at all, in another tyrosine's spot. "Furthermore," he adds, "it looks as though one of these places is a good candidate for the cation- $\pi$  interaction that Dougherty has been predicting." Another site appears to form a hydrogen bond.

**An electrical recording from an individual acetylcholine receptor in the absence (top) and presence (bottom) of a blocker chemically related to the local anaesthetics. When the trace is up, the ion channel is open and flowing at a rate of about 10 million ions per second. In the top trace, the channel opens and closes normally. In the bottom trace, blocker molecules bind within the open channel, stopping the ions' flow. Each chasm in the trace corresponds to the binding of an individual blocker molecule, which then dislodges and the flow resumes. Each trace is about a tenth of a second long.**



These concepts—and the benzene motif—apply to drug design as well. When something other than the intended messenger binds to the receptor, one of two things can happen. The drug may bind so well that it blocks the receptor permanently—spilling Super Glue on the Velcro, as it were. The channel never opens, and the trace becomes flatter than Iowa. Or the interfering molecule may fit well enough to bind, sort of, but not well enough to bind perfectly. The drug wobbles in and out of the receptor's grip, causing the channel to bounce open and closed like a screen door on a tight spring. Now the trace looks like the Badlands of South Dakota—all spires and chasms. Or, as Lester puts it, "they have a lot of flicker."

"Pharmacologists and chemists have traditionally had access to the enormous power of synthetic organic chemistry," Lester notes. "One can make a large number of organic chemicals and test how they affect the function of the neuron. So over the years, classical pharmacology has developed a large number of highly specific drugs, and also specific hypotheses about how these drugs interact with their receptors. But we have not had the structural tools to test these hypotheses."

So Lester, Davidson, and Dougherty are keen to install unnatural amino acids in many other proteins. Lester's group works with a broad spectrum of what he calls "excitability proteins"—the molecules that give a cell the ability to send and receive chemical and electrical messages. These include the neurotransmitter receptors and ion channels described above, plus the neurotransmitter transporters, which shepherd individual neurotransmitter molecules across the cell membrane rather than opening a floodgate as ion channels do. This is a very big field indeed—in fact, it's the entire grain belt. At the most fundamental level, these molecules regulate what gets into and out of nerve cells; on an intermediate level, they give the heart, the diaphragm, and every other muscle in the body their marching orders; and at the highest level, they are the molecules of thought. At every level, these molecules bind to drugs. Some are good—anti-epileptics such as Dilantin; antipsychotics, including Thorazine; the beta-blockers Atenolol and Inderol, which control high blood pressure; antidiabetics such as Glucotrol; and even the lowly "water pills," or diuretics, that go by names like Diuril and Clotide. Some aren't—cocaine, which binds to a dopamine transporter, and LSD, which does the same to a serotonin receptor, spring to mind. And some are a little of both—morphine and other opiate painkillers, including

heroin; the benzodiazepines, of which Valium is the most notorious; the list goes on and on. Knowing more about how they bind is essential to improving existing drugs, designing new ones, and curbing the abuse of others. The collaboration's methods allow systematic, molecular-scale investigations of a broad range of binding phenomena.

"All of these insights into binding interactions are critical for the pharmaceutical industry," Dougherty remarks. "I've seen reports, for example, of a drug that worked OK and had an anion in it that was assumed to interact with a cation at the receptor. They then made a new drug where the anion was replaced by a benzene ring, and sure enough, it worked very well. And the cool thing about that is, you've taken an anion, which is very water-soluble, and replaced it with a benzene ring, which is very water-insoluble, so that you've massively changed the drug-distribution properties." This drug will really want to leap out of the bloodstream and into the cell's greasy membrane.

Benzene binding to cations certainly isn't the be-all and end-all of molecular recognition—there are other well-known factors at work, and probably many unknown factors as well—but the biologically active form of many, many important molecules contains a positive charge, and there's an awful lot of greasy stuff in the cell. Not just in nerve cells—molecular recognition recurs throughout the immune system, and in enzymatic processes generally. And nature is lazy—once it finds something that works, the same trick reappears over and over again in different guises in seemingly unrelated systems. (From this point of view, benzene's quadrupole providing the negative charge to which a hydrogen bond can form is simply another variation on the theme.) Dougherty waxes lyrical about the possibilities. "It has been proposed in, or documented in, a wide variety of protein structures. It's been proposed as a catalytic force in reactions that involve the creation or destruction of a positive charge. For example, Dave Stauffer and Alison McCurdy [PhD '94] have been looking at models for an extremely important class of gene-regulatory reactions that involve the transfer of a methyl group from a cofactor called SAM [S-adenosyl-methionine] to DNA. SAM is a cation, and we believe cation- $\pi$  interactions are involved in that mechanism. It has been proposed that the cation- $\pi$  interaction is involved in cholesterol biosynthesis. The really neat thing here is that you have a way to recognize charge, but not with charge—with grease, in effect. It really opens up possibilities for molecular design." □