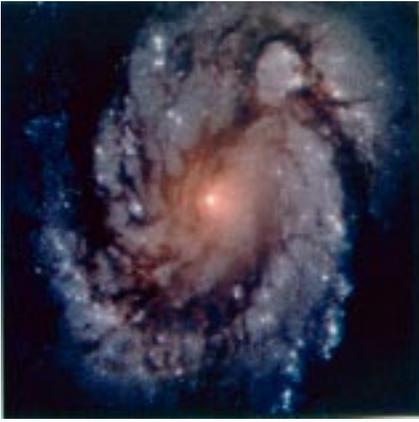


# SCIENCE ON THE FRONTIERS

Commentaries on education and research in the service of science literacy

## Table of Contents

INTRODUCTION	1
I. EXPERIMENTS IN EDUCATION	
Getting Science from the Closet to the Classroom <i>John Rigden</i>	3
The Ever More Pressing Problem of Science Literacy <i>Leon Lederman</i>	6
The Kindergarten-University Connection <i>Paul Saltman</i>	11
The Evidence Never Lies: Student Detectives in the Laboratory <i>Lawrence J. Kaplan</i>	14
II. INNER SPACE TO OUTER SPACE	
From Quarks to the Cosmos <i>Leon Lederman</i>	17
Observing the Early Universe <i>George Smoot</i>	28
The Comet Crash on Jupiter <i>Eugene Shoemaker</i>	41
Ancient Oceans and Ice Sheets on Mars <i>Robert G. Strom</i>	45
The Clementine Mission to the Moon <i>Eugene Shoemaker</i>	56
Cosmic Catastrophes and Planet Earth <i>Christopher Impey</i>	61
III. DISCOVERING THE CHEMISTRY OF LIFE	
Heavy Metals in Concert with Life <i>Thomas V. O'Halloran</i>	69
Sex, Drugs and Violence in the World of Insects <i>Jerrold Meinwald</i>	80
Clean Fuel from Solar Photochemistry <i>Harry B. Gray</i>	93
Advances in the Application of Array Detectors <i>M. Bonner Denton</i>	103
IV. INSIGHTS IN BIOTECHNOLOGY	
Proteins: Folding, Form and Function <i>Brian W. Matthews</i>	116
The Discovery of RNA Catalysis <i>Thomas R. Cech</i>	122
The Human Genome Project: Implications for the 21 <sup>st</sup> Century <i>Leroy Hood</i>	128
Peeking at the Blueprints: Genes and Modern Biology <i>David Galas</i>	136
PHOTOGRAPHY AND ILLUSTRATION CREDITS	144



## SCIENCE ON THE FRONTIERS: Commentaries on education and research in the service of science literacy

*Science on the Frontiers* is a collection of presentations on wide-ranging topics that convey much of the excitement of contemporary science. The speakers are the scientists themselves, distinguished investigators and widely recognized authorities in their fields who have received a wide variety of honors including the Nobel Prize. These illustrated lectures were first presented at an annual conference for high school science teachers and their mentors—college and university scientists. These yearly conferences were held in Tucson, Arizona, and were hosted by Research Corporation as part of the Partners In Science program. This program, cosponsored by other foundations and corporations, provided opportunities for teachers across the country to spend two summers working in research projects on local campuses.

Section one of *Science on the Frontiers* offers "Experiments in Education" by distinguished educators, including Nobel laureate Leon Lederman, who has helped to transform science and math education in Chicago schools.

In section two, Inner Space to Outer Space, the universe is explored with physicist Leon Lederman and cosmologist George Smoot, who explains that we are within hailing distance of what Smoot calls "a remarkable new model of how the universe was created, how it evolved and how it changes." Much of the groundwork for this model was laid by physicist Lederman and his colleagues who labored over many decades to construct what is called "The Standard Model" of elementary particles and forces.



A fragment of Comet Shoemaker-Levy 9 hits Jupiter.

Coming down in scale to the level of Earth and our Solar System, the late Eugene Shoemaker gives an exciting account of the discovery of Comet Shoemaker-Levy 9 in 1993, the fragmented body that entered Jupiter's atmosphere over a period of five and a half days. Mars' spectacular global topography and history are vividly described by Robert G. Strom, and a return visit to the Moon by the spacecraft Clementine—a survey that subsequently discovered water—is discussed by Shoemaker. Finally, astronomer Christopher Impey of the University of Arizona talks about a subject frequently in the news and on movie screens—the chance of a catastrophic asteroid collision with planet Earth.

Section three of *Science Frontiers* brings readers to the chemistry that drives processes on Earth. Thomas V. O'Halloran of Northwestern University dis-

T. EISNER



A spider inspects, but ultimately rejects a moth protected by chemical compounds.

cusses his discoveries in the roles of transition metals in cell metabolism, while Jerrold Meinwald of Cornell relays the fascinating story of how insects use chemistry to survive. Exciting progress toward practical solar-powered water-splitters for deriving hydrogen and oxygen, and hydrogen-oxygen fuel-cell automobiles are reported by Harry B. Gray in "Clean Fuel from Solar Photochemistry." And finally, Bonner Denton explains recent advances in chemical analysis and the detectors that make them possible.

Section four, Insights in Biotechnology, explores the frontiers of the human genome and genetic engineering, the microscopic end of a voyage of discovery. Here, beginning with the mechanisms of protein folding and RNA catalysis, are visions of a future in which—to quote Leroy Hood—"we will understand how human beings are constructed and will be able to correct mistakes in the basic plan that leads to disease."



## Getting Science from the Closet to the Classroom

Adapted from a 1991 presentation by John S. Rigden

*John Rigden is director of special projects for the American Institute of Physics. A prolific author, he is former editor of the American Journal of Physics, and researches the history and philosophy of science and education.*

*Included among the many books he has written and edited are: Hydrogen: The Essential Element, Physics and the Sound of Music, Rabi: Scientist and Citizen, Physics in the 21st Century, and the Macmillan Encyclopedia of Physics.*

**I've been asked many times** over the years what I do. I say I'm a physicist and then people ask, "What is physics?" I used to be very embarrassed by that question because the reply can be difficult, but now I don't mind because I have an answer. I tell them that physics is a basic science and like all sciences, it is a way of conversing with Nature. When a science is very young, like animal behavior, the conversation is simple and straightforward. Jane Goodall asked the question, how do chimpanzees behave? She trekked off to Africa, placed her eyes and ears in the right place and observed chimpanzees, and pretty soon Nature answered her question.

Edington asked, "Nature, does light travel in a straight line?" Now physics is very different from animal behavior, and the question was posed within a framework that is very highly developed, conceptually rich and abstract, and you might wonder why the dickens it was asked in the first place. Well, Einstein had this beautiful theory that linked mass and space in such a way that the space around a massive object was affected by the presence of that mass and, in fact, when light travels through that space the prediction was that it would not travel in a straight line. More than that, the theory told us exactly by how much light should bend. So, in keeping with the theory, Edington asked, "Nature, does light bend as it passes by a massive object, and oh yes, by the way, by how much?" and then he took his instruments to Africa.

Although both Goodall and Edington went to Africa, not all of Nature's answers *come* from Africa; that's a coincidence. But in Africa in 1918 or thereabouts there was a full eclipse of the sun and, during that moment of darkness, Edington with his sensitive instruments measured the apparent position of the star as light from that star passed the sun. And Nature said, "Dr. Edington, light is bent as it passes by the star." And the amount was the right amount. And Edington was happy and started looking for another question to ask Nature. Now, that's a far different description of science than we *normally* convey to our students.

The essence of science is this give-and-take, this conversation I have just described, and we call it "basic research." In the final analysis, basic research is a private affair. Now no research is complete until it is made public, until the results are published. Further, research is sometimes a collaboration where animated discussions drive day-to-day activities. Nonetheless, basic research has always been a private endeavor, and in recent decades it has become even more private in its character. A research project begins as an idea in the mind of an individual—often an elusive idea that appears rather sharply, only to fade as the scientist tries to grab hold of it intellectually, to comprehend it and its implications.

The idea itself is the product of a succession of deeply personal choices. First, what kind of science, life science or physical science? Second, what type of problem? Does it concern bulk matter, individual particles, plants, animals, organic or inorganic matter? Finally, from the array of problems that might be selected, the scientist—an individual—makes a choice. It remains a profoundly private affair even after the idea is born, even

after the research program is launched, even after several individuals are involved and hard at work. In research, each participant, whether working alone or in a group, is engaged in a personal struggle to understand the implications of the day-to-day results. Everyone in research has experienced the titanic struggle to grasp an insight, that by some intuitive sense, is known to be lurking in the data.

Why is the resonance line of the deuteron so much broader than the resonance line of the proton? It shouldn't be that way, but it is. What's going on? In that broadened line of the deuteron, there was a clue: the deuteron had a new property called a "quadrupole moment" and because of that quadrupole moment, theorists had to revise entirely their ideas about the nature of the forces that act within the nucleus. They're no longer central forces, they're complex tensor forces, and in the development of nuclear physics, this was an enormously important step, all because of a broadened line.

Science is a private endeavor insofar as every researcher not only struggles to understand the data but also privately hopes to formulate a seminal idea that will bring a breakthrough in the research. But even after the research results are published the process of research remains in significant ways a private affair. You know that the research literature has always been concise and impersonal. But if you look at *Physical Review* and compare papers written in the 1920s and '30s with papers written in the 1980s and '90s, you will see a very big difference. In the earlier era, you will learn more details about the research that was done. You may learn about some blind alleys that were followed, and even something about the emotional state of the researchers, how pleased they were with a particular result, how excited and surprised.

However, in the 1980s and the 1990s, a steel curtain is pulled down around the persona of the author; nothing is revealed. You don't even learn a lot about the research. All the relevant details have been left out, and unless you are engaged in research that is almost identical, the paper in and of itself is of little value. So even the published paper remains a private activity. Finally, the paper reveals nothing about the researcher's longer-range plans. How does this fit into next year's work? What are the aspirations of the researcher vis-a-vis this work five years from now?

In its essence, basic research pits the individual against the unknown. And it is private. In recent decades research activity has become even more private because, to an extent greater than ever before, it has become a nurturing of self. The individual scientist wants to influence a discipline by getting his or her ideas accepted, and the stakes

are high. Hanging in the balance are recognition, tenure, patents, grants, the National Academy, prizes—there's a whole array. The cold fusion episode was merely a blatant example of what is always there: the nurturing of self—that the rationale for the course is based on criteria internal to the body of knowledge itself. The internal approach is also a private approach because the purpose and design of the course arise from the body of knowledge and in the classroom that is the private knowledge of the instructor.

A marvelous example of the internal approach occurs in the most widely-used textbook for the calculus-level introductory physics course. The first sentence of chapter two, the first students read because it's the first they're tested over and given homework problems on, reads as follows: "A change in position of a particle is called a displacement." At this point, I could say "*quod es demonstratum*." For indeed, the internal approach is demonstrated. No reason whatsoever is given for starting with displacement or with a particle. But the physics instructor knows that displacement is the archetypal vector and that the vector is going to be needed to describe motion next week and the week after that. Furthermore, the teacher knows that the dimensionless particle, although abstract and uninteresting, is much simpler than, let's say, the change of position of a hammer. So we start with the particle.

The instructor knows it's better to start with a unique case, but what does the student think? "Why on earth am I reading a sentence, 'A change in position of a particle is called a displacement?'" The student must wonder why this is an appropriate place to start. By some intuition, the student knows that physics is one of the most basic sciences and that, through the tools of the physicist, the basic nature of matter has been learned. They know that physicists built the atomic bomb and that physics has altered the world they live in. Nonetheless displacement is where the course begins and with that the race is underway. One concept follows another and does so for reasons internal to the body of knowledge and known privately by the physics instructor. With the internal approach, the scientist-teacher stays within the body of knowledge, becoming a high priest who knows and commands the knowledge, an authority who cannot be challenged. With the internal approach, the classroom is a stage, the teacher is a master performer and the teaching process does not nurture others as much as it nurtures the authority and ego of the instructor.

I have argued that the research scientist functions in a mode that is inherently private. I have suggested that the teaching mode we adopt, the internal approach, is like-

Everyone in research has experienced the titanic struggle to grasp an insight, that by some intuitive sense, is known to be lurking in the data.

wise inherently private. This is the "closet" in my title, "Getting Science from the Closet to the Classroom." As long as we remain in this closet, we are nurturing ourselves, rather than our students. So how do we get from the closet to the classroom? This is not a trivial question. As we move between the privacy of the research laboratory or as we move from the privacy of our syllabus—by the way, I love that expression, "the privacy of our syllabus"—into the public arena of the classroom, a Gestalt switch is required. There is a profound difference between the nurturing of self and the nurturing of others. Nonetheless, all of us know a few individuals who move between the closet and the classroom and do so with dignity and charm. So we know it's not a forbidden transition.

In order to get science from the closet to the classroom, we must cut through the mystique that pervades the science classroom. Listen to a student talking to the professor of her calculus-level introductory physics course. "I do not know what's going on," she says. "You just plunk these things on the board. Where did Newton's laws come from? I have to know where things come from. Where did Newton get the laws from? He didn't just pull them out of the air. Where did they come from?"

The textbook rendition of science that pulls Newton's laws out of the air makes the instructor look like Mr. Wizard and inevitably the mystique follows. Why not make some judicious choices and in a few select cases reveal to students the process involved in science? We could show them where some of the pivotal ideas came from, how they originated and why they have proved so valuable. This is why the experience you have had, coming out of your research partnerships, is so crucial. By virtue of this research experience, you know science in a way that transcends the textbook. You know how illogical and how messy the pursuit of knowledge and understanding can be. You can, if you choose, bring your research experience to bear on your teaching and demystify some elements of classroom science.

To continue the transition from the closet to the classroom, we must modify the basic assumption that conditions much of what happens in the classroom. This assumption is spelled out in the first sentence in the Feynman lectures. It reads as follows: "This two-year course in physics is presented from the point of view that you are going to be a physicist." Consistent with Feynman's comment, introduc-

tory science courses are typically designed and implemented as preprofessional courses and as such they are means to a distant end. For the student, there is little or no purpose inherent in the course. How do we establish course goals which give the students that needed sense of purpose? Clearly you are all aware of ways to bring science from the darkness of the closet to the light of the classroom. You can bring your research experience to bear on your teaching.

In the very first lecture Feynman says, "I assume you're all going to be physicists. But before he's five minutes into the first lecture he is talking about physics in a very broad and general way: what physics is. He tells his students that physics is a subject where the test of all knowledge is the experiment. He goes on to say that the laws of physics are often the wrong ones and we have to find the right ones. For example, we once thought that mass was a constant, that it never changed. "We wrote equations in the following way to be consistent with that. And then,"

**Why not make some judicious choices and in a few select cases reveal to students the process involved in science?**

Feynman says, "we found that mass changes with speed. But it only has significant effects right up near the speed of light. Now," he continues, "you can use this inaccurate law and ignore the inaccuracies and it'll be all right for science, but philosophically it is the wrong

law. It is not merely inaccurate, it is absolutely wrong and our entire view of the world changes even though the mass changes slightly." These are provocative ideas that invite student participation.

Just a few minutes later Feynman reveals himself. He asks, "If, by some cataclysm, all scientific knowledge was destroyed and only one sentence passed on to the next generation of teachers, what statement would contain the most information and the fewest words?" Feynman gives his answer, "I believe it is the atomic hypothesis that all things are made of atoms, little particles that move around in perpetual motion, attracting each other when they are a little distance apart, but repelling upon being squeezed into one another. I believe that," he says.

All right, Feynman, believe, but I can formulate my own response to the question. I can leave another sentence to the next generation of teachers. I can believe something different. With this seductive question and with Feynman's personal answer, he is out of the closet and squarely in the classroom.

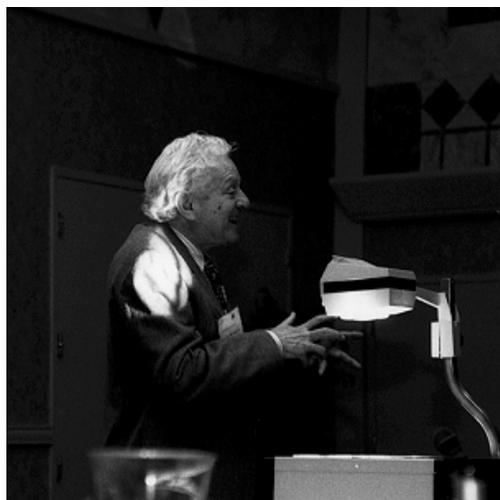
We face the challenge to make the same transition.

Leon Lederman participated in the discovery of the K-meson particle and the non-conservation of parity during muon decay and, with Melvin Schwartz and Jack Steinberger in 1962, the identification of the muon neutrino for which they received the Nobel Prize in Physics in 1988. In 1977 Lederman led the team that discovered the bottom quark.

Lederman received his Ph.D. in physics in 1951 from Columbia University where he spent the next 28 years, serving as director of the Nevis Laboratory from 1961 to 1978.

As the second director of Fermilab (1979-1989), Dr. Lederman oversaw construction of the Tevatron, the first superconducting synchrotron and initiated a wide range of educational programs and symposia.

Leon Lederman is currently Pritzger Professor of Science at the Illinois Institute of Technology; vice-chairman of the board of the Illinois Math and Science Academy, and co-chair of the board of trustees of the Teachers' Academy of Mathematics and Science in Chicago.



## The Ever More Pressing Problem of Science Literacy:

**“We will fight on the beaches...”**

Adapted from a 1997 presentation by Leon Lederman

*The causes of science illiteracy—the superficiality, misinformation, ignorance, and downright hostility toward science we encounter at virtually all levels of school and throughout adulthood—are numerous enough so that rooting out the problem will require a complex and sustained effort. But this is a war well worth fighting, and it is winnable as long as we choose our battles well, employ intelligent strategies and hang in there for the duration. As Winston Churchill exhorted his compatriots: “We will fight on the beaches, we will fight in the cities, we will never surrender!” (from “Getting High School Science in Order” by Leon Lederman.)*

**Before I talk about education,** I want to talk about the state of science today. A new study on the status of R&D reflects gloom; projections of support for basic research are very dismal. Research in this country is done in three places: industry, universities and national laboratories. Industry is getting out of research—the once-great industrial laboratories are no longer doing basic research or even basic applied research, and Bell Laboratories is down to about 30 percent of what it once was. Many of these laboratories were incredibly productive, but there’s a very strong move for industry to abandon basic research altogether.

Federal support of basic research has not been very profound. If you look at non-defense research, you find the U.S. is way, way down the list compared to many other countries. This has been alarming even to editorial writers; a *New York Times* editorial talks about crippling American science, and the *Washington Post* headlines an article with “Squeeze on Science.” Recently CEOs of the sixteen major U.S. high-technology corporations took a full-page ad in the *Washington Post* pleading with Congress to support basic research. Unfortunately, they only took one ad.

Superimposed on this depressed situation are awesome possibilities. There are emerging technologies which could add a trillion dollars to the gross national product over the next decade: superconducting materials, optoelectronics, artificial intelligence, high-tech ceramics, the whole range of molecular biotechnologies and so on. There are many of these emerging technologies, and it’s not at all clear what fraction will be captured by U.S. industry. The Japanese have taken an interesting approach. An official report of the Japanese government reviewed the Japanese economy with disappointing projections, and suggested that the way to reverse the situation would be to increase support of basic research. As a result official Japanese policy is to double the country’s research budget over the next five years.

Way back in 1922, one of the world’s great futurists, H.G. Wells, wrote that “Human history becomes more and more a race between education and catastrophe.” That is certainly true today. I want to talk about two particular activities in education, but let me first set the scene.

### A national initiative in education?

In the rhetoric about education and education reform, especially among state governors and presidents, you hear over and over again words like “we need a change and it must be ‘bold,’ ‘radical,’ and ‘break the mold.’” I want to raise an issue that might be controversial, certainly should be controversial. Our founding fathers in their almost infinite wisdom gave responsibility for education to states and localities and, therefore, the federal role is relatively minor. It includes supporting education research, and importantly, providing a bully pulpit. But how could the founding fathers have anticipated the importance to the nation of a science-literate workforce, and a science-literate citizenry? Now we don’t depend on South Dakota to defend the United States; we give that responsibility to the federal government. Should we not at least examine the proposal that a national initiative in education is called for?

This came to me after attending several so-called “renaissance weekends,” interesting gatherings of CEOs of major corporations and government officials and journalists—wise people. After listening to the various problems facing this country, I made a list. Every one of these problems, major problems, ultimately has education as a vital component of its solution. You could make your own list: decay of the cities, crime, drugs and the increasing gap between the rich and the poor in the United States. (We’re catching up with Brazil which has a ratio of 100:1, the average income of rich versus poor, the top versus the lowest levels. In the Pacific Rim, incidentally, the ratio is about 5:1.)

Other problems include those of the environment; the problems of energy sources which are benign for the environment; endemic diseases like AIDS; the whole problem of population, a major problem which was important enough to collect the leaders of some 130 nations in Cairo in 1994; industrial accidents which include oil spills; air safety; dwindling research budgets, which I already mentioned; racism and gender bias; junk science; the worldwide growth of fundamentalism, from the Taliban rulers of Afghanistan to the hardliners in Israel and the Arab world to creationism in the United States, and so on.

Despite the importance of education in solving these problems, the U.S. is not doing very well, this from international studies of science and math achievement. One global report called the problem one of focus. Let me address the issue in a general way. Visualize a circle: when I get into educational debates, I listen to many strong opinions about the origin of problems which begin with preschoolers, persist throughout life, and revisit the next generation. Some say the problem starts in preschool because that’s where the attitude of

children towards school is shaped: or the problem is in K–8, the problem is in high school, the problem is in the colleges—or the problem is that the general public is glued to the television set. I decided to finesse this debate by drawing a circle and saying, as in an elementary electricity course, that if you don’t have a complete circuit, no current will flow. You need to fix all of these problems, because the graduates of our high schools and colleges come the full circle and become voters and parents.

You might have access to these voters and parents for a general program on “public understanding of science,” and you can use TV and radio and op-eds and museums and malls, and you might even put science stories on cereal boxes. Then there are movies which portray scientists either as nerds (“Honey, I shrunk the kids”) or monsters (“Tomorrow we destroy the world”). That’s Hollywood.

### Reforming Chicago schools

So you have all these problems, and what I want to do is spend a little time on the good news, which is that they are not insolvable. Back in 1989 I had just moved into the city of Chicago from the comfortable suburbs to become a professor at the University of Chicago. The city passed a dramatic school reform bill which decentralized Chicago schools, giving us 540 little corporations, each with a CEO and elected board of directors. The CEO is the principal, still called “the principal,” but the board of directors has control. It can fire the principal if he or she doesn’t “make money,” that is, improve student and educational performance.

In the euphoria of all this, listening to colleagues who had been involved in passing this school reform bill, it was very clear that the school system needed some friendly intervention. One of the major problems was the poor preparation of primary school teachers, K–8, but mostly K–6, for teaching math and science. There was a perception that these primary school teachers were horribly underprepared, and it turned out to be true: it’s not *possible* to exaggerate their lack of preparation. It’s not the fault of the teachers, because we have found that they will work enthusiastically to become better teachers if you give them opportunities. In order to attract candidates to teach primary school, teachers’ colleges practically posted signs saying we’ll protect you from math and science, just come. And so they were protected until they were classroom teachers, and suddenly at two o’clock in the afternoon, full of their own insecurities, they had to teach math or science.

We thought we could help these teachers, so we created something called the Teachers’ Academy for Math and Science, a freestanding institution for professional enhancement. We got support from the Department of Energy, a

Every one of these problems,  
major problems, ultimately  
has education as a vital  
component of its solution.

curious place to get support, but we did. We have 24,000 teachers for 410,000 students, and—to recite a litany typical of the cities—68 percent of the students come from families under the poverty level, and they scored in the lowest one percent on any of the national tests. Chicago has crimes and gangs and drugs and old buildings and multi-ethnic backgrounds and poor parental support, you name it. Nevertheless, we thought that if we could help the teachers, they would be at least one component in a solution.

### **After extensive teacher training, progress**

We developed an in-service program with a staff, and we use all the new pedagogy and a hands-on, inquiry-based, process-rich curriculum. We steal programs from everywhere, without any bias. We take teachers out of classes, replace them with substitutes, and give them a lot of science; we also work with them on weekends. We found out that all of this pedagogy must be distributed over a minimum of three years, more likely four, and costs about \$10,000 per year per teacher. As I said, we teach them during class time, summers, weekends, evenings. We do this systemically and get the parents and local community groups involved.

Assessment and evaluation is a new profession, but we've used as much as anyone could ask. One measure is the State of Illinois math exam, a statewide exam that's been given for many years. It has a maximum score of 500, and the state average is about 270. City schools began to catch up in a dramatic way after two years of work with us. There's a long way to go; we started with the worst, the poorest schools in Chicago; we deliberately took the schools with the highest poverty levels and the worst reputations for academic achievement.

The results show an improvement after the schools have been in the program for at least two years, but it's even more dramatic after three years. My dilemma is that we really need another four or five years to see if all the things that we're not doing will overwhelm the things we are paying attention to. Hands-on science is very engaging. I've seen teachers get the children to calm down when they say, "Kids, if you don't settle down, we won't do science." And they sit up, and it's quite a thing to hear these youngsters with their street English say, "Hey man, what's your independent variable?"

I want to discuss the high schools next, because we all share a professional interest in secondary education. You know that a consensus is emerging on national standards for math and science. The math standards have been around for a while; the first national standards for science were the AAAS benchmarks, and these were reiterated by standards written by the National Research Council of the National Academy of Sciences. So we are getting a good

consensus of what kids should know in second grade, in fourth grade, in sixth grade and so on. Another trend is the increasing number of school systems that are beginning to insist on three years minimum of science and math. The City of Chicago adopted this standard in principle a couple of years ago. New York City adopted the standard, and the state of Pennsylvania requires three years. I think the trend is very encouraging, and we're heading towards a system in which all high school students take three years of math and three years of science as part of a core curriculum. This is not for *scientists*, this is for *citizens*—all high school graduates. The Chicago schools realize that they don't have the teachers or the laboratories to meet this standard, and they hope to phase it in by the year 2000.

### **Three years of science—in a coherent order**

My main point: if we require three years of science and math, and if nationally accepted standards of math and science exist, then why not seize the day and create a coherent science curriculum? Let's call it Science I, II and III to try and smooth over the disciplinary categories. And let's try to make it coherent, because you now know that there are going to be three years of science. All too often in today's schools—some huge fraction of the schools—students begin with biology. Since they have had no chemistry or physics, the biology has to be qualitative with lots of memorization, stultifying for a subject as dynamic and exciting as modern biology. I will fully cede the second half of the 20th century to biology; the first half was physics, of course, and chemistry is there all the time.

When today's kids finish biology, they put a period at the end of the exam and it's forgotten. Some of them will go on to take chemistry, and then certain areas of biology might begin to make sense, but because they had no chemistry when they took biology, the knowledge gained is less than it should be. About 20 percent go on to take physics, mostly as seniors (the ones that are going to shrink the kids or destroy the world).

There's another way of doing it that makes sense and which corresponds to the way science works. Science works in a hierarchy. It's a pyramid with mathematics at the base. Physics requires mathematics and is second (mathematicians, on the other hand, don't care whether physics exists or not; their discipline isn't dependent on physics). So physics sits on top of mathematics as a necessity. Chemistry sits on physics because chemistry involves the interaction of two atoms to make a molecule, and atoms are the bag of physicists. Chemistry depends on physics.

Modern high school biology—not the qualitative biology which is very beautiful and should be emphasized in K–8—depends on molecules like DNA. How can you study biology and not mention DNA? And how can you understand

DNA if you don't know what a molecule is? And how do you know what a molecule is if you don't know what an atom is? So the pyramid of science is suggestive of the way we should rearrange our high school curriculum to start with ninth grade physics.

Ninth graders still don't know much algebra. If we do impose this pyramid scheme, it will put pressure on the primary schools to teach some algebra in eighth grade; Science I will then be mostly conceptual, mostly physics, with excursions to chemistry where it is relevant. But these classes will hand over to Science II—which will be mostly chemistry—kids who know what an atom is, and who will understand how an atom is held together. They will also understand the process by which two atoms decide to lock onto—or avoid—each other, which is called chemistry. Then in the study of molecules, which kids will become familiar with from chemistry, they will be prepared for molecular biology—gene structure, proteins, DNA and so on. The pedagogy is that they can begin to use physics in the study of chemistry, and both physics and chemistry in biology. And with this, you have coherence, you can talk to one another. We must have physics teachers talk to the chemistry and biology teachers, and even know a little bit about those other subjects.

Biology teachers might say, "O.K., today we'll learn about photosynthesis. Let's bring that physics teacher in to remind us what a photon is. And then our chemist will tell us about photo-induced chemical reactions, and then we'll talk about photosynthesis." You begin to stress the unity of the sciences, the conservation laws which are valid in all subjects, and symmetries, vibration theory and so on. You really have a three-year subject which not only covers the science in a pedagogically sensible way, but leaves you time to talk about science and society. This could include material on the qualities of science; the process of science; the delicate balance between skepticism and respect—important because these are citizens. We are training people to be voters and citizens, and to be intelligent decision makers. We want them to think scientifically.

### Promoting education reform

As activists we formed an organization which we call ARISE, American Renaissance In Science Education. We had a meeting last September, and now have a steering group with officials from the National Academy, the AAAS and the National Science Teachers Association. We're beginning to think about how to change a system—biology, chemistry and physics—which was established in 1893 by a committee of ten, chaired by a Harvard president. Maybe that committee's logic was that the course order should be alpha-

betical. Seriously, I don't know what the logic was, but 100 years ago it probably didn't matter. It certainly matters now, and this change has been resisted for 100 years. We have to respect the difficulty in getting this done in all the high schools in the United States; we even found resistance among the physics teachers visiting Fermilab last summer because they didn't want to teach ninth graders.

In the fourth year of high school students can, if they are college bound to do science or engineering, take advanced placement courses in any of these subjects, or even begin an advanced placement in the third or second year if the high school is qualified to offer it. One can go to a liberal arts college with enough science so that the college can't get away with "rocks for jocks" or whatever passes for the minimum science requirement at too many institutions. At some schools you can get a bachelor's degree without changing any of the misconceptions you had regarding science as an entering freshman. Or you can go to work. And if you go to work, you might want to take courses in technology, computer science and all the things that will help you. With a base like this there are possibilities in earth science, environmental science, all kinds of interdisciplinary subjects which could compose an important curriculum. That's the ARISE program.

There are going to be obstacles. I've made a list of the obstacles, because you have to understand how hard it is to make a revolutionary change. A 100-year record of resistance to change, that's obvious; it requires a strong, enthusiastic, visionary, youthful, well-financed program. It's going to be expensive: you need computer terminals, probably new labs and new books. You need continuous professional enhancement, you need time for teachers to talk to each other. And we've got to get the parents involved, because their biggest objection will be, "If you make this change my kid won't get into Yale," and you'll say, "Well, that's lucky."

One of the first things I did was to get an advisory group to verify that all this was not a plot by a physicist to torture ninth graders by making them study physics. I got 25 Nobel laureates in biology, chemistry, physics, all the science subjects in which they give Nobels, to certify that this makes sense. I'm now working on the customers of the high schools—prospective employers—the CEOs of major corporations, and maybe a few military people.

Let me address the subject of the universities and their role in this reform. Perhaps the key to fixing the educational system lies in the structure of the universities, blending the disciplines. I have lot of respect for physics as a discipline, and also for chemistry and biology, but I don't like walls. Instead of walls there should be shrubs and little

We are training people  
to be voters and citizens,  
and to be intelligent  
decision makers. We  
want them to think  
scientifically.

paths between disciplines so you can go comfortably back and forth. Unfortunately, that's a very unrealistic vision of the university which is crushed by conflicts between the soaring imagination of gifted scholars and the rigid, unyielding turf established by departmental politics.

### **Community support is vital**

We can't meet the goals of fixing the sciences in the high schools or the elementary schools without making lots of other changes. We can't affect priorities unless we can get teachers, parents and school administrators on our side, and ultimately we need the general public as well. And I think we can't fix the natural sciences without addressing the humanities and the social sciences, a recognition of unity of knowledge. The two cultures gap is only getting worse, yet in this process of changing the high schools we have to incorporate the social sciences and the humanities. When you study history you study the kings of England, and you study Napoleon and Genghis Khan, yet you never study Faraday. But Faraday changed the lives of people on the planet more than Genghis Khan or all the kings of England rolled into one. He had a bigger influence, but this fact is somehow not taught in history class.

If you agree with me that education is key to solving the social problems we listed earlier, then we have to raise education's priority in this country to an entirely new level. After WWII, science was great; it had learned how to make deafeningly loud noises, and it did all kinds of things in the war. There was a wave of euphoria which elevated the status of the educational system for a while, then faded. Then came Sputnik which again established new priorities for education. The textbooks, physics, chemistry and biology, improved, but five years later those books faded in their influence; you couldn't get them, and the new ones were watered down. The new books were not produced with sensitivity, and didn't explain what they were trying to do. Also, they focused on the pipeline problem rather than on raising the level of the general student. I'd like to see our priorities for education roughly equivalent to those we might establish in our small wars.

We know that there are some good schools. Why can't all of our schools be as good or better than our good schools? Perhaps Research Corporation should get involved because it will take an unfettered imagination. This country occasionally establishes national commissions, and in fact the one impaneled by Terrance Bell created the 1983 report, "A Nation at Risk." With purple prose and military metaphors, it sounded the alarm over the quality of U.S. education. It said we had committed unilateral "education disarmament," and if any country did to us what we did to ourselves, we would have nuked them. Maybe exaggeration was needed.

### **Drawing plans for action**

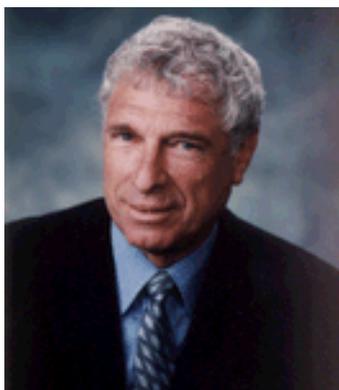
Government is probably not the place to organize a national commission like that one. It should be organized by an organization like Research Corporation, for example, which could impanel a group of 20 or so wise people: educators, scientists, CEOs of major corporations and maybe an occasional university president. The commission should take a year or two to examine the obstacles to achieving a superb education. Why can't the U.S. be number one? What are the obstacles? Are they the teachers' unions, the parents, the universities which refuse to raise their standards, refuse to hold the high schools to higher standards? Let's find out what attitudes, laws, finances and institutions need to be changed. This commission could issue a report as dramatic and more far-reaching than "A Nation at Risk." It would be a call to action, and would tell us what to do to make a dramatic change in our educational structure in order to be number one by the year 2050.

The action plan should be handed to the president of the United States. I'd say, "Mr. President, here's a plan which will give us superb education; this is what it would cost, this is who we have to convince, and this is how we should go about doing it." It's an enormous challenge, but after all, somebody invented free public education. Somebody thought of land-grant colleges, and they have had a tremendous impact on the development of the United States. Someone invented the superb U.S. graduate schools which are still the best in the world. If we give primary and secondary education this kind of priority, I think we can succeed.

I want to conclude with a story that my mentor, I.I. Rabi, always told. Rabi's little essay was called "Plato's Academy." Plato had an academy for preparing young men for entry into the world, a world perhaps as troubled as ours seems to be. Rabi thought it would be interesting to put a fresh, bright graduate with a B.A. from an Ivy League institution in a time machine to attend Plato's academy.

What would our graduate be able to tell his opposite number at the dawn of recorded history? Could this product of two millennia of human progress enlighten the Greeks on philosophy, the art of sculpture, painting or architecture? Could he demonstrate a superiority in poetry, drama, history or oratory? Would his language be more pure, his thoughts more elegant? Would he excel in athletic prowess, in grace of body? Would he be more prepared to fight for his country, to lead men into battle?

One doubts whether our 20th century grad would make a brilliant showing in these fields. Indeed only a small fraction of our students could get by the door of the academy, because at the entrance was the inscription, "Let no man enter who does not know mathematics."



## The Kindergarten-University Connection

Adapted from a 1991 presentation by Paul Saltman

*Before joining the administration at the University of California, San Diego in 1967, biologist and nutritionist Paul Saltman was a faculty member in the department of biochemistry and molecular biology at USC's Keck School of Medicine. During his 32 years at UC San Diego, he served as provost of Revelle College from 1967 to 1972 and as vice chancellor for academic affairs from 1972 to 1980.*

*Dr. Saltman received a B.S. in chemistry and Ph.D. in biochemistry from Caltech. He did his postdoctoral research at the College de France in Paris and taught at the University of Copenhagen and Murdoch University in Australia early in his career.*

*In his research in biochemical nutrition, Dr. Saltman focused on the role of trace metals, such as iron, copper, zinc and manganese in the human diet. His research led to improved dietary and supplemental strategies. Clinical applications of his work include reduction of free radical damage to hearts, prevention of anemia, enhanced physical performance, and better bone and skeletal metabolism.*

*Paul Saltman died in 1999 at the age of 71.*

**I**n 1990, Bill Aldridge of the National Science Teachers Association invited me to come back and talk about science education and a program we were doing at the University of California, San Diego. The title of the talk was "What Me Worry? Alfred E. Newman Views Scientific Literacy in America." The abstract follows: "In which the fear and loathing of science and mathematics by students and teachers, not to mention politicians and fundamentalists, will be metamorphosed to love and understanding by just saying yes to science. In which creative curricula, powerful audio/visual instruments and an avalanche of new textbooks will bring light to the darkness. And while we wait for Godot and Rumpelstiltskin, only you and I can picture the curse of ignorance in our great society."

Colleagues, the main factor limiting science literacy in America is the literacy of the men and women who teach science, kindergarten through postdoctoral. And the greatest curse is to be found from kindergarten through sixth grade. When you look at the numbers—the studies of which are endless—you find that 70 percent of the boys and girls of America are turned off to science and mathematics by the time they reach the sixth grade. Do we really want to do something about science literacy in America or is this some sort of intellectual game? What do we want to do? What are we going to do about it? Will Godot come tomorrow? That is what I'm going to talk to you about this morning.

I was coerced into this business. Bassam Shakishirri, then the director of the National Science Foundation for Science and Engineering Education, appointed me to his NSF committee. I sat there for a year or two and I listened to all the reports, many featuring elaborate curricula and interactive computers. No one, however, mentioned anything about elementary school teachers. So I said to Bassam, "Give me a million dollars and in three years' time I will metamorphose San Diego city and county schools and design a program that is operational and can be applied to other institutions. I will see to it that the best and the brightest and the most effective teachers at the University of California, San Diego, in the fields of biology, chemistry, physics and the earth sciences, take part in this program. I will take a hundred elementary school teachers and will teach them to be literate and loving of science so they can teach this marvelous integrated curriculum, the science framework which has been created by committee, someplace else." Bassam said, "You're crazy." And I said, "Give me a million dollars."

I'm going to tell you about that program, because it is my pride and my passion. I will tell you as we begin this discourse, that I've done a lot of teaching and I've won a lot of prizes for teaching and I'm an exceptional teacher. I really try very hard—I love to teach. It is as passionate an endeavor for me as is doing great research and having it published and being invited to tap dance in other arenas. But I have never had so much pleasure as I have from those three years, and I look forward to the rest of my lifetime with the same pleasure anticipated. I have never seen a more responsive and delightful group of people than the 102 science teachers that came to us on the first day.

First and foremost, let me outline the strategy we developed. It goes something like this: if you have knowledge, you have power. So the issue was to get the knowledge to the

teachers about these disciplines so they would have the power to bring them to their students. I agreed to teach the first entire six-week session on contemporary cellular and molecular biology, with a side game of nutrition because that is the subject that everybody is interested in. Little kids eat, you know. If you know everything that happens from your teeth to your tummy, you know a heck of a lot of biology. So, I was going to teach biology.

We selected Walter Munk, a world-class oceanographer from the Scripps Institute, to teach oceanography although he had never taught an undergraduate in his life, much less a grade-school teacher. Walter said, "How can I teach oceanography in five lectures of two and a half hours each?" I told Walter that I didn't want him to teach oceanography. "I want you to stand up and for those five days tell them why you are passionately in love with being an oceanographer and what it is you do. And whatever comes out, they will learn about oceanography." He gave the most brilliant series of five lectures on doing three-dimensional tomographic analysis of sound transmission through the ocean. And as he explained this wonderful experiment that he was about to do, he talked about salinity in the ocean, about temperature in the ocean, about air-ocean interfaces and meteorology. He showed movies of himself sitting on a remote island in Tasmania where he was doing these experiments while 30-foot waves are washing him off the boat—and the teachers fell in love with oceanography and Walter Munk.

We had a prominent geologist, Joe Curray from Caltech, teach geology. Gene Smith of the University of California, San Diego taught astronomy, and Richard Summerville, a distinguished meteorologist at Scripps, taught meteorology. He said, "I just couldn't possibly do this." Not only did Summerville teach meteorology, but but also wrote a textbook for elementary school kids about meteorology. He admits that it was the most passionate thing he has ever done in his life. By the way, during each course, every teacher was given the highest quality college text available on the subject and so ended up with a shelf full of the best college science textbooks.

The university scientists we found were a very diverse group of people in terms of teaching style and pedagogical methodology. Yet over the three years the elementary-school teachers became literate in the four major disciplines chosen for the program. And that was only the beginning of the program, Part I.

Simultaneously with Part I, we were taping the faculty lectures and using experiments, visual aids and visits to museums and aquariums in the San Diego area to create so-called teaching techniques courses. Out of each six-week summer session each teacher accumulated a library of loose-leaf experiments that related and correlated with

every topic taught by the scientists in a fashion that could be translated into kindergarten through sixth grade curricula. Heading up that program were a husband and wife team, Bob and Melanie Dean, who were professional educators—with a capital "E"—had taught for 45 or 50 years between them in the city and county system, and who knew the schools intimately.

During the course of these sessions, I was provided with a lesson on the labyrinth of public school education and was outraged by the handicaps the teachers have to work with. But Bob and Melanie knew about them and they could facilitate our program. They would say to a superintendent of schools or a principal, "We need your teacher to have release time." And the principal would say, "Impossible." So then Bob and Melanie would call me up and they'd say, "Saltman, call this principal and lean on him." I'm six foot five and two hundred and twenty pounds, and I can still lean very hard. So I'd make the principals a deal, an "offer they couldn't refuse."

Let me tell you what the deal was. Each teacher, during the course of this year, had to do in-service teaching of at least ten fellow teachers in their school or in their district, and that had to be put in the contract. Not only did the teacher sign it, I signed it, Bob and Melanie Dean signed it, and the principal signed it. And if they tried to renege on the deal, I put the arm on them very, very strong so they would never try it again. Every teacher would not only teach ten of their fellow teachers, each teacher would teach an average of twenty others. We would have 2,000 teachers in the county and city of San Diego involved in in-service teaching, thanks to the efforts of Bob and Melanie Dean and the scientists that would go out and help them. So that would be the second part of the program.

We had to have the best research faculty who were willing and able to communicate. We had to have people from the school system who knew the ins and outs of that system and knew what the teachers needed in the classroom in order to interpret the science, and then we needed to have a third component: the institutional ability to launder the money and to give units and accreditation and to get the certificates and to have the rooms and the labs available. We chose to do it through the University of California Extension program which provided us with the accreditation and facilities we needed.

To begin with, we sent out announcements to all 6,000 teachers in public and private schools in the San Diego city and county area, and we said, "Come to this wonderful program." We set up an elaborate screening committee to make sure we had the right demographic mix, and we were sitting back waiting for the flood of applications that would come in for the 100 available places. We just had 102 requests and so we took them all. The percentage of

responses was disappointing. Teachers take all sorts of graduate courses in the communications arts, teaching methods, etc., but they wouldn't come to do science.

I opened up my first lecture in my sweet, gentle, loving style and about four sentences into "what is modern cellular molecular biology," I could smell the fear. I could see the anxiety and hear the hearts pounding. So I looked at them and said, "Let me find out something. How many of you have had a year of college science?" Only eight hands went up, although all in the room had to teach elementary school science. When a teacher goes into a classroom and a child asks "Why is the grass green? Why is the blood red? Why is the sky blue?," a lot of physics, chemistry and biology is needed to answer those questions. You cannot say, as one of the teachers said, "The sky is blue because it's a reflection of the color of the ocean."

So I said "Relax. Learning science is a function of exposure, of playing the game, of learning language, of learning moves, of experience, of just being around, of reading." Now, I have a problem because I am an elitist. Forgive me, but I want the elitism of the NBA. I want science to be the NBA of the mind. I want the very best players playing in my game, and I don't care what color they are or what sex they are. But there is elitism, and to deny that—to not extend the same courtesy to learning that we do to sport—is terrible. I believe in equal opportunity and I believe that science should be such an exciting adventure that everybody would want to come and play. And only the very best and the very brightest would make it to the top. I didn't play in the NBA, but I loved playing basketball for Caltech and the Racing Club. And there are other places to play science in the world besides in Stockholm on the king's money. It's a great game on every level, with every player reaching for the maximum of their human potential.

Let's give everybody the opportunity to say "I'm ready to play." But they have to be willing to make the effort. Everybody in the NBA practices free throws—it's boring. Everybody has to play offense and defense—it's painful. How do you get to Carnegie Hall? Practice. That's the part of science that everybody wants to minimize and you can't minimize it. It's there. But what isn't there, what makes shooting free throws possible, is that there's never a game time for science. You are always shooting free throws. That's hogwash. We've got to have the excitement of the game in the science that we teach.

Now it's time for Phase II. Let me tell you what Phase II is because I think we are going to get the money to do it. Now the teachers who were involved in Phase I are going out into the schools, and in two week periods they're going to give their peers the same kind of scientific comprehension that they have. We estimate that, by the end of the next three-year period, we will have at least 4,000 el-

ementary school teachers directly or indirectly involved in this program with our faculty helping to do it. Those teachers were scared to death at the beginning. We were the enemy and they were the good guys and they could no more identify with me as a colleague than the man in the moon. But now I've been in their classrooms and I've watched third graders talk about and do experiments about the first and second laws of thermodynamics when the teachers didn't even know what those laws were at the beginning of the program. The second-graders are learning about DNA, how it makes RNA and protein. They are seriously studying nutrition, talking about nutrients—and not that nonsense of the four basic food groups. They're doing it! So these 100 teachers now believe that they have colleagues at the University of California at San Diego and our faculty believes there are real teachers in the public schools of San Diego city and county. They have our phone numbers and they can call us at home or in the office day or night. Do they want us to come out and be at their school? We're there. The relationship is one of collegiality. They now feel that they are a part of the science establishment.

Phase III is proceeding simultaneously with Phase II. Shakishirri told me once, "You may be able to pull it off at UCSD, Saltman, because you're bigger than the rest of the faculty and you can coerce them." That is true, and I did. Now I will tell you some exciting news. Some other institutions in California are following in our path. Stanford, Berkeley, and a graduate school, the University of California, San Francisco are all going to participate in the project. UC Davis is coming to play, and USC is going to join us in Los Angeles. Even Hawaii is going to do it.

Can it be done? You bet. But we have to do it. We have to do it not in my way, but we have to do it in the context of a group of quality university scientists taking the initiative, working with the schools and teachers, to bring the kind of knowledge and power and love to these teachers. My favorite statistic from this study: average amount of time teaching science before the program began—twenty minutes a week. They are teaching forty minutes a day now—that's an order of magnitude. They're not just teaching this out of textbooks because the books are lousy—they're teaching it from the morning newspaper because I used to have them check to see what was in the paper that relates science to their lives.

Science is relevant. But it is only relevant if you connect it to people's lives, where they are at that moment in their life span, and the world in which they live.

*It has to be done.*



Lawrence J. Kaplan is professor of chemistry at Williams College where he has developed and taught courses in biochemistry, physical and general chemistry as well as courses for nonmajors. His innovative courses include "Chemistry and Crime: From Sherlock Holmes to Modern Forensic Science."

Dr. Kaplan received his Ph.D. from Purdue University and has taught at Williams College since 1971. He is widely published in biochemistry, a frequent speaker on science education, and has received NSF grants for the development of his forensic science curriculum and for "Project Sherlock: An Interactive Multimedia Laboratory Program in Forensic Science to Enhance Introductory Science Courses."

His research interests include isothermal titration calorimetric investigation of biochemical systems and analytical techniques in forensic science.

## The Evidence Never Lies: Student Detectives in the Laboratory

Adapted from a 1993 presentation by Lawrence J. Kaplan

**A course in forensic science** is an ideal teaching vehicle because it cuts across disciplinary lines. It is easy to capture students' attention with forensic science. After all, who isn't curious about "who done it?" Who isn't interested in playing Sherlock Holmes—or Quincy, for that matter? The topics we discuss in my class are timely and truly fun for me to teach and for students to study. This is particularly refreshing after teaching some courses to largely disinterested groups of students.

In order to set the stage for discussing how nicely this course fills the needs of college and high school science, I quote from NSF's Project Kaleidoscope report called "What Works: Building Natural Science Communities:"

*Science and mathematics education succeeds whenever it takes place within an active community of learners, where students work in groups of manageable size to enhance collaborative learning and where faculty are deeply committed to teaching, devoted to student success and confident that students can learn. Effective science learning is never passive. It is active, hands-on, experiential. Effective education engages students in activity that is meaningful in a highly personal way; it is connected to historical context, to other fields of inquiry, and to practical applications of interest to students—in other words, to the reality students experience....*

The best college science courses are conceived and run in a partially investigative mode. Investigation, the natural arena for using and solidifying one's knowledge, may be manifest in open-ended laboratory or library projects made available to students.

Because my forensic science course was designed for students who do not plan to major in one of the natural sciences, many who enroll are only fulfilling the college requirement. In order to capture the interest of those disinterested students, I highlight case studies to provide them with an immediate application of the scientific and technological principles we are discussing. I do not give my class weeks and weeks of basic chemistry, organic chemistry and biochemistry and then deal with a few applications during the final weeks of the course. I usually pose a situation or case and then discuss the possible methods of resolution so that the students' interest is stimulated before the fact. This is what my colleagues call "science taught backwards."

The course begins with an introduction to forensic science. I discuss a crime scene and various types of physical evidence. Included are case readings from a forensic science text, a basic chemistry text and, of course, from Sherlock Holmes. I also read to the class from popular novels about real crimes. One recent classic is the Jeffrey MacDonald story as told in the novel *Fatal Vision*. The section read describes how the crime is discovered and relates that the military officers did not carefully secure the scene. At the same time, I introduce a comparison between the MacDonald and Sam Sheppard cases, noting how that crime scene was handled 30 years ago. And finally the Lizzie Borden axe murders are presented with a comparison of how that crime scene was managed 100 years ago. It's particularly convenient that dramatizations of these three cases are available on videotape, giving students an opportunity to see them portrayed.

The cases immediately stimulate the interest of the class. As one student said, "Where else can you take a chemistry course and read such interesting stuff?"

During our discussion of the material, I introduce the Locard Exchange Principle, "every contact results in an exchange," and quote Ralph Waldo Emerson to the effect that, "Wherever a man commits a crime, God finds a witness ... every secret crime has its reporter."

In a discussion of evidence and testimony, topics such as general acceptance of scientific tests and the concept of the expert witness are introduced. While we're discussing the qualifications of an expert witness, I present this tongue-in-cheek definition of an expert:

An expert is a person who passes as an exacting authority on the basis of being able to turn out with prolific fortitude, infinite strings of incomprehensible formulae, calculated with micromatic precision, which are based on debatable figures from inconclusive experiments carried out with instruments of problematic accuracy by persons of doubtful reliability and questionable mentality.

Throughout the course, I have surreptitiously been introducing the scientific method, the importance of clearly identified samples, the use of controls and standards, and other aspects of "doing good science." But it isn't until this point in the course that we actually begin to *discuss* science.

We now turn our attention to two very interesting cases which open the door for the introduction of some basic chemistry: the Tylenol poisoning case in which atomic emission spectral analysis was used to examine samples; and recent efforts to authenticate the Shroud of Turin using radiocarbon dating. Using the "science taught backwards" approach, we discuss the cases first and then study the procedures used to resolve them. These cases require an understanding of emission and atomic absorption spectroscopy, neutron activation analysis and radiocarbon dating which in turn demand an understanding of atomic structure.

We also spend time discussing substance abuse and its detection. Case studies include the Jasclevich murder involving the detection of curare; also the cases surrounding the deaths of celebrities Janis Joplin, Marilyn Monroe and John Belushi.

During the serology unit, we talk about classical procedures as well as the newer ones. Since we had studied proteins in the biochemistry section, we are ready to learn about antibodies and their interaction with antigens on the surface of blood cells. Of the classical procedures, we study methods of identifying a substance as blood, the factors for the determination of human blood and finally resolution of blood

type.

We spend some time on determination of other body fluids such as saliva, seminal and vaginal fluids. This leads to discussion of rape cases and the medical procedures involved in the treatment of a rape victim as well as analysis of the body fluids of a suspected rapist.

This is the ideal point to introduce the new technique of DNA profiling which has been responsible for resolving many cases. While the scientific basis is solid and the test is very accurate if done correctly, DNA profiling is controversial because of the technological problems encountered in its transition from the research lab to the service lab. Because of these problems, its acceptance is an issue in the courts.

As you see, the class moves from scientific ideas to case studies to legal aspects of forensic science. Students remain interested and even anxious to learn science because they know they will apply it to something interesting.

### **A crime is committed**

A full-scale laboratory program is the highlight of the class. The labs are based on an exploratory approach in which the students perform many of the jobs of the forensic scientist. Working in investigative units they analyze the crime scene; collect the physical evidence to be analyzed; and plan experiments to resolve the case.

One of the crime scenarios is a hit-and-run which I stage on campus using my teenage daughter as the victim. We rope off the scene of the crime, confirm that someone has called for an ambulance and collect samples for analysis. Glass and fiber samples are placed in evidence envelopes and numbered to correspond to sketches made by students. Later, in the lab, I outline the situation for them: a hit-and-run accident has occurred; we have collected samples from the scene. Since the accident two suspects have been taken into custody by the police, and two broken headlights from the suspects' cars have been turned over to our crime lab. Our job is to find any points of similarity between the evidence found at the scene and the and the two headlights. We also have a sweater that was removed from the victim, a sample of the victim's "blood" to compare with blood smears on the headlights and urine samples from the suspects.

The investigative unit must decide how to proceed. They have five weeks to work on the different aspects of the investigation and to analyze the physical evidence. Their goal is to resolve the case by reconstructing as much as they can about what happened. They are not under specific time constraints

Throughout the course, I have surreptitiously been introducing the scientific method, the importance of clearly identified samples, the use of controls and standards, and other aspects of doing "good science."

and they don't have to perform any given experiment.

At first most of the students think it is a game—and that all they have to do is figure out whether glass from the scene is from headlight "A" or headlight "B." I point out that it would be reasonable to find all sorts of glass—such as broken beer bottles and even headlights from previous accidents—at the accident scene. They set to work with the refractometer and density gradients and flotation vials, and by the end of the first lab they are confused and starting to compare data with one another to try to make sense out of their results.

Then a college security officer comes to the lab with an evidence bag containing another headlight removed from the car of a newly-identified suspect. After the class stops groaning, I remind them that this is not a game. The investigation is ongoing and our job remains the same: try to establish a similarity between the glass fragments collected at the crime scene and the suspect headlights. The students begin to realize that they really have to coordinate their work within the investigative units.

By the end of the second lab period they are discussing—even arguing about—the data and trying to make sense out of the results they have obtained. Of course this is exactly what I had in mind. Now they want to go back and redo some of their experiments or use their own procedures to analyze samples previously evaluated by other investigative units.

During this investigation, they also examine whether the suspects may have been driving under the influence of alcohol or drugs. There are many procedures for the determination of alcohol in body fluids and almost any area of science could incorporate one of these methods to help stimulate the students' interest. We discuss the pharmacological effects of alcohol and the metabolism of the drug and the laws relating to driving while intoxicated, specifically those regarding the Fourth and Fifth Amendment rights—freedom

from unlawful search and seizure, the self-incrimination privilege, due process, implied consent laws and, finally, Miranda rights and how they apply to real cases. Finally, a comparison of the "blood" smears found on the headlight glass with the blood from the victim provides another component of the resolution of the case.

### **Just the facts, ma'am**

At the end of the investigation, the units present their findings to the rest of the class. They realize by now that there is no "right answer" . . . they simply have to be confident enough of their investigation to present their findings in court!

We cover a lot of ground in one semester. Students develop an appreciation for the application of science to real-life situations and learn a lot of chemistry at the same time. When the course is over many students comment that a class they expected to be boring was, instead, interesting, and that they're encouraged to be able to read and understand scientifically-oriented material.

If I have accomplished what is indicated in the following quote from Sherlock Holmes' *A Scandal in Bohemia*, then I can assume that I have trained my

students to be observers, and that's what a scientist is—an observer and an interpreter.

I could not help laughing at the ease with which he explained his process of deduction. "When I hear you give your reasons," I [Watson] remarked, "the thing always appears to me to be so ridiculously simple that I could easily do it myself, though at each successive instance of your reasoning I am baffled, until you explain your process. And yet I believe that my eyes are as good as yours."

"Quite so," [Holmes] answered, lighting a cigarette and throwing himself down into an armchair. "You see, but you do not observe. The distinction is clear...."

**Students develop an appreciation for the application of science to real-life situations and learn a lot of chemistry at the same time.**



## From Quarks to the Cosmos

Adapted from a 1997 presentation by Leon Lederman

Leon Lederman participated in the discovery of the *K*-meson particle and the non-conservation of parity during muon decay and, with Melvin Schwartz and Jack Steinberger in 1962, the identification of the muon neutrino for which they received the Nobel Prize in Physics in 1988. In 1977 Lederman led the team that discovered the bottom quark.

Lederman received his Ph.D. in physics in 1951 from Columbia University where he spent the next 28 years, serving as director of the Nevis Laboratory from 1961 to 1978.

As the second director of Fermilab (1979-1989), Dr. Lederman oversaw construction of the Tevatron, the first superconducting synchrotron and initiated a wide range of educational programs and symposia.

Leon Lederman is currently Pritzger Professor of Science at the Illinois Institute of Technology; vice-chairman of the board of the Illinois Math and Science Academy, and co-chair of the board of trustees of the Teachers' Academy of Mathematics and Science in Chicago.

**An experimental physicist** and a theoretical physicist were traveling by car, and the experimental physicist says, "I think my turn signals aren't working, would you go out and check." So the theoretical physicist goes to the back of the car and says "Now they're working, now they're not; now they're working, now they're not."

Another remarkable scientist is quoted as saying that there are only two kinds of mathematics books: one in which you can't get past the first chapter, and the other in which you can't get past the first sentence. I tell you that I can't get past the first word, and other people have the same trouble with quarks. When you say "quarks," there is a tendency for eyes to glaze over. I teach freshman physics, and I have a lightbulb on the desk. It's connected to a foot switch, and whenever I get to something real hard, I push the foot switch, the bulb lights brightly, and I tell my students it's an eye-glaze detector.

I was told this lecture was supposed to be serious, so I'm going to address the connection between particle physics and cosmology or astrophysics. I'll include the Big Bang, the origin of the universe. There are various references such as Steven Hawking's book, *A Brief History of Time*, which was on the best seller list for 160 weeks. I also co-authored a book called *From Quarks to the Cosmos* which was on the worst seller list, but there is hope: it just came out in paperback, and it's much cheaper.

I was reading the *New York Times* yesterday, and on the front page, lo and behold, was an article titled "At the Other End of the Big Bang May Simply Be the Big Whimper." It was a report of the conference of the American Astronomical Society, and the writer selected this quote: "There will come a time when not only the sun will die, but the lights of all the stars will vanish. All that will remain in this bleak, darkened"—this will be a really funny talk and you should enjoy it—"darkened future will be an increasingly diffuse sea of electrons, positrons, neutrinos and radiation." What the story was reporting was a quantitative theory of *the future* of the universe. You're going to hear tomorrow from George Smoot, an expert on the other extreme, *the origins* of the universe. It's impossible for scholars not to try to predict the future, but we'll deal with that later.

### Science and the Greeks

I begin at the beginning of science, which took place in 650 B.C. (I wasn't there, but I am old enough to remember when the Dead Sea was only sick.) Miletus was a colony of the Greek empire, a crossroad for land traffic from the motherland to its possessions in North Africa. It was a very popular colony city on the west coast of what is now Turkey, washed by the Aegean Sea, and its well-traveled citizens were familiar with cultures on both sides of the Mediterranean. One of them, Mr. Thales, was a businessman, and he and his colleagues made two remarkable breakthroughs in thinking about the world. One was that the world could be understood by logical reasoning.

Up until Thales' time, that part of the world coped with the mysteries and terrors of nature by developing marvelous mythologies, a beautiful set of stories that accounted for the scariness of the thunder and volcanoes and earthquakes and other terrible things. Thales said let's put that aside and see if we can understand the world by logical reason-

ing; logos replacing mythos. The second breakthrough was that he and his colleagues also believed that there must be a simple, overarching principle which accounts for how the world works in all its complexity. Of course the tragic fire that destroyed the library at Alexandria also destroyed the original writings of these early scientists, so most of our information about that school comes from the historians of the time and from Plato and Aristotle. At any rate, the idea of a unifying principle rattled around the Mediterranean basin.

In 400 B.C. a gentleman by the name of Democritus tried to put it into a more realistic context. He said that while the world looks very complicated, there's something deep down that's very simple. He suggested the existence of things which are too small to see, in continuous motion in a void, and which are impossible to subdivide. "Tomos" is the Greek word for cut, and "a-tomos" denies the possibility of cutting. If you have a mythically sharp knife, Democritus proposed, and you could cut something in half continuously, eventually you'd come to a point where your pieces would be too small to be seen, where it would not be possible to cut further, a kind of reductionist philosophy. This would be basic primordial stuff. Today, of course, we use the term "atom" for the uncuttable pieces.

Now if you want to follow the reductionist road, you have to take things apart. To do that we have to invest some energy or heat, so we introduce the units of energy the physicists and chemists use, the electron volt. This is the energy you get if you connect the top knob of a one-volt battery to its case, and an electron zips from one terminal to the other. If you want to melt snowflakes, you need a hundredth of an electron volt, but if you want to do something serious and take apart a molecule, you need a few tenths of an electron volt. If you want to break apart atoms of oxygen and hydrogen, then you need one, five, ten electron volts. And if you really want to dismember an atom, you need thousands of electron volts, even up to tens or hundreds of thousands of electron volts.

### Accelerators: exercises in reductionism

We know that the atom consists of clouds of electrons surrounding the nucleus and, that if you want to zoom in on the nucleus, you need a million electron volts. And to really take the nucleus apart, you need billions of electron volts, abbreviated "GeV," because the European billion is "giga." Continuing the voyage to inner space, if you want to look inside at the constituents of the nucleus, like protons and neutrons, you need tens of billions of electron volts. The world's most powerful "taker-aparter" of things is the giant accelerator, the Tevatron, which is at Fermilab outside of Chicago. It can generate an energy of 1.9 trillion volts. Not terribly far from here is the little village in Texas

called Waxahachie, and I think that what's considered a village in Texas is about the geographical size of New York. In Waxahachie, Texas there was supposed to be a 40-trillion-volt accelerator but, sadly, that's defunct. The Europeans are constructing a machine of 14 trillion electron volts in their laboratory, CERN, near Geneva.

You can follow this reductionist road which starts in Miletus in 650 B.C. (figure 1). Originally we thought it would end in Waxahachie, but a funny thing happened on the way to Waxahachie. My map shows such way-stops on the road to understanding the world as Democritus, Copernicus, Galileo, Newton, all the way to Einstein, Feynman and so on. If you don't see your name on the map, you have to work harder. The map does include Burger King—an old joke. As a proponent of the superconducting supercollider, I gave a talk to the National Conference of Governors, at the time when six states were offering competing sites. I mentioned Waxahachie, Texas, as a remote place that was a candidate, and the governor of Texas interrupted me and said, "Y'all need to learn how to pronounce Wax-a-hach-ie. That's where it's going to be at." I explained to the governor that when I actually went there, I *really* wanted to learn how to pronounce it, so I went into a restaurant and asked the waitress to tell me where I was and to pronounce it very carefully. The answer was "Bur...ger... Ki...ng."

Back to the road. In 1687 A.D. we come to Newton's publication of *Principia*, a tremendous synthesis, and really one of the major milestones. The reason we're on this so-called reductionist road is not because it's the only way to learn about science—in some cases it leads in the wrong direction—but because we believe the Greek idea of simplicity and overarching principles has validity. And talk about overarching principles, there's nothing more all-encompassing than the Newtonian revolution, encompassing the laws of motion and the law of gravity. There are Newton's directions on using this knowledge to make predictions about a falling apple and the rising moon—it's a great overall synthesis. When Newton proposed his law of gravity, it was a bold leap. He had no firm evidence to prove that two masses attract each other through empty space with the inverse square law of force. He had no idea why the Earth should know that the Sun is there, the Moon should know that the Earth is there, acting somehow across vast reaches of empty space. It was a tremendous leap, and Newton himself was aware of the fact.

In the 18th century, people with curious names so reminiscent of electrical units—Coulomb, Ampere, Faraday—developed the theory of electric charges and current loops and the laws of electricity. Integral to those breakthroughs were chemists—and some of my best friends talk to chemists. The only trouble was that the chemists were totally

ignorant of the Greek language, so they called the smallest pieces of a given chemical element—the smallest pieces that have the properties of that element—atoms. Perhaps they called them atoms because it was thought they were not divisible. As we already have seen, the easiest thing in the world is to take apart atoms. All you need is a gas flame and atoms start coming apart. In fact, people started systematically taking apart atoms with great glee at the beginning of this century, and terrible things began to happen. Atoms were complicated. Clouds of electrons appeared to circulate around a miniscule nucleus, occupying much less than a trillionth of the volume of the atom.

This is somewhat like a grain of dandruff floating in the middle of a large lecture hall.

**Splitting the atom**

The study of the nucleus of the atom began in the early 1900s, and required probes of many millions of electron volts, considerable technology and rather large sums of money. A serious attack on the nucleus took place after WWII. The technique was to accelerate particles in machines operating at hundreds of millions of volts, aim these at an innocent nucleus, break up the nucleus and watch the way the debris scattered around the place. By keeping track of the results of these collisions, we learned how nuclei stayed together, and the nature of the forces between nuclear constituents. A picture of the nucleus shows the tiny grain of dandruff filled with protons and neutrons in very close contact. The protons all have a positive electrical charge and that should blow the nucleus apart. It doesn't, however, so there's a stronger force that's holding it together. The physicists, in their vivid, poetic language said, let's call that the *strong* force.

The program worked. By watching the protons and neutrons coming off from these energetic collisions we learned a lot about nature of the forces that hold nuclei together. But, as usual with a technique designed to solve one problem, fresh problems cropped up. New particles were created in the collision which could not have existed before. The mass of these particles was generated by the incoming energy using a formula which will be on the quiz this afternoon:  $E = mc^2$ .

The new particles didn't live very long, but they lived long enough so that experimenters were able to measure some of their properties: mass, electric charge, spin, lifetime, and many other quantum properties that we don't have to define. The particles had to be given names, and by international accord the individual names were Greek letters. And so physicists became adept at the Greek alphabet—delta, kappa, lambda, sigma—as if we understood it. Collectively, all these particles, born in the strong collisions at the accelerator, were called *hadrons*, and the collection and study of these characterizes the period from about 1948 through 1970.

I'll tell you the story about the mouse who was starving but was terrified, for although he knew exactly where the cheese was on the kitchen table, he kept to his little hole because there was an animal walking up and down outside. And then the mouse heard the animal bark. "Ah," said the mouse, "that's the dog—I can run faster than the dog." So the mouse ran out of the hole, and the cat jumped on him and ate him up. "It's important to know two languages," the cat said as she cleaned her whiskers.

Almost as depressing as the end of the universe is a

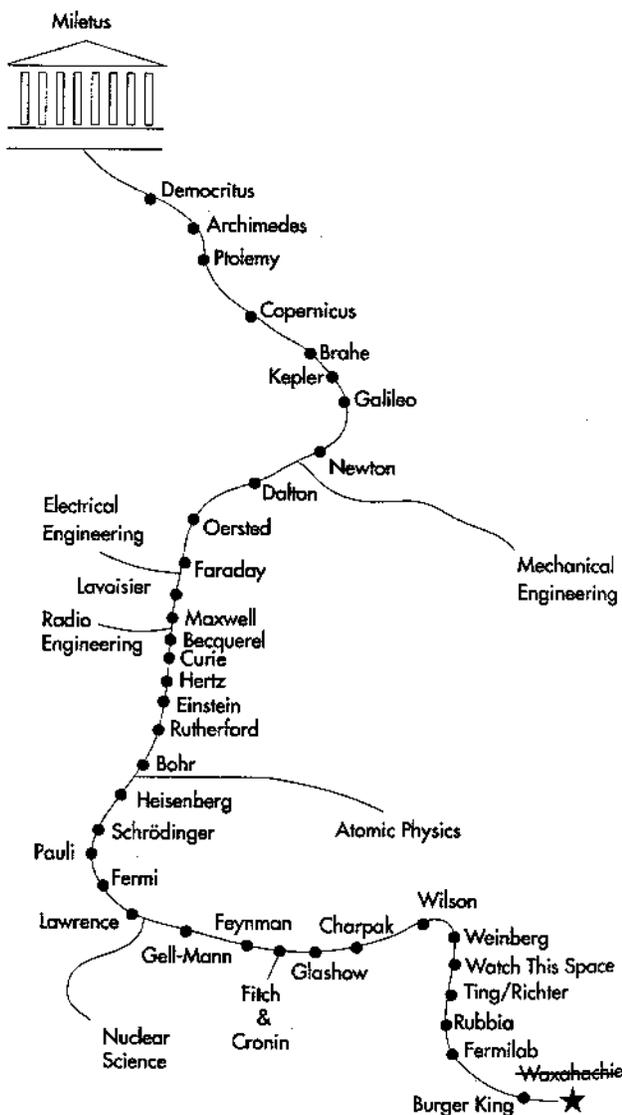


Figure 1. The road map of physics history from 650 B.C. to the SSC at Waxahachie (almost).

snapshot of our knowledge of those particles taken about 1970—tables of hadron particles. So many! It would almost be routine, you'd go to a laboratory, you'd do an experiment, you'd pass "go," you'd collect a Ph.D. The professors who sponsored the Ph.D. students would say they deserved Ph.D.s, for they had discovered this or that hadron. By university regulation there had to be an English professor or social studies faculty on the committee, and they would complain to the dean that the accelerator was some sort of factory for producing Ph.D.s. And that generated conflict between faculties and administrations, as if you needed any new issues. In my institution there's a sign in the men's room by the electric hand dryer that says, "Push to get a message from the dean."

People were discouraged because there was a great advance in our ability to read subnuclear structure and we got complexity instead of simplicity. The Greeks promised simplicity, and what could we do? Well, we took advice from our biology colleagues—they know all about complexity. Their rule is, if you don't know what to do, classify. So we started classifying particles according certain properties which I will not go into, and suddenly found interesting arrangements which suggested some kind of family relationship for those extra electric charges, neutrons and so on. And these hundred or so particles, hadrons we called them, were clustered into octets and decuplets that suggested some sort of organization. Remember the periodic table of the elements which you learned in high school chemistry, the chart on the wall with all the chemical elements? The fact we had columns of particles with similar properties was a message from the atom screaming out at these chemists that there's some organization inside the atom that's repeated. We now know that's the structure of the electronic shells.

### Enter the quarks

The message got to Murray Gell-Mann and his Caltech colleague George Zweig that maybe these Greek-letter particles were composites of something simpler. This gave rise in 1964 to the quark hypothesis. They separately proposed three hypothetical objects, which by their various combinations would account for the 100 or so particles in that Greek-letter table. What they suggested was to adjust the properties of these hypothesized three quarks, which we now call *up*, *down*, and *strange*, with a half or dozen or so of the Greek-letter particles: the proton is one of them, the neutron is another, the lambda, the sigma. You then found different combinations of quarks that matched the different particles in the table of hadrons. Hadrons are then composed of quarks, stuck together with a strong force.

The quarks, those mythical particles, are supposed to balance each other. So, for example, if a proton is two up

quarks and a down quark, the electric charge has to add up to the charge of the proton, with its unit charge. We do that in this case by calling each up quark  $2/3$  of the charge, and so  $2/3 + 2/3$  and a down quark is  $-1/3$ —you've got to learn to add fractions, it's a new skill for physicists—forget the differential equations and add fractions. But it's true,  $2/3 + 2/3 - 1/3 = 1$ ; and the neutron is  $+ 2/3 - 1/3 - 1/3$ , which is zero, and so on. By using the quark hypothesis we can explain all of these Greek letter particles as combinations of up, down, and strange quarks: the delta minus is three down quarks, the neutron is an up and two downs and so on, all the way up to the particle that has three strange quarks, the omega, a very strange particle. Before the 1964 quark model that particle wasn't known, but the family looked like it was truncated, and it was predicted to exist. In fact, the Brookhaven Laboratory found it with the help of a hint as to what its properties would be.

Implicit in the quark hypothesis was the idea that the quark was primordial, indivisible, the *a-tomos* of Democritus. These were the real, uncuttable objects. To emphasize the fact that they can't be divided, physicists think of these particles as having charge and mass and so on, but no radius. They are deficient in radius; their radius is zero. This is consistent with results of measurement. Now that doesn't mean that they're really points, it just means that we have failed, over the many years we've been dealing with these particles, to find any evidence that they take up space. All of the straightforward ways of measuring the radius of particles tell you that the radius of quarks—if there is a radius—is less than some tiny number. As we get better experiments, that number gets smaller and smaller.

We've known about other particles like the electron, the muon and the neutrino for a long time, and they're different from the quarks in one particular way: they do not have the strong force. Indeed, they *ignore* the strong force. The electron was discovered in 1897, and this year marks the 100th anniversary. I think a party is scheduled—dancing in the streets of Tucson—to celebrate the event. These other particles were also given Greek letters, and collectively they were called *leptons*. And again we cannot find any evidence for size after many years of experimentation, especially with the electron. We think they may well be points as are the quarks. Today we believe that matter is made up of quarks and leptons. If we find something that is not made up of quarks and leptons, it's very easy for us to say that it's not matter, but something else. And, indeed, we will come to that something else.

Well, theorists were overjoyed; they had a wonderful model. They had three quarks, up, down and strange, and three leptons, the electron, the muon and the neutrino. They were happy theorists, and said, "three is a deep and

wonderful number, and there must be a deeper connection between quarks and leptons, because there are three of each." Well, it didn't work out that way. In 1961 the lepton class was augmented by a new neutrino, one sort of like the old one, but different, and closer in characteristics to the muon. It was discovered at Brookhaven, and so there were now four leptons and not three. "Four is nice too," said some theorists, "but wouldn't it be charming if there was also a fourth quark?" And in 1975 out of respect to the theorists' wishes, the fourth quark was discovered in the SLAC accelerator at Stanford. That's on the West Coast, I might add, and it sometimes confuses Easterners because the beaches are on the wrong side of the ocean.

In the '70s the fifth lepton was discovered, which was given the name "tau," and the fifth quark, which was called "bottom." Now why, when you find a particle, do you call it "bottom?" Well, for a long time the symmetry in the organization of these particles had been apparent. They had a strong tendency to come in pairs. And in fact the electron had its own neutrino which grew a subscript after a 1961 study. I was part of that experiment, and 25 years later people in Sweden sent for us to present us a Nobel Prize and tell us what a nice experiment that was. The muon was discovered to have its own neutrino, so these things came in pairs, and were coupled together. It appeared, then, that the fifth quark must have a partner, so we named it "bottom" quark. Another one would soon be discovered and it would be "top."

Now because all of this is a bit heavy, let me tell you that in Chicago there are a lot of bars, and in one, at about five in the afternoon, two gentlemen were sipping scotch. "Do you mind answering me a question?" one asks. "Where do you come from?" "I come from Scotland," the other answers. "Well, I come from Scotland too," says the first. "What part of Scotland, if I might ask?" "I come from the west," the second replies. "Oh, I come from the west too, and what city?" "Edinburgh." "Edinburgh? I come from Edinburgh!" "In Edinburgh, what street did you live on?" And while this is going on, the bartender goes to the phone and says, "Send a cab, the MacTavish twins are drunk again."

So we have up and down, quarks with partners, and they make protons and neutrons. We have the charm and strange quarks which are exotic particles. And then in 1977, a group at Fermilab discovered a new quark and called it "bottom" in full expectation that its twin, MacTavish, would be found in a few years. In fact it took 17 years; Fermilab discovered the top quark definitively in 1995. Even before the top quark was discovered, there was already an organization of particles reminiscent of the periodic table. These are now called the fundamental particles, because they are points; or at least that's a convenient working hypothesis

Quarks	u up	c charm	t top	$\gamma$ photon	Force Carriers
	d down	s strange	b bottom	g gluon	
Leptons	$\nu_e$ electron neutrino	$\nu_\mu$ muon neutrino	$\nu_\tau$ tau neutrino	Z Z boson	
	e electron	$\mu$ muon	$\tau$ tau	W W boson	
I      II      III					
<b>Three Generations of Matter</b>					

Figure 2. The table of fundamental particles.

(figure 2). We classified them in three generations. The first generation included up, down, electron-neutrino and electron. This accounts for 99.99 percent of the matter that we know about in the universe. But if you walk through Fermilab or the other big accelerators in the world, they generate charm, strange, neutrinos, and muons with a fair amount of facility. Because these particles tend to be heavier, they're hard to come by. They play a much smaller, but vital role in our understanding of the world in which we live. The third generation, and for many reasons the last generation, we believe, consists of top, bottom, tau neutrino and tau. These particles are even heavier and more puzzling; why nature makes these "Xerox copies," I don't know.

A Fermilab public relations picture (figure 3) shows a top quark and an anti-top quark produced in a collision of protons and anti-protons at the accelerator. This excited *Time* magazine, which produced a new word called "Gotcha," and played up the curious properties of the top quark—that it's incredibly heavy, and that finding it took 17 years. The heavier a particle is, the harder it is to generate in the accelerator; although we thought the bottom quark was the heaviest of all basic particles, the top quark puts it to shame. It has a mass 34 times heavier! The heaviness or the mass of these particles is a key concept we're beginning to come to grips with.

### Socializing by elementary particles

The next question is what is the social relationship of the particles which I've introduced you to. How do they get along? Earlier we talked about Newton's universal law of gravity. The concept hasn't changed much except for

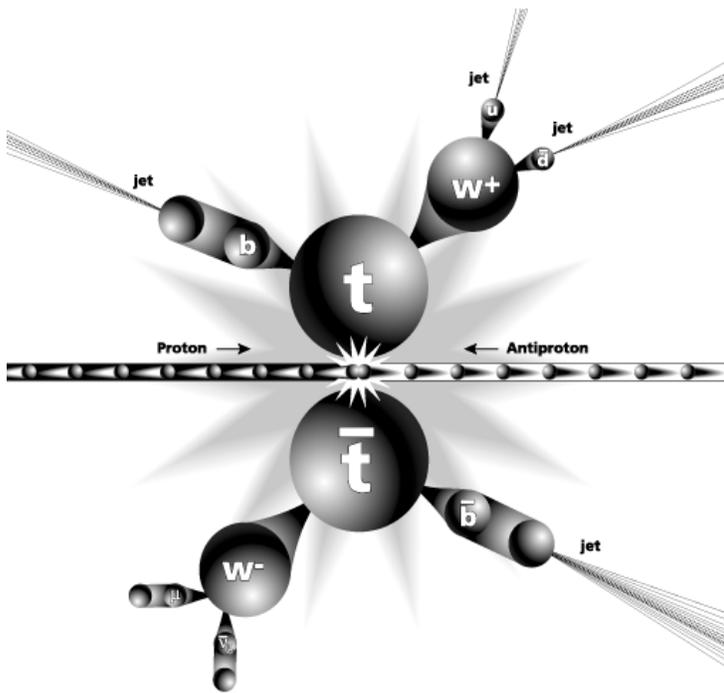


Figure 3. Artist's representation of a particle collision—a proton and antiproton collide at nearly the speed of light.

Einstein's 1916 General Theory of Relativity, a theory of gravity. And we talked about the electromagnetic pioneers: Faraday and Maxwell who established the mathematical description of the ELECTROMAGNETIC force. This includes Coulomb's law and the principles of motors and generators and TV and light, stability of matter and force. In the 20th century two new forces were defined. One was the STRONG force which generates nuclear bonding; it generates the bonding between quarks in the nucleus. The other is radioactivity, which was discovered 100 years ago. We recognized that radioactive explosions of nuclei and particles—particles don't live very long—are generated by the WEAK force; it's a 90-pound weakling. Studying the forces as well as the particles, especially the strong, weak, and electromagnetic forces, is the province of the accelerator laboratories. Gravity really can only be studied by astronomers and people who look for gravity waves. It's extremely difficult to do experiments on gravity; all you can do is look around and try to observe.

These four forces behave very differently, they reach out different distances, and they have different mathematical behaviors, but there's something they have in common: a similar mathematical idea which introduces the notion of symmetry. Symmetry is something you know about from art, from architecture and sculpture. One of the more abstract symmetries is called *gauge symmetry*. The evidence

of the commonality of these forces, their respect for gauge symmetry, implies how they actually are transmitted in space, something which puzzled Newton. The answer that modern physics gives is that the forces are transmitted by yet another class of particles. If I have a proton in one place and an electron in another, and they're coming together and trying to decide whether they should form a more perfect union, there's an exchange of photons, like an exchange of vows. The photons go back and forth at velocities always less than the velocity of light, and by so doing they create the force that is Coulomb's law. And now all the forces, the three major forces we know about, have force-bearing particles. Again in poetic splendor, we call these particles that carry forces, "force-bearing particles." (They're sometimes called *gauge bosons*, but not by us.)

The electromagnetic force is carried by individual photons which were demonstrated to actually exist in 1923. The weak force, the subtle force, is carried by three very heavy particles, W's and Z's. These three particles were discovered in Europe in the early 1980s, and their properties are still being measured. And then the strong force is carried by particles called *gluons*, obviously. (What else would you call them?) And in the 1980s all these force-carrying particles or gauge bosons had been documented and studied, and their properties are still grist for research.

Let's summarize: the matter that makes up the universe is composed of six quarks and six leptons, and they were discovered from 1897 to 1994 when the top quark was revealed. To complete this picture, the strong force is carried by gluons, the weak force is carried by W's and Z's, and the electromagnetic force is carried by photons. And gravity—well, we have a name for it, *graviton*—but we have no experimental evidence that gravity is carried by gravitons. There are attempts being made to try to detect these things, but unlike electrons which are very easy to produce and then produce photons in turn, we know of no way to arrange the production of a graviton.

In the history of accelerators, the first one was a linear accelerator, now called the Leaning Tower of Pisa. It was built vertically, and it was the accelerator from which Galileo dropped a skinny graduate student and a fat graduate student, and they hit the ground at exactly the same time. I'll skip from that to the Fermilab accelerator in Batavia, a town that's so small that its zip code is a fraction. It's 30 miles west of Chicago, and it's not the end of the world, but you can see it from there.

The Fermilab accelerator is a circular machine with a radius of one kilometer, completed in 1972. An aerial view shows a part of the underground accelerator, the graceful high-rise, and a rather spectacular project in ecology: an extensive tall-grass prairie restoration which includes a herd of wild buffalo. The laboratory is gracefully centered

on the site. The first director of Fermilab was Robert Wilson, a professional sculptor and physicist. He designed the entrance gate, which is called "Broken Symmetry." You can see the Chicago skyline from Fermilab, and an aerial view reveals the ring of the main machine with small accelerators that feed into the big accelerator (figure 4). We might compare the ring of Fermilab to Stonehenge and conclude that the Druids knew how to build accelerators. In going through the historical literature, the cost of Stonehenge was \$8 billion on the basis of man-years, because it required 1000 people working for 100 years. If we allow \$50,000 salaries with fringe benefits, that comes to about \$7 billion; the other billion was for ropes and other things for pulling the stones upright.

If you go down into Fermilab's accelerator tunnel (figure 5), you see the original machine completed in 1972 with conventional iron and copper magnets. In the 1980s we built the superconductor version of the accelerator, literally shoe-horning everything in with the old machine. The old one accelerated particles to 400 billion volts; the new one goes almost to 1000 billion volts, but uses just one-quarter as much electricity.

### Particles are elusive

There are various ways of detecting particles: one is with an obsolete instrument called a bubble chamber which doesn't exist any more. Tiny bubbles form on the trail of the charged particle shot into a liquid-filled chamber—a proton at 400 billion volts smashes into liquid hydrogen inside the chamber, for example. When it smashes into another proton, because  $E = mc^2$ , it beautifully produces a whole array of new particles. We later define these particles by studying the tracks in detail, including those made by neutral particles and particles which are unstable and explode in route, generating new particles. There are literally millions of pictures from bubble chambers, all taken to make measurements and learn things about particles.

Neutrinos are a nice subject. Neutrinos rarely interact with other particles, and a famous poem points out that they come to us when the sun is out and when it's 3 a.m., because they go right through the Earth. If you want to have a 50 percent chance of stopping a neutrino, you need something like a path through a light year of lead, but neutrino collisions are produced in a detector that weighs only a few hundred tons. The reason is that billions and billions of neutrinos are going through the detector, and you get lucky, and you institutionalize that luck in some sense.



Figure 4. The large ring, Fermilab's main accelerator, is a four-mile underground tunnel that is home to the Tevatron, the world's highest-energy accelerator. The ring at the bottom left is the Main Injector, currently under construction and due to begin operation in 1999.

Thousands of such events give us some idea of the properties of neutrinos.

Another particle detector is a modern Geiger counter called "CDF" for Colliding Detector Fermilab. Built by a consortium of ten universities, three national labs, and all of Italy and Japan, it's three stories high and weighs 5000 tons. The CDF has about 100,000 sensors, which are cabled to a home-built computer, and it calculates what's happening and generates a picture for the benefit of the dumb physicists. This whole thing is enclosed in a powerful magnetic field which allows you to measure the energy and momentum of various particles. The computer tells you what kinds of particles they are and what happens when they collide.

### The Standard Model in brief

With that digression, we go back to the story. The Standard Model, so called, with its particles and forces is elegant, and you can go to some university bookstores and get the Standard Model on a T-shirt. It leaves out some information, but it does summarize in a compact way a huge amount of data going all the way back to the "leaning tower." If you want to criticize the Standard Model, you can argue that aesthetically there are too many particles and too many forces. Why do you need six quarks, six leptons and 12 force-bearing particles? And this is not to men-



Figure 5. Inside the tunnel at Fermilab is the Main Ring with the second ring of the Tevatron directly below it. The superconducting magnets forming the second ring operate at liquid helium temperatures and accelerate particles to 1000 billion volts.

tion that gravity is somehow not included, and we don't even talk about the anti-particles because they're dangerous and annihilative. There's an aesthetic objection to the Standard Model, it looks complicated, and it takes about 20 parameters, which are determined experimentally to show us that the model allows us to understand how things work. Then there is another pathology: if I try to predict what will happen in a domain in which particle collisions add yet another factor of 10 or 100 in energies, this model predicts nonsense; probabilities greater than one, things that can't happen. So there's something wrong with the equation, something that has been left out. The situation is called the *unitarity crisis*. It raises the issue as to whether there are simpler, more unified, structures that describe the universe. The Standard Model, we believe, is an intermediate and a not a final step.

Successive efforts to unify the forces took place in the 1970s and 1980s, and Einstein spent the last 30 years of his life trying to unify the two forces he knew about, gravity and electricity. He cut himself off when developments in the field rendered his efforts fruitless; he did not include the other forces and the quantum theory that was involved in gauge symmetry.

Gauge symmetry was successful as a theory, and further thinking about unification took place. A hypothetical field was proposed, which I'll call the Higgs field, because that's what everybody calls it. You start out with, say, four particles of zero mass. Field theorists love to have zero-

mass particles, and they hate particles with mass when it comes to their mathematics. You start out with all the particles on equal footing when they all have zero mass, they all get along, and it's a happy family. Then the Higgs field comes in, and the Higgs field makes three of these particles extremely heavy, and ignores the photon. And suddenly we have two forces instead of one: the electromagnetic force and the weak force. And in keeping with this idea, people like Glashow (actually there is nobody like Glashow) were able to predict the masses of the W's and the Z's, which were then found in the 1980s.

### Astronomers look up, physicists down

Let me change topics and tell you a little bit about the cosmos. Astronomers are busy now. When we were building our accelerators and looking down, they were building telescopes and looking up through various devices. They were looking at gamma rays, X rays and infrared using the Hubble Space Telescope, space-based observatories and rocket satellites, and they made a huge amount of progress in understanding what's out there. We found that the universe is a dynamic place, not a pattern of fixed stars with an occasional comet swishing by. There are exploding stars, double stars that dance around one another, things called quasars and pulsars, and undoubtedly there are black holes. And the universe is filled with something called dark matter. The universe is expanding, and it's bathed in a sea of radiation which is called black body radiation. Black body radiation simply means that it has a characteristic distribution of energy over the different wavelengths. This tells you that the universe is 2.7 degrees above absolute zero. We know a lot about the abundance of the elements, hydrogen, helium, lithium and deuterium, not only in our own solar system, but—because of the spectral analysis—in the distant stars. And we know something about the large-scale structures such as galaxies.

A conclusion drawn from this vast array of data that is being collected is something called Big Bang cosmology. It's a little like the Standard Model of particles. The standard model of cosmology, however, is about an event approximately 10 billion years ago. We're getting to know this number better; the universe started ten billion years ago, give or take a few million years, and all matter presently observed in the universe at that time was compressed to some submicroscopic point at incredibly high temperatures and pressures. That's the key to the connection between inner space and outer space.

The expanding universe was discovered in 1929 by Hubble when he noticed that all the stars he was able to observe were moving away from him. It wasn't anything

personal. He discovered that the stars nearby had low velocities, and the stars far away had high velocities. He was able to plot a graph showing the distance and speed: the furthest stars were moving faster. If the points fall on a straight line, it told us that all of the stars coexisted at one particular point at some particular time. That time is called Hubble time. The graph also told us that the reason a star is close to us is that its velocity is low; the reason one is far away is because its velocity is high. This all fits together leading to the conclusion that the universe is expanding. Now is it expanding from Hubble? No, and I'll explain.

### Metaphor for the expanding universe

This is the most famous metaphor in all of science. Think of yourself as a two-dimensional person, maybe on a two-dimensional world, which happens to be the surface of a sphere. You have north, south and east, west, but up and down don't exist. You're a two-dimensional person, there's no up or down, and at the same time, the universe is expanding like a balloon. What are the properties of such a universe? First of all it has no edge and you can't fall off because wherever you go you're still on the surface. There's no center, because you're on the surface, and all points are equivalent. In fact, if you sit in one place, you'll see everyone moving away from you as the sphere expands. But if you sit in a different place, they're moving away from you also. So Hubble may be here, but there's a second observer way over there called Knubble who looks up and sees exactly the same thing; everything is moving away from her. There's no special vantage point.

This is a metaphor if you like, for our three-dimensional universe: we don't have an edge, we don't have a center and the universe is expanding. The ancients who worried about the edge of the universe asked, if you could go to the edge of the universe and throw a javelin, where would it go? We don't have an edge, and there's no problem throwing a javelin; however, it might hit you in the back of the head.

Well, here's the bottom line: if the Big Bang model is correct, all of the matter in Tucson, all the matter on planet Earth, all the matter in our solar system, in the Milky Way galaxy, and in the far distance—about 100 billion suns in our galaxy and at least 100 billion galaxies as far as our telescopes can see—all of that matter coexisted in a space which was smaller than a proton. The universe would have had a hot, dense beginning, hot enough to decompose matter into its primary components.

Well, suddenly astronomers started hanging around Fermilab, and we'd say, "What are you doing here? Aren't you an astronomer?" And they would explain that in the beginning the universe was quarks and leptons, and they wanted to learn more about this primordial soup and the

fundamental forces in order to model the evolution of the universe. So you had quarks and leptons smashing into each other in a hot, dense soup, but expanding and cooling (figure 6). And gradually as the universe expanded and cooled, two quarks were able to combine to make a meson. With more cooling there was enough accumulation of quarks to make protons and neutrons. Then eventually, protons and neutrons coalesced into nuclei. And as things cooled even more, an electron might have come into being and stuck to other particles to make something called an atom.

So that's the connection between particle physics and cosmology, and you can't model the universe without understanding this. On the plot of time and of size, the Big Bang happened at a point, and the universe has been expanding in time, and we are now at ten billion years. There are three possible futures based on different assumptions. The expansion of the universe, due to some primordial explosion, is slowing because gravity is an attractive force, and is trying to pull everything back together again. If gravity is strong enough and there's enough matter, eventually this expansion will reverse, and we'll have what disrespectful astronomers call "The Big Crunch"—all the matter will come back together again. But don't worry, the time scale is rather long.

Another prediction, however, is that the universe will continue to expand faster and faster forever, and as it expands it will cool. It's now 2.7 degrees, although there are local places like Tucson where it's hotter, but on the average it's 2.7. It can go down to .00001, however, giving the kind of heat death eloquently described in the recent article in *The New York Times* mentioned earlier (page 6).

The third possibility is the asymptotic universe, mathematically determinate. There's an equation with the parameter of omega: if omega is 1, the universe expands at an even slower rate, but it continues to expand forever. That's the most popular picture. The time when the action stops was written out in trillions in the *The New York Times* article, and it turns out to be  $10^{100}$  years. This should not concern the federal government; it's a very long time scale.

We don't know which of these scenarios is right, but the betting is that there's not enough mass to close the universe. Then there are more elegant theories. We don't know the answer yet, but it's knowable which path the universe is on; we may know within the next decade or so. Another view is that if we start anywhere on a logarithmic scale of energy and temperature, we can go backwards or forwards in time; if you know the physics you can trace things in either direction. At one point the universe cooled, quarks combined to form neutrons and protons and these in turn produced hydrogen, helium, deuterium, the light

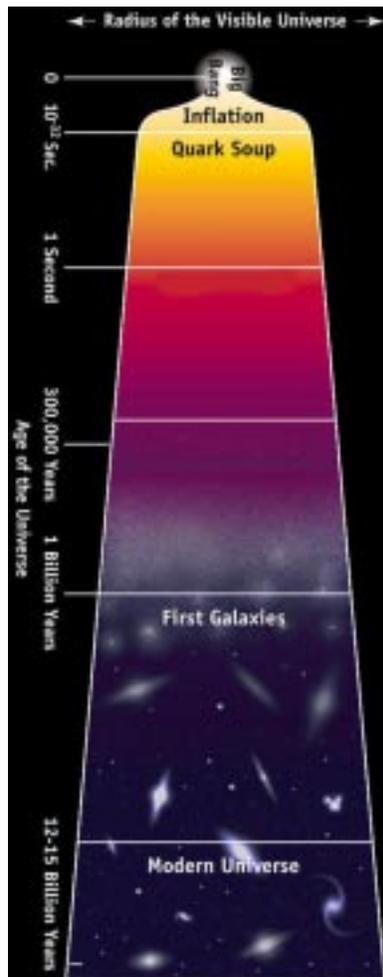
Figure 6. Illustration traces evolution of the universe from Big Bang and inflation to quark soup and formation of first atomic nuclei ( $10^{100}$  seconds after the Big Bang). Decoupling and recombination mark 300,000 years in which protons and atomic nuclei roamed a sea of photons, electrons and neutrinos. At last, matter and radiation began to clear and temperature dropped; photons traveled freely through the universe, decoupled from matter. About one billion years after the Big Bang, the first stars and galaxies were born.

elements. Nucleosynthesis is part of the test of the Big Bang theory. Before this time, incidentally, we didn't need chemists, because there was no chemistry. Galaxies formed a bit later, the solar system formed four billion years ago, and further along the time line there will be world peace, and the Cubs will win the World Series. If you look at an accelerator, it acts like a time machine. If the accelerator at Fermilab is running now, the collisions are characteristic of a time about  $10^{-12}$  seconds after creation. The particles are behaving like a microscopic copy of the early universe. And the supercollider, if it had been built, would have advanced us closer to the beginning.

We still have some prominent mysteries. Dark matter is a puzzle. For many years the cosmologists have been aware that most of the matter in the universe has not been identified. In other words, it reveals itself by its gravitational effect on galaxies and stars, but it doesn't give rise to light. Without light as a clue, there's a great deal of speculation about what dark matter is, and it could be that it is simply made up of neutrinos. But if dark matter is neutrinos, neutrinos have to have a finite mass at rest; indeed many experiments at Fermilab and other places are aimed at trying to see whether neutrinos are dark matter. The connections between cosmology and particle physics are very tight, and each discipline continues to address the other's questions. We think now that 90 percent of the matter in the universe is dark matter, but it may be new kinds of particles, perhaps exotic ones. And theorists, since they come in late in the morning, have to say, "I just thought of a new particle."

### Higgs: mysterious, ghostly

That's the inner space-outer space connection. Accel-



erators zoom down to ever smaller sizes, and telescopes zoom out. Astronomers, astrophysicists and cosmologists try to understand the origin and evolution of the entire universe. The Big Bang is generally accepted. The universe began about ten billion years ago in this explosion at  $t = 0$ . Originally it was very small, hot and dense; matter was decomposed into quarks, leptons, W's, Z's and maybe things we don't even know about. And it expanded and cooled, quarks combined, and then in the midst of it there was this mysterious Higgs. I want to leave you with the Higgs, which is a very ghostly kind of thing, as a puzzle (figure 7). This is a way of trying to make the world simple, without going inside electrons and quarks. We think these are points, but they may not be; there may be little people inside them. Experimentally, however, we're very happy with the notion that they're points, because particles that took up room would give a picture that wouldn't fit with Big Bang cosmology.

The Higgs field is the thing that complicates matters. In the beginning of the universe, there was something called the inflationary phase. The way cosmologists understand the universe is that there was a phase in its evolution where

it expanded rapidly in a trillionth of a trillionth of a trillionth of a second, going from an atom to a golf ball in size. And that was generated by a kind of Higgs ghost, and there are Higgs children, grandchildren, great grandchildren. Again it was the Higgs field; here there was perfect symmetry, only one force and one particle. Then the Higgs appeared and that gave gravity a separate character from the other three forces. So you have three forces combined, strong, weak, electromagnetic, and gravity.

Another Higgs came in at a lower energy scale and separated the strong from the electroweak. And then finally, at a much lower temperature, the electromagnetic and the weak were separated. And so, today, we have these four forces and six quarks and six leptons. And it's all the fault of Higgs. And the supercollider, the motivation for the supercollider, was to address the Higgs particle. If you want to know more details, there's the book\*, which is in paperback. It's only \$12.95, and it tells you everything I told you but in much greater, more boring detail.

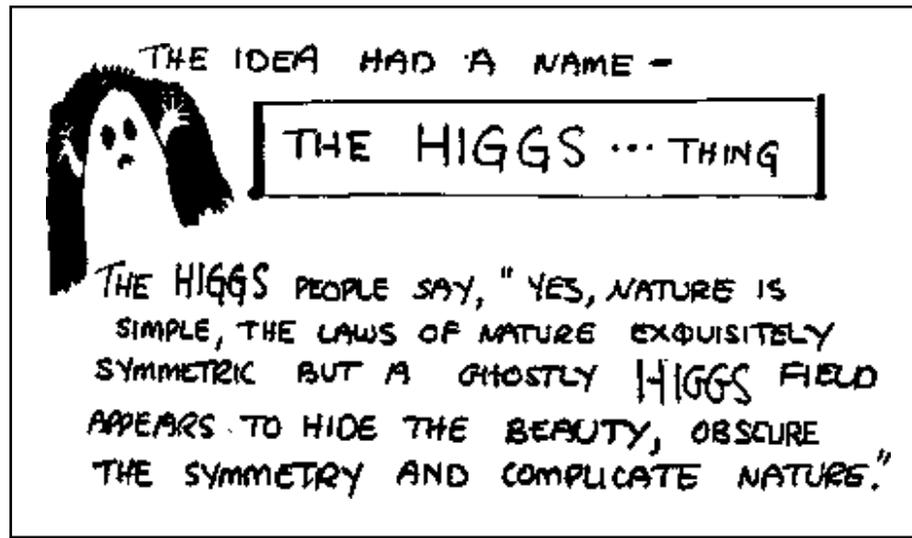
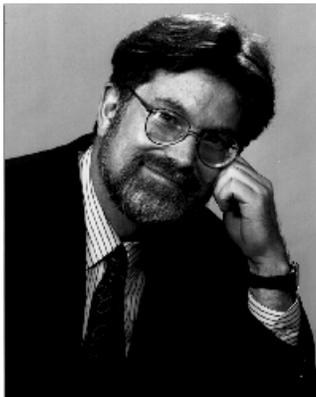


Figure 7. The Higgs field—as illustrated by Leon Lederman.

\* Leon Lederman with Dick Teresi, *The God Particle: If the Universe is the Answer, What is the Question?* New York: Dell Publishing, 1994.



*George Smoot works in experimental astrophysics and observational cosmology and is most famous for his research on cosmic background radiation. In addition to his research at Lawrence Berkeley National Laboratory and the University of California Space Sciences Laboratory, Dr. Smoot also teaches physics at the University of California, Berkeley.*

*Using instruments carried by balloons, on U-2 spy planes, and finally by satellites, Smoot has spent over 20 years examining faint radiation remnants of the Big Bang.*

*In 1992, Smoot confirmed that the long-sought variations in the early Universe had been observed by the COBE (Cosmic Background Explorer) team he led. The COBE satellite mapped the intensity of the radiation from the early Big Bang and found variations so small they had to be the seeds that grew into the galaxies and superclusters of galaxies that populate today's universe.*

## Observing the Early Universe

Adapted from a 1997 presentation by George Smoot

**Cosmology is the study** of the universe as a whole: how it came into being, how it developed, where it came from, and where it's going. What's the most successful theory of all time? I use this to tease my theoretical physicist colleagues. It's a theory of the universe we owe to the Greeks, but it's called the Ptolemaic picture because Ptolemy later corrected it. Its original form dates almost 2,000 years before the Big Bang theory. It's a reminder that all societies have had explanations of who they were, where they came from, and their place in the world.

The Greeks were the first to make systematic observations of the world around them to try to explain what was going on. They actually had several different models of the universe, including one that's very close to the Copernican system, the system that we teach students today. It was championed by Aristarchus, but it was ruled out, not because of religious or cultural reasons, but because of observation.

At the time, people said that if the Earth goes around the Sun, then our view of the stars should be affected by parallax. Here are the stars, out at some reasonable distance. If the Sun is here and the Earth is here and if it's moving around the Sun, we should see changes. When you're on one side then the angle to the stars would be different than when you're on the other side, for the same reason you have depth perception. The Greeks looked very carefully, and they saw no parallax. Aristarchus was not ready to give up his model, and decided it must mean the stars were very, very far away.

They let matters slide for a while, but a few years later Aristotle proposed that things moved in circles, the Sun, Moon, planets and stars, with the Earth at the center. This model took care of parallax, because we were in the center, and so the angles were always the same to everything else. Everything moved in circles, and in fact the model had a series of concentric crystalline spheres.

The Greeks knew the Earth was a sphere, and in fact they measured the radius of the Earth to an accuracy of about 10 percent using the simple geometry you learned in high school. Eratosthenes, a Greek who lived in Alexandria in northern Egypt in the second century B.C., knew that on June 21 at Syene in the far south of Egypt, the sun shone directly vertically down a well. The sun made a 7.2 degree angle at Alexandria, and the distance from Syene to Alexandria was measured to be 5000 stade (one stade equals 500 feet). So using simple trigonometry, or plane geometry, he calculated the radius of the Earth. The Greeks had actually calculated the size of the Moon and its distance from Earth using simple arguments: basically by looking at the size of the shadow of Earth during eclipses, figuring out what the geometry would be, knowing roughly the angular size of the Sun, and observing how the shadow changed.

The Greeks were good observers, and they chose the Aristotelian model for a reason: by using it they were able to explain the motions of the planets with fairly high precision. It was the model that reigned for nearly 2,000 years (slightly corrected by Ptolemy and others).

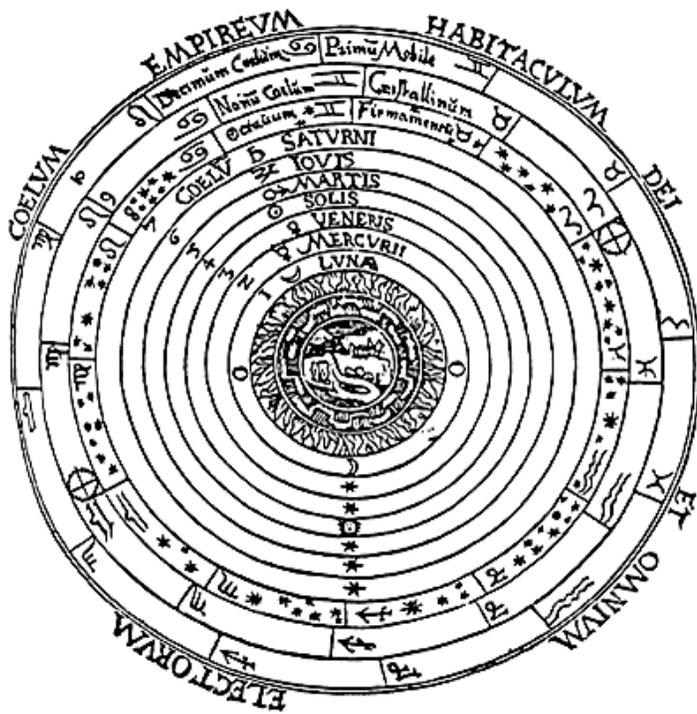


Figure 1. A 16th century illustration of the Ptolemaic system. The Earth is at the center and the “dwelling of God and all the elect” at the perimeter.

### The Copernican revolution

About 400 years ago, however, our world was shaken. Copernicus, the leading astronomer of the day, was called in by the pope in 1514 and asked to reform the calendar because it was inaccurate. Copernicus told the pope he would do it, but there was one problem bothering him—the motion of the planets which could not be accounted for by the Aristotelian model. This, in fact, was something Aristotle had pointed out—the motion of the planets had to be understood for science to be placed on a firm footing. At any rate Copernicus studied the motion of the planets, and as you know it got to be such a controversial topic, in disagreement not only with the Aristotelian view but Christian theology, that he didn’t publish it until 1543, the year he died. The Sun was the center of the universe, he maintained, and the Earth orbited it along with the other planets.

Some years later Danish astronomer Tycho Brahe provided evidence that supported Copernicus. Although not the first to do really precise work, he was the first to realize that he would have to break those crystalline spheres to change the model. And he discovered a comet, made observations, and showed that the comet came from outside the sphere of Mars and passed inside the sphere of

the Sun. It was actually passing through the crystalline spheres that the Ptolemaic model called for.

Tycho observed changes in the sky, and that things were not the way that people expected. Changes in theory were called for, but they weren’t taught. You know how conservative traditional schools can be. In those days, they kept teaching the Earth-bound Ptolemaic model even though there were alternatives available. Only when you went to the university and took advanced astronomy did you hear about alternative theories. What it really took was for Galileo to come along and overturn Aristotelian physics.

### Theology, culture and the Ptolemaic model

When change did come it was revolutionary, because the Ptolemaic model had been thoroughly absorbed into the culture. Here was Earth, and in Aristotle’s view things that were heavy and base fell to Earth. Those things were humans in the religious view; if you were going to be wicked and be condemned to hell, you had to be on Earth. The further *out* you went, on the other hand, the more holy and spiritual you were. Obviously God lived up there (figure 1). Everything was thus satisfactorily explained, and now those damned scientists and astronomers came along, and tried to overturn it all.

This turn of events was in keeping with much of the history of mankind. People have always had a mythology about the creation of the universe, how they came into being as people, and their place in creation. Indeed, we’re now in a period of transition between these kinds of models. But there was a revolution in the 16th century, as astronomers turned their telescopes to the heavens, they found that the Greeks, perceptive as they were, had made mistakes. Using simple trigonometry to estimate the distance to the Sun can be difficult. When you have a half moon, sunlight strikes the Moon and the angle to the Earth is a right angle. If you know the distance to the Moon and can measure the angle to the Sun between the point of the Moon and the Sun, you can use trigonometry and calculate the distance to the Sun. The angle turns out to be extremely close to 90 degrees, so a small error makes a big difference. The Greeks got a distance that was too short by a factor of 20. I suspect that they fudged the data because they couldn’t believe the Sun was that far away. They knew the Sun was much bigger than the Earth, in fact had done calculations, but for it to be that much bigger than Earth was perhaps not credible at the time.

It was confirmed at the beginning of the Renaissance that the distance to the Sun was wrong, and people started turning their telescopes to the heavens and measuring the

distances to the planets more carefully. Every time they made a measurement, the distance to the planets was further than predicted, first by a factor of 20, and then more. Then people started measuring the distance to the stars. Before Galileo turned his telescope to the heavens and saw many more stars than one could see with the naked eye, people had actually tried to estimate the distance to the stars by memory. A French astronomer, physicist, scientist and mathematician devised a technique for measuring the distance to the nearest star and chose Sirius for a test. He would go out at night and stare at Sirius for hours until he'd memorized how bright it was, and then he'd go in his house and go to sleep. He'd wake up at noon the next day and make a series of holes in his roof. Each day he put a different size hole in the roof, until he found one with exactly the same size and same image brightness as Sirius. If it weren't for the fact that Sirius was so much brighter than the Sun, he would have gotten approximately the right distance.

Once people had telescopes and new methodologies, they began to understand scale. During the 16th century, estimates of distance from the Earth to the stars increased by a factor of about 10,000. This was a theological problem, it's harder for God and the angels to look down on man when they're 10,000 times farther away; there's a psychological difference. In 1576, the pope issued a new canon. No longer was God thought to be in one place, but as infinite and all-powerful. He made stars that extended out to infinity. This also turns out to be a problem, but before I discuss it, I want to digress briefly. Just keep in mind that the pope had stars running out to infinity.

### The current model of the universe

Today's model has the universe expanding following what we call the "Big Bang" (but we don't mean a big explosion). The universe is expanding and evolving, and it's cooling down as it does so. If you looked back roughly 15 billion years ago, you would find that the universe was pretty much a uniform soup and, as it expanded and cooled down, that soup coalesced like milk curdling and created stars and galaxies and clusters of galaxies. These things evolved over time so that if you were able to look back to earlier times, you would see the first generation of galaxies, quasars, and lumpy things that are just forming. The universe over time is like an embryo, developing more and

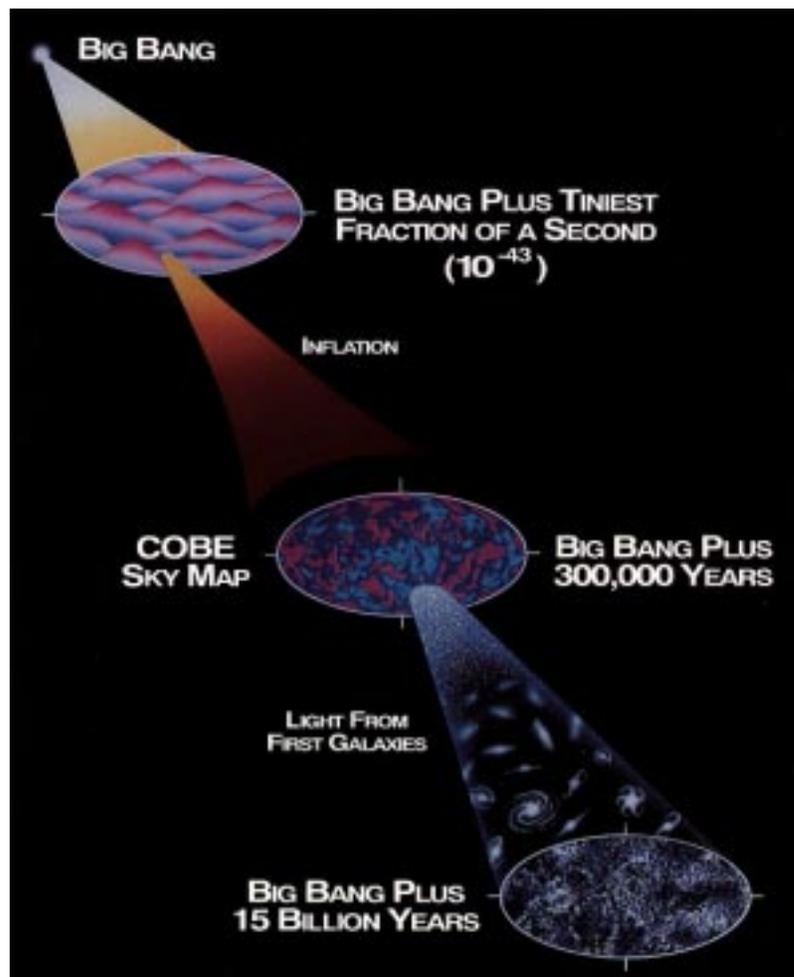


Figure 3. Artist's concept of the development of the universe. The top left image shows a small, hot universe just after the Big Bang. The center map uses one year of COBE satellite data to represent a time less than half a million years after the Big Bang, while the bottom shows conditions 15 billion years after.

more features (figure 3).

Roughly five billion years ago our own solar system formed, and we think that it's made of second- or third-generation materials, not the primordial material of the Big Bang. We've already been processed through the stars that created the elements we're made of. We've gone through a second generation which makes planets and life and everything else possible. We have this view of an evolving and changing universe which is very different from the static, infinite universe that the pope called for in 1576.

The Big Bang model was one in a number about 50 years ago, became a leading one in 1965, and the dominant candidate in the last ten years. In the next decade we'll get to a point that will be quite startling: we will be able to test

this model with precision. Dr. Lederman (see Chapter XX) told you that particle physics has a Standard Model, and it has been studied very carefully and tested to a great extent. Now people are looking for tiny variations in that Standard Model, at the level of 1/10th, 1/100th of a percent, parts of 1000, parts of 10,000, to see if it's going to break down. But we think we really understand how the simple particles like electrons and quarks interact with each other, and you can set up experiments and test them very precisely.

We're going to be at almost that point with cosmology in the next ten years. We have now visualized, and in fact launched in some cases, experiments that are going to test in different ways our model of the Big Bang to about the 1 percent level. And once we've tested the model, we're going to be able to extract from that same data a set of 10-15 parameters that will describe the universe in a very clean, thorough way. How is this possible?

### **The essence of the universe**

Just imagine that you need a new suit and the tailor is an alien from outer space. In order to get a good fit—he's never seen a human being before—he needs to measure every square inch of your surface area. If you bring in a human tailor, he measures your waist, sleeve length and shoulders. He has a model of a human, and by making a reasonable set of measurements, namely your waist, chest, inseam and arm length—perhaps six to ten measurements—he will be able to go back and make a suit that fits you extremely well. If he wanted to, he could make a dummy that looks like you. In the same manner, we're going to reduce the essence of the universe to ten numbers, and we're going to know those numbers to about the 1 percent level. This will tell us about the universe because we have a model, just as the tailor has a model, and because we know certain laws of physics.

If I know the ingredients of the universe at its beginning and the important laws that govern its evolution, and if the original situation is simple enough, I can make predictions on how the universe will evolve. I can't predict exactly where Mars will be, but I can predict that there will be stars and galaxies and roughly how many of each, what the life cycles of the stars and galaxies are going to be, and how fast the universe is expanding.

In short, we should now have a remarkable new model of how the universe was created, how it evolved and how it changes, and our model will have an incredible amount of precision. The analogy that I want you to embrace is the tailor—the suit may not be a perfect fit, but it will be pretty good.

In 1576, however, the pope had just given us an infinite universe that went on forever, and that was the be-

ginning of the new cosmology and other dilemmas. If there are an infinite number of stars going on for infinity, and it's been that way forever, then we're faced with what's known as Olbers' paradox. If the universe goes on forever and it's been there forever, and if you look out at the night sky, then eventually your line of sight will end up on a star no matter where you look. It doesn't take very long before you realize the entire sky should be as bright as the surface of the Sun, or even brighter. The next question is, why aren't we fried?

There are two solutions: one is to assume that the universe has only existed for a finite time, so the light from really distant stars hasn't yet reached Earth. The speed of light is only 300,000 km per second, and if the universe has only been around for a few seconds, you will only see out a short distance. If the universe has been around for a billion years, you will only see out a billion light years. The second solution is that because the universe is expanding, it's shifting the frequency of the light. If it's expanding rapidly enough, light is moved out of the visible into the infrared; if it's expanding even faster, it's moved down to microwave and so on.

Although both these theories still existed in 1960, both required an expanding universe; in one case a universe that had existed for a finite time, in the other a universe that was expanding and would continue to exist forever. There had to be more distinguishing features.

Back in the 1600s and today, however, it was—and is—evident that when you go out at night you don't get fried. And even if there is light down in the infrared, the total amount of energy striking the Earth from the universe is much lower than what would be needed to cook you, from microwaves all the way through to gamma rays. We know the atmosphere doesn't boil off. So Olbers' paradox was actually a key one, because it indicated that the universe must not be either static, infinite or uniform, but that one or more of those things must not be true. It also tells us that the total amount of information coming to the Earth is finite, but that's another story.

A picture from the Cosmic Background Explorer (COBE) satellite shows that we live in a spiral galaxy (figure 4). You're going to ask how we got this picture of our own galaxy, how far we had to go. The answer is, we don't live downtown; the nucleus of our galaxy is very bright, and we live way out in the sticks. If you live near a big city, you can always see the bright lights of downtown late at night, and that's what you see in a COBE picture. If you're a galactic astronomer, you're really impressed, because the picture is actually the first good view of our galaxy. It is not quite in the optical, but in the infrared which gets through dust better than visible light. Instead of dark regions where there are dust and gas clouds that obscure the stars, infra-



Figure 4. This photo of the nucleus of the Milky Way Galaxy shows “downtown.” Earth is “out in the sticks.”

red reveals streaky brown regions. You can see the starlight leaking through behind them. If you’re a cosmologist and you’ve studied Olbers’ paradox, you’ll be interested in the dark areas. We can see outside our own galaxy because we live far from downtown and the stars around us are not very important.

### Cosmology focuses on space between stars

To do our cosmology we must focus on all of the darkness out there. What we see is essentially no radiation in the visible or infrared between the stars. And if we look for any intervening material—there should be detectable absorption if there’s any intervening material—we find that most of space is a vacuum. Matter is clumped in our galaxy, as far as we’re concerned. It’s clumped in the stars or in the big clouds. Those clouds are ordered in a very peculiar way into a flat plane or into a ring around a central bulge. There are arguments about what might be going on, but what’s clear is that ordinary matter, the kind of matter that you and I are made of, is very much clumped in terms of our own galaxy.

The vacuum between us and any star is a better vacuum than anyone has achieved on Earth. We talk about these clouds being condensations of matter, but a lot of them have 100 atoms per cubic centimeter. If you should ever get to 100 atoms per cc in a laboratory vacuum, you could call the press in and make a big deal out of it. If you were able to look out and peer through the dust a bit you would see the nucleus of our own galaxy. Although we’re still in the galaxy, we’re two-thirds of the way out, and what’s behind us is much less dense than what lies towards the

center. This allows us to do cosmology in a very straightforward way.

If we are part of a galaxy then, the next question is: are we living in an island universe? We can solve the problem of the dark night sky if it can be shown that we live in an isolated system of stars. Well, in the 1920s Edward Hubble decided to attack this problem. He wanted to understand the distribution of what were then called nebulae, how they operated, and how they behaved in the night sky. Eventually he discovered a technique for estimating how far away they were, and that there were a number of nebulae that were at least ten times as far away as typical stars in the galaxy. That meant that they were outside the spiral structure that we live in.

Hubble founded a field called extragalactic astronomy, and one result is the Lick Sky Survey, a survey of the northern hemisphere, now 25 years old. It shows the location of a million galaxies.

The distribution pattern shows that the galaxies have a funny structure and they’re not uniformly spread across the sky. The pope wasn’t quite right, for the stars aren’t uniformly distributed, although they might appear that way if you go to a really large scale. Galaxies are often together in groups and large clusters. The Coma cluster near the north galactic pole has a couple thousand galaxies in it. Even clusters of galaxies are often in larger groups of galaxies, and there could be clusters of clusters of galaxies. There also seems to be a string, or a connection of clusters of galaxies in this region, while in other regions there are almost no galaxies at all. If you look carefully, you find huge regions, sometimes 50 million light years across, where there are no galaxies. And right next to some of these in a volume of ten million light years across, they’ll be 2,000 or 4,000 galaxies. There’s grouping and clumping, and it gets more and more complicated as we study it. It’s clear that there’s neither uniform distribution nor just dark, and that somehow there’s a mechanism behind these structures we see in the universe.

Every time we look on a larger scale, we see larger-scale structures. How is it all put together? That is the question we must answer if we want to explain how the universe evolved and how it works. We’ve surveyed up to four million galaxies. The photos show there are clusters, and the clusters are grouped, and there are voids. But now the voids are beginning to fill in, and we can see galaxies behind them. One needs a three-dimensional plot, and I have one that includes observations done in New Mexico by the Center for Astrophysics from Harvard (figure 5).

Work on this plot was started by Margaret Geller. John

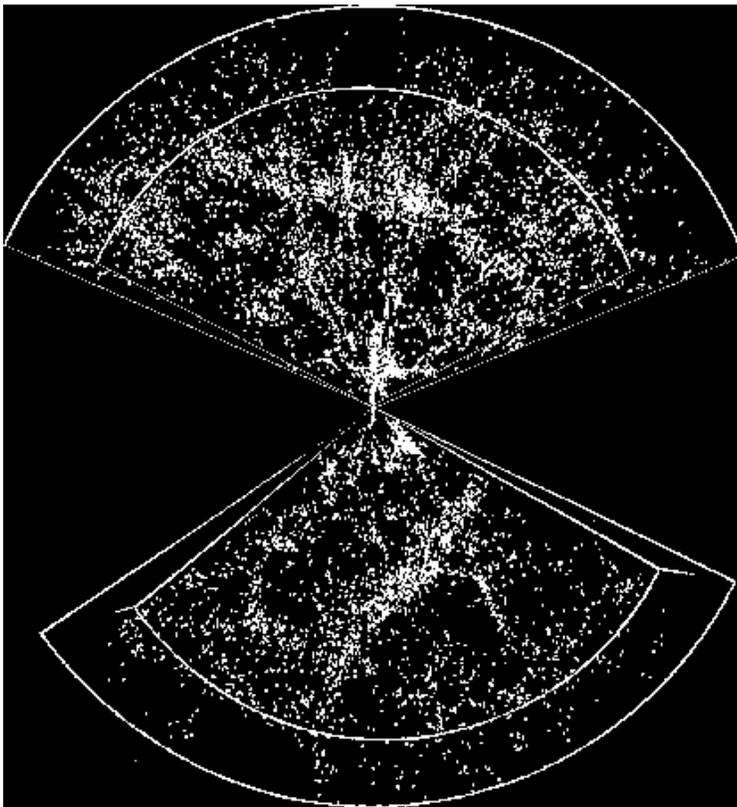


Figure 5. Large-scale structure in the universe is revealed by this plot of the northern and southern galactic hemispheres (note the large-scale patterns like the Great Wall stretching across the northern hemisphere). Each of the 9,325 points represents a galaxy. The Earth lies at the center.

Huchra and their students. Basically they took a series of pictures with a telescope as the Earth rotated. As you continue taking these pictures you get a wedge of the sky, and one series of wedges seems to show the stick-figure of a man. There are also elongated structures, known colloquially as the “fingers of God” because they’re pointing at us. We now know that the reason they look like structures is because the method used to estimate distances was based on either the red shift or intensity. When you have a cluster of galaxies and you’ve estimated that some galaxies are different than others in the cluster or have speeds relative to other members, you have a cluster that gets stretched out along the direction in which you’re estimating the distance. This happens because, while you can get angles very accurately, you can’t estimate the third dimension very accurately. So the “fingers of God” are just an experimental error, and the stick-figure of the man is simply anthropomorphic.

### The “Great Wall”

Another incredible structure is a big flat sheet that has been given the imaginative name, “the Great Wall.” The Great Wall is a big plane of superclusters of galaxies, and it’s at least 300 million light years across. That means that if the universe was fairly uniform in the beginning, you had to move this thing at close to the speed of light over the whole history of the universe in order to create it. So it was an awesome engineering job. The voids are much clearer when you see them in three dimensions. There are these bubbles in the universe, and they look like foam.

So we’ve gone from Olbers’ paradox, to knowing that the universe isn’t static, to discovering there are galaxies outside of our own with a complicated distribution in the sky. But there’s another thing: Hubble wasn’t satisfied with discovering galaxies and starting a whole new field of astronomy; he wanted to know where the galaxies were going, so he decided to measure their motions. He moved from Arizona and New Mexico to Palomar on Mt. Wilson to make observations of galaxy brightness versus recession velocity. The technique he proposed to use? The Doppler shift.

My explanation of the Doppler shift for high school students is the following: I have a Doppler dog, and he’s trained to bark very uniformly—bark, bark, bark—once a second. As long as he’s sitting still, uniform circles of sound come from him, and wherever you are, you hear the sound bursts come to you at the same speed. Now if the Doppler dog runs and barks at the same rate, his mouth moves forward and the center of each of those concentric spheres changes. His first bark was at one point, his second at another and so forth. And you can see that if he’s moving toward you, the barks are going to reach you more rapidly. And if he’s running away from you, they will be further apart. If you knew he’s trained to bark once a second and you measure the arrival rate, you can tell his speed along the line of the sound you are hearing. You can’t put a dog in every galaxy, but fortunately we have trained atoms, and they all have characteristic frequencies. If you look at the frequency shifts of spectral lines from atoms, you can tell how fast they are moving.

### The Hubble Law

With the data he collected, Hubble estimated the distance to the galaxies in millions of light years (figure 6). The recession velocity of the galaxies, as represented by the fractional increase in their spectral lines or redshift, was 0 to 1/3 percent. There were actually two sets of data, because you can treat galaxies as either individual galaxies or as clusters, but both were close together. So, when Hubble

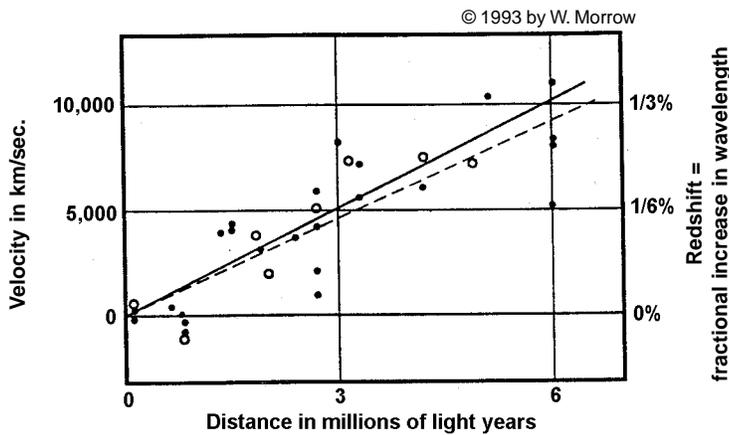


Figure 6. Hubble's 1929 velocity versus distance plot. His two sets of data are represented by the black dots and solid line (the Hubble Law applied to individual nebulae), and dashed line and circles (the law applied to groups of nebulae).

looked at this data, did he say "we have a new cosmological law?" The data were best fit by a straight line, now called the Hubble Law. Well, I think the data are adequately fit by a straight line because you don't have good enough data for anything better.

Fortunately the Carnegie Institution thought Hubble was a great guy and hired Milton Humason to help him out. Two years later, by 1931, they had measured out many times as far to 100 million light years, and up to almost 7 percent redshift or speed of light. With that many data points, a straight line was a pretty good observation. So the conclusion in those days was that the nebulae (they still didn't call them galaxies) and the entire universe were in recession. They were moving apart. This led to the Big Bang theory that the universe is expanding.

Unfortunately "Big Bang" is a misnomer, and I'm going to try to persuade you that it's just a catchy phrase. The idea of the Big Bang is that there was a huge explosion, things flew out, and those moving fastest went furthest. When you measure their distance and their velocity, you find a linear relationship—and that's exactly what you would get in an explosion or a race. If we race a tortoise and a hare and measure the distance they cover in a certain unit of time, and each has a constant velocity and the tortoise is moving half as fast as the hare, you will get the same relationship whether you're talking about one unit of time, two units or three units of time. There will be a simple linear relationship, and it is important to know the distance scale in order to know the velocity. Indeed, this is the major problem we still have in modern cosmology. If everything was crunched together at the beginning and there was a giant explosion, then you could explain what

Hubble saw that gives this relationship.

Well, I told you the Big Bang theory was wrong, so now I'll tell you what I think is right. We've got to go back to the ancient Greeks, actually to the Egyptians, because geometry was very important to them. They lived by the Nile, and needed the Nile floods to replenish the land. They had to get the crops planted fast enough so that they could eat and so the Pharaoh could collect taxes. This required rapid survey of field boundaries so crops could be planted. Geometry in ancient Egypt was an honored profession for improving surveying. Eventually Euclid condensed everything down to five rules called Euclidean Postulates. For

the next 1800 years, people tried to show that Euclid's fifth postulate was not necessary, and so a lot of mathematicians wasted a lot of time. The fifth postulate, I'll remind you, is that parallel lines stay a constant distance apart, and the sum of the angles in a triangle is 180 degrees.

In the 1820s, Gauss began to speculate about a geometry in which parallel lines could diverge. If you think about a funnel, for example, lines spaced out uniformly around the bottom will diverge as they travel outward. One consequence to this is that the sum of the angles of a triangle or a funnel will be less than 180 degrees (figure 7).

Eventually, Gauss, Lobachevsky and Bolyai independently came up with a hyperbolic, or equivalently negatively curved, geometry. Gauss was so convinced that it was true that he hired surveyors to go out and measure the sum of the angles of a triangle between mountain peaks that were 100 kilometers apart. Of course they were close to 180 degrees when you correct for the curvature of the Earth and just measure straight-line distances. But Gauss was convinced this might be the real geometry of the universe. Though it seems obvious, the idea of converging parallel lines (e.g., on the surface of a sphere) took a while. It wasn't until Georg Riemann came along and developed the theory that allowed us to calculate geometrical curvature that we got some idea of what a closed universe might look like. A sphere is a good example: if you have positive curvature, the sum of the angles in a triangle will be greater than 180 degrees.

Riemann was convinced that he not only understood geometry, but that he understood all physics, because all physics in those days was electrostatics and gravity. What happens with electrostatics and gravity? Well with gravity, two things attract in proportion to mass. And with electrostatics, opposite charges come together and same charges move apart. Riemann thought he could geometrize all of physics but he was wrong because he forgot about the fourth dimension: time. Then Einstein came along with

the theory of special relativity, and eventually general relativity. Einstein came to the realization that there is a relationship between the density of mass plus energy in space, and the curvature in space-time. A massive object like the sun will warp space-time and cause light passing near it to follow a curved path.

The general theory of relativity was published in 1915 and tested in 1919 by an eclipse. An eclipse was needed because observations of starlight passing near the Sun

would otherwise be impossible. With an eclipse, it was evident that Einstein's theory was right: the position of the stars close to the rim of the Sun appeared to shift as compared to observations of the same stars without the Sun. The amount they were further apart was very small, but it was a measurable quantity. People were able to measure the position of the stars when they were on one side of the Sun, and when they were on the other side of the Sun, and show that in fact light curved, and it curved by the right amount. Einstein's original prediction, which was made in 1915, didn't take time into account. He did the rough calculation and it was off by a factor of two, because he only took the bending of space into account; taking space and time together, you get the full factor of two. The full factor of two was the curvature of space-time; when you allow space and time to have curvature, you allow them to have a scale. That scale can change with time.

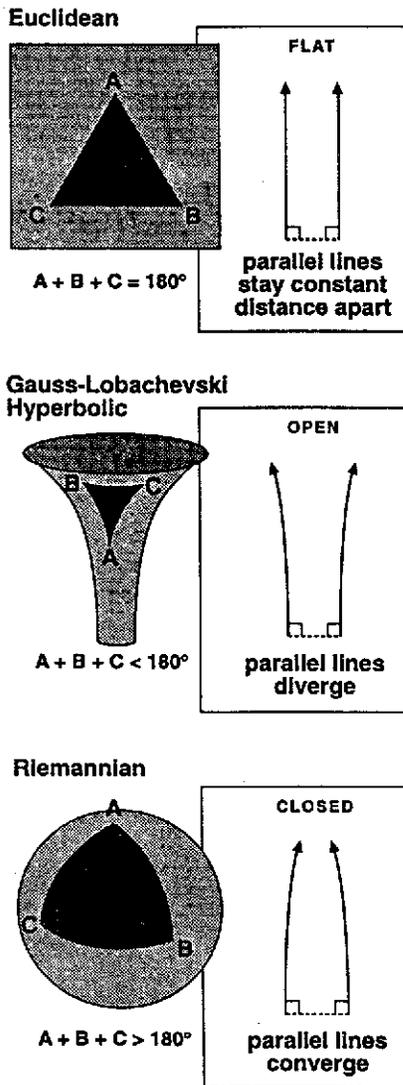


Figure 7. In contrast to Euclidean or flat-space geometry (top), the concept of curved space was proposed by 19th century mathematician Carl Gauss. Both he and Nikolai Lobachevski independently developed an open geometry (center) in which parallel lines diverge and the sum of the angles of a triangle are less than 180 degrees. The idea of converging parallel lines awaited Georg Riemann, who calculated geometrical curvature, a hint of what a closed universe would look like.

### The Expansion of Space

We now have a different view of how the universe is expanding; it's not due to a giant explosion, but because space itself is expanding. Imagine that space has only two dimensions (it is too hard to visualize it with three). And imagine that the spatial dimensions of space are like the surface of a balloon. We ride along a spot on the balloon representing our galaxy, which is being blown up at a relatively constant rate. As you can imagine, any two spots on the balloon will move apart. Depending on where you are, you'll still find a relatively linear relationship between the distance between these two spots and the rate they appear to be moving apart. So that the rate they're moving apart is just their separation multiplied by the fractional rate at which the balloon is expanding. You get exactly the same results as with the Hubble law, but it's a much different result in terms of psychology.

What's happening is that the space between the galaxies is increasing and growing, rather than the galaxies themselves moving out into space. The wrong view is that space already exists and the galaxies are being blown up into it, either on a balloon or in an explosion. What's really happening is that space is being blown up and the galaxies are going along for the ride. All of the problems—things moving at the speed of light—disappear in this view because locally one galaxy isn't moving relative to any other galaxies around it. It's only when you look over the long distances that space is expanding—like a rod that's being heated expands. Nothing local moves very fast, but the ends move apart fairly rapidly. Remember, it's space that's expanding.

### The universe and the structure of matter

We also have to study the structure of matter. The history of physics starts with matter at low energies, adds energy to it, and breaks it apart to see what it's made of. For the history of the universe, it's the opposite. You start with a hot universe and put together building blocks. At each epoch, those building blocks are bound with much less energy than the previous. As the universe cools down, you get more and more fragile objects.

Water, as an example, makes some very complicated and fragile things—snowflakes, snowmen, ice houses, ice sculptures and so forth. When you have frozen water as a solid, heating it up gains a symmetry; that symmetry is what we call a liquid. If you heat water up even more into steam, you have the molecules bouncing off each other. If you look at it, you realize that water is made up of two hydrogen atoms and an oxygen atom. And if you add more energy to the system to break it apart, you get individual atoms; looking at the atoms you find that each one is made of electrons and a nucleus. The nucleus is made of protons and neutrons. Now you've got a really pretty picture: everything you know of in the universe is made of three things—protons, neutrons and electrons.

This would explain everything, but life didn't come out this way, because now we know things are made of quarks, too. And if we look down inside a proton and a neutron, we know they're made out of quarks. But the nice thing about it is that there is a hierarchy—the energy levels are usually an order of magnitude or several orders of magnitude apart. And so, in each regime, you're dealing with one class order.

The universe starts out at a very high energy. It was very simple, because quarks interact in a very simple way; protons in a more complicated way, atoms even more so, and liquids and solids are even more complicated. As the universe cooled down it made more and more complicated structures, building up those levels of complexity. The history of science starts with the complex and progresses toward the simple: the world, is made of molecules, the molecules are made of atoms, the atoms are nucleons and electrons, and the nucleus is made of baryons, and the baryons are made of three quarks. That's the layered structure that Professor Lederman discussed in his talk. But in our cosmological view of the universe, we must start with those quarks, and work our way up until we make the Earth.

Our idea of Big Bang cosmology is to start out with quark soup at the beginning and build to make protons and neutrons. We then burn the protons and neutrons to make the light elements, which are going to form atoms. Those atoms form clouds, and the clouds form galaxies and stars. Finally those stars are going to burn the light elements

into heavier elements. Then the stars generate the elements that we're made of. In the beginning there were just quarks and electrons, which built up to the things that we see today. That's the whole process of how we think the atoms and the materials that we're made out of were generated.

Bringing us to this stage were Einstein with his general theory of relativity, and George Gamow and his coworkers, Ralph Alpher and Robert Herman, who worked out the idea of how the elements formed. There was an alternative theory still around during this period, and people didn't understand the structures of the nuclei before the 1940s. That's why we didn't have the Big Bang until physics was far enough along so that we could understand it.

These were the pioneers who gave us the foundations for a complete theory of how the universe was created and formed. The high-energy physicists working with accelerators are learning about quarks, leptons and how they interact, and how things fit together. We were all so disappointed when the SSC didn't get funded because the SSC corresponds to the energy level at  $10^{-12}$  seconds, micro-microseconds after the Big Bang. That's the point at which we think a big phase transition occurred in the universe, the time at which baryons started to form and particles gained mass.

### A message from the universe

Another thing happened back in the 1960s, a very funny thing. Somebody had the bright idea that you could save money by launching satellites and using them to relay telephone calls. Bell Labs hired Arno Penzias, now head of the successor to Bell Labs, and Robert Wilson, a scientist there, and asked them to measure the brightness of the sky and the background noise, because satellites in those days were tiny and weak, and didn't have strong signals.

Penzias and Wilson made systematic measurements and found that the brightness of the night sky was about 20 degrees. Part of it due to the atmosphere, part of it due to the signals from our galaxy, and part of it due to their receiver. There were 3 degrees out of the 20 degrees that were left over. Now most people would have just swept that under the rug because they didn't understand it, but Penzias and Wilson studied it carefully for a year, and concluded that since the signal didn't vary regardless of where they looked, it either came from their instrument or from the universe as a whole. They went to the trouble of proving it came from the universe as a whole, and 12 years later won the Nobel Prize for being so careful and precise.

What they had found was the light from the early times. The universe was much denser and much hotter in early times, and it had to be as hot as the Sun, or even hotter—

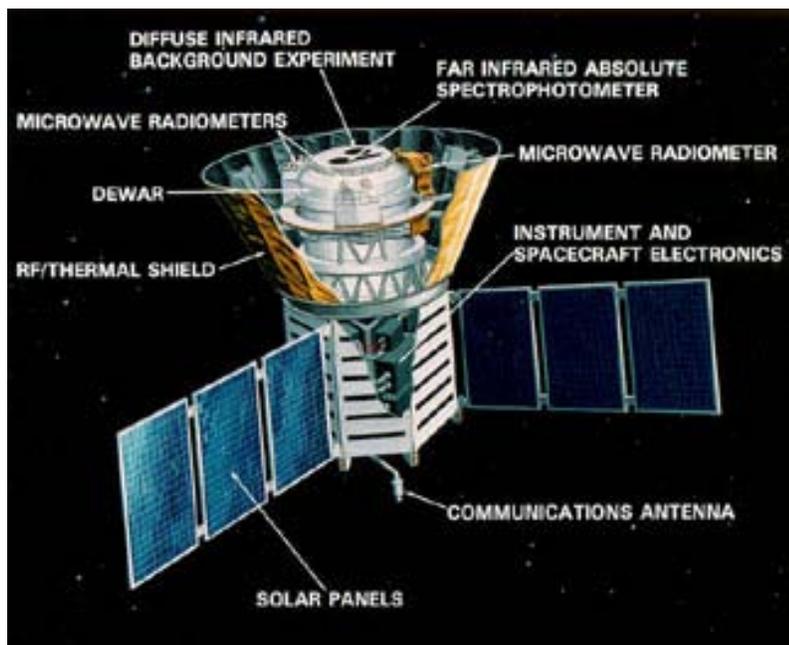


Figure 9. The first satellite devoted to cosmology, NASA's Cosmic Background Explorer (COBE) was developed to measure the diffuse background radiation from the early universe. Launched in 1989, it carried three experiments to search for and gauge background infrared radiation; to precisely map microwave background radiation, and to compare the spectrum of this radiation to blackbody.

a million degrees or more—in order to cook the elements. There would be some radiation left over from that high temperature, and that relic radiation would be cooled by the expansion of the universe. Its wavelength was stretched and therefore its temperature was stretched as the universe expanded. If the universe wasn't expanding, the night sky would be as bright as the Sun. That expansion had moved the relic radiation from the hot phase of the Big Bang down into the microwave region, right at the region in which they would want to make cellular telephone calls, only they didn't have cellular telephones in the 1960s.

That's how they discovered relic radiation coming from the early universe while exploring the possibility of satellite links. After they published that finding and there was confirmation, I realized that it was like light from the Sun: if you're careful with your instruments you can make a good image of the Sun. And if we were very careful we could make a picture of the early universe, and find out what it looked like when it was on the order of 10,000 to 100,000 times younger that it is today. That became one of my primary goals in science.

### Cosmic Background Explorer Satellite

Other people had this realization too, and eventually

three different groups proposed what became the Cosmic Background Explorer Satellite (COBE), NASA's first satellite dedicated to cosmology (figure 9). We would make a picture of the early universe to measure the thermal spectrum, and to look for light coming from the first generation of galaxies. Our satellite included three experiments with all the instruments stored inside a shield. We were well above the atmosphere, and were able to make very good measurements of the whole sky.

We cooled our equipment down to the same temperature as the radiation, 3 degrees Kelvin above absolute zero; liquid helium was needed for some of the equipment. Any fingerprints or other contamination would be absorbed onto the cold equipment, so we had to make sure everything was clean and wear "bunny suits" while working. We wanted our satellite to be on the terminator line so we

could look away from the Earth; so that we could always rotate the spacecraft so it looked away from Earth; and to keep it at 90 degrees to the Sun to shade it. And then we rotated the whole thing like a chicken on a barbecue to keep it a uniform temperature, and that also let us scan circles in the sky. We covered a lot of the sky on every orbit, and nearly half of the sky every day. After six months we swept the entire sky, and now we have four years of data.

### Measuring the temperature of the universe

A 1979 look down at the Earth with an earlier satellite shows a huge, but not surprising temperature variation: from 220 to 300 Kelvin. It's hot at the equator, no big surprise, and especially hot in the deserts; and of course it's cold at the poles. These measurements were made with equipment similar to the COBE instrument. They were checked by what we call "ground truth." You have weather stations all over the Earth, so if you want to check your satellite results you call up Buenos Aires and ask for the average temperature for 1979. You can get a grid of points to calibrate a map, and you know whether you have the right temperature or not. With COBE we had a problem: we don't have ground stations all around the universe, and so we had to test our equipment very carefully before and after launch to make sure we were getting the right values. It ended up taking about 20 people 20 years to make our picture of the universe.

At a wavelength of six millimeters approximately, the universe shows a diffuse uniform glow, and is different from what you see at any other frequency. There is a mean

temperature of 2.73K, or 2.728K to be more precise. There's hardly any variation at all, it's a very uniform sky. That's because the universe on a really large scale is very uniform. If you look at the early universe, it has hardly anything at all in it; it's just a uniform soup. And yet the modern universe has a complicated structure, so our goal is to measure the properties of the early universe that allowed

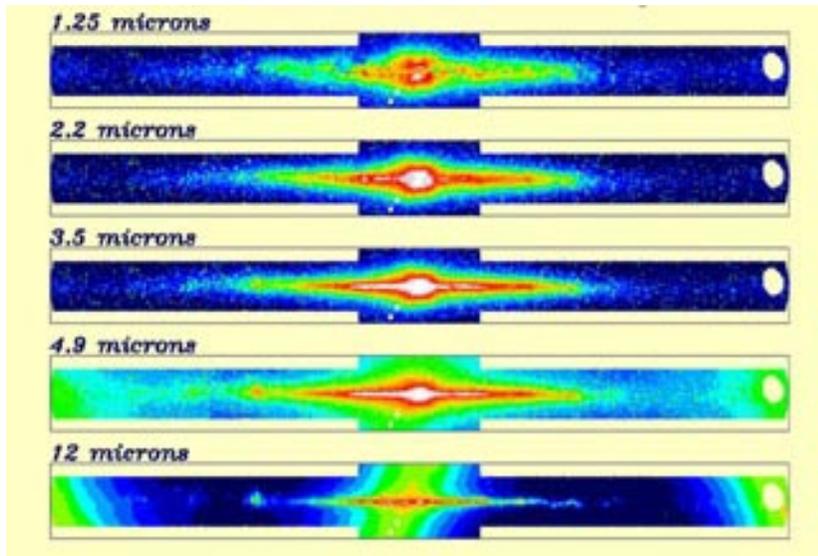
these structures to develop. Only at this wavelength—and it's really different from every other wavelength band, from the gamma rays down to radio—has nature provided a window that allows us to look out and see the early universe with incredible precision.

Pictures taken in the other wavelength bands (figures 10 and 11) show we do live in a spiral galaxy. Now we can

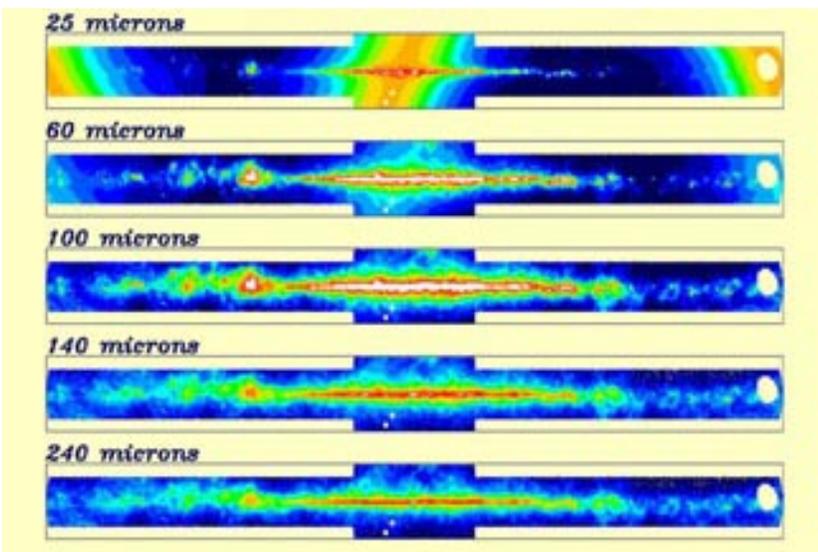
look through a series of pictures taken at different wavelengths. The stars in the central bulge are around 3,000 degrees and, because they're burning very slowly, they show up in images as a white puffy cloud. If you go down a factor of ten in wavelength to look at something with a mean temperature of 300K, roughly the temperature of the Earth, you can see the plane of the galaxy showing up and one spiral arm that encompasses Earth. There is also a structure with a temperature slightly below 300K—dust in the solar system. There's a lot of dust in the solar system because there are a lot of asteroids and comets. When the asteroids orbit and comets come in near the Sun, the asteroids grind against each other to make dust, and comets heated by the Sun spew out dust. In our studies we find it's roughly 50-50: at least 30 percent due to comets and the other part due to asteroids.

If you go down another factor of ten in wavelength to 30K, you see the dust in our galaxy showing up very strongly. The Large Magellanic Cloud and the Small Magellanic Cloud show up, and there's a big cloud in the galaxy relatively near us where stars are forming that shows up pretty well. Then there is the planet containing solar system dust. You can see the Moon, Jupiter and the other planets in this plane. This is a very different picture than you see in the microwave, because the microwave is so uniform.

You can't go down another factor of ten in wavelength because that would be 3K, but you can go down by a factor of three. A picture in this range shows cirrus dust which looks like the high, thin clouds that are highest in the atmosphere. These are the clouds that are the highest in the galaxy so they're the furthest from the stars and the coolest. These wispy



Figures 10 and 11. COBE images from the Diffuse Infrared Background Experiment . At 1.25-3.5 microns, are stars within our spiral galaxy; images show both the thin disk and central bulge of the stars. Our Sun is in the disk(right), about 28,000 light years from center. At 4.9-25 microns, emission is largely from dust in our solar system; the bands are due to streaming dust grains from colliding asteroids. Images at 60-100 microns are dominated by "infrared cirrus," the wispy-looking features resulting from dust between the stars in the Milky Way. At the longest wavelengths (100-240 microns), more infrared cirrus is seen, this from cosmic as well as closer sources.



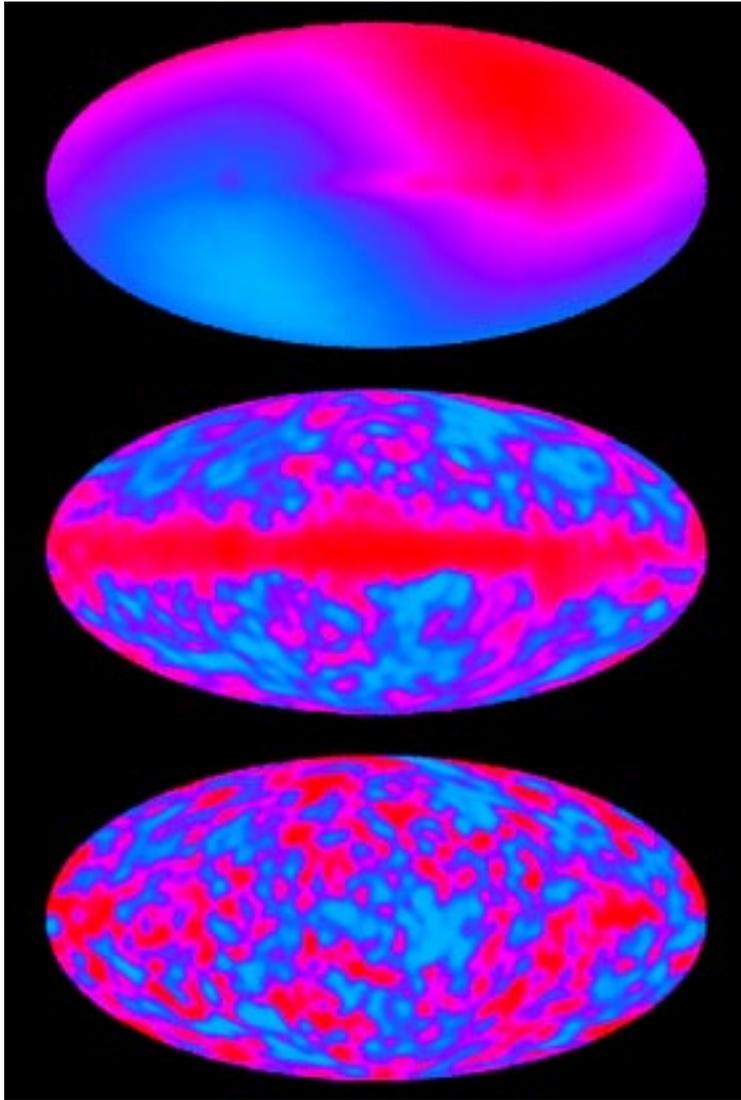


Figure 12. Microwave background maps from COBE. The top map includes the dipole and galaxy. The dipole, a smooth variation between relatively hot and relatively cold areas from the upper right to the lower left is due to the motion of the solar system relative to distant matter in the universe. The dipole has been subtracted from the middle map; both the dipole and galaxy have been subtracted from the bottom map.

clouds show up in regions at the edge of where the stars are. You can see what the galaxy looks like, and that roughly 100 times as much radiation comes from the plane of the galaxy than comes from far off the galaxy.

Let's look in more detail at this radiation from the early universe. We have precise measurements over this whole long wavelength range, and can see that the spectrum is very, very precisely thermal, and that means that it really was thermal at the beginning. On an early map we see a

region that's a little warmer than average and a region that's cooler than the average by a part in 1000. The galaxy shows up in its more intense regions at about a part in a thousand of the intensity of the microwave background (figure 12).

We believe this stronger intensity is due to the motion of our galaxy and our solar system relative to the cosmic microwave background radiation. It's like driving in the rain, when you're driving forward your windshield will hit more raindrops than your rear window. In the case of radiation, you increase the energy of the photons coming toward you, and you collect more, so it looks hotter. You're decreasing the energy behind because you're escaping; photons are not catching up with you quite as fast, and you collect fewer of them. So you see a temperature variation, which explains everything, including the galaxies.

We believe that our galaxy is moving at about a part in a thousand of the speed of light. We know our galaxy's rotating, and that its speed is roughly two-thirds of a part in a thousand of the speed of light. But this is in the opposite direction, so that means our whole galaxy is moving at slightly more than a part in a thousand of the speed of light, and that's pretty strange because none of other galaxies around us—we're in a group of about 14 galaxies—have a very large velocity. That means the entire group is moving at about a part in a thousand of the speed of light. It turns out that the group is moving off towards a very large cluster of galaxies in the distance. It is given a very unusual name, "The Great Attractor," because when the clusters were discovered people started looking for a great attractor that was pulling our

whole group of galaxies toward itself. They finally found it and the name stuck.

You can see the galaxy in a picture of the early universe; you see the spiral arms and the blackened center, and there's structure along the galaxy. Off the galaxy are regions that are cooler, and regions that are warmer, and unfortunately the contrast in such a picture isn't very great. We're talking about parts of 100,000, so the universe is incredibly uniform. That's about a factor of ten more uniform than a billiard ball. If you pick up a billiard ball, you don't feel any structure. If you could pick up the early universe you wouldn't feel any structure. You can see the seeds for those galaxies and clusters of galaxies on the Great Wall, these parts in 100,000 variations that we're seeing. Now we want to go back and measure these extremely precisely on all angular scales. The early universe is very simple,

when things only change by one part in a thousand, linear theory works great. You can understand the universe very well if you do precise measurements, and most of the information we're going to gain in cosmology will be through precise measurements of these variations in the early universe.

If we shift the color scale of our picture of the early universe, we can emphasize the very large regions that are cooler, and the very large regions that are hotter. What you see is a very long wave that is cooler than the average, and then smaller bumps, also cool showing up on it. It turns out that this is characteristic of what we call a scale and variant spectrum. There are equal size bumps on every scale, and they cover roughly equal fractions of the sky. So there are roughly equal, very large-scale structures, medium-scale structures and small-scale structures, and you'll have a long wave of something that's hot or cold and then the bumps that are the same size will show up on top of them. You'll have regions where there will be a whole lot of clusters together, because the clusters will come from small-size bumps, but they're going to be riding

on a middle-size and a long-size wave. We think we understand it, and it's the only theory that makes sense: otherwise the universe would be filled with things bigger than a Great Attractor, and everything would fly around the universe and come crunching together. Or if it was tilted the other way, the universe would be full of small black holes and you'd see them too. So it turns out that scale and variant is pretty close to what we need in order to explain how the universe came into being.

That's why we're confident we're on the right track. Our theories are developed to the point where we have a model and we can really go out and make these observations and learn the parameters of the early universe. We have a concept of an advanced version of the Big Bang which includes a thing called inflation, and NASA and the European Space Agency have both approved and begun missions to make the high precision maps that we need. The NASA launch should be around the year 2001, and ESA will be around 2004.

There's an exciting future ahead.



*First of the discoverers to spot Comet Shoemaker-Levy 9, Carolyn Shoemaker joined husband Eugene in the search for asteroids and comets after their children were grown. In 1980 she accepted a position as a visiting scientist with the Astrogeology Branch of the U.S. Geological Survey and has been an adjunct research professor of astronomy at Northern Arizona University since 1989. As of late 1994, she had 32 comet discoveries to her credit, more than anyone alive.*

*Science writer David Levy developed an interest in astronomy at an early age and is the author of *The Quest for Comets*, published in 1994, as well as 11 other books on astronomy. As an amateur astronomer, he has discovered 21 comets, 13 of which were in collaboration with Dr. and Mrs. Shoemaker. Levy got to know the Shoemakers in 1988 when they were tracking a comet he had discovered. (For biographical data on Dr. Eugene Shoemaker, see page 56, *The Clementine Mission to the Moon*.)*

## The Comet Crash on Jupiter

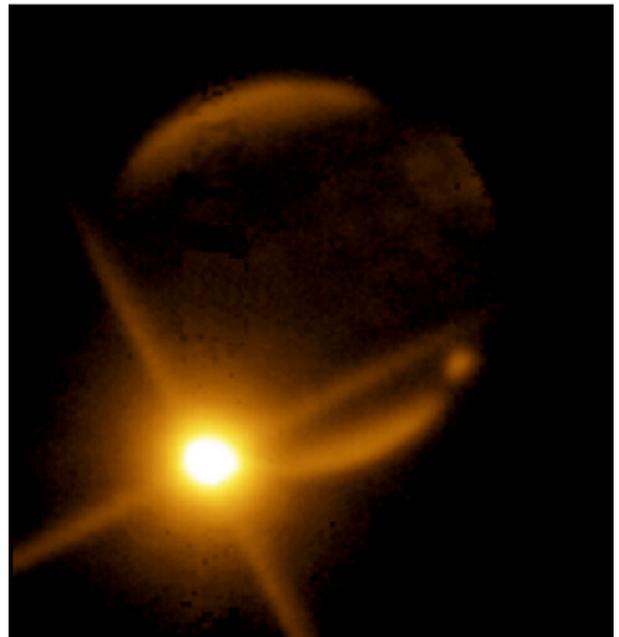
Adapted from a 1994 presentation by Eugene Shoemaker

**I want to share with you** what has to be the most exciting scientific event in my life—and I've been privileged to participate in a lot of exciting events in the exploration of the moon and the planets. I started studying impact craters nearly 40 years ago, including Arizona's "Meteor Crater," and other craters on the earth and moon.

A daydream I had was wouldn't it be fun to watch one of these craters made; to actually see the impact? I had in mind a small meteorite, but big enough to make a crater, maybe in the outback of Australia where no one would get hurt. I'd be nearby and rush over and map the crater. I didn't analyze that daydream, but if I had, I would have realized that my best chance of seeing a comet or asteroid hitting a planet would not be on the earth, but on Jupiter. Nevertheless, Carolyn and I and David Levy, who were the discoverers of this comet, were extraordinarily lucky; an impact of the kind that we saw happens perhaps once in a millennium. So even though the odds were very much higher of seeing an impact of this sort on Jupiter rather than on Earth, we were still very lucky.

What I've been calling "the gang of four" includes the telescope, David Levy, and my wife Carolyn, whom I call "Old Eagle Eye." It was really her remarkable ability to discover very faint objects on photographic film under the microscope that's been the heart of our program at Palomar for the last dozen years. She had discovered 28 other comets, and her eyes are very well trained. Seeing something, as all of you know, is not just a matter of visual acuity. It's having all the right information there in your brain to know what it is that you're looking at.

David Levy is a resident of Tucson and has his own private observatory in his backyard. He has independently discovered eight comets visually, and has been a very important member of our team ever since he joined it in 1989. Since 1982 we've taken more than 20,000 photographs of the sky. Working one week each month with the 18-inch



Impact of Fragment G of Comet Shoemaker-Levy on Jupiter, seen 12 minutes after impact. Impact site A is seen faintly on the opposite limb of the planet.

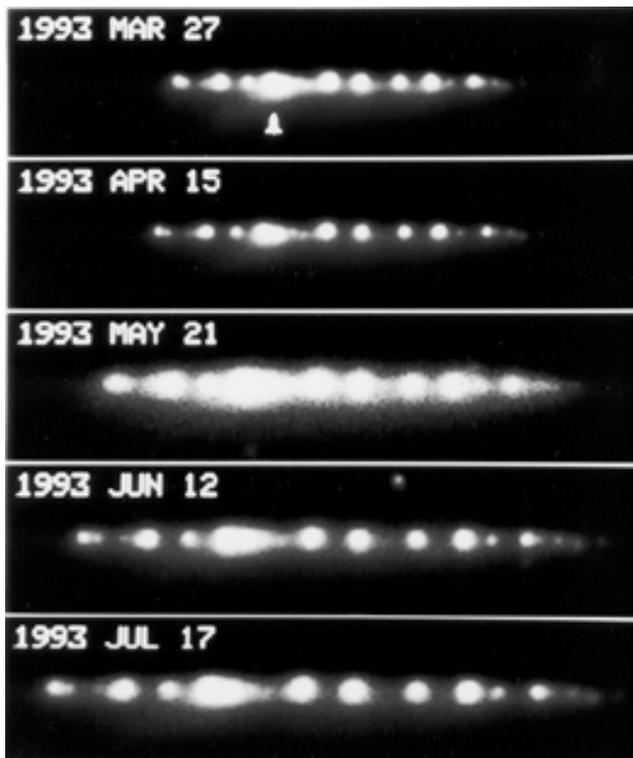


Photo series traces evolution of “the string of pearls” from March to July, 1993. As the string got longer and longer, observers counted as many as 21 nuclei including the very faintest.

Schmidt telescope camera on Palomar, we cover, on average, an area equal to about one-tenth of the of the heavens on each observing run.

### Light-struck film revealed secret

We had rotten luck with the weather in the winter of 1993. We got about one good night out of seven in January, about one hour out of seven nights in February, and had accumulated a large supply of hypersensitized film. This supply of film was accidentally exposed to light and only marginally usable, but we decided to use some on March 23, 1993 when the sky was partly overcast and we didn't want to squander good film on a bad sky. One of the exposures we took happened to have Jupiter in the field. When I developed it, I thought it was a bad part of the film with an artifact on it, but it was really Jupiter. A day later, Carolyn, after she had finished scanning a set of films we took on the first good night of our observing run, began to look at the light-struck film of March 23. She had always been lucky in discovering comets while scanning films taken from near Jupiter. She sat up in her chair, peered intently into the microscope and said, “I don't know what I've got here, but it looks like a squashed comet.”

What she saw looked like a fuzzy object floating above the stars. It was four degrees away from Jupiter. The rest of us rushed over to have a look. We reported it to the Minor Planet Center by e-mail, but we only had small-scale images and by now it was snowing—no chance of another observation. Then David Levy remembered that a close friend of ours here in Tucson, James Scotti, a young astronomer at the University of Arizona, was going to be on the Spacewatch telescope at Kitt Peak that night. Scotti has been an important member of Tom Gehrels' Spacewatch team, which has a program to search for planet-crossing asteroids. Jim was very skeptical when we called him; he commented that “it sounds like you've got an artifact” due to the reflection of stray light in the telescope.

Carolyn was sure that it was a comet, but it was a comet like no one had ever seen before. It had this funny elongated shape. Most comets are nice, round, fuzzy objects. Doubts aside, Jim agreed he would look at the object, and was able to do so since the storm we were under had not yet reached southern Arizona. He made a series of images, and meanwhile we made precise measurements of our images. A couple of hours later we called Scotti up again, and we had an obviously excited guy on the other end of the line. “Have you guys ever got a comet!” he exclaimed. We now had our confirmation.

Scotti immediately notified the Minor Planet Center in Cambridge, Mass. of his observations, and the next day an electronic message went out to many observatories around the world announcing the comet. Scotti came down off of Kitt Peak and one of his images was posted on the bulletin board at the University of Arizona's Lunar and Planetary Laboratory. There was a lot of excitement, hubbub and discussion, and the next night Wieslaw Wisniewski went up on Kitt Peak and made an image with a 90-inch telescope. In this image you could see about 11 discrete nuclei in the comet. The following night, another image was made with the 88-inch scope on Mauna Kea by Jane Luu and David Jewett, who gave the comet the name “string of pearls,” and followed it for four months. From our view on earth, the string of pearls got longer and longer. Jewett first counted 17 discrete nuclei. Later, by including the very faintest nuclei, Jewett counted 21 that persisted for four months.

The discovery of this squashed comet, a comet that had been catastrophically broken up into about 11 sub-equal nuclei plus a constellation of smaller fragments, was itself an unusual discovery; no one had ever seen a comet broken up this way before. About two weeks after the discovery, Brian Marsden, director of the Minor Planet Center, announced that the comet was in orbit around Jupiter. We guessed that it had gotten too close to Jupiter and had been broken up by the planet's tidal forces. This turned out to

be correct: the real surprise was that the comet was in orbit. Two months after the discovery, another hundred or so observations were in hand, sufficient to refine the orbit, and on May 22 Marsden sent out an electronic circular stating that there was a good two-thirds chance that the comet was not only in orbit, but was going to hit Jupiter on its next close approach to the planet in 1994.

### Comet impacts on Jupiter

When we got the news that there was probably going to be a hit, I was in the darkroom and I could hear David Levy reading this message from his portable computer. David says the next thing he heard was a lot of slamming of lids in the darkroom, and I rushed out to see what he had on his screen—literally pushed him aside so I could read the message—and was ecstatic that we were actually going to see an impact. Carolyn's reaction was different: "Oh no, not *my* comet!"

Our initial enthusiasm was tempered by the calculations which indicated that the comet was going to hit on the back side of Jupiter. But fortunately, as time went on and the orbits were refined, things changed. The calculated impact points kept marching toward the limb, but didn't quite make it to the front side. The final impact sites were just around the edge of the planet as we saw it from Earth. The next step was to figure out what would happen on impact. A number of people, (in Tucson, at Caltech, at the Sandia National Laboratory, at NASA, Ames and elsewhere) were trying to calculate what would happen when the comet hit. I joined forces with colleagues, experts on nuclear explosions working for the Department of Defense, to carry out a separate set of calculations. A one kilometer diameter body with one gram per cubic centimeter density, close to the expected density of the comet, would plunge into the atmosphere and start to come apart. We calculated some of it would go down 90 kilometers below the cloud tops. Then the shocked material of the comet and Jovian atmosphere would form a fireball that would start to expand and blow back out the entry tunnel. At 60 kilometers per second it would take less than two seconds to pass below the cloud tops to where it would completely come apart and start to expand. In five minutes the expanding fireball would rise to a height of about 800 kilometers above the cloud tops before collapsing in a hot "splash."

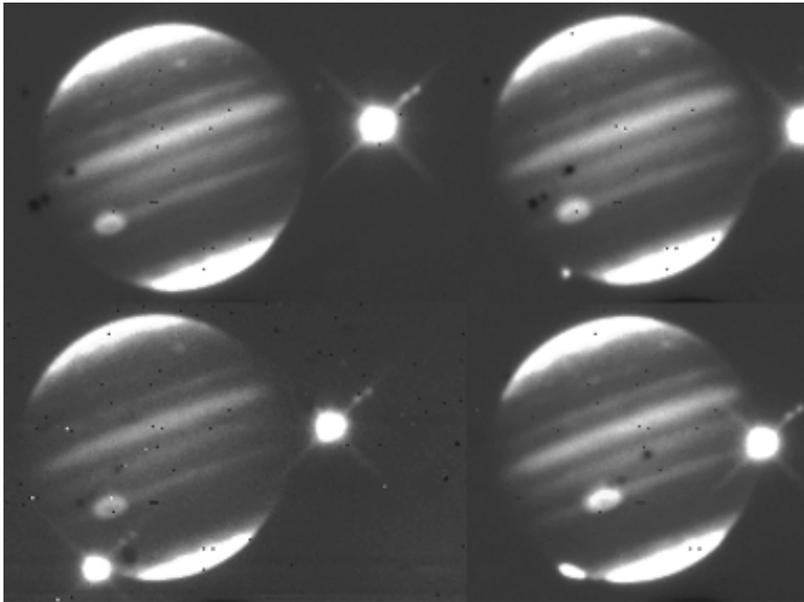
This was the first set of calculations that carried out the dynamic behavior of the plume for this long a period of time. I had those figures with me, and was able to present them on the evening when nucleus "A" hit Jupiter. In fact, NASA had set up a press conference that evening, and we knew that something had been seen in Spain and Chile, but we didn't know just what. We hadn't seen the data



Dr. and Mrs. Shoemaker view the impact of the first fragment at the Space Telescope Science Institute. Carolyn Shoemaker wears a "string of pearls" —a gift from her husband in celebration of the comet she discovered.

from the Hubble space telescope, and I'd gone through this very sequence of slides presented here, while downstairs the team from the Hubble was anxiously receiving the data. We were all holding our breath because we really weren't sure how big these nuclei were. I had no sooner finished going through my litany, however, when Heidi Hammel, the principal investigator for the Hubble images of Jupiter, came marching into the press conference auditorium with a huge grin on her face. She was waving a picture taken by Hubble about 90 minutes after the impact. Unbeknownst to me, the team had all been tipping on champagne downstairs, and she brought up a half-empty bottle of champagne, and we celebrated with her.

The Hubble images taken of the limb just after impact showed not only the portion Jupiter that is dark, but also part of the shadow of Jupiter cast out into space. And they showed the top of the plume rising above the shadow of Jupiter in three minutes. Three minutes later the plume had risen to about its maximum height. It hovered there and then it started to spread out, fall down and flatten out into a pancake, as our calculations predicted it would. I felt pretty happy when I saw that, but the calculations weren't quite that good. It turns out that the height of the plume was higher than we had calculated by a substantial factor; it actually rose up to about 3000 kilometers above the cloud tops. Clearly what happened was that the comet nucleus had come apart higher in the atmosphere than we had calculated; more energy was getting into higher regions of the atmosphere. That allowed the plume to rise higher than predicted in our initial model.



Time-sequence photo shows the impact of the first fragment of Comet Shoemaker-Levy 9 on Jupiter, seen at the lower left of the planet. The fireball flamed to its maximum brightness in the third frame—the total time elapsed between impact and fading was about 20 minutes. The bright object to the right is one of Jupiter's moons, Io; the fainter oval in the southern hemisphere is the Great Red Spot. This image was taken from Calar Alto Observatory in Spain.

The entire "string of pearls" took five and one-half days to hit Jupiter. Meanwhile, besides Hubble, many telescopes around the world observed these events. The impact of nucleus "G," one of the brightest, was observable from Hawaii to Australia. A telescope at the South Pole was able to observe a majority of the impacts. The tremendous flash of heat produced when the impact plumes fell back down on Jupiter's atmosphere was well observed by these ground-based telescopes. Another method of observation was with spacecraft; the Galileo spacecraft, on its way to rendezvous with Jupiter, was able to look directly at the impact sites. Whereas from earth the impact sites were slightly below the horizon, you could see the impacts directly from Galileo.

A lot of radio astronomers were watching Jupiter during this time period. During impact week the decimetric radiation from Jupiter gradually built up, rising to a crescendo of about twice its usual value by the end of the week, and then faded away. Evidently the impacts produced more charged particles spiralling around magnetic field lines in the Jovian magnetosphere. An interesting observation was captured about 45 minutes after the impact of nucleus "K;" two auroral spots appeared in the northern hemisphere near the longitude of the impact site; evidently the impact event triggered the deposition of charged

particles from the magnetosphere at lower than normal latitudes (the normal aurorae are higher).

When the impact sites were observed in the ultraviolet with the Hubble Space Telescope, we saw much more prominent features than were seen in visible light; the dark aerosol clouds were exceedingly dark in the ultraviolet, and that's one of the clues to their composition. In visible light, the clouds were a dark, reddish-brown color. These colors are characteristic of complicated aromatic hydrocarbons, the kinds of compounds that are made in the laboratory by passing electric discharges through a reducing mixture of hydrogen, ammonia and methane. The real question is, was the dark stuff derived mainly from highly shocked constituents of the Jovian atmosphere or did it come mainly from the comet itself? We know that Comet Halley has the right starting constituents to make some of this kind of "organic gunk" as we call it, a

messy mixture of hydrocarbons and other things like amino acids in trace amounts.

### Evolution of the impact clouds

In the days following the impacts, the dark clouds were smeared out by upper atmosphere winds. For the first time we had visible tracers in the stratosphere. It was the first experiment, so to speak, where nature provided a means to study the circulation pattern in the upper atmosphere of Jupiter.

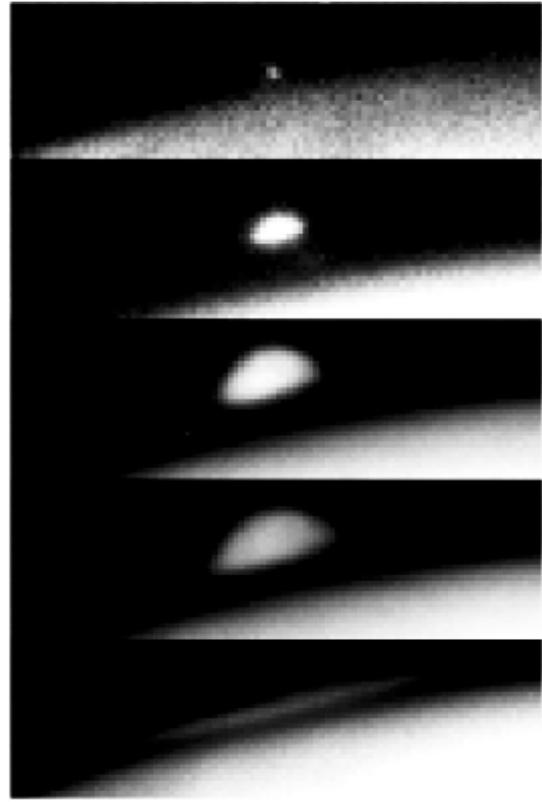
One of the most important things has been to study the long-term evolution of these clouds. Hubble images of of the "D"-"G" impact cloud complex showed how it was gradually smeared out by the upper atmosphere winds over the first month. The optical depths of these clouds were several, which means that more than 90 percent of the sunlight was absorbed as it went through. Optical depths can be measured by looking at features in the underlying normal ammonia cloud tops. The dark stuff was thinning out but was still there when Jupiter disappeared into the glare of the sun in September 1994. When Jupiter emerged from this glare in January 1995 there was a more or less continuous dark band still there, but not as dark as it had been. How long would this band remain?

On Jupiter the material was being spread out over an area enormously larger than Earth, and it was still fairly dense months after the impact. If a single fragment like one of the larger nuclei of Comet Shoemaker-Levy 9 were to strike Earth, and if the constituents of the outer dark cloud are largely from the comet (that's a big "if," we don't know that), the impact probably would produce a dark cloud covering the entire earth. Such a global cloud would

have a profound short-term impact on the climate. In fact, it would dramatically lower the temperature all over the earth for a period of some months. It could produce an impact "winter" if it happened in the summertime of the northern hemisphere, the kind of catastrophe that could lead to the loss of much of the world's crops and possibly the loss of a significant fraction of the population.

These events on Jupiter have been our first direct indication of what the global effects would be of impacts by objects like Comet Shoemaker-Levy 9. Our best guess is that the largest dark clouds were produced by comet nuclei that were no bigger than about one kilometer in diameter. If we've learned anything from this, it is that climatic effects of such impacts on earth might be even more severe than we've thus far imagined.

Right: A plume from impact of Fragment G, partially obscured by the shadow cast by the planet's limb, rises high above Jupiter before spreading out. The plumes provided visible tracers, a means to study the circulation pattern in the planet's upper atmosphere.





*Robert G. Strom specializes in planetary geology and has been teaching and doing research at the University of Arizona since 1963. Beginning with the Apollo missions to the Moon, Prof. Strom has been a member of many NASA science teams including the Mariner missions to Venus and Mercury, and the Voyager missions to Jupiter, Saturn, Uranus and Neptune.*

*His present research interests span the solar system and utilize data from many planetary spacecraft: ancient oceans and ice sheets on Mars; the impact cratering record on solid bodies in the solar system, and the surfacing history of Venus. Along with his current teaching and research duties at the University of Arizona Lunar and Planetary Laboratory, he is also a member of the science team for the proposed Messenger mission to Mercury.*

## Ancient Oceans and Ice Sheets On Mars

Adapted from a 1998 presentation by Robert G. Strom

**A comparison between** Earth and Mars is a good place to start in making the case for ancient oceans and ice sheets on Mars. Mars is about half the size of Earth, and receives about 44 percent as much sunlight. Mars is, of course, further away from the Sun than Earth, and the surface temperatures are very cold, about minus 180° or so on average. Mars' rotation time is similar to Earth, about 24-1/2 hours, and its axis of rotation is tilted at about 25° which is also similar to Earth. So Mars does have seasons, but they are about twice as long as Earth's because its orbit around the Sun is about twice as long.

Surface water, which has implications for life, existed on Mars between half to a billion years ago; in the recent past, geologically speaking. Another provocative item is that Mars is the only planet in the Solar System with features similar to those on Earth. It is very Earth-like in its geology and the distribution of certain land features, some of which we will discuss today.

Another similarity to Earth is that Mars has ice caps, but they are mostly frozen carbon dioxide. Mars' atmospheric pressure is only 7 millibars, a very tenuous atmosphere, made up of over 90 percent carbon dioxide which freezes in the polar regions during the winter. There is also water ice: when the CO<sub>2</sub> ice melts there is a residual ice cap of H<sub>2</sub>O. Presently these ice caps are much, much smaller than Earth's, occupying perhaps 2 percent of the surface of the planet.

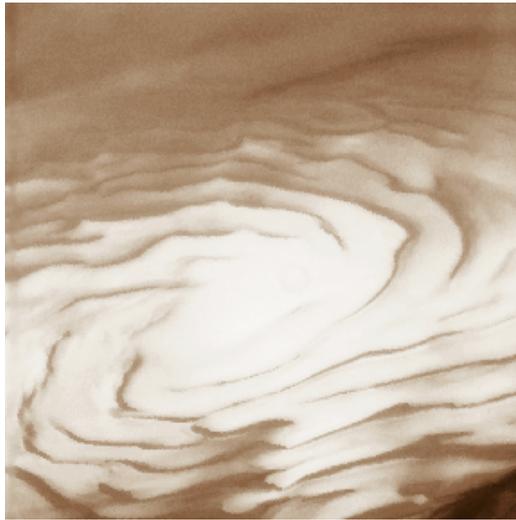
A mosaic picture of Mars from Viking shows the polar caps, and the one at the south is water ice, not carbon dioxide (figure 2). Color photos reveal an orange-red color which is why we call Mars the "red planet." Why is Mars red? The answer is that it's oxidized, rusted. The reason, just as it is on Earth, is water, and in fact Mars is a water-rich planet. Most of that water is located in its subsurface at the present time, and I'll show you evidence of that. Mars is a planet of extremes; it has the largest channel system in the Solar System, the largest volcanoes and the largest canyons. In this discussion, I'm going to focus on those channels.



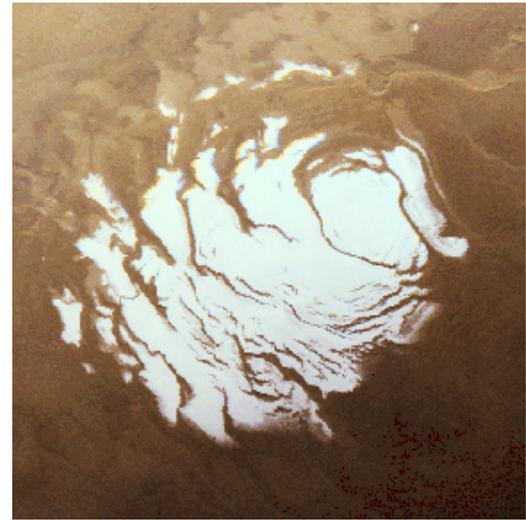
NASA

Figure 1. Mars is about half the size of Earth

Figure 2. North (left) and south polar caps. Mostly carbon dioxide, the ice caps are smaller than Earth's, occupying about 2 percent of Mars' surface.



NASA/LPI



NASA

### Ancient Mars a Watery Planet

A picture of what Mars may have looked like in the past (figure 3), shows a lot of water in the Northern Plains. The Northern Plains are a lower area on Mars, and they average about 2 km below the heavily cratered highland region. A large canyon system, Valles Marineris, may have been filled with water, and there are layered deposits, sedimentary in origin, that may have been laid down by water. One thing that's missing from the picture is probably a large ice cap. The oceans are the likely source of the water in the ice cap, just as they are on the Earth.

At the present time Mars does not have an ocean, but it may have looked something like figure 3 in the past—this is an artist's rendition, and should be taken with a grain of salt. Mars' atmosphere is now too thin for there to be an ocean, so the presence of oceans—and if there were oceans there were probably ice sheets—in ancient times implies an atmosphere that was quite different. There was a lot more carbon dioxide—which you all know is a greenhouse gas—possibly from 1 bar to 4 bars, enough to allow water to be stable on the surface. Certain features that occur only at high latitudes on Mars resemble glacier features on Earth that only occur at high latitudes. Areas on Mars may have looked something like Earth's Alpine-type glaciers, but its ice sheets probably covered a much larger part of its surface.

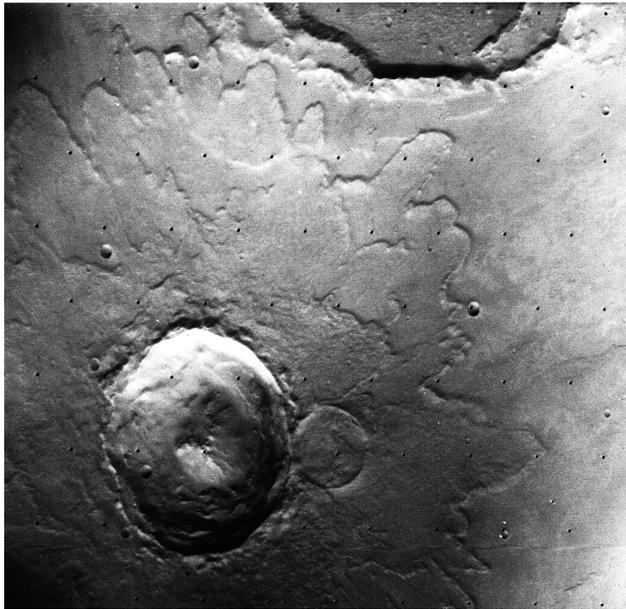
Lower areas in the Northern Plains may have been the sites of oceans. One low circular area is Hellas, a large impact crater about 2000 km across, and I'm going to focus on Hellas and Argyre, another impact crater where we think glacial land forms are present (figure 4). In addition to an ocean, the Northern Plains has a very large plateau called the Tharsis Uplift. It is the site of three major volcanoes, and Mons Olympus, the largest mountain in the Solar Sys-

tem, is associated with the plateau. The canyon system, Valles Marineris is yet another feature associated with Tharsis, and to the east is a region of large outflow channels. These outflow channels occur elsewhere on Mars, but there is a concentration of them in this specific region. All of them, almost without exception, discharge into the Northern Plains.

Geologic and topographic maps show that the Northern Plains are about 2 km lower than the average surface; in addition to the Tharsis Uplift, volcanoes and a large canyon system, there are areas called chaotic terrain. This region is the source of the outflow channels, but the source of the water itself is in the subsurface; these are not runoff channels as we know them on Earth. There are different types of channels on Mars. The big ones are outflow channels and are concentrated in one area. Then there are other types of channels



Figure 3. The discovery of possible ancient shorelines suggests to some scientists that the release of water associated with the formation of outflow channels may have led to the temporary formation of oceans as seen in this painting by Michael Carroll.



NASA

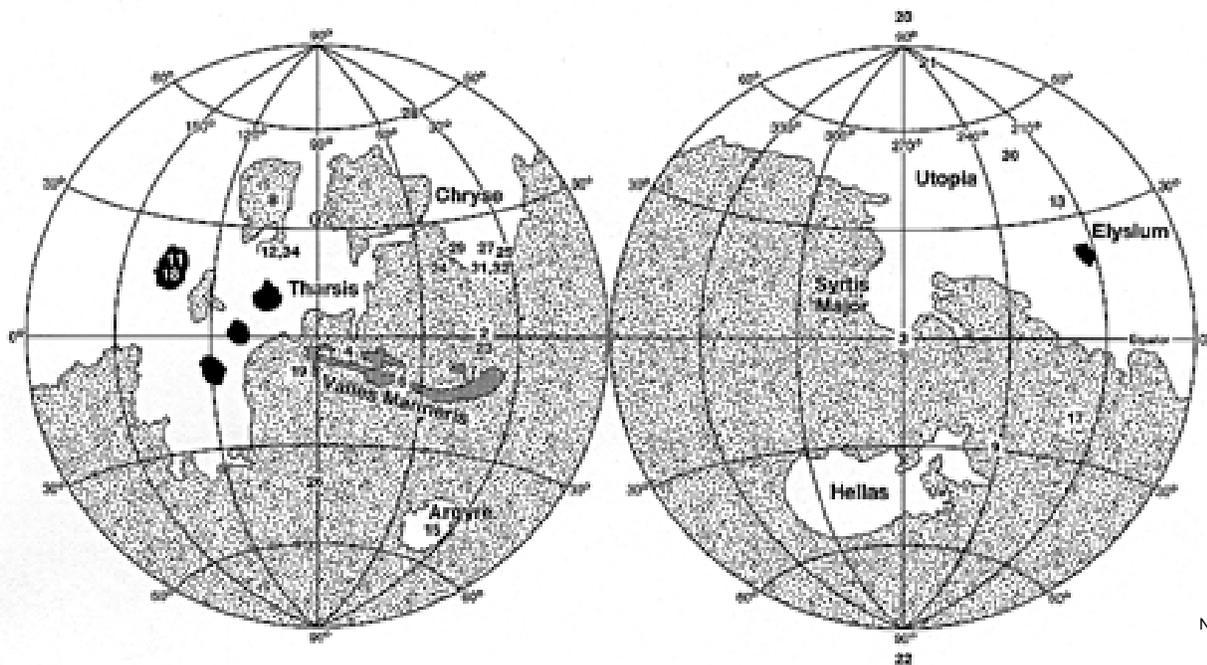
Figure 5. Yuty crater in Chryse Planitia shows the many overlapping lobes that form when an impacting object rapidly melts ice in the subsurface and the ejected material flows along the surface. The crater is 18 km across.

called valley networks, and also different types of valley networks. Valley networks are located primarily in the highland regions and are due to a process called groundwater sapping.

**The Evidence for Subsurface Water**

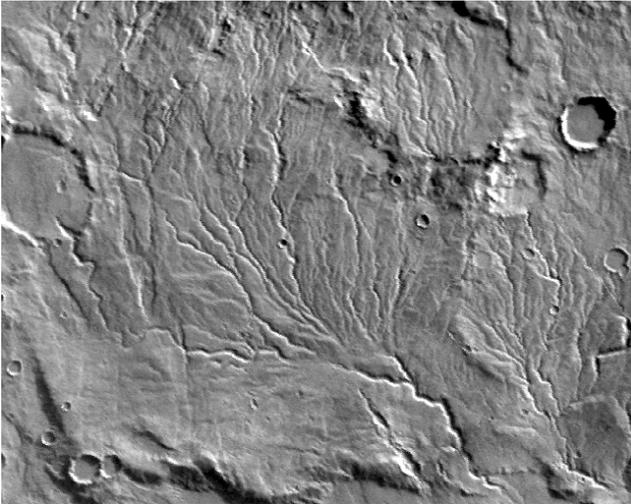
Four main lines of evidence point to subsurface water on Mars. One is flow ejecta deposits around fresh craters. Without exception, every fresh crater that displays an ejecta blanket shows morphology that indicates a flow of this ejecta (figure 5). This is unique in the Solar System, for there's no other planet that shows this type of morphology. It is almost certainly the result of impacts into a subsurface water/ice layer. These flow ejecta deposits are found almost everywhere on Mars' surface, an exception being the summit of Mons Olympus. Here we find ballistic ejecta blankets, very similar to those seen elsewhere in the Solar System.

Mars' unique ejecta blankets are due to impacts penetrating into a subsurface water/ice layer, the mixing of that water with the ejecta, and then flowing out. An impact event would generate heat as well as mechanical energy in excavating a crater; the heat would melt the subsurface ice, and water mixed with ejecta would flow across the planet's surface. Laboratory experiments on Earth demonstrate the same type of ejecta flow in a water-saturated target, but I want to remind you that we see this phenomenon only on Mars.



NASA

Figure 4. Map showing approximate location of major topographical features.



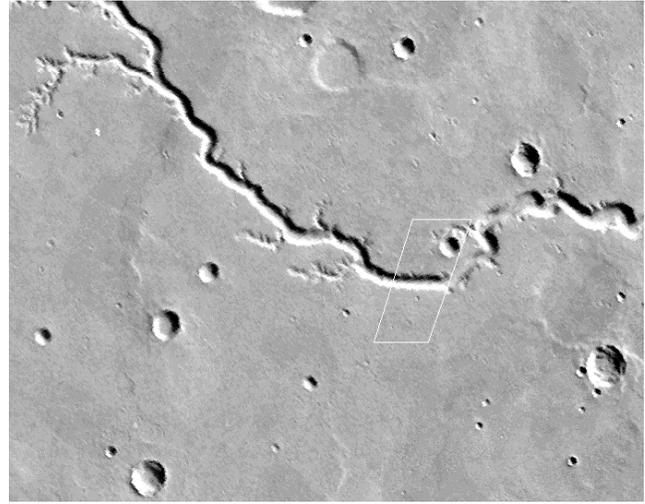
NASA/LPI

Figure 6. Valley networks do not show evidence for catastrophic flooding as outflow channels do. It is thought that they were carved primarily by groundwater flow and so bear a closer resemblance to drainage systems on Earth where water acts at slow rates over long periods of time. The area shown is 200 km across.

In addition to the outflow channels, there are Mars' "valley networks" distributed primarily in the highlands on some volcanoes near the Northern Plains (figure 6). These channels are formed by water in a process called "groundwater sapping."

A view of Valles Marineris shows tributary valleys with blunt ends, what we call amphitheater ends. These are huge things, and for perspective, the Grand Canyon would be a small part of the huge tributary channels to the main canyon. The Earth in the Colorado Plateau has the same type of morphology—amphitheater ends. As groundwater seeps out, it undermines the overlying rock and it collapses to form these amphitheater, blunt-end channels. This is exactly what we see on Mars, and everybody agrees it's due to groundwater sapping. All of this indicates a global distribution of subsurface water, so Mars is, in fact, a water planet—you just can't see it because it's below the surface.

When this was first proposed people didn't understand that there had to be a process to keep the action going. That's a profound statement, because it means that the groundwater table on Mars had to be continually recharged. The way it happens on Earth is simple—it just rains and rains and recharges the groundwater system. How does it happen on Mars? Maybe rain. There's another way of recharge through hydrothermal action, but it's very unlikely to happen on a planet-wide basis, so it may have rained on Mars as well.



NASA/JPL

Figure 7. Valley network. With fewer tributaries than typical terrestrial drainage systems, it is thought that the valleys were carved primarily by groundwater sapping rather than runoff from rain. Water flow in this valley would have been from left to right. This image of Nirgal Vallis was taken by the Mars Global Surveyor from an altitude of 800 km.

### The channels of Mars

One view of a valley network (figure 7) shows tributary channels much like runoff channels on Earth, except that they have the blunt ends and a morphology more similar to channels formed by groundwater sapping. These are concentrated in highland regions of Mars. There is another type of channel with one example that looks identical to the Grand Canyon, and another large valley network shows a quite different morphology, more intricate than others. One channel of this enormous network is about 2000 km long—the Grand Canyon isn't even close in size.

Yet another type of valley network is more like that associated with Valles Marineris; these networks are almost certainly due to sapping, and the amphitheater ends are prominent. The evidence for a flow of water is unambiguous—there's a delta at the end of one which indicates deposition in water. You can see other channels on Mars associated with this deltaic form: there's no question that water flowed down the valley onto the plains. There are channels at the boundary to the Northern Plains as well, very fine channels, very small, and there is also a deltaic deposit. The interesting thing about the delta at the end of one channel is that it has a steep scarp at the edge. The only way such a form evolves on the Earth is if a channel flows into water, indicating that the material was probably deposited in a standing body of water, possibly an ocean or at least a large lake.

The other type of channels on Mars, the really big ones, are called outflow channels as we mentioned previously,

and they originate from chaotic terrain. Figure 8 shows the chaotic terrain and the channels going off in one direction. The source of the water is in the subsurface, and when that water pours out, the surface collapses inward to form this chaotic terrain. The water flows out into the Northern Plains, and we have views which show a portion of a channel with hydrodynamically-shaped erosional forms. Sometimes craters acted as barriers to the flow; the water flowed around them forming a streamlined shape (figure 9).

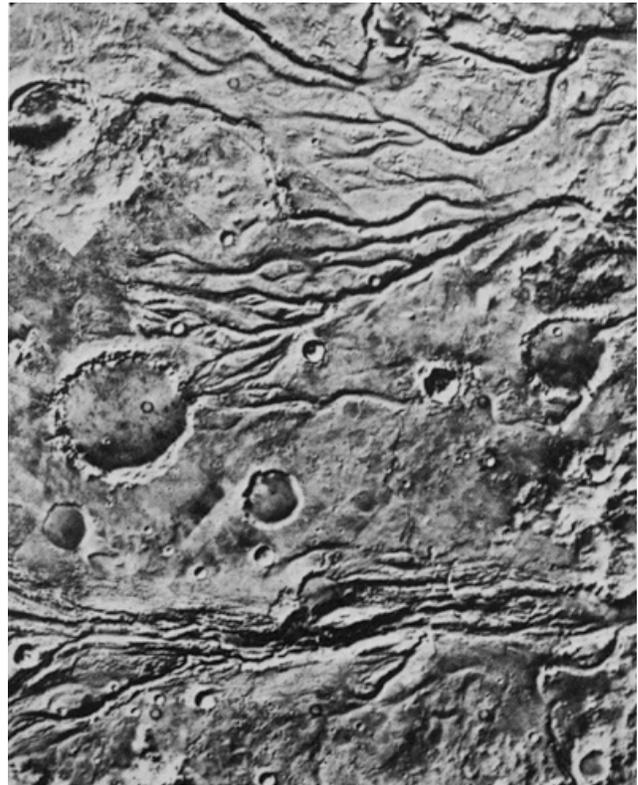
A map of the channels shows the chaotic terrain and the large canyon, Valles Marineris; the chaotic terrain shows up as stippled areas and the channels as flow lines. The channel deposits go out from them. To try to put it in perspective, just a portion of the channel deposits would have completely inundated Arizona and many adjoining states. That will give you an idea of the magnitude of the flow that must have issued through those channels. All of that water was dumped into Mars' Northern Plains.

### Water discharge a "Biblical event"

Consider what that means as far as discharge rates were concerned. You can estimate them from the shape of the channels, their depth and their slope. The estimated discharge rate in this case was enormous, and the estimated discharge rate from the outflow channels in cubic meters per second, not per minute or day or hour, was something like 1-10 billion; 10 billion is probably more realistic. Now for comparison, the discharge from the seven largest rivers on the Earth, the Amazon, Congo, Yangtze and so on, is only 370 thousand. In other words, the estimated discharge rate on Mars was about 3,000 to 30,000 times greater than the combined rate of Earth's seven largest rivers. This was not due to rainfall; this water is all disgorged from a large subsurface reservoir in a catastrophic event. The trigger was probably volcanism, and I'll return to that later.

What does this mean for the size of the Northern Plains ocean? We don't know exactly how big it was, but we can make estimates based on the contours. Zero contour would be similar to sea level on Earth, although there is no sea level on Mars. There are -1 to -2 km contours up in the Northern Plains. At the zero contour, the depth of the water would be about 1.7 km, and the equivalent amount of water would cover the entire planet to a depth of about 450 m. In other words, if you could spread that water uniformly around a flat Mars, it would cover the entire planet to about 450 m. The -1 contour would be about 1.1 km deep, -2 to about 0.7 km and so forth.

If you can estimate the discharge rate, you can also estimate the time it took to fill the Northern Plains. Based on the contours, the filling rate would be 1 billion at the bare minimum, and 10 billion would be the more realistic. At the zero contour interval filling would take about 11 weeks;



NASA/LPI

Figure 8. Outflow channels emptying in to the Northern Plains of Chryse Planitia. Channels flowing out of chaotic terrain with abrupt beginnings and no tributaries suggest that the water that carved them was released under great pressure from underground.

Below, closeup of outflow channel originating from chaotic terrain. In outburst flooding the enormous discharge from the collapsing aquifer results in chaotic terrain.

NASA/JPL/MSSS





NASA

Figure 9. Craters and other landforms were obstacles to powerful floods and consequently experienced modification as water was channeled around them. Flow ejecta can again be seen around the crater to the right.

at -1 it would take five weeks, and at -2 it would be ten days. It was a truly biblical event; remember these discharge rates are the bare minimums and probably unrealistically low. Filling could have taken anywhere from 11 weeks to ten days. It was a catastrophic release of water, huge walls of water, and there has been nothing comparable on Earth.

A drawing of the probable size of the ocean shows a line that has been interpreted to be a shoreline. It was proposed by two groups of people, one at the USGS in Flagstaff, the other at the JPL in Pasadena. A second line has also been interpreted as a shoreline; in other words there were probably at least two oceans at different times, a large one earlier and a smaller one later. They may also have been episodic; at different times in Mars' past there may have been large outflows that formed oceans of various sizes. There's good evidence, based on the crater density, that these channels did experience multiple episodes of flooding.

### Ice sheets on Mars

Now I want to talk about the other side of the coin, ice sheets on Mars and the evidence for them. When you look at features in the southern hemisphere at 45 degrees latitude, you see floor features and some sinuous ridges which we'll explore in detail. There's a large impact crater called Argyre, about 1000 km in diameter; the interesting thing

is that the southern rim has a morphology that is very different from the northern rim, indicating that some process has affected the southern rim and not the northern. This, in a nutshell, is the glacial geomorphic associations that we think we see on Mars. These are totally different land patterns that are all traceable to glaciers, and we think we recognize them in Argyre. There are more of these patterns east of Hellas, more in Hellas and in the Northern Plains. There were probably two ice sheets, one in the north and one in the south.

Back on Earth, the Wentworth Glacier in Alaska shows landforms that have analogues on Mars. The glacier has striations and scours which are called fluted drifts. They are scours on a glacial deposit called a ground moraine, and are caused by the movement of the glacier. Another feature is a ridge which is called an esker; it is due to the flow of water underneath the glacier. This flow when the glacier retreats leaves a ridge because sediment is packed up around the sides and floor of the glacier.

An outwash plain includes holes filled with muddy water which are called kettles. Kettles form as the glacier retreats, leaving behind chunks of ice buried in the ground moraine. When that ice melts the holes are filled, on Earth, by lakes. You've all heard of the land of 10,000 lakes in Minnesota; it's actually the land of 10,000 kettles, all from the last ice age, the Pleistocene. Finally, there is a ground moraine, a mound-shaped sediment deposit left behind by a glacier as it recedes. There are various types of moraines, but that's Geomorphology 101.

A high-resolution picture of the sinuous ridges in Argyre shows what we think are eskers (figure 10). They have a geometry which is identical to eskers on Earth, and their sizes are not too dissimilar. The surrounding mountains have extremely sharp crests which may be what are called arêtes caused by glaciation; they are very common in the Alps. The northern rim of the Argyre crater is very rounded and subdued compared to the southern rim. There's quite a difference in morphology, so something has shaped the southern rim that has not acted to the north, which is at lower latitudes. We think the thing that did the shaping was ice.

Most people in the planetary community believe they see rock glaciers in the photos. These are just sloughing off of the mountains and slowly creeping down onto the plains. Rock glaciers also occur on Earth, usually associated with glaciation areas, and a lot of them have recently been cored. To the surprise of everybody the cores of these rock glaciers are ice. Are the cores of those on Mars ice as well? Maybe.

A detail of some of those sinuous ridges shows an interesting thing; they often anastomose—divide and join again. This is characteristic of eskers and is very difficult



Figure 10. These sinuous ridges in Argyre may be glacial landforms known as eskers. (Arrow does not indicate eskers.)

to explain by anything other than flowing water. It is uncharacteristic of sand dunes; they don't anastomose. Hydrodynamic processes do anastomose on Earth and, presumably, also on Mars. These ridges were laid down by water, but they're high. The best explanation we've got is that these are, in fact, eskers. Along the top there's a trough, and this is also characteristic of eskers on the Earth.

Let's compare the geometry of some sinuous ridges on Mars, those in Argyre, with the pattern of eskers on Earth. In fact you can quantify this to a degree by measuring the amplitude of the meanders against the wavelength. When you do this for Earth eskers and for Martian sinuous ridges, you find they're identical.

Another feature which is very interesting are little mounds with a crater at the top. What are these things? They're certainly not volcanoes, but they may be pingos, another glacial feature. Pingos are very common in the arctic regions of Earth. The ground ice expands and forms domes on the surface; the centers then collapse into craters. Mars may have pingos, ground ice residuals from past glaciation.

**Outwash plains from water flow**

I mentioned outwash plains and there are some in Argyre. There is an impact crater that has been breached,

and one can see the hydrodynamic bed forms that are certainly due to the flow of water. Where did the water come from? Everything points back accusingly to the large crater. It was probably filled by part of a glacial ice sheet that melted, and the melt water flowed out of the breach. There's no chaotic terrain in the crater, and perhaps meltwater from an ice sheet flowed out and kept going to form a visible hydrodynamic bedform, the Martian equivalent of an outwash plain. There is a terrestrial example in Iceland, an outwash plain with kettles.

The distribution of the various Martian features that we interpret as glacial are an arête area, eskers, scoured terrain, a kettle field, and the outwash plain. These occur in an ordered sequence as eskers, scoured terrain and kettles; a similar sequence occurs on glacial terrain on Earth.

The Hellas Basin on Mars is over 2,000 km across, and the topographic contours are from 4 to -5, so it is a deep area. There's a huge difference between maximum and minimum elevation, so it's very steep in some regions. If you superimpose on that some features that we see in Hellas, you detect scours, grooves cut in. The steepest part is more scoured than down below, and there are ridges which may be eskers, are there are channels. There are teardrop-shaped mounds which may be like drumlins on Earth—shaping the moraine by movement of the ice. There

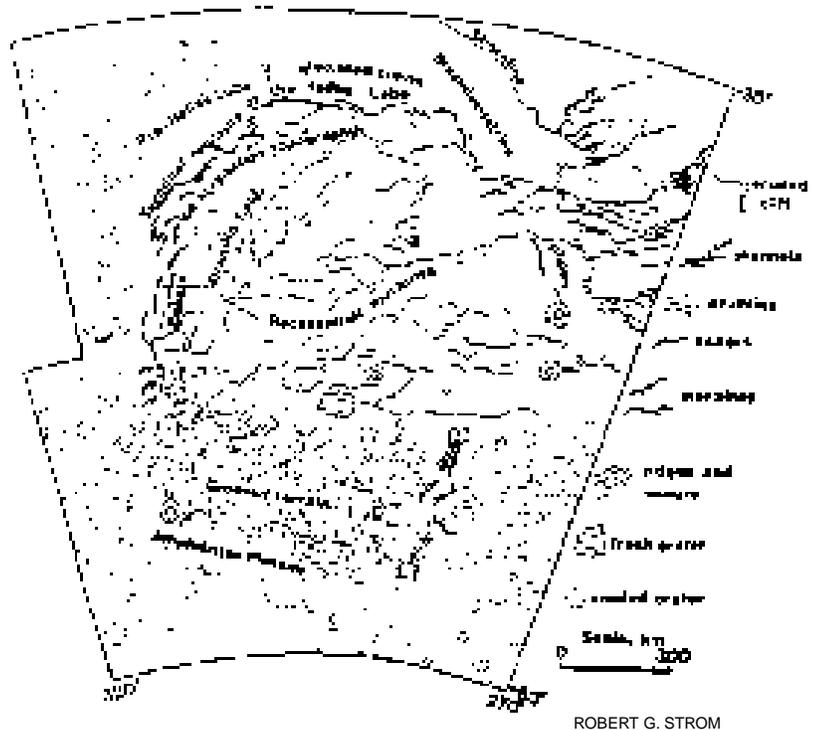


Figure 11. A geomorphic map of Hellas showing the sequence of glacial features.

is an area which was probably a proglacial lake, and ridges which we think are terminal moraines. So there's a whole ordered sequence of geomorphic forms which are identical to glacial forms on the Earth, and which have the same ordered sequence (figure 11).

You can construct a drawing of the way a glacier might have flowed on Mars showing the terminal moraine and what would be called a proglacial lake on Earth. It would be enormous, not a glacier, but an ice sheet. One can see a dark line that would have been the northern limit; we don't know what most of the forms are at the south pole because we don't have high resolution imaging of this area yet; perhaps Global Surveyor will fill that in. Given the extent of the ice sheet in the southern hemisphere, however, it would have covered 10 to 20 percent of the surface of Mars. We think the source of the ice was the ocean, as on Earth. These events can be dated on a relative time scale by the distribution of craters on these glacial plains and also in the ocean. The relative dates of the features on the glacial plains and in the oceans are the same.

I want to go for a minute to the northern hemisphere. Dark areas seen there may be glacial forms, and there is the possible ocean and a series of channels in the Northern Plains. The complexity of these channels makes them very unlike rainwater channels or sapping channels on Earth or Mars. They have ridges down the center, and we think they are what's called tunnel valleys, also a glacial land form. A picture of a tunnel valley in Sweden, filled with water and with an esker down the center, is similar.

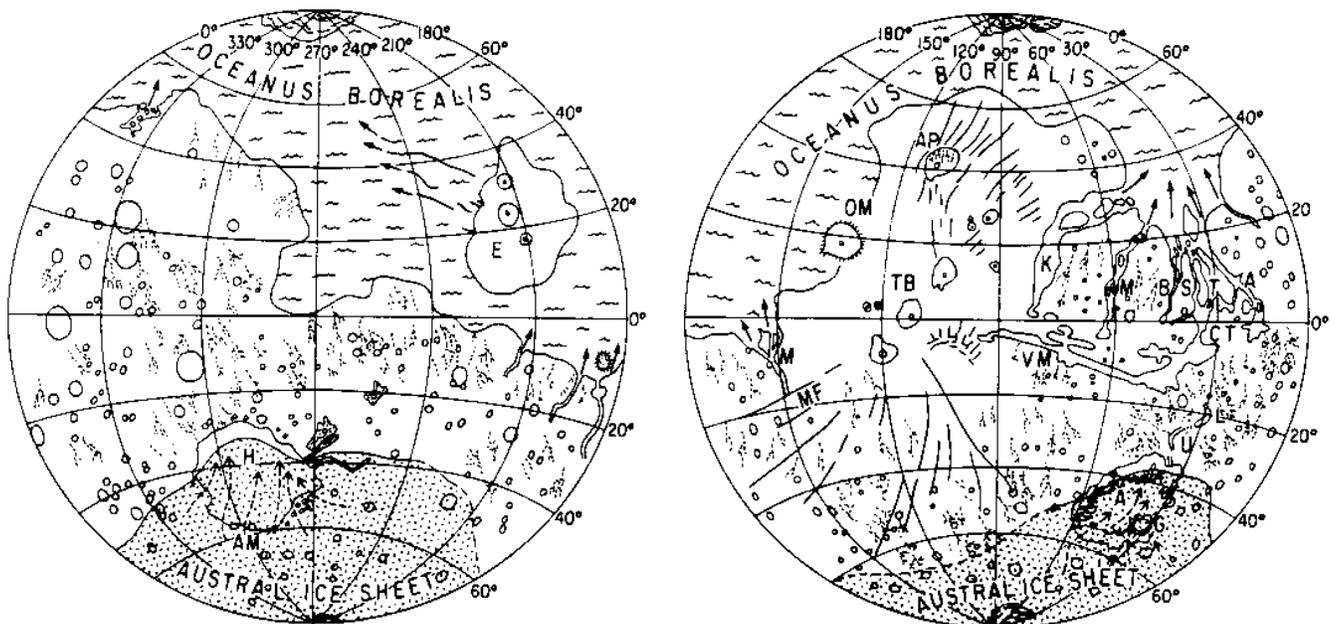
We think that's what we're seeing on Mars—tunnel valleys with eskers down the center. Next to these features are ridges, fairly closely spaced, and it's likely they are the moraines.

**Moraines point to ocean site**

I'm not a glacial geomorphologist and neither are my coauthors, so we called in an expert. Some of our interpretations of the glacial features on Mars are controversial, but when we showed him these pictures, which we thought were moraines, he said they were not only moraines, but *De Geer moraines!* De Geer moraines are formed in water when a glacier flows into a lake or ocean. Remember, this is the Northern Plains, so what have you got? Perhaps the site of an ocean. It was a blockbuster of a discovery.

A map of that region shows the possible De Geer moraines and the tunnel valleys. This is in just one area in the Northern Plains, but wherever you see this type of valley, you always see ridges in the center of them, and you always see these types of moraines. Their morphologies are similar to those on Earth. A region of tunnel valleys up in Canada from the last ice age shows an intricate pattern very similar to what we think are tunnel channels on Mars.

Some researchers, Henry Moore and others from the Pathfinder team, believe they've identified conglomerates on Mars. Some rocks may show numerous sockets where pebbles resided, protrusions which look very round, and other cobble-like things that may have fallen out of the sockets. A conglomerate is made up of rounded pebbles,



ROBERT G. STROM

Figure 12. Maps of Mars showing the past extent of the Northern Plains covered by ocean, the austral ice sheet, flow directions of outflow channels (arrows), and glacial landforms (dashed arrows). The dashed lines indicate the general pattern of valley networks.

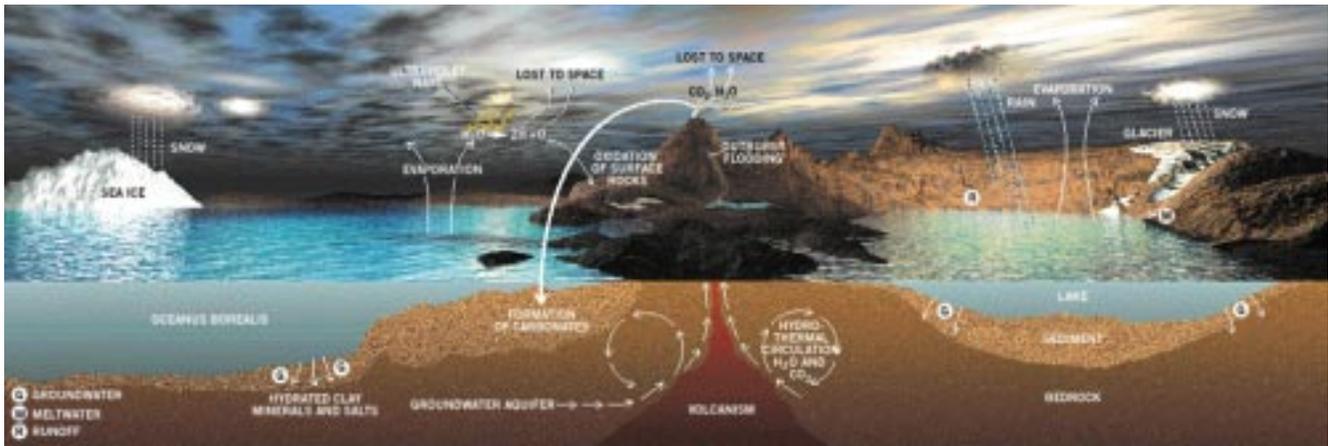


Figure 13. Model of the water cycling mechanism which may have formed oceans and ice sheets on Mars. Volcanism and resulting heat flow caused hydrothermal circulation of groundwater and triggered outburst flooding of water and carbon dioxide. The resulting water vapor and  $\text{CO}_2$  in the atmosphere would have precipitated to form snow, glaciers and ice sheets at high latitudes, and rain at lower latitudes that recharged the aquifer. (Illustration from “Global Climatic Change on Mars,” by Jeffrey S. Kargel and Robert G. Strom. Copyright ©1996 by Scientific American, Inc. All rights reserved.)

and that rounding can be due to one of two things: one is the flow of water. You could get a conglomerate of these rounded pebbles due to stream action, and remember these are at the mouth of an outflow channel.

The other way rounded pebbles are formed is by glaciation—and you can form glacial conglomerates as well. Where do these rocks come from? They were probably washed down from the highlands. One guess is that they were formed in the highlands’ valley networks by stream action, which means that there had to be running water on Mars. You don’t have that today, so that would mean the atmosphere was different. Or conglomerates may have been formed by glacial action.

A drawing of what we think Mars looked like during one of these wet and warmer epochs with the ice sheet helps sum up (figure 12). It shows the Austral ice sheet, and there was probably a second that produced those northern features. We’re calling one of the oceans Oceanus Borealis, the northern ocean. It filled primarily the Northern Plains, and we can see the outflow channels which produced that hypothetical ocean. There are little lines marking valley networks where there has been some fluvial action.

### Outburst flooding triggered by volcanic activity?

I’ve been describing the model that we’re proposing, and remember that all of this may have been episodic. Mars’ features are ephemeral, they come and go depending upon mechanisms like the outflow channels. The model appeared in the November 1996 *Scientific American*. It is a model of the mechanism by which we think

the oceans and ice sheets formed (figure 13). We think that it was triggered by volcanic activity: the volcanism came up in the subsurface and reacted with the groundwater. What happened when hot lava came in contact with groundwater? It started boiling, and that may have led to the outburst flooding.

There’s also a geophysical model that says that mantle overturn could have resulted in episodic volcanism extremely intense and regional in nature, and that it may have been the trigger. Episodically it would have led to outburst flooding, produced oceans, and evaporated them. At the same time this stuff came out, probably what came with it was  $\text{CO}_2$ . In other words, it may have been more like a soda than like the water we know on Earth. If that was the case, a lot of  $\text{CO}_2$  was released into the atmosphere; some models predict that there could have been from one to four bars of  $\text{CO}_2$  released by this process. That would warm the atmosphere so you could have had standing bodies of water, but it would still be cold, so there might have been evaporation of  $\text{H}_2\text{O}$ . It might have rained, in fact, in low latitudes; snowed in high latitudes to produce ice sheets in both polar regions. Eventually, the precipitation would have seeped into the ground to become groundwater again.

A short synopsis of the evidence shows that there are diverse features at high latitudes. By high we mean greater than  $\pm 35$  degrees from the equator. These features have morphologies, geometries, scales and ordered distribution. They’re virtually identical to glacial features formed on Earth during the Pleistocene ice age. Second, there are shorelines and other marine features that have been iden-

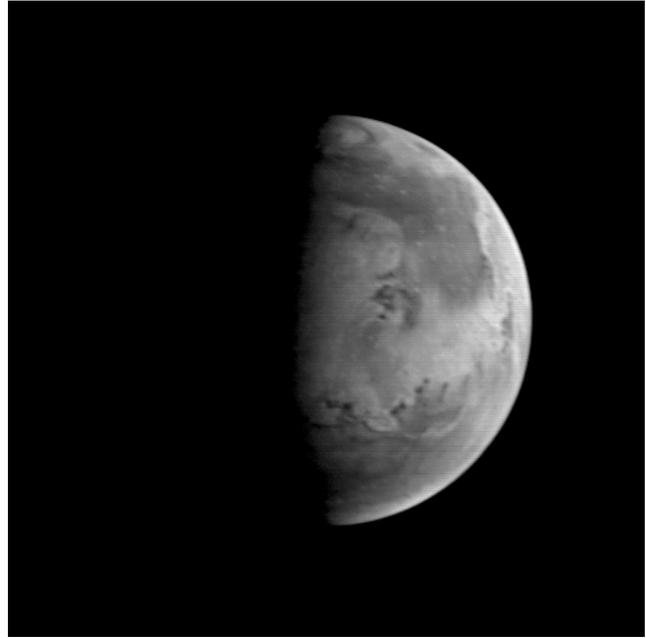
tified, not just by us, but by others, particularly the USGS and JPL. They occur at the margins of the Northern Plains which suggests that there was a shoreline and in fact an ocean there.

The ice sheets probably covered about 25 percent of Mars, and the oceans probably filled the Northern Plains relatively late in Martian history; we don't know how late, but perhaps a billion to 200 million years ago. All of this may have been episodic. Climatic conditions during these times had to have been vastly different than they are today: denser atmosphere and temperatures high enough to allow liquid water to exist on the surface of Mars.

The water originated in the subsurface—that's important. It was not rainwater. It existed in the subsurface, and it may have been released in catastrophic flooding caused by volcanism heating that water. There's no question that there were large amounts of water released catastrophically. A little of this water was lost to space, maybe 5-10 percent, but the rest probably still exists, residing in the subsurface.

How did the water get to the subsurface? Mars is heavily cratered, and when you form craters you fracture the surface, so that surface is probably very porous and permeable. As a consequence it's likely the ocean didn't last very long; in fact we'd have trouble explaining what we see today if it lasted for many millions of years. There would have been so much erosion that we wouldn't see any craters. The water resided on the surface for a length of time we're not sure of, and then disappeared, probably into the subsurface.

There's some recent evidence for this: one piece is the conglomerates from Pathfinder that I mentioned, and another is an extremely flat Northern Plains. This information came from the Global Surveyor within the past couple months. They did a profile through the Northern Plains and up onto a volcano, and got excited about the volcano and the implications for a flat Northern Plains. The surprise is that they are extremely flat, flatter than any surface anywhere in the Solar System except some places on Earth. A flat surface, like Lake Bonneville, is produced by sedimentation from standing bodies of water. This very



NASA/JPL

The Mars Orbital Camera took this picture in August 1997, when the Mars Global Surveyor was 3.52 million miles away. Immediately to the south of the large dark marking stretching from the right center, is the brighter area of Chryse Planitia. Many outflow channels are located here, as well as the Viking Lander I and Pathfinder landing sites. New images and data from the Global Surveyor will greatly contribute to the identification of water- and ice-formed features on Mars.

flat surface that we found on the Northern Plains is further evidence that they were caused by sedimentary deposits laid down by a large body of water.

That's the evidence; some of it, not all of it. We think it is very exciting, but I must say it is still controversial because it paints a picture of Mars that is completely different from the one we had ten years ago. But our picture has great implications for life: if life started on Mars we can gain insight as to its origins, and begin to answer the basic questions that we've been asking for hundreds of years. How did life originate? To understand life on Earth, go to Mars.



*Eugene M. Shoemaker started his career with the U.S. Geological Survey in 1948. He investigated meteorite impact craters, the structure and history of the moon, and developed geological mapping of the moon and organized the USGS Branch of Astrogeology in 1961. Dr. Shoemaker received master's degrees from the California Institute of Technology and Princeton University, and his Ph.D. from Princeton.*

*In a long history with NASA, Dr. Shoemaker was acting director of the Manned Space Sciences Division, a principal investigator for Project Surveyor and the Apollo lunar landings, coinvestigator for Project Voyager, and most recently science team leader for the Clementine Mission. He also remained a Lowell Observatory staff member. Codiscoverer of Comet Shoemaker-Levy 9, his most recent research was on impact processes in the solar system and the effects of large-body impacts on the evolution of life.*

*Dr. Shoemaker died in a car accident in July of 1997 in the Australian outback where he was studying impact craters.*

## The Clementine Mission to the Moon

Adapted from a 1995 presentation by Eugene Shoemaker

**On January 25, 1994** a Titan IIG rocket, built from stages taken out of silos near Little Rock, Arkansas and Tucson, Arizona, took off from Vandenberg Air Force Base. The Titan was used to send a spacecraft into a polar orbit around the moon, our first return to that body in about 20 years. We went back to do a job that it would have been nice to have done before we sent the astronauts there, which was to get the broad reconnaissance needed to understand the context of the scientific results that were obtained from the samples brought back years ago. It was my good fortune to lead a team of dedicated scientists who helped guide the mission and who have made a preliminary interpretation of the extraordinary flood of data.

The spacecraft was relatively modest in cost and size—about 2 meters high. On the bottom was a directional antenna that was used to send the data back to earth. The critical part of the spacecraft was a series of cameras mounted on a beryllium optical bench; this slab of beryllium not only served to align the cameras, but also as a heat sink, particularly for the infrared cameras.

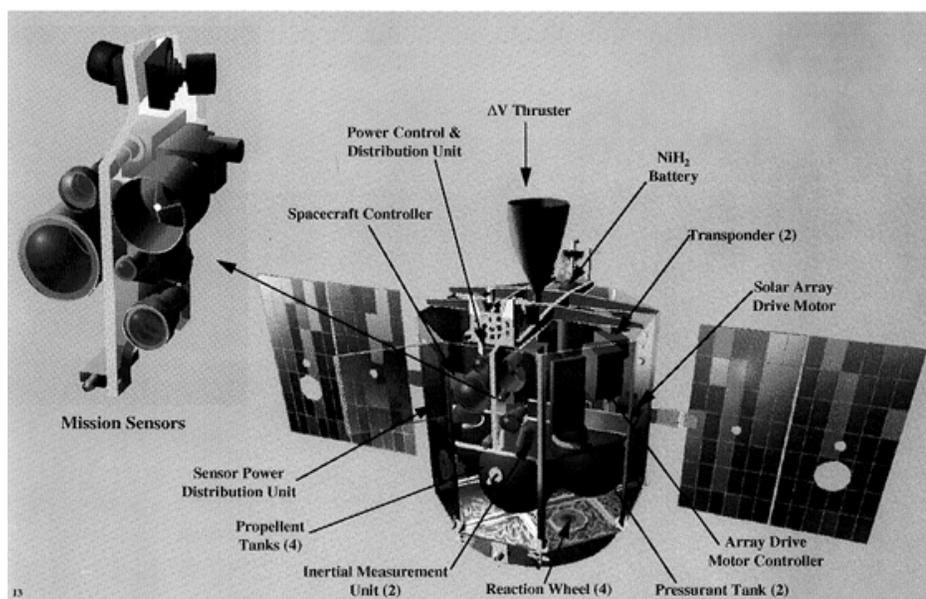
These cameras were developed at the Lawrence Livermore National Laboratory, and represented the state of the art—meaning *lightweight*—two years ago. The UV-visible camera incorporated a CCD electronic imaging system and had a series of filters, but the whole thing, including electronics, weighed just 500 grams.

Other cameras included high resolution, near-infrared, and long-wavelength infrared cameras. The last was sensitive over a narrow wavelength band centered around 8.5 microns, and was a heat-sensing instrument set at about the peak of blackbody radiation for heat coming from the moon. The near-infrared camera, was sensitive in regions of interest in identifying the mineral composition of the lunar surface.

A Nd-yttrium-aluminum-garnet (YAG) laser, sent out a pulsed signal used to determine the distance from the spacecraft to the lunar surface. The signal was then received by a telescope. Also included was a little charged-particle detector to monitor the environment of the spacecraft, and two other lightweight cameras, 400 grams apiece, which were star trackers pointed in different directions. The star trackers



Image of the Moon taken by the Star Tracker camera onboard the reconnaissance craft Clementine. The glow over the horizon is the solar corona; craters on the left are illuminated by "earthshine," light reflected off Earth and onto the Moon.



A compact vehicle designed specifically for moon mapping, Clementine was a testbed for a number of sensors and lightweight spacecraft components. The spacecraft was built and operated by the Naval Research Laboratory.

can recognize stars down to about magnitude 6; a map of the sky was "hard-wired" into the cameras, so that they could determine the direction in which the spacecraft was pointed and provide updates every 10 seconds.

In part, Clementine was intended as an engineering test mission. The spacecraft was built at the Naval Research Laboratory in Washington. Three different computers were on board. A lot of state-of-the-art components were being tested for use in various spacecraft.

In addition to testing components, a primary mission goal was to do some really serious science. We had intended originally to spend two and a half months studying the moon, which we did. When the spacecraft escaped from the moon, it was to go into a couple of phasing loops around the earth and then intercept an earth-crossing asteroid, Geographos. Unfortunately, we lost the spacecraft before it got to Geographos, this because a software error had depleted the supply of attitude-control propellant.

The spacecraft was put into an eccentric polar orbit about the moon. The orbit was eccentric for two reasons: we wanted the spacecraft to be able to escape the moon to intercept Geographos, and we needed five and one-half hours to get all of the data back to the earth. The sleeper in all of the science was the laser ranging data. They were one of the scientific returns that were easiest to interpret, and we got very extensive results. In addition, two million digital images were returned from the moon. We com-

pletely mapped the moon with two of the cameras, getting complete coverage in 11 spectral bands—one of the major goals of the mission.

We weren't sure how the laser ranging system (LIDAR) would perform, but it met all of our expectations. The laser was fired every second or so along the orbit, and 20 percent of the outgoing pulse signals were received, giving us about 70,000 points on the lunar surface. We actually have the orbits determined to within 10 meters or better in terms of radial distance from the center of mass of the moon, so we know where the spacecraft was at any given moment. The precision of the timing of the pulse was about 40 meters. You can take that data for all 70,000 points

and make a map of the moon's global topography.

### Mapping the Moon's topography

What you see on a color-coded, shaded relief map is a huge range of elevation with more than 13 kilometers difference from the high points to the low. Clearly visible is the largest preserved giant impact crater that we know of in the solar system. It's called the South Pole-Aitken Basin; it was recognized from scarps seen in earlier photography taken by satellite, but we had no idea of its shape. The floor lies at an average depth of 12 kilometers relative to the rim. It's 2500 kilometers in diameter—in contrast to the radius of the moon which is only 1740 kilometers.

When you know both the topography and the gravity you can learn some basic new things about the structure of the moon. We measured the gravity by very accurate tracking of the spacecraft. Small accelerations of the spacecraft along its orbit, reflects variations in the gravitational tug of the nearby parts of the moon below.

It is possible from the LIDAR data to compare the topography going over two large impact basins, the Orientale Basin and next to it the Mendel-Rydberg Basin, with the gravitational acceleration of the spacecraft as it flew over them. The Orientale Basin, at its lowest part, is flooded with basaltic lavas which are reflected by a bump in the gravity called a free-air anomaly. That load of basalt has bowed down the crust so that there are negative anomalies on each side. In the basin there is no flooding of basaltic lavas, no extra load on the crust.

The remarkable thing is that when the spacecraft flew over the Mendel-Rydberg Basin with 6000 kilometers relief—a big hole in the moon—the spacecraft hardly no-

ticed it. There was very little deflection of the orbit. That is an important statement: it means that there must be dense rocks from the lunar interior that have been lifted up under the center of the basin. The material that has been lifted up has a higher density than the average lunar crust; the excess mass in this dense material just about compensates for the missing mass in the hole at the lunar surface. And we find this to be true for essentially every large basin on the moon. The large holes in the lunar surface are compensated by dense material that has welled up from deeper in the moon.

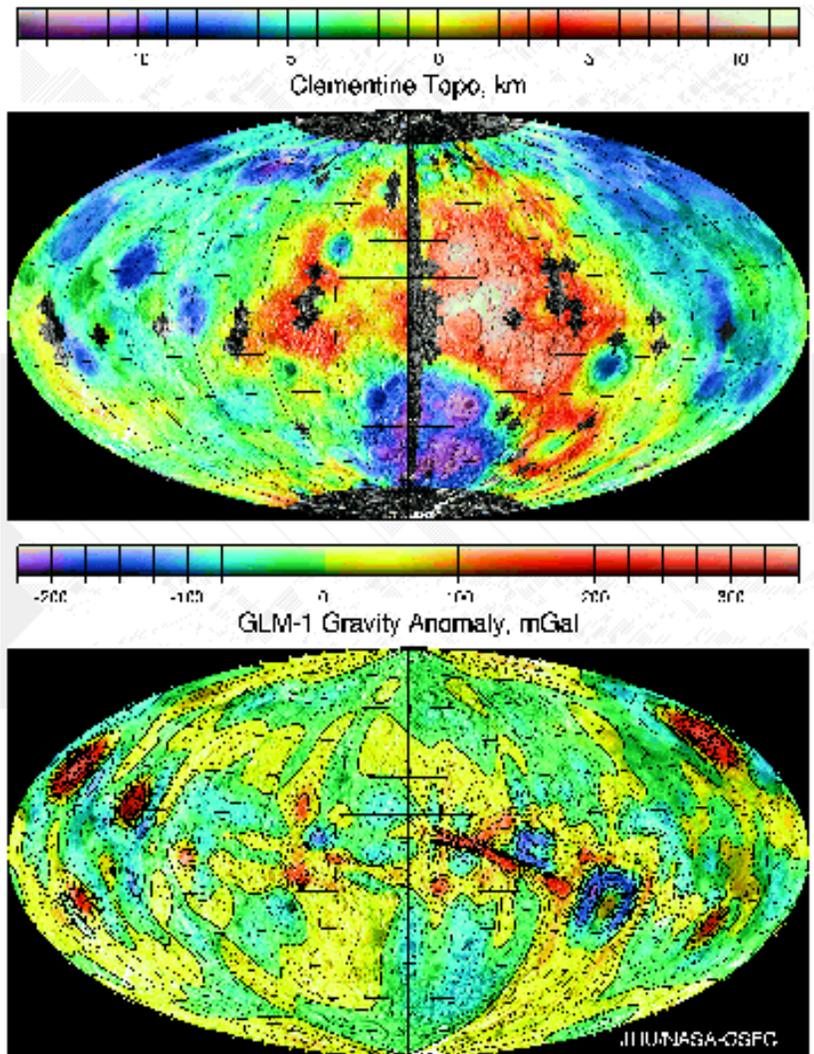
You can calculate the missing volume in a given basin from the topographic data, and if you know the density of the crust you can calculate the missing mass and can make a correction to the observed gravity. By making this correction from the topography and adding that correction to the free-air gravity anomaly, we get what is called the Bouguer gravity anomaly which reflects variations in the density in the moon's material subsurface.

Next, if you make an assumption about the contrast in density between the lunar crust and the underlying lunar mantle, you can calculate the thickness of the lunar crust. A global map of the thickness of the lunar crust has been derived from the global Bouguer anomaly. This is a first-order result from the Clementine mission. The lunar crust is about 12 percent of the volume of the moon. The next step in analysis of the topographic and gravity data will be to look at variations in the strength of the lunar crust and how it has changed with time.

### Imaging in 11 colors

Turning to Clementine's cameras, they looked at the moon in 11 colors, sharply-defined colors as determined by interference filters on a filter wheel that could be rotated into the optical train of each principal camera. The colors were chosen by analysis of the spectral reflectance of the moon rocks and of samples of lunar soil. We chose the colors so that we could identify the major mineral constituents of the lunar surface. Important were the spectral bands of the minerals pyroxene, plagioclase, olivine, and ilmenite. We wanted to be able to determine the proportions of the different minerals so that we could identify and map the distribution of rock types.

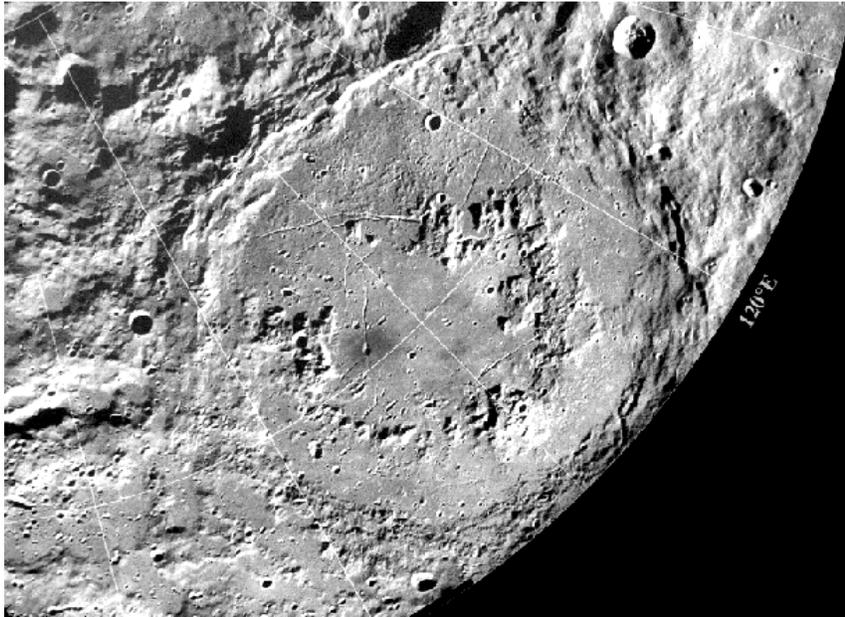
There's a bit of a trick to mapping rock types because the lunar surface is mostly covered by a soil of fine-grained particles that have been produced by prolonged bombard-



The Moon's topography was accurately detailed by mapping with laser ranging system. Some 70,000 surface points were measured to a high degree of accuracy, and revealed a 13 km difference between highest and lowest points. Free-air gravity anomaly, measured by precise tracking of the spacecraft, results from varying composition and densities of moon's surface.

ment by small meteorites. The pounding by meteorites has ground up the lunar surface and converted part of it to glass. The surface has also been impacted with high-energy particles from the sun. All those things have altered the optical properties of the lunar surface. So if you look at the soil, instead of seeing the pronounced mineral absorption features that are found for fresh lunar rocks, you see a subdued-looking reflection spectrum. The name of the game is to try to identify very subtle features in that spectrum.

Although most of the data from the two million im-



The Schrödinger multi-ring basin is the second-youngest great multi-ring basin on the moon, one of the freshest, least-modified, giant impact craters of its kind.

ages are still being interpreted, it has been possible to take some preliminary looks at the data. Using the ratios of only three colors, a global view of the moon, made at three different rotations separated by 120 degrees, shows pink regions where there is very fresh rock; orange regions representing the highlands of the moon and mostly made of the mineral plagioclase; blue regions which are basaltic lavas; a light blue color indicating a high-titanium basalt; and an indigo color, which is a low-titanium basalt. (These artificial colors are merely a way of presenting differences in the color ratios.) You can see differences that reflect the rock types plus the effect of weathering.

One of the interesting things that Dr. Paul Lucey, one of our team members, has done is to break the lunar surface down into a sequence of heights corresponding to the laser-determined altimetry. He's found a relationship between color, elevation and mineral content. There is a variation in the composition of the exposed lunar crust corresponding to the different heights.

There's been a long-standing theory that much of the moon was largely molten (a magma ocean) after it was formed, and in the course of freezing, plagioclase feldspar crystals floated to the top to make the crust, and then the rest froze from the bottom up. From the Clementine multispectral data we can say that the vast bulk of the high region, which corresponds to the thick part of the crust, is indeed a very low-iron anorthosite, a rock composed al-

most entirely of plagioclase, which is precisely the crustal composition it should be if there was once a magma ocean on the moon and the crust was formed by this process. So we have a beautiful confirmation of the theory. In addition we can say that the amount of plagioclase probably indicates that the moon cannot have come from the earth. The bulk of the composition of the moon, when all the plagioclase is mixed back in, does not match what you would get if you pulled off a piece of the early earth, which was once a prevailing theory of the moon's origins. The moon probably came from the mantle of a very large body that collided with Earth.

I'm going to talk about one small region of great interest: the Aristarchus Plateau. Most of this plateau is mantled

by material which we have long suspected was produced by fire-fountaining and violent volcanic activity that covered the surface with ash. Purplish colors shown in a color ratio map correspond to mare basalts, and the red corresponds to the basaltic ash. Other colors are seen in the ejecta blanket of the large impact crater Aristarchus on the southern edge of the plateau.

The walls of Schröter's Valley, a giant sinuous rille in the center of the plateau, have a color ratio that corresponds to freshly exposed basalt. The walls of the valley are formed by talus from basaltic lava flows, and for the first time we have an estimate of the thickness of the ash that's on top of the basalt; it's only a few tens of meters thick. The whole plateau is built up of basaltic lavas with a thin veneer of ash on top.

Aristarchus is one of the youngest, large fresh craters on the moon. Very fresh rock is exposed on the central peak, and when we look at the spectral reflectance we find that it is made up of almost pure anorthosite; is it a block of the old crust or is it a later intrusion in the crust? A spot on the walls of the crater is made up of olivine-rich rock. When we look at a large deep crater like Aristarchus, which is a large drill-hole into the crust, we see a variety of rock types, which means that the lunar crust is heterogeneous and has had a complicated history. The task ahead is to work out the detailed structure of the crust to achieve the best insight into the geological evolution of the moon.

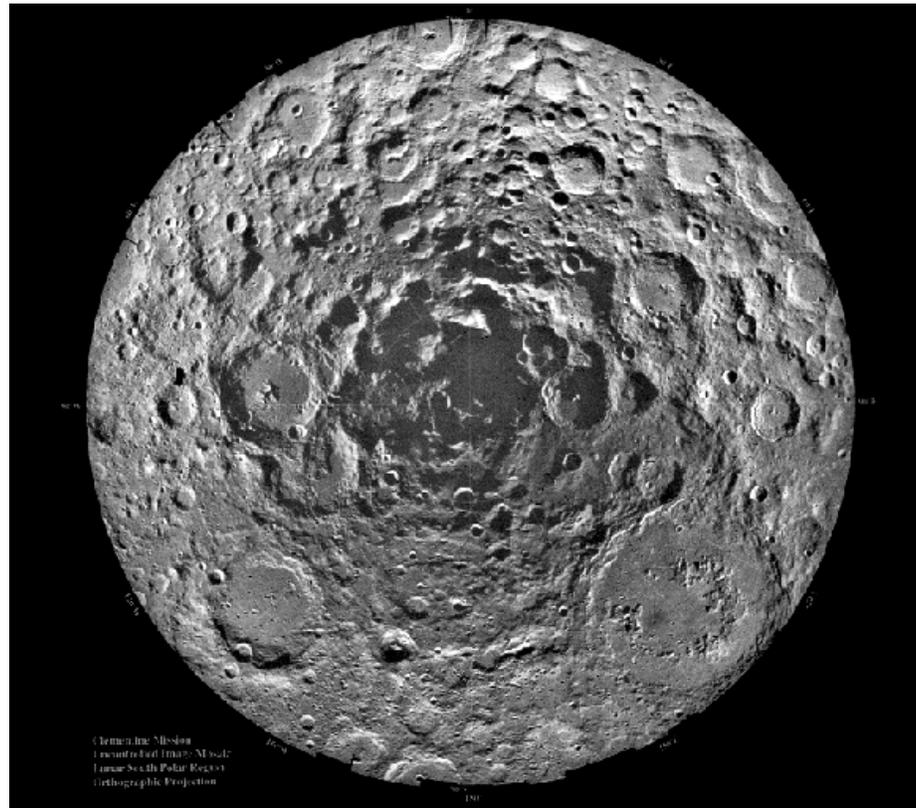
### The Schrödinger Basin

In closing, I will tell you about a feature in a region of the moon near the south pole which was not well-observed

before the Clementine mission. The Schrödinger multi-ring basin is the second-youngest great multi-ring basin on the moon, one of the freshest, least-modified, giant impact craters of its kind. Among the things we see on the basin are secondary craters formed by flying debris thrown out of other giant basins; some are formed by debris flung out of the giant Orientale Basin which is about 90 degrees away. That's how we determine the age relationships—we know the Schrödinger Basin is older than Orientale, because Orientale secondary craters are superimposed on Schrödinger, but if we actually count all the craters on the surface of each basin and compare them, the two are very close to the same age.

In the center of the Schrödinger Basin is a peak ring formed during the collapse of the initial crater. The floor of the transient cavity that was punched into the moon during impact was forced upward by the converging flow of material from the collapsing outer walls to form a central peak. In a basin as large as Schrödinger, the center of this rising peak subsequently collapsed to form a central depression. The smoother terrain in the floor of the central depression is underlaid by shock-melted material, and there are also a few patches of mare basalts formed much later and other volcanic deposits formed long after the crater was formed.

A beautiful little volcano, one of the largest of its kind on the moon and ringed by a dark deposit of ash is present in Schrödinger. When we look at the distribution of tiny impact craters on its rim, we can date the volcano rather accurately. It is substantially older than the basin, but younger than the nearby mare basalts. A fresh impact crater on the rim of the volcano reveals material that is even darker than the rest of the dark ash deposit and also quite different in color. The spectral characteristics of this freshly exposed ash show us that there is an iron-bearing mineral



The moon's south polar region, only half of which had been previously visible from Earth, is dramatically portrayed in this mosaic of images made during the Clementine Mission.

in it, so it's not pure glass. The ash probably has crystals of olivine or pyroxene. Not only is the freshly exposed material darker, it is redder than the soil on the rest of the ash deposit. We think that, with a prolonged period of bombardment of the surface, material from the surrounding melt sheet gets spread around by many small impacts and mixed with the soil on the ash deposit which is what determines the color of the dark soil.

One can combine all these observations to produce a geologic map that shows geologic units of different composition, origin, and age and which gives the chronological sequence of the units, determined by counting the small craters and from the relationships of superposition of the units. The task ahead is to make a new, detailed geologic map of the whole moon which will enable us to unravel the evolution of the lunar crust.



*Astronomer Christopher Impey of the Steward Observatory at the University of Arizona has research interests in quasars and the discovery of ultra-faint galaxies, and is the only observer to have extragalactic projects approved in each of the five observation cycles of the Hubble Space Telescope. An active and enthusiastic teacher, his emphasis is on presenting astronomy with a humanistic approach, especially to nonscience majors.*

*Dr. Impey has also been involved in setting up a summer program to train high school science teachers at the University of Arizona, and is coauthor of *The Cosmic Journey*, an introductory astronomy text. A native of Edinburgh, Scotland, he is a graduate of the Universities of London and Edinburgh, and was a British Science Research Fellow at the University of Hawaii, and a Weingart Research Fellow at Caltech.*

## Cosmic Catastrophes

Adapted from a 1995 presentation by Christopher Impey

**T**he “cosmic catastrophes” I normally investigate are events that take place in the centers of galaxies. Quasars are assumed to be powered by supermassive black holes, perhaps a billion times the mass of the sun, and the physics that goes on near the centers of galaxies is extremely interesting and somewhat catastrophic. The other kind of “catastrophe” that I’m interested in is the current model of the universe, the big bang, which was the catastrophe to end all catastrophes.

Today I want to talk about cosmic catastrophes that take place closer to home, in our own solar system and in the history of our own planet. I want to step back a little and give a broader context to some of the spectacular events that you’ve been hearing and reading about in the last few years; to give a sense of the role of these objects that fly around in space. The role they play in the history of life on earth is a subject of interest to a lot of scientists.

Let me start with a picture of what the solar system might have looked like four and a half million years ago when the solar nebula was collapsing and the sun was first igniting. At that point there was a lot of debris being swept up by the process of accretion and forming the planets. We don’t know the details of how this took place, but the point is, that there was a lot of stuff left over from the process of star formation. There were chunks flying around through space, and they were still very hot, heated by intense radiation from the newly-formed sun. There were chunks of every conceivable size—from the size of grains of dust to the size of small cities. And while most of these pieces of hot rock eventually slammed together and formed by accretion the planets we know and love in our solar system, a considerable fraction of the debris was left over, just as in the natural process of star formation.

In a sense we should look at star formation the other way around, as a highly efficient process. This is because now, four and a half billion years after this chaotic beginning, we’ve ended up with 99.9 percent of the material of the entire solar system in the sun, deposited in that one glowing star. So the debris—all the planets combined—is a tiny fraction, one part in a thousand of that mass. In fact all the debris forming the comet cloud and the asteroid belt and meteors form about one part in  $10^5$  of the mass of the solar system. The debris in a statistical sense is quite negligible. Thus we can look at star formation as being a very efficient process, because it is overwhelmingly effective in collapsing into a fusion-burning ball of gas. But the leftovers are of course interesting, and although they don’t amount to much statistically, their impact on the history of the solar system is important.

To bring this home we go to another hypothetical view, to a very interesting time on our planet 65 million years ago, known as the K/T (Cretaceous/Tertiary) Boundary, when it is speculated that a very large object hit the earth. It is one of the greatest scientific detective stories of this century, an interdisciplinary effort that has involved astronomers, planetary scientists, geologists, and paleontologists. We can speculate that the quiet earth sitting there 65 million years ago was the recipient of a very unwelcome visitor landing at an extremely high velocity. There is still considerable debate over the effects of a collision of this magnitude. The impact, the conversion of

kinetic energy, would do several dramatic things: It would almost immediately vaporize a large fraction of the ocean in which it falls. The penetration to the base of that ocean would vaporize instantaneously a trillion cubic tons of the sedimentary layer. This translates into carbon dioxide and a vast amount of rocky particulate matter and debris thrown high in the atmosphere, there to be carried by the trade winds and encircle the earth. The other effect would be the generation of hundreds-of-meters-high tidal waves that would travel around the earth many times, devastating coastal regions.

This event was prodigious enough that there is extreme interest in understanding it, and estimating how often these events happen. We would like to know how rare these really big hits are, and to piece together the effect that such catastrophes have on our solar system; how they affect the very interesting chronology of the earth which in four and a half billion years has generated us, intelligent life forms who care about our survival. The first calculations of this effect were first done in the context of the cold war, but similar computer simulations are now being used to calculate the impact of dramatic climate change from an impact.

### Origins of a catastrophe

In the impact 65 million years ago, somewhere between one-third and one-half of all the species on earth were obliterated. The dinosaurs were among the many casualties of a very catastrophic event. How does this kind of thing happen astronomically speaking? What can possibly cause or trigger impacts? We know that there are a certain number of earth-crossing bodies that have the potential to impact the earth. There is also an enormous reservoir of rocky material sitting in the Oort cloud, at a distance of hundreds of astronomical units, that can be triggered into orbits that will bring it into the central solar system.

As an example of a mechanism that might do this, we can project from the known positions of stars in the vicinity of the sun that their orbits will bring them in close proximity to the Oort cloud at times in the future (figure 1). Computer calculations show when various nearby stars will approach the sun and, if you do the dynamics, you can show that a close passage of a nearby star will indeed

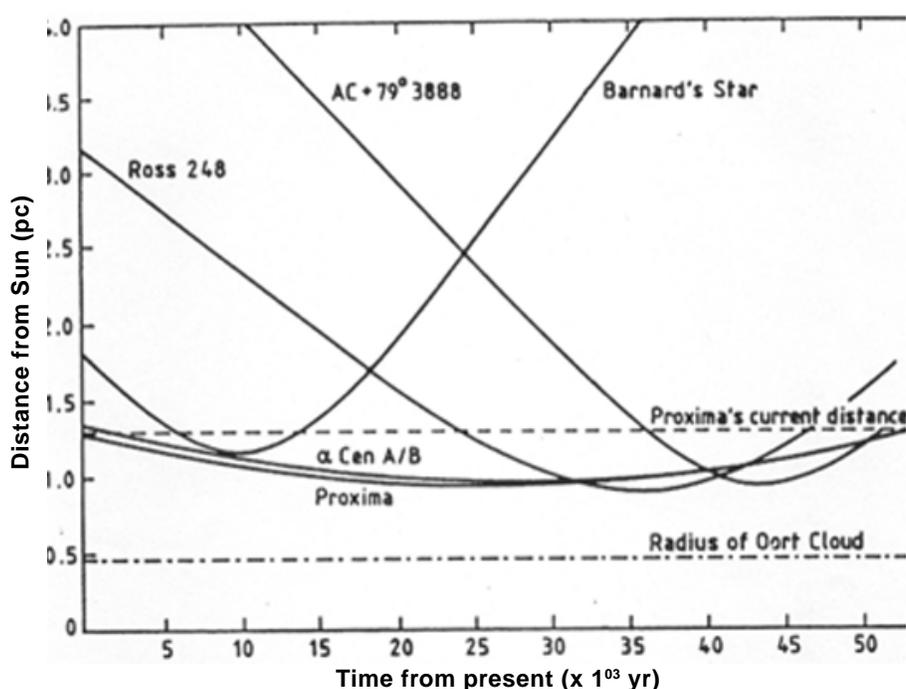


Figure 1: The close approach of the six closest stars over the next 50,000 years; note the cluster of encounters beginning about 30,000 years from now. Proxima's trajectory assumes a bound state about a Cen A/B; that of Ross 248 assumes the Coravel value for radial velocity of  $-79.2$  km/sec.

cause a perturbation of the Oort cloud, and may well trigger a pulse of objects in earth-crossing orbits. This is an emerging field, because the computing power necessary to calculate these orbits with any certainty is prodigious; super-computers are required. But it is clear that, on time scales of thousands of years, we can have quite large fluctuations in the amount of debris that comes into the inner solar system. Also it is evident that we can, in principle, start to predict such fluctuations if we know enough about our solar environment.

Let me talk a bit about the nature of a catastrophe. When an object traveling tens of thousands of miles per hour is instantly stopped and its kinetic energy is instantly converted into heat, a fantastic amount of energy is released in a short amount of time. A fast-traveling rock becomes both liquefied and vaporized at the moment of impact, and will send out a shock wave and a pulse of liquefied rock that will travel outward. When the pulse has traveled a certain distance, it will eventually solidify again, leaving a rim. Finally, there will be a reverse wave that travels inward, eventually producing a little blip in the middle. It's directly analogous to the pattern seen in high speed photography when a drop of liquid hits a liquid surface.

The devastation that we're talking about is truly amazing compared to the kind of events we're used to. The two nuclear devices used over Hiroshima and Nagasaki were rated in only tens of thousands of tons of TNT. The largest nuclear bomb, a fission device used in the Bravo underground explosion in Nevada, was at the level of tens of millions of tons of TNT, and that's about the level of the largest natural disasters that we're aware of in this century. In these terms, the amount of power associated with the K/T Boundary Event 65 million years ago is almost off the scale, something like a trillion or several hundred trillion tons of TNT. The point is clear, even if we're off by a factor of ten, that it represented a truly phenomenal amount of energy (figure 2).

**Cosmic intruders**

The K/T impact is a catastrophe the like of which we really cannot imagine, and yet such things do happen. If you go back through the history books you find many pieces of evidence, some anecdotal, some documented, which clearly link less dramatic visitors in the sky and devastation on the earth. You will find evidence of cataclysms, and although it's always difficult to discern their nature, you'll find strong and cumulative evidence that they can happen on human time scales. And there's a whole field, called eschatology, concerned solely with the end of the world. It's a somewhat disreputable field for conventional scientists to deal with because its bound up with mysti-

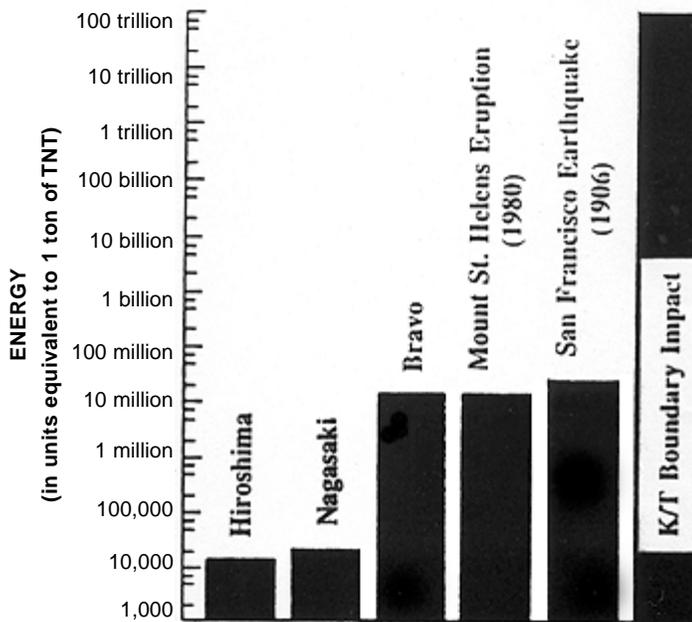


Figure 2

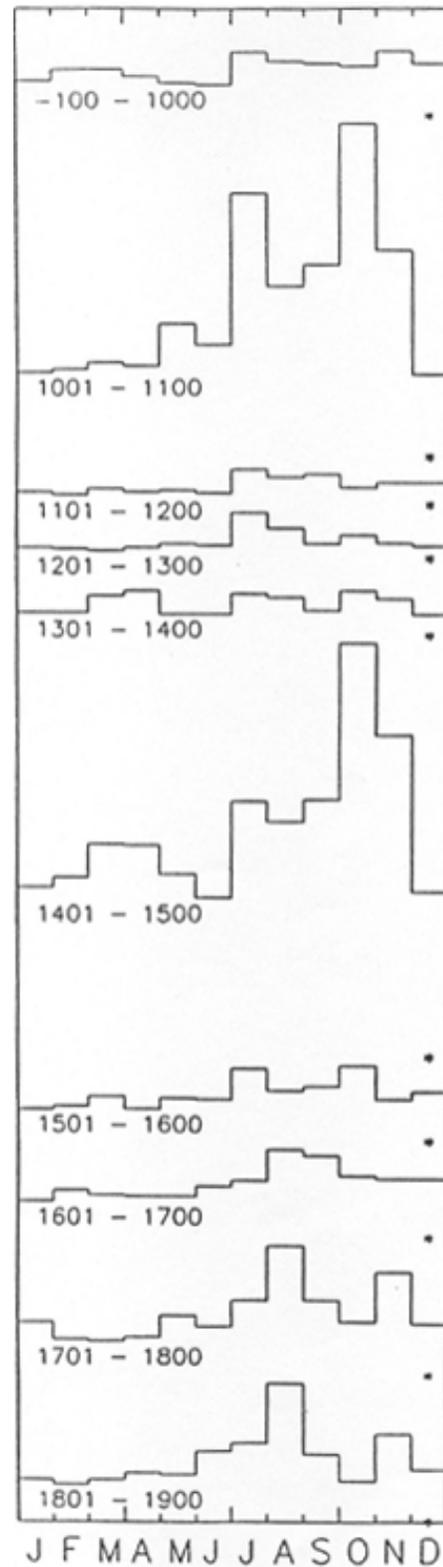


Figure 3

cism and religion, things we're not supposed to talk about as scientists, but it gives us interesting clues.

For most of the past 2000 years people in the West have been hacking away at each other and invading each other's countries, but luckily the Chinese have been carefully counting events in the sky—fireballs, meteors, large astronomical events—and giving us systematic data that we can compare from one epoch to another. A cumulative plot, divided by centuries, of the monthly occurrence of fireballs or transient events in the sky reveals two things: a time line with patchy evidence before the year 1000, and that the occurrence of these events is by no means random, but is significantly peaked (figure 3).

Whatever is causing these astronomical events has a periodic mechanism with a time scale of hundreds or thousands of years. We have an astronomical reason to believe that pulses of devastating events could occur on a tens-of-thousands of years schedule, caused by pulse tides of nearby stars, but there's also a shorter time scale. I don't think you can necessarily believe some of the peaks, but a plot reveals a periodicity of 500 years (figure 4).

The people who look at the interplay of human myth, religion, and science will be fascinated by the potential implications of the peak at the time of Christ. When you divide the record by centuries, it appears that there were enormous pulses in the Dark Ages, in medieval times and in the Renaissance. And the final thing you notice, the most informative, is that the events are not uniformly distributed by month.

There are clear peaks in August and November that correspond to our best-known meteor showers. The speculative interpretation of this data is that the Taurid Stream, the debris that is associated with the Taurid Meteor Shower,

leaves a trail which on some occasions has caused large pulses of events like the 1908 Tunguska fireball in central Siberia. This flattened almost 2,000 square kilometers of pine forest. These events have occurred through history, and the Taurid Stream is responsible for the meteor showers in those months when the earth passes through it. This is still at the speculative level, but it's very provocative.

### A mechanism for periodic extinctions

Let me give you yet another time scale of events that happen, "positive catastrophes" insofar as they aid understanding. This time scale is based on a compilation of a variety of pieces of evidence, and it links events that effect the earth and the history of life on earth, and the geological record. So far we have talked about pulses of fireball events—Tunguskas—which occur on hundreds-of-years time scales, and we have an astronomical mechanism for pulses of incoming chunks of rock on a tens-of-thousands of years time scale. But there's another very important astronomical measure, a much longer one, and potentially more devastating, and that's the time scale of the earth and the Sun's passage in and out of the plane of the Milky Way. The solar system travels around the Milky Way in a slow, circular orbit that takes about 225 million years to complete. The earth and the solar system have been around the Milky Way about 40 times in the history of the earth.

An important consideration is that stars in the disc of the Milky Way are not in perfectly circular orbits, and they show a vertical modulation. Earth and the rest of the solar system oscillate up and down as they traverse this long orbit. The time scale of this vertical oscillation is about 32 million years. That means that the entire cosmic environment of the solar system changes as we travel in and out of the plane of the Milky Way where the dust and the debris from star formation are found. And if you do the gravitational calculations, you can show that the tidal force induced by that oscillatory motion could trigger perturbations in the Oort cloud, or Earth might just pass through debris. We don't know the details, but this is clearly a physical mechanism with potential for causing cosmic catastrophes. You can create a time line and include a variety of pieces of observational evidence including mass extinctions. There is some spacing on the order of 225 million years in the data on mass extinctions.

Then there are enormous catastrophic changes on a geological time scale of the flood basalt—a violent change in the level of the oceans (figure 5). These are reflected in the geological record, and line up pretty well with the mass extinctions. If you turn to the fossil record, you find a thin layer of iridium which we believe is extra solar material that does not appear naturally in the earth in such a con-

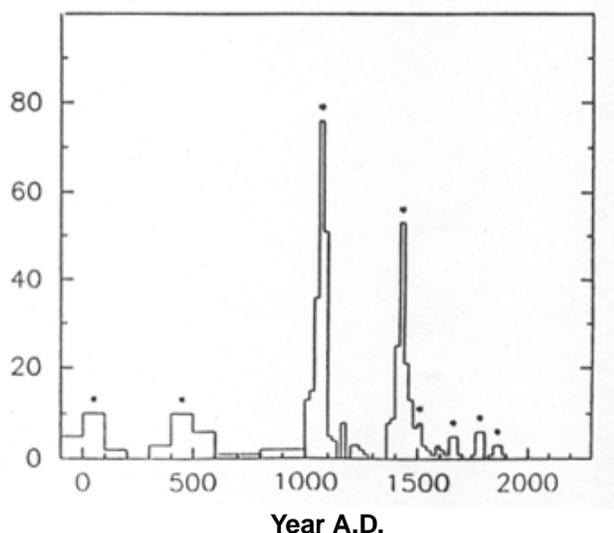


Figure 4

centration. It must be deposited externally from space, and is therefore evidence of some external mechanism.

This iridium layer is found in all places around the world, and that is one of the keys to the K/T detective story. The iridium layer has been found in the same time spans as the mass extinctions from the fossil record, and the flood basalt from the geological record. The last piece of evidence is the existence of tektites and microtektites, tiny, glassy nodules of volcanic material caused by intense heat, also part of the geological record. They could exist near any volcano, but in some layers of the geological record you find these beads deposited in places where there aren't volcanoes, and deposited in different places in the world at almost exactly the same geological time.

Each of these pieces of evidence has some uncertainty attached to it. Taken as a cumulative whole, however, they are strongly suggestive of a correlation between massive impacts and the life history of the earth; between global change as indicated by the sea levels, and between the violent deposition of energy; and of an external mechanism indicated by the iridium layer. It's a very provocative combination.

**Effects of impacts of different sizes**

The least certain aspect of this work is the periodicity. Is there really a 32-million-year cycle for which we would find a natural explanation in terms of our solar motion? In this emerging subject it does lead us to ask about the effect of these catastrophes on the history of life on this planet.

Let me just give you a thumbnail sketch of what the impact of objects of a certain size would be on the earth and on its living things. We know that objects hit; we have

the Tunguska event, and evidence in the historical record which is very hard to interpret, but which indicates that such things happen. In 1178 the monk Gervase of Canterbury saw a lunar explosion. We don't know what he saw 1000 years ago, but there's some speculation that it was the formation of one of the moon's younger, fresher craters that caught his eye. We have the magnificent spectacle of the Comet Shoemaker-Levy 9, and we know that in 1975 the Taurid Stream slammed a bunch of rocks into the moon some years after magnetometers were put in place by the Apollo missions. We know that debris of all sizes must exist, and that occasionally we can observe it.

What are the frequencies of different impactors? At the level of something about 10 meters across, we have the impact of a Hiroshima blast, and the incoming rate of those is about a 100 per year. That might worry you, but you have your friendly atmosphere protecting you from ultraviolet rays and effectively demolishing things that are 10 meters across. (These high atmosphere blasts were first detected by early warning radars in the 1960s and for a tense time were suspected of being Russian nuclear tests.) Go up in size by a factor of 10 to something about 100 meters across, equivalent to 15 million tons of explosive, and you might get one impactor per hundred years. These event rates are highly uncertain, but they're good to an order of magnitude. Now one 100-meter rock per 100 years makes you think of something like Tunguska about 100 years ago. Luckily for all of us, it fell in one of the most uninhabited places on the earth, for 2000 square kilometers were leveled, the atmosphere was girdled with debris for many months afterwards, the shock waves were felt in the U.S., and the sky was lit up by the material for some time.

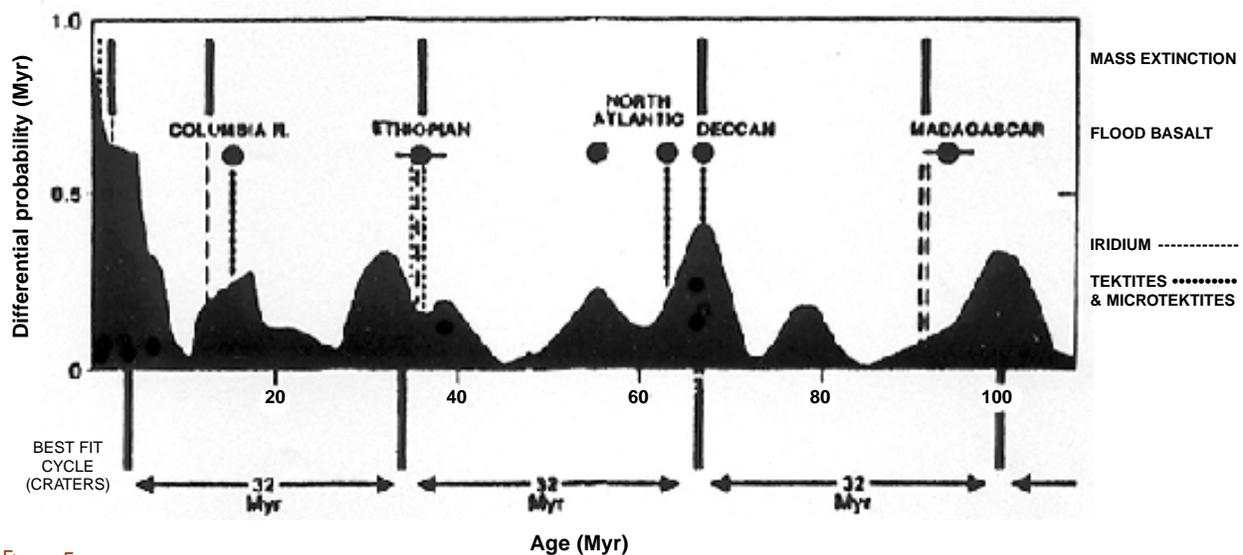


Figure 5

Go up a factor of ten more, to a kilometer-size chunk, you've got about a billion tons of high explosive in terms of the impact—about equal to all the world's nuclear arsenals taken together. The implied event rate is very uncertain, maybe one per 100,000 to 500,000 years; rare on a time scale of a civilization, but not on that of a species. Such a collision would have the effect of a global nuclear war. It would leave a 10 kilometer crater, would level an area the size of Belgium, and would make 100-meter waves that would encircle the earth. Indeed, it would probably lead to the kind of nuclear winter scenario that would cause massive crop failures and worldwide famine. Even though the impact itself might not kill that many people, a significant fraction would die from famine.

Let's take a 10-kilometer impactor as a reasonable upper boundary for anything that might hit the earth, and at that point you are talking about a global catastrophe, an event rate that's indeterminate, maybe one per billion years, one per hundred million years. Such events may have occurred in the epoch of massive bombardment when material was being swept up to form the planets, but since then not at all. The destructive power of a 10 kilometer impactor would make a 100 kilometer crater, would trigger global earthquakes, 1000-meter-high tidal waves, and would almost certainly destroy the food chain on both the land and in the ocean, such as to wipe out all life (figure 6).

**Protection against impacts?**

Luckily, in the post-cold-war world our friends at Los Alamos and Livermore have taken it upon themselves to protect us from these devastating impacts. Big studies have been done at the national labs, one at Livermore called "Operation Chicken Little" which involved a very complex strategy. Some of the grandiose ideas involve antimatter beams, ion beams, and the redeployment of our nuclear arsenal to protect us from the potential impact.

This has a little whiff of pork about it, because, in fact, we can protect ourselves much more cheaply. Rather than a military buildup for a potential threat that may never happen, we just need an inexpensive three-step project. First we need to build a telescope of 2-3 meters. With that and a set of small telescopes you can pretty much look for, and catalog down to a relatively small size, all the potential earth-crossing objects there are. Then you get time on some supercomputers and calculate the orbits and trajectories; at this point you have several years of advance warning before a potential impact. Only if your supercomputer tells you that within certain tolerance limits the earth may be subject to a possible hit, *only then* do you bother to build a nuclear warhead and deflect the menace.

A three-phase program isn't initially very expensive at all. There is plenty of warning of large debris, but you have

IMPACTS			
All planets form 1% of mass of Solar System			
All debris forms 0.001% of mass of Solar System			
SIZE	POWER	RATE	EFFECTS
10 m	Hiroshima	100 per yr	Fireball, many seen by Early Warning System
100m	15 m tons explosive 1000 Hiroshimas	1 per 100 yr	Tunguska (1908), 2000 km <sup>2</sup> leveled, shock-wave measured in USA, night sky lit by material thrown up
1 km	1 b tons explosive World's arsenals	1 per 500,000 yr	Equivalent to global nuclear war, 10 km crater, big country levelled, 100 m waves, nuclear winter leading to crop failure and famine.
10 km	Catastrophic	1 per billion yr	K/T Boundary (65 Million yr), 100 km crater, global earthquakes, km waves, disruption of food chain, extinction of 60% species, all animals over 50 lb.

**How to Stop Them...**

1. Monitor with telescope network for Earth-crossers
2. If one is imminent, supercomputer for accurate calculation
3. If it will hit, deflect if with megaton warhead (after prediction will have few years to prepare)

Figure 6

to be smart about dealing with it. You don't want to destroy it because all the fragments will continue on the same trajectory and will hit the earth. The goal is to deflect it with a carefully planned nuclear explosion so it misses the earth.

One of the arguments that has been used to attract funding is that the danger of an impact is statistically at an interesting level. A graph of the probability of dying by certain causes in the U.S., plotted against the number of people killed, shows that despite society's obsession with violent crime, you're more likely to die in a car accident. The fear of violent crime is partly an irrational fear, and not based on statistics. You really don't need to worry about dying from botulism, and it rarely wakes me up at night worrying (figure 7).

But the point is, if you look at a graph you'll see that the individual average probability of dying in a global catastrophic meteor impact is almost the same as the probability of dying in a plane crash. Think about how much the government spends on making planes safe, and then think about the paltry amount of money we want to spend on an event that has almost the same probability of happening. Well, this problem is bogus—not the statistics, for averaged over a significant time span the probabilities of individual death by these means are identical—because the

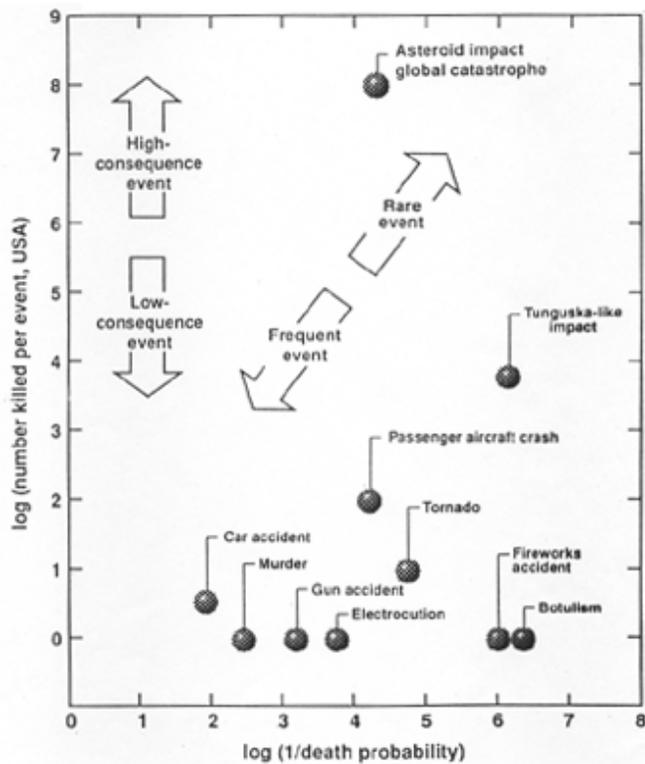


Figure 7

actual death rate in this kind of event is zero. It has never happened in recorded history. So this is a statistical argument, and what precautions you take depend on whether you feel lucky or not.

People do worry. I was taken by Thomas Pynchon's book, *Gravity's Rainbow*, set in WWII London, with German supersonic rockets raining down with very little directional control. They would suddenly take out a city block, giving the book's hero a devastating, evocative nightmare. In the dream, a V2 is about to hit his head as he walks through the streets of London, destined to be at the very space-time point with the rocket. The moment of impact is when the head of the cone is about to touch his head and will kill him, and he's obsessed by it. I'm not sure our worries are legitimate, but I think calling a project "Operation Chicken Little" is *asking* people to worry.

In case this does disturb you a bit, let me give you a couple of practical pieces of advice to avoid being hit by the smaller stuff. The first thing is, obviously, don't fly. Once when I was flying I didn't have a good novel and must have been bored, so I sat there with an envelope and I tried to calculate the danger. I tried to work out on an envelope, given the rate of debris and small particle impacts, the chance of being hit while you were in a plane. That's not a good thing to do while you're flying. After I'd sort of scared myself a little, I tried to calculate the chance

of anyone in history having been hit and killed by a chunk. So I worked out the rate of smallish chunks hitting the earth, I worked out the effective surface area of people on the planet, and I worked out the probability of anyone in recorded history having been hit. And it's an interesting number: about one in ten billion.

In the history of the planet there have been about ten billion people, so we know who it was. It was that woman in North Carolina who got hit about 30 years ago, Mrs. Hodges, so she was the unlucky one. The second bit of advice is remember the peaks of activity during the Taurids. In August, September and November you should probably spend a lot of time in a bunker under your house. The third suggestion is really obvious—just sleep standing up. The chunks that come down vertically are the ones you have to watch out for because they penetrate most deeply through the atmosphere. The woman that got hit was lying in her bed. If you sleep standing up you minimize your cross section, so it's a really smart thing to do.

Let me put in context the importance of debris hitting the earth. The earth has a very violent and chaotic history, and not just in the early period. After earth formed it was a chaotic place, there was a period of bombardment, the oceans had barely condensed out of steam, the land surfaces were almost molten. Interestingly, life formed almost as soon as it possibly could. Within 400 million years after the formation of the earth, we have the first evidence of life. The oldest evidence of life dating back to about 3.9 billion years ago is not fossils or single-celled organisms, but is based on the imbalance between the normal and the radioactive isotopes of carbon, an indirect indicator of metabolism activity. The earth was barely habitable in those days, so life wanted to form, life formed as soon as it could.

**The evolution of life on Earth**

Some people argue that life had multiple starts on the earth, and that some streams of life were obliterated by heavy bombardment. We infer that life is tenacious. We also know that life has a long progression towards complexity, intelligence, and higher evolution. Humans sit at the pinnacle of this tree of life as the natural product of eons of time and the inexorable workings of natural selection. It sounds very comforting: we're here because we're meant to be here; evolution has been a motor that has been driving in this direction for billions of years.

This is the conventional view, a cone of increasing diversity where natural selection prunes away the branches, and where we, the only species capable of altering the global ecology of our planet, sit at the end point of the evolution. But that's not the way the history of life looks when you get down to the actual fossil record (figure 8). All these false starts in the early days of the planet, and the erratic

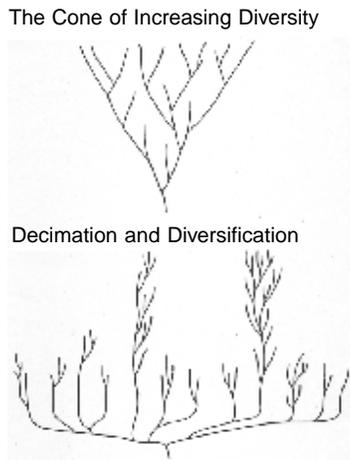


Figure 8

progress since then, tell a very different history. There was, for example, an explosion of life in the oceans of the earth in the Precambrian era 500 million years ago, and then many of these life forms were obliterated subsequently in a very short period of time. The fossil record also shows that for about one and a half billion years not much happened, that multicelled organisms took an awfully long time to evolve. In fact, if they had taken five and a half times longer to evolve the sun would have died before humans evolved, and we never would have had the chance to sit here and think about all this.

It's evident that the inevitability of intelligence is very controversial, but regardless of that, we see a very interesting history; bursts of evolution, sudden rapid changes, and then decimation. There is no paleontologist that could have looked at the nature of these life forms and predicted which would survive to become reptiles, lizards, mammals, primates or humans. In other words, there's no clear advance plan of evolution because, through catastrophes, nature has obliterated some of the more favorable routes to evolution. Evolution is not a smooth path, and there's no inevitability to complexity or intelligence.

Many people still think that humans took over from the dinosaurs—mammals took over from reptiles and lizards—because they were better adapted. That's not true, because dinosaurs and mammals coexisted for 100 million years on this planet with no sign whatsoever that mammals were to be ascendant. In fact, for all that 100 million years, it was the reptiles' world, and mammals fit into the little evolutionary niches. And then, *wham!* Cosmic intruder, cosmic catastrophe, nature rolled the dice, the dinosaurs were unlucky, mammals were lucky, so here we are to talk about it. I think it gives a very different spin to the nature of evolution of life on earth. And we're really only still beginning to get a sense of what that all means.

Let me finish with the original detective story. The evi-

dence—the tektites, the iridium layer, the implication of a sudden change in the fossil record—was all very provocative, but for many years people have asked about the smoking gun, the point of impact. Well, it's most likely to be in the ocean, and it turns out that we have found the crater. Many research groups around 1980 were pushing the hypothesis of a cosmic intruder causing the K/T extinction, but the vital pieces of evidence were actually missing for a number of years because the Mexican oil company PEMEX had done core sampling in the Yucatan basin and had lost the samples. So the geologists couldn't look for the physical evidence that would give them a sense of where this crater was (figure 9).

The detective story has been played out over 10 or 15 years, and since science is an inductive business, I'm not going to tell you the answer is guaranteed, because there's still active debate. The excitement is building now because there is evidence that a crater off the Mexican coast ties together all those correlations I mentioned earlier. If you go down to the Yucatan, you'll find the Ring of Cenotes, a chain of water-filled sinkholes etched in the rugged terrain. You can imagine the outer rim is made up of the drop-plet craters left over from the big impact itself. Swim in one of the cenotes and you can be a part of Earth's violent history, a witness to the effects of a cosmic catastrophe.

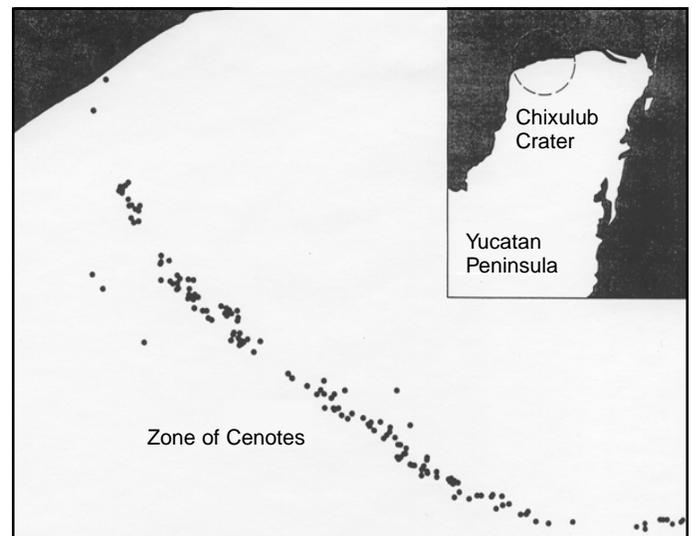


Figure 9



*Thomas V. O'Halloran's research at Northwestern University deals with a curious paradox: how our cells handle the highly toxic metals essential to their normal functioning. A broad-gauge chemist and molecular biologist, he uses a variety of tools and specialists to help explore the mysteries of biology and chemistry of transition elements. He focuses on metals involved in human health and disease such as copper, zinc, iron and such toxic metals as mercury and cadmium.*

*In addition to being a professor in the departments of chemistry and biochemistry, molecular biology and cell biology at Northwestern University, O'Halloran is director of research in the Radiation Medicine Institute of Evanston Hospital and director of the Molecular Oncogenesis Basic Research Group at Northwestern's Robert E. Lurie Cancer Center (he also has an interest in platinum anticancer drugs).*

*Dr. O'Halloran is a graduate of Columbia University. Following an NIH post-doctoral fellowship at MIT, he did further work at Cold Spring Harbor, joining the Northwestern faculty in 1986.*

## Heavy Metals in Concert with Life: Chaperones are important, but are they easily fooled?

Adapted from a 1999 presentation by Thomas V. O'Halloran

**I**t is surprising how little is known about the cell biology of the transition metals. Living systems respond to heavy metal stress at several levels, depending on whether the stress involves metal starvation or overload. This is true for both essential metals such as copper, or excessive levels of a nonessential metal, such as mercury. In what amounts to a fascinating interdisciplinary detective story, we have been studying the mechanisms of these processes using methods from the fields of physical inorganic chemistry, molecular biology and cell biology.

All creatures need transition metals in a variety of enzymes, yet even minute amounts of some of these elements can be quite toxic to cells. In several cases we have a pretty good idea how cells handle excess quantities of metals. When concentrations build up, certain proteins can bind the metals into clusters and thus prevent toxic effects, but how cells deal with the tiny quantities required in enzymes is more mysterious.

Copper, the third most abundant transition element in humans after iron and zinc, is critical to the activation of a variety of proteins. The functions of these metalloproteins range from electron transfer and oxygen transport to insertion of water, oxygen or main group atoms into a substrate. On the other hand, the release of copper ions from proteins can be toxic, even at low concentrations; cells apparently avoid this situation by a variety of mechanisms. In this presentation we consider the chemistry involved in the cell's control of copper.

Recently the scientific community has identified a few proteins, the "metallochaperones," that help the cell (1) achieve a proper concentration of metals; (2) distribute metals to the sites where they are needed; and (3) prevent dangerous reactions and adventitious chemistry that can cause damage to the cell. Our theme is thus the tension between organic life and reactive inorganic compounds.

Let's look at transition metal chemistry. The early alchemists grouped the elements lead, mercury and tin together because of similarities in their properties. Mendeleev later organized these and dozens of additional elements into groups with similar chemical properties. In the modern version of his table, transition metals are in the center and



The O'Halloran research group at Northwestern, from left: Tom O'Halloran, Caryn Outten, Christopher Fahrni, Robert Pufahl, Wayne Outten, David Huffman, Tracy Rae, and Chris Singer.



Figure 1. Representation of the metalloprotein carbonic anhydrase shows the zinc-binding site at center (small round sphere), similar to the copper-binding site in proteins with very different functions.

clearly share many attributes. They exist as ( $\text{MnO}_4^{2-}$ ,  $\text{WO}_4^{2-}$  exceptions) charged cations in solution under biological conditions, and many of these ions can have similar ionic radii, so they tend to bind to similar classes of ligands. Indeed, the ligating atoms that bind metal in a particular protein are found in four types of amino acid side chains: histidines, cysteine, glutamic and aspartic acids. Less frequently we find tyrosine and methionine coordination. Nature generally uses an array of two to five of these coordinating side chains in the majority of metalloenzymes.

How each protein selectively acquires its appropriate metal ion is not a trivial question. Take for example a couple of well-studied metalloproteins, the zinc-containing carbonic anhydrase (CA), and the copper-containing hemocyanin. Carbonic anhydrase (figure 1) speeds up the process of  $\text{CO}_2$  solubilization to form the bicarbonate buffer that is critical in every cell to maintain the osmolarity, the turgor of the cell, and to help maintain its shape. The zinc-binding site in CA (center, figure 1) is very similar to the copper binding site in proteins that have a very different function, such as hemocyanin, a blood protein in lobsters that carries oxygen. It is also very similar to copper enzymes that are key to processing neurotransmitters.

### Which metal goes where?

How does the cell, once it gathers these different metals, know which one to put into which enzyme-active site? A traditional view in the field of bioinorganic chemistry has been that metal selectivity is due to very sophisticated

chelating properties of the individual apo-proteins. In this scenario, apo-proteins are thought to poise the exact orientation of these side chains to match the precise ionic radius and electronic preferences of the functional metal ion, let's say  $\text{Zn}^{2+}$ , and to discriminate against all others, such as  $\text{Cu}^{2+}$  or  $\text{Fe}^{2+}$  ions. In some of these cases there is very little difference in ion radii. The proteins are thus viewed as highly specific chelating agents with finely-tuned kinetic and thermodynamic properties that have been selected through evolution to bind only one type of transition metal ion. In this model, each apo-protein as it is produced in the cell simply selects the correct metal ion from the cytoplasm.

There are several problems with this perspective, especially when you consider the bigger enzymes. If you could see the coppers in the crystal structure of ceruloplasmin, for example, they would be buried very deeply in the folds of the protein. Not only would the right metal ions need to diffuse in, but there should be some means to prevent other ions from getting there first. I will argue that biology does not ensure that a given metalloprotein acquires the correct metal cofactor by virtue of the protein's selective chelating properties. There is another solution to the problem of getting the right metals to the right site.

Before discussing how metal ions get to target proteins, let me put this issue in a numerical perspective. We've done many measurements of the quantity of metal ions in cells. The total number of metal ions per cell is easy to measure actually, but the quantities still amaze me. The following numbers are for a mouse fibroblast cell, which is a generic, nondifferentiated cell. We find that there are  $6 \times 10^8$  atoms of Fe per cell. Now  $10^8$  atoms is a very small number of atoms; if you divide it by Avogadro's number, you can tell me how many orders of magnitude less than 1 mole of iron it is. It's a small number, but this cell is a small body. If we consider the cell as a bag of a simple biological buffer, and we divide this number of atoms through by Avogadro's number and the volume of the cell, we find the molarity of the metal ion in the cell. We get numbers that are in the millimolar range!

If any of you have worked with enzymes or proteins or DNA, you will be horrified to think of millimolar quantities of iron and copper in the same solutions as your biopolymers. These metals can denature, hydrolyze, or oxidize proteins at micromolar and nanomolar concentrations. Thus, biochemists routinely sprinkle EDTA, a small chelating ligand, in their buffer to tie up the metal and try to prevent it from reacting with anything.

Millimolar levels of total iron and copper are, by far, more than you would find as a trace contaminant in a biochemical buffer solution. This is a high concentration, but there are two aspects that make this dilemma less con-

founding. One is that biology does its chemistry and carries out living processes by isolating chemical reactions in specific internal compartments. There are lipid bilayers that separate components in the mitochondria, for instance, from the aqueous solution that makes up the cytosol. Other compartments include the golgi, the endoplasmic reticulum, and peroxisomes. It is too early to make generalizations, but as methods become more sophisticated, I suspect we will find that essential metal ions are highly concentrated in these and other, yet to be discovered, subcellular compartments.

### Metal concentrations in cells

Millimolar levels of intracellular transition metals are less ominous if one realizes that this does not correspond to the concentration of the free metal ion. No one has been able to directly measure a value for the intracellular concentration of free metal ions which are not bound to any cellular component. I hope to convince you that we have a good idea of an upper limit to free metal concentration, at least for copper.

Before diving further into the chemistry of a cell, let's consider the concentration of metals in the ocean. The concentration of iron in seawater is extremely low because iron in the oxidized state is insoluble. Zinc, with a molarity of 0.10 micromolar is more available. We see that in our blood we have several orders of magnitude higher iron concentrations than does seawater. Individual cells have a concentration of iron yet again another order of magnitude higher than blood. Therefore, cells concentrate elements from the environment. How do they do that?

A series of copper transport proteins that have been found in the last six years. The human proteins were first found by the genetics community, including groups at Columbia University, UCSF and the Murdock Institute in Australia. The first copper transporter discovered in humans was associated with Menkes disease. It was later shown that a very similar protein is associated with Wilson's disease. Similar disorders of copper metabolism are known in sheep and in humans. From that work we began to achieve a better understanding of cellular metal transport. A variety of divalent cation transports had been cloned in bacteria before this development, so we don't want to overemphasize the human biology contribution, but that work in 1993 really brought a new wave of activity to this field.

Nomenclature can sound like jargon, so let me take a minute to define some terms. One could say that any protein that gets a metal from one place to another is a "transporter." In general, however, the field has moved to the point of thinking of membrane-bound proteins as transporters. Today we will consider another class of proteins

that act as diffusible metal receptors within the cell to move the metal ions to and from target destinations, the "metallochaperone."

The Menkes disease protein and the Wilson's disease proteins are large copper transporters, about 1500 amino acids in length. They go in and out of the cell wall, a lipid bilayer that is about 40 angstroms wide. The lipid bilayer is greasy, and variations of it make up many of the intracellular barriers that separate one compartment from another in the cell. These copper transporter proteins intriguingly include a series of metal-binding motifs with a pair of cysteines. Two of the copper chaperones I discuss here have similar motifs, and the chemistry of these cysteines is central to the function of both the copper chaperones and the transporters.

### Two copper chaperones

The two copper chaperones I will focus on are called Atx1 and CCS. I will discuss the chemistry of chaperones isolated from single-cell fungi, baker's yeast; however, several of these conclusions can readily be extrapolated to human copper chaperone proteins. Likewise, yeast cells have a membrane-bound copper transporter very similar to the Menkes and Wilson's disease proteins in humans. This protein, called Ccc2, ultimately puts copper into an

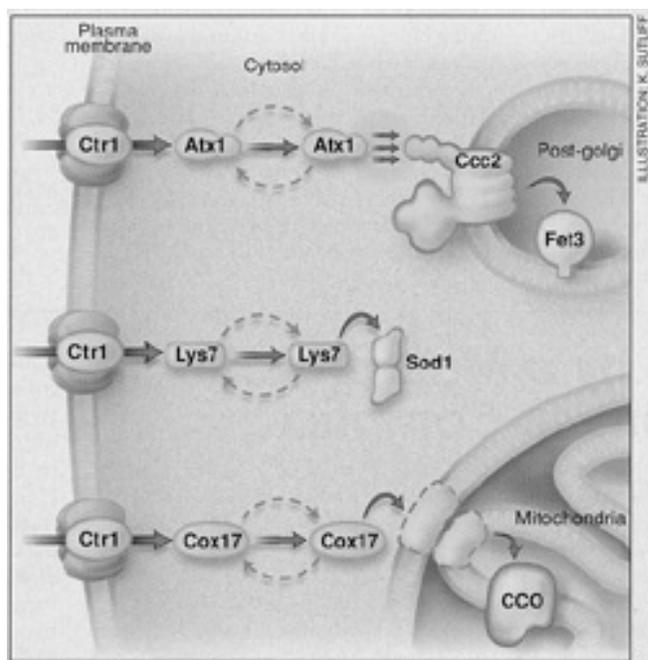


Figure 2. Chaperone Atx1 plays a key role in distributing copper to a cellular compartment containing Ccc2. (Illustration reprinted with permission from *Science* 1997, 278: 817. Copyright 1997 American Association for the Advancement of Science.)



Figure 3. Interdisciplinary collaboration was essential in solving the complex chaperone puzzle: Northwestern crystallographer Amy Rosenzweig (left) solved the crystal structures of the chaperones; Jim Penner-Hahn from the University of Michigan, utilized synchrotron-based spectroscopies to study metalloproteins; yeast geneticist Val Culotta (below) of Johns Hopkins cloned Atx1 and CCS.



enzyme required for iron metabolism in yeast. My colleagues and I were able to show that Atx1 plays a key role in distribution of copper to the compartment containing Ccc2 (figure 2). As for another copper chaperone, yCCS, which we will discuss later, it doesn't take copper to a compartment, but rather to an enzyme that works in the cytosol. The CCS chaperone directly inserts the copper into superoxide dismutase (SOD). There is also another possible Cu-chaperone, Cox17, which I won't discuss in detail. It is thought to give copper to that key energy generating enzyme, cytochrome oxidase.

As I take you through this detective story, I want you to realize that I'm a bioinorganic chemist and molecular biologist; however, that breadth is not sufficient to nail down this problem. Our collaborations with a yeast geneticist, a spectroscopist and a crystallographer, all professors, allowed rapid progress. The geneticist is Val Culotta of Johns Hopkins University, a fabulous yeast geneticist, who, incidentally, trained a Westinghouse scholar a few years ago. The spectroscopist is Jim Penner-Hahn, a master of synchrotron-based spectroscopies at the University of Michigan. Amy Rosenzweig, a new colleague of mine at Northwestern, solved the crystal structure of these chaperones (photos, figure 3). There are also several talented graduate students and postdoctoral fellows whose discoveries I will discuss (photo, page 69).

Val Culotta first cloned the Atx1 gene as a high copy

number suppressor of oxidative stress. This was a puzzling property and she needed to know more about its normal function. Her genetics provided the first critical clues in this biological mystery. She found that the amino acid sequence of the Atx1 protein was similar to a protein involved in mercury resistance protein. Atx1 was also similar to domains of the Menkes and Wilson's disease proteins.

What is a mercury resistance protein? In bacteria, there is a small protein called MerP that sits outside the cell proper, in the periplasmic space. Its role is apparently to hand off mercury to a transporter that gets the cation through the greasy lipid bilayer and eventually into an enzyme that can reduce Hg(II) to its elemental form. Elemental mercury is volatile and evaporates from the medium, thus protecting the bacterium. Val tried to see if the Atx1 gene would confer mercury resistance in yeast but found immediately that Atx1 didn't protect the yeast against mercury poisoning. My group had been studying related proteins, so she called me and said, "let's think about what to do with this protein." Slowly we began to make the connection with copper and iron biology and carried out a series of experiments that ultimately led us to characterize Atx1 as a copper chaperone.

### Protein structure solved

Bob Pufahl, a NIH postdoctoral fellow in my group, purified and loaded mercury and copper into Atx1, using these results we were able to characterize several key aspects of its chemistry using spectroscopic tools. David Huffman, another postdoctoral associate, and Amy Rosenzweig were able to crystallize Atx1. Amy then determined the structure of the Hg-bound form of Atx1 at better than 1 angstrom resolution. Amy skilfully accomplished this feat using direct methods. Atx1 is a small protein, only 72 amino acids, but it's the largest one that has ever been solved by direct computational methods. David then showed that Atx1 could transfer Hg(II) to CCC2 (figure 4).

What Amy was able to show was that the mercury was bound to two cysteine sulfurs, in a loop between an alpha helix and a beta strand. Furthermore, as we look at the crystal structure we found that the two helices and a four-strand beta sheet were very similar to the structure of MerP. This common unit has been conserved through evolution.

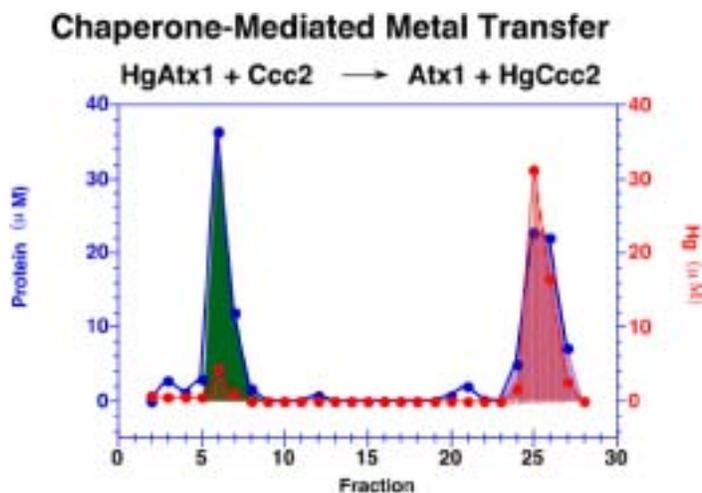


Figure 4. Chaperone-mediated transfer of mercury from Atx1 to Ccc2 was charted using spectroscopic tools.

In fact, none of us in this field have been able to get copper to crystallize with these proteins. The NMR group that solved the structure of a related domain of the Menkes protein chose to use silver ions instead of copper. This is part of the reason for my subtitle, "Chaperones are important, but are they easily fooled?"

We've been able to put a variety of metal ions into apo-proteins in the test tube, so how does the cell know which metal to put into which chaperone? Well, a hypothesis that we tested was that Atx1 was loaded with copper (by still unknown mechanisms) and then acted as an escort to deliver Cu(I) to Ccc2, the yeast homologue of the Menkes disease protein. We thought that it would then dock and there would be a handshake, followed by metal transfer. In the latter step, Cu(I) would exchange ligands. The chaperone's cysteines would release the ion to cysteines in the acceptor. Ultimately, Atx1 would be released and engage in another cycle. The copper would ultimately be moved into the vesicle where it could diffuse into the active site of the final target, Fet3 (figure 5).

### Establishing copper binding mechanisms

It's easy to say that Atx1 binds copper ions (most any protein will), but we actually showed that its preferred oxidation state is Cu(I). That told us for the first time that the important oxidation state of copper in intracellular transactions is Cu(I). I'll now tell you a few of our tests of this mechanism. Establishing a mechanism in vitro and showing that it works in the cell are two different things: chemists can put any metal into any protein, but that doesn't mean that has anything to do with the way the cell does it. Likewise, based on the phenotype of a mutation, geneticists can postulate a hundred mechanisms. But only when

we combine the genetic and chemical results and make rigorous tests back and forth can we establish a reasonable mechanism.

I would like to focus for a moment on two issues: does Atx1 actually transfer copper, and then what's the chemical driving force for this step? Is this thermodynamically downhill? Is the equilibrium constant  $10^2$ ,  $10^4$ ? Is it 1? To answer this we excised one of the target domains of Ccc2 using genetic engineering methods and purified it. We had reason to believe the isolated domain would have a similar structure as a separate unit to its structure when attached to the entire protein.

### Transferring copper between proteins

David Huffman mixed the chaperone protein and the acceptor domain and looked at the distribution of copper between them. A surface representation of the crystal structure showed the positively and negatively-charged regions on the surface of the protein. It was clear that this chaperone protein had a vast abundance of *positive charge*. The acceptor domain to which Atx1 gives copper is predicted to have, if it folds the way we think it does, a significant excess of *negative charge*. This difference in charge allowed a simple separation of the mixture using ion exchange chemistry.

David loaded one protein with copper, mixed this with the apo-form of its partner, and then asked what is the distribution of metal ion between the two? He worked out an analytical method to determine the equilibrium constant. The copper Atx1 elutes quickly because it doesn't have much anionic character (figure 6). Initially we started

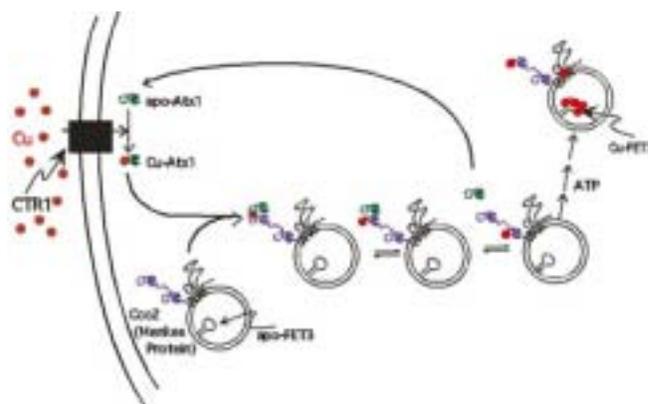


Figure 5. A hypothesis was that Atx1, loaded with copper, would escort it to Ccc2, where there would be a handshake followed by metal transfer. The Atx1 would be released to engage in another cycle, and the copper would be moved into the vesicle where it could diffuse into the active site of the final target, Fet3.

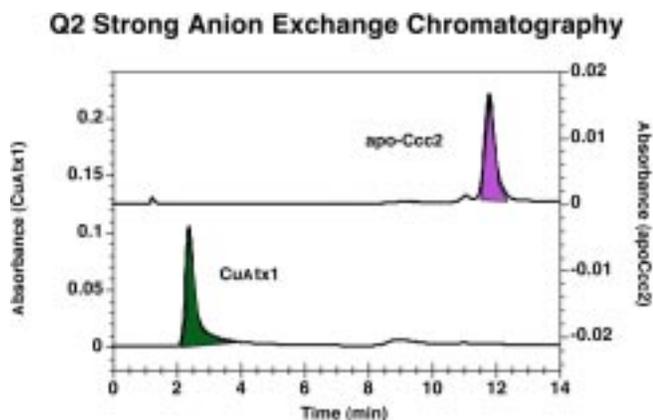


Figure 6. Copper in Ax1 elutes quickly. Initially all the copper was in the Atx1 peak. After mixing there was a distribution in the two proteins, with about half of the Cu(I) moving from the acceptor to the donor.

off with all the copper in the Atx1 peak. After mixing there was a distribution of copper in the two proteins. About half of the Cu(I) was moved from the acceptor to the donor, and it also occurred extremely rapidly, and with small equilibrium constant of about one. This means that in a mixture, we'd get about a 50-50 distribution at equilibrium of copper between the two proteins.

What is the driving force for the chaperone protein to diffuse and release its copper to its particular partner? How is the specificity achieved? The answers to these questions are not in; however, we are testing a model in which the chaperone protein docks in a specific orientation with its target (figure 7). In this arrangement of donor and acceptor ligands for copper the activation barrier for metal transfer is much lower. Based on this model we predict that a rate of transfer to the acceptor protein is much faster than it would be if copper were to be transferred to a non-partner protein.

Enzymes work not by perturbing the equilibrium position of a given reaction, but by accelerating the rate at which equilibrium is achieved, and we think that's exactly how these copper chaperones work. We have proposed a chemical mechanism that allows the chaperone to hold on tightly to the Cu(I). Once the protein docks with its partner in the proper orientation, however, we envision that the chaperone rapidly releases some ligands to the metal ion and transfers it to the acceptor domain, even though that acceptor domain doesn't have a much higher affinity for copper. A rapid movement of the copper ion back and forth between partners, and only between partners, ensues.

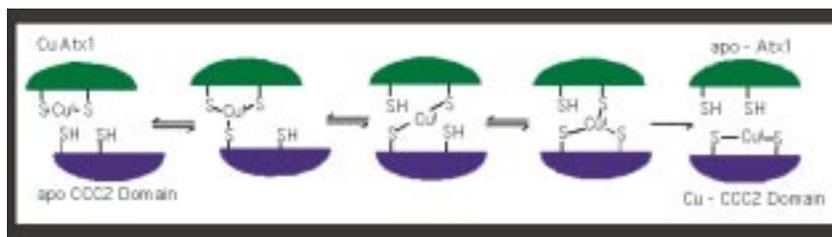
### ATP drives copper trafficking

In this model the overall driving force for Cu-trafficking is ATP (adenosine triphosphate, the nucleotide important in biological energy metabolism) hydrolysis. The Menkes protein is a pump that uses ATP hydrolysis to move the cation through the lipid bilayer. Remarkably, mercury, which binds tightly to sulfur compounds, is readily transferred by the chaperone in the same way (figure 4). This was very reassuring to us because we were using mercury as a mimic for copper. It also gave us a first and important clue as to why mercury is particularly toxic. We think that it gets in on these copper transport pathways, and then finds its way to active sites where the cell is expecting a redox active metal. Mercury can't do redox chemistry, and so it then disrupts the function that Cu normally carries out.

I want to go briefly into a new story on the other copper chaperone, yCCS. This protein directs copper to superoxide dismutase (SOD). When I started teaching, I did my demonstrations at the beginning of the lecture because I wanted to get them out of the way: I could then get into the meat of the lecture. Well, of course, people would nod off and went to sleep right after the demo. So I started waiting and saving them for the middle of the lecture; this is why I will take a detour for a minute to talk about SOD: what is SOD, and why is it so important?

It turns out that SOD is antioxidant protein. Mutations in this protein, however, give it a more ominous overtone, and place it at the heart of a neurological disorder in humans, specifically amyotrophic lateral sclerosis or ALS. The substrate of SOD is the superoxide anion, the one electron reduction product of  $O_2$ . It is reactive and can be destructive to metabolic enzymes, so cells rely on SOD to catalyze the disproportionation of superoxide to  $O_2$  and hydrogen peroxide. SOD accelerates this reaction ( $H^+ + 2O_2^- \rightarrow HOO^- + O_2$ ) by about four or five orders of magnitude. Enzymes are not affecting the equilibrium position; the copper enzyme SOD is affecting the rate at which equilib-

Figure 7. How does a chaperone protein diffuse and release its copper to a particular partner? This model shows chaperone proteins docking in a specific orientation with target. In this arrangement of donor and acceptor ligands for copper, the activation barrier for metal transfer is low, permitting a fast rate of transfer.





Tom O'Halloran combines chicken blood and hydrogen peroxide which produces water and dioxygen but results in a "foaming, milkshake-like mess." (Figure 8).

rium is attained. Copper ions in solution can achieve this without the assistance of a protein scaffold, so it is not a difficult chemical reaction to carry out.

The latter fact leads to a conundrum about the role of the enzyme and debates in the field as to what was the true function of SOD. If simple copper complexes can catalyze this reaction, why does the cell need SOD (much less so many copies of this specialized enzyme)? The answer to that question depends on the availability of copper ion in the cell, and, as I will show you, our best evidence indicates that there are no free copper ions in cells under normal conditions. But as a prelude, I want to discuss a puzzling aspect of SOD: it converts one dangerous molecule into another, namely superperoxide into peroxide.

### How reactive is hydrogen peroxide?

Based on our every day experience, we know peroxides as reactive materials, and these should be just as dangerous to a cell as superoxide. Peroxide is used to bleach hair. It can also be used to kill bacteria. You pour it in wounds, and it bubbles vigorously. Why does the cell make this toxic substance? The cell's response is to build another catalyst, an enzyme called catalase. I have to do a little demonstration that illustrates its function (figure 8). If I were giving a general chemistry lecture, the students would wake up at this point: not only is there a break in the routine, there is the distinct possibility that the professor will make a fool of himself. While we have fun with these, I always slip

some demonstration questions onto exams.

This demonstration uses chicken blood, and blood contains the enzyme catalase, another metalloenzyme. It takes hydrogen peroxide, and like SOD, catalyzes a disproportionation reaction. In this case it takes peroxide to water and  $O_2$  ( $2H_2O_2 \rightarrow 2H_2O + O_2$ ). I'm going to add a solution of 30 percent hydrogen peroxide. This concentration is about a tenfold higher than a solution you might apply to an open wound to kill infectious organisms. What's going to happen when the catalase and hydrogen peroxide mix? Oxygen will be evolved, and the gas will expand rapidly. We know that the free energy change is typically negative for production of gas from a condensed phase material, therefore some heat evolution is possible as well. You can guess what's going to happen—we get a foaming, milkshake-like mess.

This is the type of thing that my ten-year-old son, Victor, just loves to do in the kitchen with whatever he can find. This mess is pinkish-white, and it not unlike what you would get when you take egg whites and whip them up. When my son does this, some of the egg white albumin, which is similar to the serum albumin (an abundant constituent of blood), is physically denatured. I say some of it, because with Vic at the whisk, the egg whites often end up escaping, ending up anywhere but the bowl!

What is happening here? Heat and oxygen gas are evolving—you can feel that the test tube is a little warm—causing among other things denaturation of the albumin proteins. The foam is not blood red or nearly as pink as you might guess because the hemoglobin is in the red cells, and not as much of it is denatured; there's far more albumin nearby. Other agents can catalyze the same type of reaction. Since this is a metals meeting, I'll show you another metal catalyst for this process—and this time you might have to move back a little bit.

This is the same hydrogen peroxide stock solution, but this time I will add a classic solid-state catalyst: manganese dioxide. This is not going to dissolve, but it's going to be a heterogeneous catalyst for the disproportionation reaction. I could add other oxides, but unless they're redox-active, this reaction won't occur. It doesn't matter how much of the catalyst I add, a little is enough to create a plume of steam. This reaction is often called the genie-in-a-bottle. We typically do it on a larger scale with a teapot in the big classrooms, and the manganese dioxide in a teabag—all you do is lift the lid and instantly you get a dramatic plume of steam that goes up about 5 or 10 feet. What you see is simply evolution of oxygen gas, steam and the heat that was released with it.

### Copper: poisonous yet essential

Before getting back to the chaperones, let me fill in a



Figure 9. Joan Valentine of UCLA is another scientist at the forefront of research into copper trafficking in cells. She characterized the copper center in SOD as a "wolf in sheep's clothing".

few more aspects of copper biology. I have told you that copper is very poisonous, and indeed, one of the best antifungal agents in widespread use. The copper-containing Bordeaux mixture, for

instance, is extensively used on grapevines and tomato plants to prevent rot and fungal growth on the leaves. If you use Miracle-Gro, you are employing copper in the same role: this product gets its blue color from copper sulfate. Thus copper presents a dilemma: it is incredibly toxic, and yet is also a really useful cofactor in many enzymes, including superoxide dismutase.

Copper has often been implicated as one of the key agents in Lou Gehrig's disease (ALS). It is the chemistry of copper that has led my colleague, Joan Valentine at UCLA (figure 9), to classify the copper center in SOD as a "wolf in sheep's clothing." In her analogy, dangerous copper chemistry is typically held in check in the normal enzyme. However, this copper ion is envisioned to be at the center of trouble in the disease-causing mutations in SOD. In this model, the mutant proteins are thought to lose control of reactions catalyzed by copper. Other models are being evaluated by teams around the world, and Joan is right up in the front lines, doing some of the best research in the field. I just shudder, however, to think of these nice transition metal ions as "wolves" under normal conditions! Instead, I argue that they need some protection from other components in the cell. This protection is provided by metal ion chaperones.

A chaperone, by Webster's definition, accompanies a young person or a group of unmarried people to insure propriety. Thus chaperones are conducive to proper interactions between young men and women during courtship. But have you ever considered whom such a chaperone is protecting? Are chaperones protecting society? Or are they protecting the more innocent young person from the untoward advances of the more rambunctious elements of society? Are they protecting the individual or the system?

And so we asked, in a facetious manner, the same question of the copper chaperones. Are these copper chaperones there to scavenge and keep copper from destructive chemistry that could damage or denature other types of biopolymers? Do they provide a direct protective function for the cell? Or is it the innocent copper ion, that poor

wolf, that needs protection from intracellular chelating sites that would otherwise prevent the metal ion from being bound by the apo-form of SOD? Is the chaperone's function to guide and protect the metal ion or to protect the cell?

To explain what my colleagues and I found, I need to provide you with some background on the newest copper chaperone and the partner for superoxide dismutase, namely CCS. CCS is significantly bigger than the Atx1; it has three domains. One of these is similar to Atx1. Based on amino acid sequence and proteolysis experiments, we've shown this to be a three-domain structure (figure 10). It also has a region that is similar to its target, SOD, and finally it has a third domain that is essential for the CCS function. The last domain is small and has two cysteines that can bind Cu(I). I won't go into the possible mechanisms, but I will describe an astounding aspect of cell biology of that we have deduced from studies of this copper chaperone.

In this line of experiments we first asked, "how does this chaperone work? Does it just simply transfer a metal ion to the target?" These questions could not be directly resolved by genetic experiments, so we purified the CCS protein and its target, SOD. We then removed all metal ions from SOD, making the apo-form of the enzyme. Next we loaded Cu(I) into CCS, mixed it with apo-SOD and looked for enzymatic activity characteristic of the Cu-loaded SOD. We did this two ways: in one case we included copper chelators (such as ethylenediaminetetraacetic acid [EDTA]) in the reaction, and in the other case we left it out (figures 11 and 12).

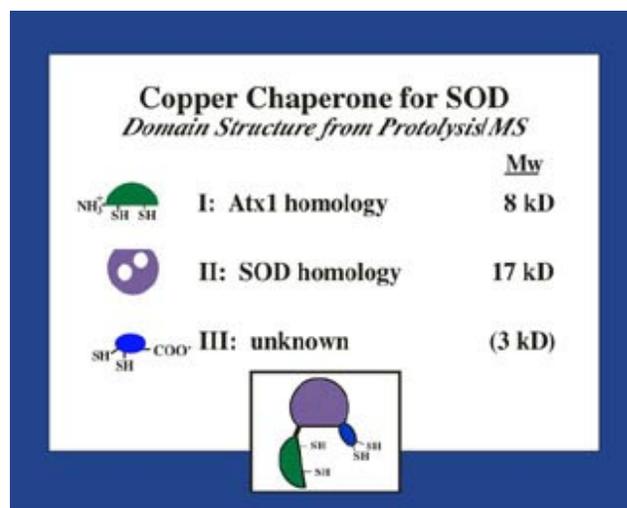
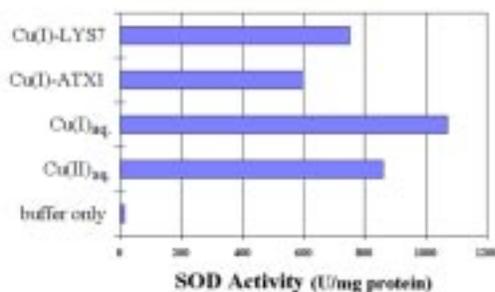


Figure 10. The newest copper chaperone is CCS, which has three domains, one similar to Atx1, one similar to its target (SOD) and the third is not similar to other known protein sequences.

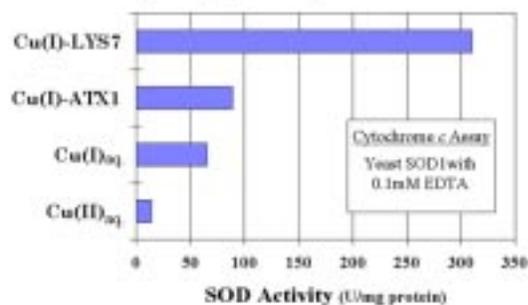
**In Vitro Activation of Apo-Superoxide Dismutase  
by Various Copper Sources (no EDTA in sample)**



(T. Roe & R. Puffikl)

Figures 11 and 12. Adding a copper chelator (EDTA) to CCS protein and the apo form of SOD provided a way to test how the CCS chaperone transferred copper to its SOD target.

**In Vitro Activation of Apo-Superoxide Dismutase  
by the Copper Chaperone LYS7**



(T. Roe & R. Puffikl)

The idea was to test whether the chaperone has to dock directly with the SOD. If the chaperone first releases copper ions into solution, the copper scavenger can prevent it from activating the enzyme. We were amazed to find in our first experiment without EDTA that the chaperone did not work better than simple copper salts. However, when we included EDTA, which kept the concentration of the free-metal ions below about  $10^{-16}$  molar, we found that the copper-loaded chaperone was a significantly better Cu-donor than any other source of copper. This was an early result, and as we refined the assays the controls have gone down to negligible levels of activity and the chaperone-induced activation has stayed very high. The genetic studies of Culotta and coworkers had told us that CCS was necessary. The biochemical studies revealed that CCS could do this reaction directly but that it was only necessary when copper concentrations were severely limited.

Since CCS is always necessary for activation of SOD in the cell, these results led us to the hypothesis that free

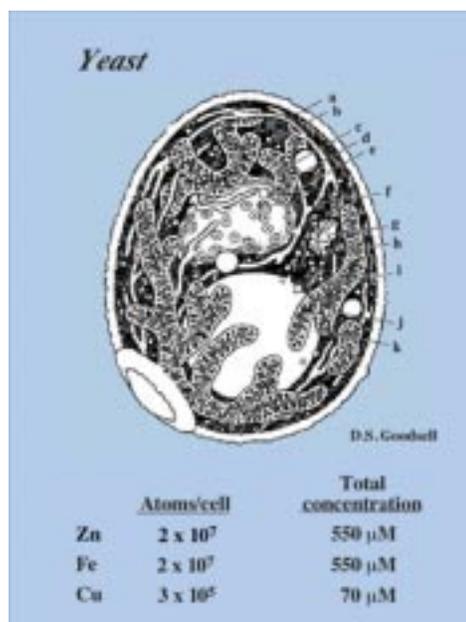


Figure 13. Free copper is severely limited in the cytoplasm of a yeast cell amounting to less than one free copper atom per cell, even though the concentration of total copper is about 70 micromolars (iron and zinc shown for comparison).

copper concentration must be severely limited in the cytoplasm of the cell. To test this and other possibilities we first measured the total number of atoms of copper per cell. These are some of the numbers I mentioned before: total copper is about 70 micromolar in the yeast cell, and the total SOD protein concentration is about 10 micromolar (figure 13). When you kill the chaperone gene, knock it out completely (this is the Culotta experiment that I mentioned earlier), you still find the same number of SOD molecules per cell. You also find the same number of copper atoms, but none of that copper is present in the enzyme. When the chaperone gene is eliminated from yeast, we find the number of SOD molecules stays the same, but now there is at least two orders of magnitude less of the copper-loaded form of SOD. In fact, we can't even detect any active Cu,Zn-SOD: we relied on a carefully determined detection limit to set the concentration of Cu-SOD at less than 300 molecules per cell.

Not only can free copper activate SOD, but this enzyme has an extraordinary affinity for copper. So we asked ourselves, if this free SOD is sitting around in the absence of the chaperone, and we know an upper limit for the ratio of Cu-SOD to total SOD, can we estimate an upper limit for the concentration of the copper in the cytoplasm of the cell? This is a question for general chemistry students. With the estimated upper limit value for that ratio, we

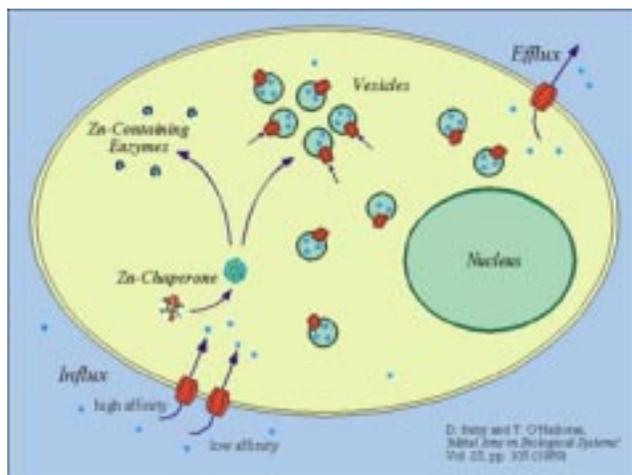


Figure 14. Cytoplasmic chemistry of zinc is an emerging field: vesicles that store zinc are abundant in neurons, and special neurons have been found in the brain that release zinc into the synaptic cleft.

obtain a copper concentration that is less than  $10^{-18}$  molar.

### No free copper in normal cells

If we take the volume of the yeast cell and divide this through—as suggested by my former mentor, Steve Lippard—we find that this “concentration” of copper corresponds to about  $10^9$  atoms of copper per cell. I am happy to consider this to be less than one atom of free copper per cell. This tells us that free copper ions are not frequently present in the cytoplasm normal cells. It further suggests that there is a significant overcapacity for chelation of copper in the cell. There must be multiple processes that bind the copper and prevent it from ever being randomly available.

For those of you who aren't swayed by simple thermodynamic arguments, I offer a kinetic consideration. If you estimate the rate at which copper would bind to SOD and plug in two limiting possibilities, we calculate that at most one or two of the 70,000 copper atoms in a yeast cell dissociate and become free ions in the lifetime of a yeast cell. This is far less than the amount required for the pool of apo-SOD proteins.

The results of these “back of the envelope” calculations frame the results I've been telling you about this morning. They also explain why copper chaperones are conserved from yeast to humans. Based on these results, I argue that the role and primary function of the chaperones is to protect copper from the host of competing binding sites in the cell. The copper chaperones then find the target, deliver and directly insert this essential cofactor it to the correct apo-protein. In collaboration with Culotta's group, one of the proteins that competes for intracellular copper has been identified, namely metallothionein.

Here are some other conclusions we've come to: chaperones really don't protect the cell from copper toxicity in a direct manner, there are other proteins that work with the chaperones. Once the Culotta group clones these, we will begin evaluating their role in copper trafficking. Before copper can be a “wolf,” it must be a wolf cub: the chaperone's job is to protect the cub so that it can get to its functional site in a protein and do its job. There's very strong evidence now for this concept of no free copper in the cytoplasm, and I would suggest to you that this situation isn't unique to copper. It may well be true for zinc and iron as well: the cytoplasm of the cell is not a buffer with metal ions available for the taking. I anticipate that every event of delivery of cofactors, particularly reactive ones, is a multistep process with several players.

### Chemistry deconvolutes biology of metals

The cytoplasmic chemistry of zinc is just emerging. There are vesicles that store zinc (figure 14), and these are abundant in neurons. Others have found what they call “zincergic neurons” in the brain that may release zinc into the synaptic cleft. The cell biology of transition metals such as copper and zinc is in its infancy. The discipline of chemistry is absolutely central—it plays a central role in deconvoluting the biology of metals. The photo (figure 15) shows a chelating ligand that only gives off light when it binds with zinc. When you put this in cells as a chemical probe, it lights up specific organelles. We think these are vesicles that pick up zinc, transport it, and use it in various specific processes in the cell. We are now starting to localize these processes in the cell.

This is where the perspectives of physics and materials science communities are valuable; it would be wonderful to take a cell like this, which is 10 microns, map out the

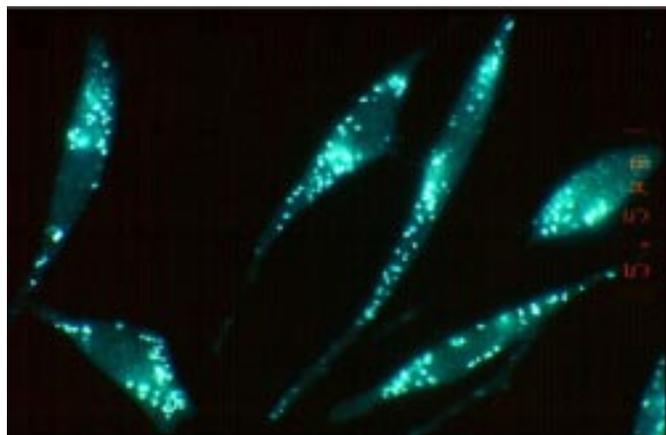


Figure 15. Microphotograph shows results of a chemical probe that fluoresces when it binds to zinc. Put into cells, it lights up organelles that are thought to be vesicles that use zinc in cell metabolism.

elemental distribution, and define the places in the cell where metal concentration is changing. We're beginning to do some experiments like this at Northwestern.

I'd like to thank you again for inviting me to share some of my antics, demonstrations and some discoveries that my colleagues and I have recently made. This detective story is still in the early stages of development so don't take these mechanisms as the complete solution. Rather, encourage your students to test them and take them apart.

### Selected References

- Pufahl, R.A. and C.P. Singer, K.L. Peariso, S.-J. Lin, P. Schmidt, V. Cizewski Culotta, J.E. Penner-Hahn, T.V. O'Halloran, "Metal Ion Chaperone Function of the Soluble Cu(I) Receptor, Atx1," *Science*, 1997, 278, 853-856.
- Valentine, J.S. and Gralla, E.B., "Delivering Copper Inside Yeast and Human Cells," *Science*, 1997, 278, 817-818.
- Culotta, V.C., Klomp, L.W.J., Strain, J., Casareno, R.L.B., Bernhard Krebs, and Gitlin, J.D. "The Copper Chaperone for Superoxide Dismutase," *J. Biol. Chem.* 1997, 272, 23469-23472.
- Klomp, L.W.J., Lin, S.-J., Yuan, D.S., Klausner, R.D., Culotta, V.C., and Gitlin, J.D. "Identification and Functional Expression of HAH1, a Novel Human Gene Involved in Copper Homeostasis," *J. Biol. Chem.*, 1997, 272, 9221-9226.
- Pufahl, R.A. and O'Halloran, T.V. "Mechanisms of Copper Chaperone Proteins" in *Metals and Genetics*, B. Sarkar, Ed., New York, Plenum Publishing Corporation, 1999.
- Rae, T.D., Schmidt, P.J., Pufahl, R.A., Culotta, V.C., and O'Halloran, T.V. "Undetectable Free Intracellular Copper: the Requirement of a Copper Chaperone for Superoxide Dismutase," *Science*, 1999, 284, 805-808.
- Lippard, S.J. "Free Copper Ions in the Cell?" *Science*, 1999, 284, 748-749.
- Portnoy, M.E., Rosenzweig, A.C., Rae, T., Huffman, D.L., O'Halloran, T.V., and Culotta, V.C. "Structure-Function Analyses of the ATX1 Metallochaperone," *J. Biol. Chem.*, 1999, 274, 15041-15045.
- Rosenzweig, A.C., Huffman, D.L., Hou, M.Y., Wernimont, A.K., Pufahl, R.P., and O'Halloran, T.V. "Crystal Structure of the Atx1 Metallochaperone Protein at 1.02 Å Resolution," *Structure*, 1999, 7, 605-617.
- Lamb, L.L., Wernimont, A.K., Pufahl, R.A., O'Halloran, T.V., and Rosenzweig, A.C. "Crystal Structure of the Copper Chaperone for Superoxide Dismutase," *Nature Structural Biology*, 1999, 6,
- Schmidt, P.J., Rae, T.D., Pufahl, R.A., Strain, J., Hamma, T., O'Halloran, T.V., Culotta, V.C. "Multiple Protein Domains Contribute to the Action of the Copper Chaperone for Superoxide Dismutase," *J. Biol. Chem.*, 274, 23717-23725.



## Sex, Violence and Drugs in the World of Insects: A Chemist's View

Adapted from a 2000 presentation by Jerrold Meinwald

Jerrold Meinwald's research has covered a broad range of topics including molecular rearrangement mechanisms, the synthesis and reactions of highly strained ring systems, organic photochemistry, natural product structure and synthesis, stereochemistry of inhalational anesthetics, and insect chemical ecology.

Some of his most recent research has been with the synthesis of biologically active compounds. For example, recent reports of antiviral activity among the lucibufagins—firefly defensive agents, which Dr. Meinwald discovered more than a dozen years ago—have prompted him to study the total synthesis of these steroids. His lab is also developing chiral probes into the mechanisms of anesthesia and its side effects.

Besides teaching organic chemistry and chemical ecology, he initiated a course, "The Language of Chemistry," which introduces non-science students to the basic ideas of chemistry by examining some of the roles played in nature by organic chemistry.

After receiving a Ph.D. from Harvard, a DuPont Postdoctoral Fellowship brought Dr. Meinwald to Cornell where he has been a teacher and investigator since 1952. He was a founding member of both the International Centre of Insect Physiology and Ecology in Nairobi, and the Cornell Institute for Research in Chemical Ecology.

**T**onight I am going to tell you some fascinating stories of chemical communication and of chemical defense mechanisms as related to the lives of insects. I will describe insects' use of chemistry for defensive purposes and for communication purposes, and finally delve a little into politics and discuss the urgent need for conservation. There is a piece of Italian Renaissance pottery that depicts God surrounded by the animals he's just created. A careful look, however, reveals a very serious omission: there are no insects in the picture. That's rather curious, because if you look at the numbers of described species that live on Earth, over half of all these species are, in fact, insects. A pie chart (Figure 1) of all known species reveals that the arthropods, including not only insects, but also millipedes, arachnids and crustacea, comprise about two-thirds of all the described species on our planet. Vertebrates are a hardly significant portion of the pie diagram. So there is something about the design of insects which has made them enormously successful.

Insects have the greatest species diversity and exist in greater numbers than any other creatures. Some of them are household pests, some are disease vectors, some are agricultural pests, but most lead lives about which we know almost nothing. There are many reasons why they've flourished, including a number of biological adaptations. But one underappreciated reason for their success is the fact that insects are very good organic

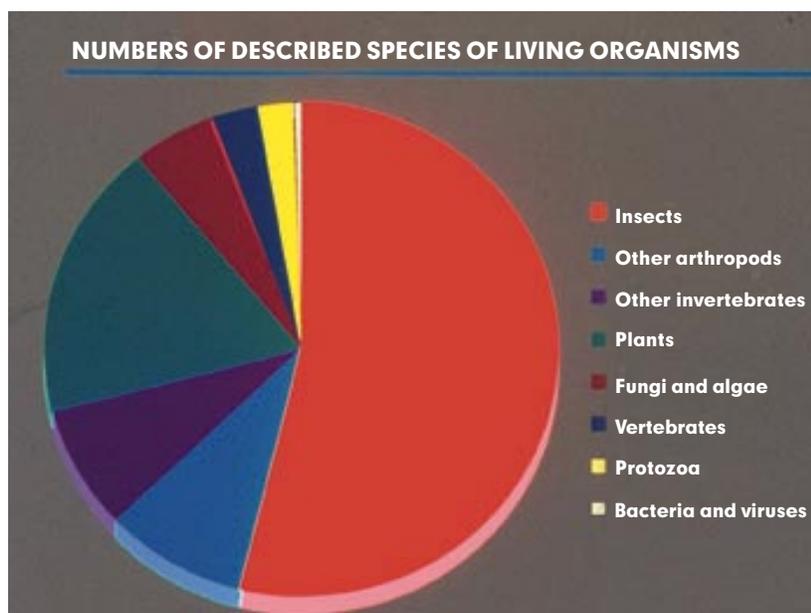


Figure 1. Thanks to successful survival strategies, arthropods—including insects, millipedes, arachnids and crustacea—comprise about two-thirds of all the species on our planet.

Figure 2. The whip scorpion or vinegaroon (right), *Mastigoproctus giganteus*, has evolved a simple but effective chemical defense over millions of years. (Photo [and hand]: Thomas Eisner)



Figure 3. Pinching any leg of the whip scorpion causes it to release a powerful defensive spray containing 84% acetic acid, 11% water and 5% caprylic acid. (Photo: Thomas Eisner)

chemists. They have two particularly useful survival skills: one is that they can defend themselves chemically, and the other is that they can communicate very well by sending chemical signals.

### A scorpion's simple but effective chemistry

The whip scorpion is a native of Arizona (Figure 2). In the Southwest, these spiders are called vinegaroons and they're very interesting creatures. I want to tell you about this scorpion for several reasons: one is that the photo of this arachnid includes the hand of my colleague Tom Eisner. Almost all of the work that I will talk about tonight is collaborative with Tom, whom I've worked with for over four decades at Cornell University—he's the biologist half of our team and I'm the chemist half. The vinegaroon is a stingless scorpion with a remarkably simple but effective chemical defense. For the purposes of this talk it will be considered an honorary insect.

From a turret at the abdominal tip of the animal, it can spray a defensive material, which we characterized in our first collaborative research. If one of these whip scorpions (Figure 3) is affixed to a steel rod, and an appendage is pinched, it will release a well-aimed and noxious defensive spray. The scorpion can do that around twelve times before the supply of chemicals runs out, and no matter which leg is pinched, it will spray in the right direction. It's actually quite



easy to study the chemistry on this sort of secretion. A test tube is put over the scorpion's rear end, a leg is pinched, and the defensive material is collected.

I mentioned that these scorpions are called vinegaroons, and they're really *super*-vinegaroons. The vinegar that we put on salads is 4 to 6 percent acetic acid, however, the major constituent of this defensive spray is 84% acetic acid combined with a small amount of water (11%). Interestingly, we prepared an 85% acetic acid solution in water and tried to see if it was effective as a spray against insects. The answer is that it was not. In fact, it's been discovered that in delivering pesticides to insects, a component must be added that will help the active components permeate an epicuticular wax layer, and that's what this whip scorpion has done. In its defensive strategy, this ancient animal has developed a very clever trick which has to do with adding 5% of a longer chain fatty acid, caprylic acid, which is lipophilic—it dissolves the epicuticular wax layer that protects insects from outside chemicals. This is an animal that's approximately 400 million years old, and to survive it has learned this one trick: the very simple chemistry of adding an agent that will get an acid through a wax layer. It is a simple and lethal defense.

One is not going to get a very excited response from organic chemists by characterizing acetic acid. But as far as the insect world goes, there's a lot of simple chemistry that is very effective in meeting important needs. We've looked, for example, at millipedes who make hydrogen cyanide defensively; it actually serves as a very good defensive agent. Once again, from an organic chemist's point of view, hydrogen cyanide is not tremendously interesting, although find-



Figure 4. The squash beetle, *Epilachna borealis*, an agricultural pest that lives on zucchini plants. (Photo: Thomas Eisner)

ing a way to store hydrogen cyanide in the form of an innocuous precursor is an important challenge that this millipede has met. However, on the other end of the scale is a squash beetle, *Epilachna borealis*, which has a very sophisticated chemical defense.

#### Beetle pupa employs complex chemistry

These beetles live on zucchini plants, they're an agricultural pest (Figure 4). The particular chemical defense I want to describe is not of the adult, but of the pupal stage, which is a stage between the larva and the adult. The pupa is especially in need of defense, because it can't run away from



Figure 5. Above, the pupal stage of the squash beetle is effectively protected by the secretion on its spiky defensive hairs. (Photo: Thomas Eisner)

Figure 6. A novel class of defensive compounds is contained in the poisonous droplets at the end of each hair of the squash beetle pupa. (Photo: Thomas Eisner)

predators the way an adult could. The pupa (Figure 5) has glandular hairs over most of its surface. At high magnification, these glandular hairs can each be seen to support transparent droplets at their tips (Figure 6). A scanning electron micrograph of the hairs (taken by Maria Eisner) is quite dramatic, and was featured as a cover picture in *Science* (17 July 1998) when our work on these pupae was published. The hairs are really fearful-looking defensive structures. If an ant comes along and tries to bite the pupa, the ant gets a mouthful of those droplets and is immediately repelled.

What are these compounds that the pupa is making to defend itself? The chemistry of the droplets turns out to be enormously complicated. An early attempt was made to carry out a high performance liquid chromatographic (HPLC) separation which normally separates minute amounts of chemical components very nicely one from the other. But this defensive material is such a forest of compounds that it just showed one smear that did not resolve the separate components in a useful way. Finally, we combined this chromatographic technique with selective ion monitoring in a mass spectrometer (MS), and we were able to search through the forest, and in ways that I won't explain, we could discern that there was a group of components with molecular weights of 626, 640, 654, 668, and 682 (Figure 7), another group with molecular weights in the 800s, etc. And, indeed, when we searched by molecular weight we found discrete sets of compounds coming out of the chromatograph, with higher and higher molecular weights.

We soon realized that the beetle was making what is called a "combinatorial library" consisting of many hundreds of closely-related organic compounds all derived from three closely-related molecular building blocks which were combined to form sets of very large rings. The rings are made



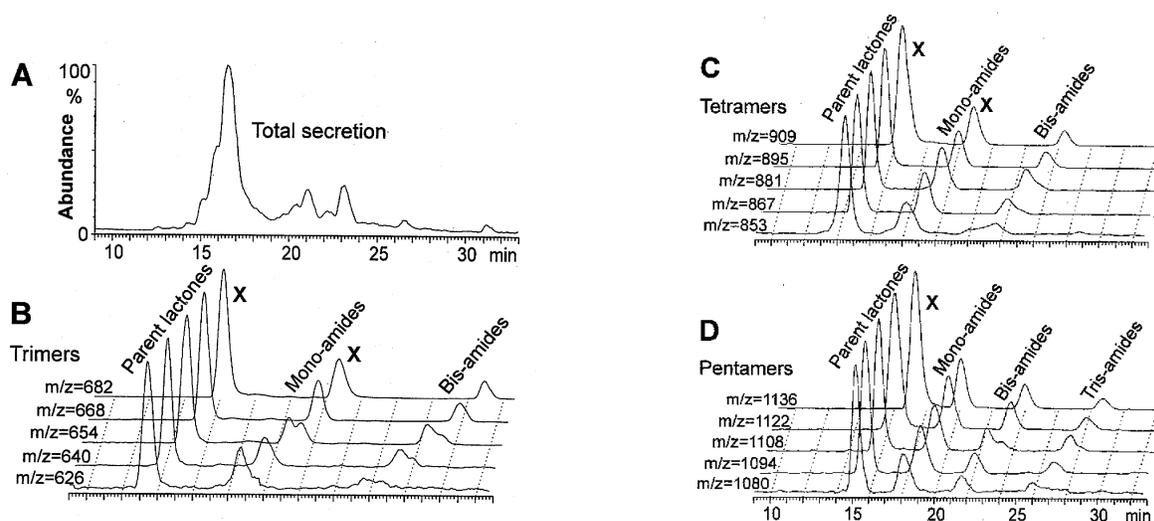


Figure 7. HPLC/MS data characterizing *Epilachna borealis* defensive secretions.

from three hydroxyamino acids—the dominant ring size is 42 atoms, produced by combining three 14-atom chains. There are components that use four building blocks, others that use five building blocks (Figure 8), and the ring sizes go up to about 200-membered rings. These are probably the largest “necklaces” of atoms found in nature, other than some cyclic DNAs. So these very unusual molecules are what these particular squash beetles are using to defend themselves. We’ve synthesized several of these rings, and they constitute a genuinely novel class of defensive materials.

I want to tell you one more story about insect defense, then I will shift gears and look at an entirely different area. This is an older story, but it’s interesting because it says

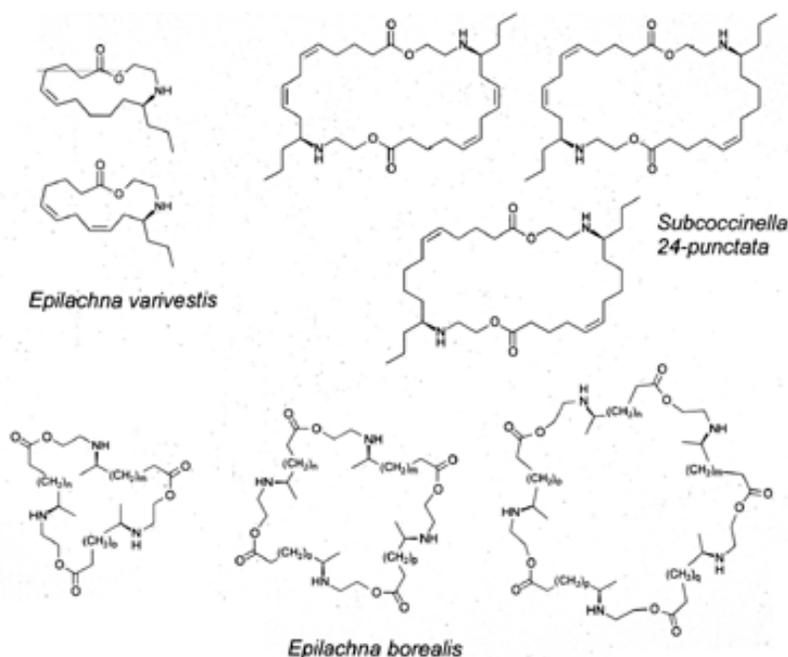
something about the way chemistry can lead to the understanding of interactions that are going on in nature.

### Bad taste allows blatant advertising

A photo of the paths of flying male fireflies was taken by a colleague from inside a Bangkok hotel room. He just left the camera shutter open, and caught the light trail of individual males flying along. The photo shows how the firefly males turn their lanterns on and off in flight. The photochemistry of that light production has been well worked out, but I want to relate an even more interesting aspect of these signals: their role in firefly mating behavior. These are courtship signals, and female fireflies recognize males of the same species by the dark periods in between flashes. The way firefly courtship works—which I’m sure all of you are quite curious to know—is that the males identify what species they belong to by this coding. The females sit in obscure places and when they see their males’ signal, they respond with a coded light signal. The males then fly to the females, they mate, and it’s all very romantic.

However, Tom Eisner pointed out to me that this sort of advertising is surely a dangerous way to live, because as a flashing male flies around, he announces his presence to possible predators. So Tom speculated that fireflies must somehow be exceptionally well-defended: they don’t sting, they

Figure 8. Unusual large-ring molecular structures of the defensive compounds of some species of ladybird beetles.



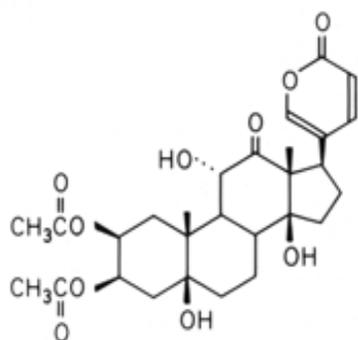


Figure 9. A lucibifagin, one of the chief protective steroids found in firefly haemolymph.

don't bite, and they don't scratch. So how do they defend themselves? Tom's guess was "I bet they taste lousy." To prove that hypothesis, he took a thrush, an insectivorous bird, fed it dozens of different species of insects and the thrush happily ate them. However, when he tried to feed it fireflies, the thrush would have nothing to do with them. They apparently tasted awful.

Since the thrush experiment confirmed his suspicions about firefly defenses, Tom suggested, "Why don't you see what's in fireflies?" Well, that's not quite a well-enough defined research proposal, because in looking at a whole organism, we would expect to find thousands upon thousands of compounds, so who knows which ones to care about? Tom therefore proposed that we prepare a firefly extract, and carry out thrush bioassays on specific extract components. When we took *Photinus pyralis*, extracted the whole animal with an organic solvent, and did some chromatography, a large array of components came through the chromatograph. Several of these components were particularly bioactive. An NMR-spectrum done on the isolated compounds showed them to be very similar, one to another. If even a half microgram of any of these compounds was put on the surface of a mealworm, that was enough to prevent the worm



from being eaten by the thrush. A half microgram is a minuscule amount, (below a millionth of an ounce,  $1 \mu\text{g} = 10^{-6} \text{g}$ ), so these compounds can be seen to be very powerful protective agents.

Using standard NMR and mass spectrometric techniques, the structure of the compound "C", the main compound defending these fireflies, was established (Figure 9). An organic chemist would immediately recognize this compound as a steroid, specifically, a highly oxygenated steroid. It has an additional ring attached to the "D" ring. The only known compounds in nature with a similar structure are cardiac-active compounds—compounds that slow down and strengthen the heartbeat—similar to digitalis. But these are produced mainly by oriental toads, they'd never been seen in an invertebrate before. It was a particularly curious type of compound to find as a major component in an insect defensive role, since insects are known not to be able to biosynthesize steroids.

To complete this story, I need to tell you about another firefly. This firefly, *Photuris versicolor*, (Figure 10) is somewhat bigger than the one I described earlier. For the experiment I'm about to describe, we switched from a bird predator to a spider predator, *Phidippus*, a jumping spider. These are wonderful creatures, they have eight eyes—four of them point forward and four are at the back of the head. These are spiders that don't build webs to catch prey. When they spot a potential victim, they observe its position. They then move some distance, again visually fix its coordinates, and having taken two readings, they jump. So, these are fearful predators. If this spider jumps onto a larva of *Photuris versicolor*, the larva will be immediately eaten. On the other hand, these spiders will not eat the fireflies we have discussed before, *Photinus pyralis*, which are chemically protected. They may touch them, but they won't eat them because they are repelled by their defensive steroids. However, lab-raised *Photuris* larvae and adults are eaten readily. The chromatographic analysis of the blood of one of these fireflies shows that, indeed, they're devoid of protective steroids, so it makes sense that they are so promptly eaten. Interestingly enough, however, if adult *Photuris* are collected in nature, some of them have protective steroids in their blood and some don't. And when they do have protective steroids, the compounds sometimes have different structures—it's a very mysterious situation. After this discovery, we realized the biology of these fireflies was going to be very interesting.

Figure 10. The firefly, *Photuris versicolor*. Literally a femme fatale, the female of this species has discovered the advantages of learning a foreign language. (Photo: Thomas Eisner)

### A multilingual predator

To understand the entire story, we have to know something about mating behavior of *Photuris*. The female *Photuris* does exactly what aforementioned unmated females do; she hangs around at night and watches for males of her species. When a male of her species signals, she responds, he flies to her, and she mates. However, unlike other species, after she's done that, she *still* hangs out evenings. But in contrast to her earlier mission, she now hunts for foreign males. When she spots them—these are males who are laden with protective steroids—she answers foreign males and says "Here I am." Amazingly, she can read their code and she can answer in their language. When a male then flies to her expecting to mate, she jumps onto him (Figure 11) and eats him—a remarkable way of earning a living! She's already mated, and now she attracts foreign males and gets nutrition for herself and for the next generation.

What interested us was that although the male firefly was protected against predators such as birds and spiders, he was clearly not protected against this predatory female! The female *Photuris* has apparently developed a taste for these defensive compounds. If the blood of a *Photinus* male is analyzed it is seen to be full of defensive steroids. If the blood of the female *Photuris* is analyzed before the feeding, it is devoid of these defensive compounds, but after feeding it contains the two main defensive steroids in the prey (Figure 12, compounds C and D from *Photinus pyralis*). So by eating the male, the female has obtained not only a nutritional benefit, but has also gained protection for herself—she is rendered inedible for the rest of her life. It's really a very neat interaction: the protective chemicals are made by one species, but a female of another species has learned how to acquire them, and so doesn't have to do any defensive chemistry on her own.

### Butterflies communicate with chemistry

The next story I'll relate is about communication. A cartoon I once found shows two German scientists trying to discover something about insect communication. As they observe, they exclaim in amazement, "They're speaking French!" That's a romantic thought, that insects might be speaking French to each other, but what they're really speaking is organic chemistry. The organic chemistry I'm going to tell you about is not the well-known phenomenon of female moths being able to attract males from immense distances, but quite a different story.



Figure 11. Protected against birds and spiders, the male *Photinus* firefly is defenseless against a predatory female *Photuris versicolor* who acquires his defensive compounds to protect herself and her progeny. (Photo: Thomas Eisner)

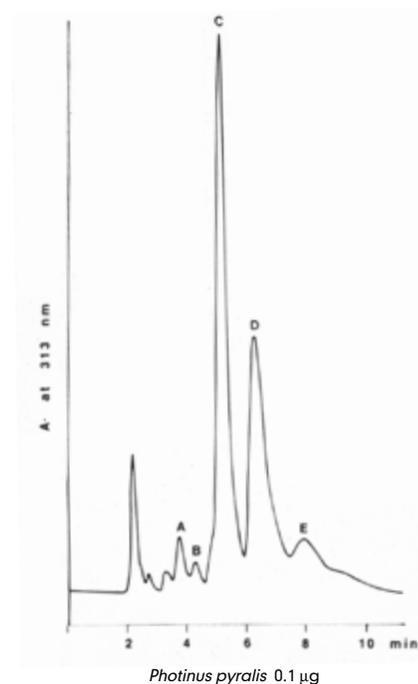


Figure 12. HPLC analysis of compounds in the blood of fireflies: C and D are the main defensive steroids.



Figure 13. Tropical butterfly *Lycorea ceres* provided the first evidence that its male courtship pheromone is made from plant toxins. (Photo: Thomas Eisner)

This study began over thirty years ago with a tropical butterfly called *Lycorea ceres* that lives in Trinidad (Figure 13). It was already known in the 19th century that male butterflies of a related species have organs tucked into their abdomens which seem to play a role in courtship. If the abdomen is squeezed a little bit, these lovely structures called hair pencils are extruded (Figure 14). Unfortunately, the courtship behavior of *Lycorea ceres* in Trinidad was not easily studied because they court far above the canopy of the trees and it's difficult to watch the behavior. Nevertheless, it was easy to get those courtship organs, examine the chemistry, and, we hoped, do the biology later.

When the compounds in the hair pencils were extracted, we found some long-chain compounds similar to female pheromones from moths. We also found a very unusual-looking structure built out of two 5-membered rings, and containing a nitrogen atom (Figure 15). It was a new and relatively simple natural product, but it didn't look like any known animal metabolite. When we first found this compound (named "danaidone") in the mid-1960s, it struck us that it bore a structural resemblance to a very widely distributed family of plant toxins called pyrrolizidine alkaloids. So here was an odd-looking structure in a butterfly which may or may not have had a biological role, and which may or may not have been related to the plant alkaloids. It was not at all clear what this finding meant.

Figure 14. The hair pencils (below), delicate structures on the courtship organs of male butterflies, are brushed across the antennae of the female butterfly who will only mate if their chemical signal wins her approval. (Photo: Thomas Eisner)



Somewhat later, we were able to collaborate with a biologist, Lincoln Brower, at that time at Amherst College, who studied the courtship of a related insect which was much easier to observe, the Florida Queen butterfly. Lincoln Brower showed that these animals, which also have hair pencils, used the hair pencils in courtship. The male pursues the female, and as he flies, he everts his hair pencils and brushes them across her antennae. In response, she settles down on some greenery and folds her wings; he then hair-pencils a little longer and, finally, she mates with him. It appears that the female decides whether or not to mate with the male on the basis of this signaling. We looked at the hair-pencil chemistry of the Florida Queen, *Danaus gilippus*, and we found the very same compound, danaidone, to be present.

We also discovered, totally by chance, that when the male butterflies were raised in a very large greenhouse area Tom Eisner built in Florida for these studies, the males courted avidly, chased females, and stroked the females with their hair pencils, but to no avail. Totally unexpectedly, the females would have nothing to do with these captivity-grown males. When we looked at the chemistry of the males, we found that danaidone was absent! The males were chemically deficient, and females would no longer mate with them. We had synthesized this compound, so we could apply it to the courtship organs of the deficient males and re-release them. They then courted normally and the females responded well. We therefore knew that danaidone was a biologically important compound, although we had no idea why it was missing from the homegrown males.

### **Institute provides impetus for insect studies**

The chance to study this problem further came totally out of the blue, as the result of a fortunate interaction between Carl Djerassi and Tom Odhiambo. Tom Odhiambo was a Reader in Entomology at the University of Nairobi who had studied with Sir Vincent Wigglesworth in the U.K., one of the great insect physiologists of his generation. In 1967, Odhiambo published a paper in *Science* discussing why so little research in biology, chemistry and physics was going on in Africa. He put forward the very interesting hypothesis that it was very hard to do science in Africa because the African view of nature was different from the standard Western view of nature. Nature was something that one stood in awe of, respected and perhaps worshipped. Under those conditions, nature becomes less an object of study and more something to be admired from afar.

At the same time in the United States, chemist Carl Djerassi became interested in how the pursuit of chemistry could be encouraged in Africa. Djerassi proposed in a 1968 article in the *Bulletin of Atomic Scientists*, that if a specific subject were chosen that would be really important in Africa, and could be optimally studied there, one could put together an international team of scientists, build a lab to study that subject, and develop science that would be relevant to the African setting. The field Djerassi suggested was insect chemistry. When Djerassi published his paper putting forth this view, Odhiambo immediately wrote to him suggesting a collaborative study of insect physiology and chemistry. Djerassi had just founded a company called Zoecon, which was involved in developing new techniques for insect control, and so there was a great meeting of minds.

### **International team heads new center**

The undertaking they imagined eventually evolved into the International Centre for Insect Physiology and Ecology (ICIPE). Two chemists—my good friend, Koji Nakanishi and I were invited to serve as research directors. Other founding research directors included: biologist, Carrol Williams from Harvard; Dietrich Schneider from the Max Planck Institute for Behavioral Physiology; Jan de Wilde from Wageningen, the largest agricultural school in Europe; Martin Lüscher from Switzerland; Herbert Roeller from Texas; George B. Craig from Notre Dame, an expert in mosquito genetics; Raquel Galun from Israel; and John Pringle from Oxford. All of these scientists were contacted by either Djerassi or Odhiambo and asked if they would like to participate in setting up a new institute: they could choose any insect-related research, could collaborate, and set up what Djerassi hoped might evolve into a "Weizmann Institute of East Africa."

We managed to find quite a few organizations to support the center, in particular the United Nations Development

Program, which contributed about half of the necessary money for building and maintaining the lab. One thing that particularly interested the sponsoring foundations about this lab was not only that it might become a useful science nucleus in East Africa, but also that it was to be managed by a consortium of the national science academies of the countries from which the participating scientists came. This center was the first institute anywhere where representatives of the U.S. National Academy of Sciences, the British Royal Society and counterparts from Germany, Switzerland, Sweden, and Holland all sat on a governing board that planned the work of an international laboratory.

### **African butterflies provide solution**

One of the things that Dietrich Schneider and I were able to do together at ICIPE was to study the African Monarch butterfly, *Danaus chrysippus*—a close relative of the Florida Queen—often seen sitting on milkweed plants near our lab in Nairobi. It was by studying this species that we learned much more about insect communication. The African Monarch males also have hair pencils, use them in courtship and make the same signalling compound, the same male pheromone, as the two butterfly species discussed earlier. They make huge amounts of this pheromone; we extracted 5 mg from only sixty-seven males. Pheromones are often produced in microgram or nanogram quantities, and sometimes only in picogram quantities. So 5 mg was a comparatively large amount, and that made it a very easy pheromone to work with from a chemist's viewpoint.

The discovery of the same signaling agent in the African butterfly and the Florida Queen led us to solve the problem of what was going wrong with the production of danaidone in our Florida homegrown males. That problem was actually solved by Heidi Schneider, who noticed that in the open fields, hundreds of African Monarchs could sometimes be seen sitting on *Heliotropium steudneri* plants, and seemed to be sucking up juices from these plants. Upon careful observation, we saw that it was only males who were doing this behavior, and doing it very purposefully. If frightened away, they would come right back to the plants and start feeding on the juices again.

It was clear that there was something in *Heliotropium steudneri* that the males very much wanted. We analyzed the compounds in an extract of this plant, and found—what do you know—they contained a known alkaloid called lycopsamine. One doesn't need to be an organic chemist to see that the molecular architecture of this alkaloid (Figure 15) is very similar to the architecture of the pheromone danaidone; the same two 5-membered rings and a nitrogen atom where the rings join. If we supplied these plants to African Monarchs in captivity they made their pheromone; if we didn't, they did not. Even if fed only the pure alkaloid,

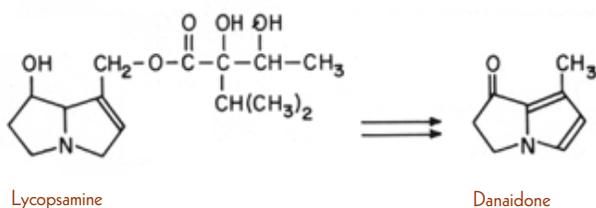
*Heliotropium steudneri*

Figure 15. Lycopsamine acquired from *H. steudneri* is transformed into danaidone, the courtship pheromone isolated from the African Monarch butterfly.

they would still make that same pheromone.

So the answer turned out to be that the precursor from which this essential compound is made is, in fact, a toxic alkaloid from a plant source. This butterfly has apparently become drug-dependent in order to reproduce, and acquisition of the alkaloid has become so beneficial to them that they have evolved a very complex chain of behaviors in order to gain the chemical they need. The males have to fly out to nature, identify the right plant, take in the juices, and do a little bit of chemical synthesis in order to make the pheromone. They then present it to a female and say, "Here it is," and she says, "Good." If a male does not make this chemical, then the female is just not interested.

### Other advantages of dining on toxins

Well, Tom Eisner suggested that we might better understand this system if we studied a species closer to home. We picked a species of moth called *Utetheisa ornatrix*, chosen because it lives on toxic plants that are laden with pyrrolizidine alkaloids, alkaloids that have the same signature of two 5-membered rings. In fact, *U. ornatrix* larvae burrow into the pods of *Crotalaria spectabilis*, a legume, and eat the beans, the part of the plant richest in alkaloids (Figure 16). This is a highly toxic plant for animals; in Australia, entire herds of sheep have been known to die as a consequence of ingesting these plants. But the caterpillars love them.

Is there an advantage to subsisting on a poisonous plant? We found that these moths are able to sequester the plant alkaloids from their food. If the adult *Utetheisa* is caught in a spider web, the spider jumps out and seizes it. The spider then tastes the alkaloid and immediately jumps back and cuts the web's threads, releasing the moth (Figure 17). So if an organism can live on a poisonous food, first of all, there are not so many other species who want to eat that food, and secondly, it can protect itself by sequestering the distasteful poison without having to do any chemistry on its own.

As a control experiment, Tom Eisner discovered that we

could also grow these moths on a diet of pinto beans, a human food which does not contain any alkaloids. If we put a pinto bean-fed moth in a spider web, the spider immediately jumped out and ate it. Tom was quite happy with this outcome; we didn't need any sophisticated statistical calculations to show that this was a significant result. So these creatures are really protected by feeding on this legume because of its alkaloid content. Interestingly enough, when the female *Utetheisa* lays her eggs, they're also totally unpalatable to insect predators such as ladybird beetles. When we analyzed *Utetheisa* eggs, we discovered they contain about one-half percent by weight of the plant alkaloid. So the eggs themselves are chemically protected by the mother; she acquires the alkaloid for herself and then puts it into the eggs.

We thought that was a great discovery until I came across a cartoon which showed a mother hen pointing to her basket of eggs, decorated for Easter. She'd obviously been drinking, and the caption said, "This year they've all got liquor in them!" In fact, this cartoonist had exactly the same idea about incorporating chemicals into eggs as *U. ornatrix*: eat a toxic material, and if it's put in into the eggs, they can be chemically protected.

We wanted to learn more about this, and wondered how the eggs would come out if the mother moth had protective alkaloids but the father didn't. And, as one might guess, the eggs came out protected. We did it the other way around as a control experiment: suppose the father is chemically protected but the mother grew up on pinto beans, what kind of eggs are then produced? When the eggs from the unprotected mother and protected father were fed to the beetles, the beetles wouldn't touch them. So we analyzed those eggs from an "alkaloid-free" female, and discovered that they were still protected by the alkaloid!

### The gift that keeps on giving

When we examined what had happened, we found that the alkaloid had been given by the male to the female through the mating process. Males are able to put the protective alkaloid into a spermatophore; when they transfer the spermatophore to the female, she can share the father's chemical protection and use it. In fact, if you measure a number of males' alkaloid contents and measure how much alkaloid they transfer to females, it happens, not surprisingly, that the very well-endowed males give a large "nuptial gift" to the females, while the males who are very poorly endowed give only a very modest gift of alkaloids.

It turns out that this chemical transfer is a valuable behavior for these insects. If wild *Utetheisa ornatrix* females from the field are analyzed, some are chemically well-protected and don't really need to mate with alkaloid-protected males in order to protect their eggs. But other females, for one reason or another, are not chemically protected; they



Figure 16. The larva of *Utetheisa ornatrix* feeds on beans of the toxic *Crotalaria* plant that offer it chemical protection. (Photo: Thomas Eisner)



Figure 17. A spider inspects—but ultimately rejects—the alkaloid-protected *Utetheisa ornatrix*. (Photo: Thomas Eisner)

didn't gain access to the seed pods or absorb the alkaloids properly. In that case, their eggs will very likely be eaten unless they can find a male who's going to give a large nuptial gift and protect the offspring. This is a matter of life or death to the next generation: the chemically unprotected female must be able to distinguish males who have ample alkaloid stores from those that don't if she is going to pass along her genes successfully.

#### Female moths skillful analytical chemists

How does the female do that? She has to be a good analytical chemist. We heard earlier today, in the exciting talk by Bonner Denton, about the importance of analytical chemistry: this again is analysis. The moth's courtship lasts only about five seconds, so she has very little time in which to determine whether a courting male will offer chemical protection. So we devised some female choice tests. We put a female along with an alkaloid-containing and an alkaloid-free male together in a small chamber and noted which male she mated with. This experiment showed a strong preference for the protected male over the unprotected male. Interestingly enough, even if the female doesn't really need the alkaloids—if she grew up on *Crotalaria* seeds and is perfectly well protected—she has exactly the same bias. So all the females, whether they're protected or not, are programmed in the same way to select the male who can protect the offspring.

How in the world can she figure that out in a five-second courtship? It turns out the male moth advertises. He flies up to her, he whips out these beautiful courtship organs and brushes them across her antennae. And on these organs is a chemical signal. That chemical signal is a molecule which looks very much like our butterfly compounds, and one can

see the same pyrrolizidine alkaloid theme that we saw in the alkaloid the *Utetheisa* has been eating. We noted that only a male who had ingested these protective compounds was able to make this signal, and the female mated preferentially with males who had the pheromone—the female actually avoided males who didn't give the signal. So this signal is like an affidavit: it says, "I'm the guy you want, and here's the evidence!"

These experiments are interesting because they demonstrate the use of a chemical signal whose actual meaning can be read by chemists. It can easily be seen that this male pheromone is not an arbitrary molecule; it is a molecule which says by its structure that the bearer of this messenger compound is well-defended. We know it because the male can only be defended if he acquired the poison from the plants, and he needs the plant toxin to make his courtship pheromone.

There's a bit more to that story that Tom Eisner has investigated very recently. This further study is about sexually-transmitted defenses. A group of unprotected, pinto-bean-fed females were confined with two kinds of males. They could mate with chemically protected males that had been fed on *Crotalaria*, or with males fed on pinto beans. Now, I've already explained that females who are chemically protected produce eggs which are chemically protected. However, when we looked at the fate of the mother herself after she mated with the proper (i.e., protected) male, it turned out that spiders would not eat her; she's also protected herself as a result of having mated. But if she mated with a chemically unprotected male and was then put into a spider

web, she was eaten. This indicates that sex with an appropriately chosen mate is beneficial both for the mother and for the next generation.

### Life insurance for the whole family

We've looked at the original Florida Queen butterflies to see how well this system works, and it's really remarkably efficient. If a male is fed a thousand micrograms of monocrotaline, another pyrrolizidine alkaloid, he sequesters almost 60 percent of what he eats, chiefly in glands called accessory reproductive glands. When he mates with the female, he transmits about two-thirds of this alkaloid with his spermatophore; she puts over 90 percent of the alkaloid into her eggs and keeps a small amount for herself (Figure 18). The actual amounts of plant alkaloid could be measured and followed through the male's digestive tract and reproductive tract, through the female reproductive tract and into the eggs; of the original 1000 micrograms, about 345 micrograms (one-third) can still be found protecting the next generation. This work illuminates a kind of plant-insect interaction which would never have been recognized or understood if the organic chemistry had not been done to discover what was going on at the molecular level.

The story I've been telling you was the first discovered example of what Darwin called sexual selection based on a chemical criterion: the females use a chemical signal to decide whether or not to mate. The examples Darwin talked about involved males who compete on the basis of the size of horns or strength of musculature, and the female mates selectively with the winner of a contest. In the insects' case, it's more of a chemical contest: the female prefers the chemically protected male as a mate.

### Saving nature's chemical "library"

We come now to the last part of my story. What I've been relating in this section of my lecture is how insects go out into nature, look at the plant world, and find compounds that are good for them. And in fact, humans have been doing the same thing for a very long time as well. We now call this activity ethnomedicine or chemical prospecting; it involves searching our environment to find chemicals of potential usefulness.

Plants have yielded many of the most widely-used drugs. The opium poppy supplies the most important known natural analgesic, morphine. The cinchona tree, which yields the medicinal alkaloid, quinine, was the first plant from which a medicine to treat a specific disease (malaria) was ever studied. Foxglove, from which digitalis is extracted, has been known for centuries as a source of valuable cardiotonic agents.

These studies of "natural products" led to the development of the modern pharmaceutical industry. For instance,

in the early 1900s the Hoffman-LaRoche Company sold a crude extract of foxglove, and one of their main commercial products was this digitalis extract. As they became able to formulate a well-controlled version of the extract, based on a very careful study of its chemistry, they were able to produce and sell a preparation that could be used with confidence is the treatment of heart diseases. This research launched a company on its way to becoming one of the present giants in the pharmaceutical industry.

Nature is still a largely untapped pharmacopoeia. In a forward to 1996 book on beetles, the distinguished Harvard biologist and naturalist, Edward O. Wilson, called attention to biodiversity as our "most valuable, but least appreciated resource." In illustrating the value of preserving biodiversity he suggests that "a rare beetle from the Andes might turn out to be the source of a chemical that might cure pancreatic cancer." This notion provides the motivation of much current research. There's a very recent example of a tiny Ecuadorian frog studied by John Daly of the NIH. He found this frog to be the source of a natural product with a most unusual structure which has analgesic properties hundreds of times greater than morphine. However, the frogs themselves are almost certainly not the producers of that compound. John Daly has shown that many bioactive compounds isolated from frogs actually come from the tiny insects that the frogs eat. So the frogs themselves are also chemical prospectors; they're going out into nature, eating various bugs, and sequestering and concentrating those compounds that can protect them. We know so little about these chemical ecological relationships that it's hard to anticipate from what species very valuable chemistry may turn up in the future.

Unfortunately, frogs seem to be disappearing from the face of the Earth. It's not known exactly why, but it seems that viral and parasitic diseases are wreaking havoc with these populations. So it is becoming critical to somehow study all this chemistry in nature before it vanishes. Is species loss important? Well, again, Ed Wilson estimates we're losing—completely obliterating—20,000 species a year. That's a very high rate of information loss. It's hard to establish a number like that, but even if it were to be off by a factor of ten in either direction—suppose it were only 2,000 species—that would still be a devastating loss. If you go up by a factor of ten to 200,000 species, it's almost beyond imagination. When you lose a species what happens? You lose all the significant chemicals they make, all the enzymes responsible for the synthesis of those chemicals, and you lose all the genetic information that goes into making the enzymes. That's like burning up a library in which each book is unique; the information contained is gone forever.

It would be wonderful if species loss could be slowed so that we could study more of the chemistry they carry out, find out what's going on in nature and learn from it. About

ten years ago, Tom Eisner and I joined the International Society for Chemical Ecology in formulating an approach to the control of species loss. In part, we pointed out that: "... forests constitute a treasury of immense value to mankind. The current alarming rate of species extinctions is rapidly depleting this treasury with potentially disastrous consequences. The International Society of Chemical Ecology urges that conservation measures be adopted worldwide to stem the tide of species extinction, and that vastly increased international studies be undertaken aimed at discovering new chemicals for use in medicine, agriculture and industry. These exploratory efforts should be pursued by a partnership of developing and developed nations in such fashion that the benefits flow in fair measure to all participants."

We were trying to give that statement a sense of urgency, but a vastly more beautiful expression of urgency can be found in a lovely poem by Andrew Marvell, titled *To His Coy Mistress*. He writes:

Had we but world enough and time  
This coyness, lady, were no crime . . .  
But at my back I always hear  
Time's winged chariots hurrying near  
And yonder all before us lie  
Deserts of vast eternity. . . .  
The grave's a fine and private place  
But none, I think, do there embrace.

Now I don't know what results Marvell had with his coy mistress, but he certainly put his case elegantly. The fact is that our biodiversity resources really are suffering depletion, and we cannot afford to be coy about both preserving and studying them.

### Shared benefits from chemical prospecting

Our hope was that chemical prospecting research groups might be set up in developing countries where one could study taxonomy and ecology, try to preserve species and collect samples for chemical study. A university might organize the program and also train people to do the fieldwork and study the chemistry. An industrial partner would screen for pharmaceutical activities and pursue the development of promising leads. The investment in developing a new drug, even once there is a lead, is roughly \$500 million over a period of ten years. No university can make that kind of investment; that is a job for a large company. The point of such a partnership would be that if a new drug were to be discovered—a new analgesic from Ecuadorian frogs, perhaps—then some appropriate percentage of the sales would

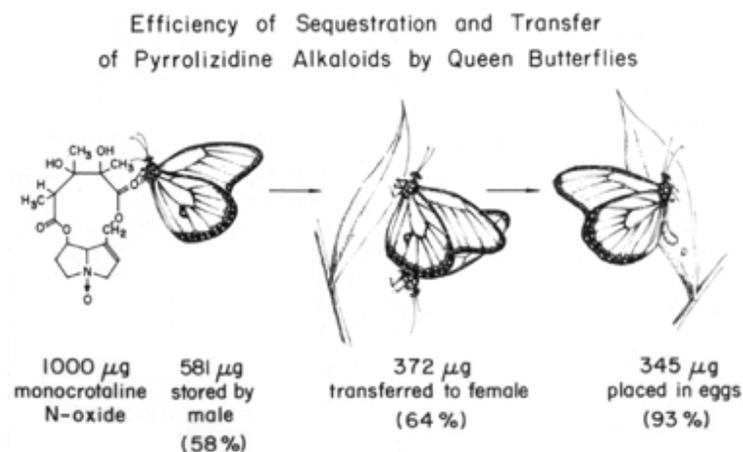


Figure 18. Protective pyrrolizidine alkaloids from plant nectars, ingested by the male butterfly, are transferred to the female during mating and passed on to her eggs. (Illustration: David Dussourd)

be returned to the relevant source country to help fund the national park system perhaps or nature preserves. The objective is to make it economically more profitable to conserve and study natural resources than to simply consume them in a destructive way. This would constitute a very efficient feedback loop, and in fact, a consortium of agencies (NIH/NSF/USAID) established exactly this sort of program about a half dozen years ago.

Currently, several of us are trying to set up a very ambitious project of a similar nature in northern Brazil. The idea is to work in the Amazon region, which is almost one-half of the area of Brazil, and which, by itself, has one-third of the world's rainforests. The species diversity in this part of the world is almost beyond belief, as is the amount of scientific research that needs to be done. At one of our early organizational meetings in Brazil, I heard a talk by the secretary for Amazonian affairs. He showed a wonderful map in which all of the major European countries were shown floating in the Amazon region—they all fit. The secretary explained that, so far, about 12 percent of the Amazon has been destroyed, an amount of land that's equal to the area of France! So the biological richness of this area is on an incredible scale, and the knowledge we could gain from Amazonian studies is incalculable.

As it happens, there is an ambitious group of scientists at the almost 100-year-old Butantan Institute in São Paulo, who would like to undertake such biodiversity-based research. The institute's main focus has been to provide anti-venom sera to treat the bites of South American poisonous snakes. Several of us, including my good friend, Koji Nakanishi, are trying to help develop this center which might work on the sort of feedback loop model mentioned earlier.

Right now, a few such programs are being established around the world, and I think it's important that more chemists get involved in work of this sort. It will require a lot of effort and imagination because each region has its own special opportunities as well as its own problems. As you know, research costs money. All of us are enormously dependent on Research Corporation, the National Science Foundation, the NIH and many other sponsors. It is these sources of support that make the sort of chemistry that I'm talking about happen. I myself am infinitely grateful not only to these funding agencies, but also to a number of very good friends in the chemical and pharmaceutical industries.

#### **Nature's chemistry a source of scientific inspiration**

I want to end by reading a few sentences from a lovely autobiography, *For the Love of Enzymes*, by Arthur Kornberg, a Nobel laureate M.D. who loves chemistry. He writes: "Chemists need to be aware that three billion years of cellular evolution have perfected molecules and molecular societies of awesome chemical sophistication. The chemist interested in catalysis, stereospecificity, polymer structure, metallo-organic reactions, surface effects and a hundred other facets of chemistry will be chal-



lenged and instructed by the myriad forms in biology." And indeed, biology provides an infinite source of inspiration for chemists. The chemist doesn't always need to be that imaginative, he just has to observe nature and try to understand it; interesting findings will surely result.

Kornberg makes another statement in his book that I find most inspiring: "Much of life can be understood in rational terms if expressed in the language of chemistry. It is an international language, a language for all time, a language that explains where we came from, what we are, and where the physical world will allow us to go."

Not only are studies of the sort I've talked to you about of potential utility, but they also have, for many of us, a certain beauty. I think what inspires many biologists and chemists is not so much that they may discover something patentable, but that they may understand something beautiful. While the beauty of nature is not often emphasized in the way we teach biology and chemistry nowadays, it is important to bear in mind that much of the thrill of pursuing science comes from the wonder that our own subjects inspire, and from the deep satisfaction that comes from understanding each new aspect of how nature works.



Harry B. Gray is the Arnold O. Beckman Professor of Chemistry and the director of the Beckman Institute at the California Institute of Technology. He is a specialist in "bio-inorganic electron transfer chemistry." He and his group are trying to create a new kind of photosynthesis using metals like ruthenium and rhenium in specially engineered metalloproteins, that would derive oxygen and hydrogen from water for fuel instead of reducing carbon dioxide to food and producing oxygen.

Winner of the National Medal of Science and the Gibbs Medal among many other awards, the Kentucky native received Ph.D. from Northwestern. He also counts twelve honorary degrees among his honors, those from Northwestern, Columbia and the University of Chicago among them.

At the 1999 Partners conference, Harry Gray made a strong case for the roles of inorganic chemistry (plus a bit of organic) and science education in saving the planet.

## Clean Fuel from Solar Photochemistry

Adapted from a 1999 presentation by Harry B. Gray

**What I'm going** to talk about tonight is of direct relevance to all of you because I want you to go back to your classrooms and recruit your very best students to go into the field of chemistry. We need them to solve certain problems, literally to save the planet during the next century.

What's the challenge? We've got to stop burning hydrocarbons. I've been saying this for 20 years, and I've talked myself into really believing it. We've got to stop burning hydrocarbons for energy and producing carbon dioxide in the process. This is a very poor use of hydrocarbons; we need them for materials. Most people have been talking about global warming as the reason that we've got to stop burning them, but that's only part of the story. As Mendeleev said right after he invented the Periodic Table, it's crazy—burning hydrocarbons is just like burning \$100 bills.

There's no doubt that in the next 20 or 30 years the production of oil is going to start dropping and we're going to be faced with an immense energy crisis. Some of you will be around to see that crisis if we don't do something about it. We really have to get smarter and use clean fuels. We have to replace dirty fuels with clean fuels that don't contribute to global warming; the clean fuels are electrons, or electricity if you like, and hydrogen.

There's no doubt that we've got to convert this planet to electricity and hydrogen, and we only have two real choices. You can talk about windmills all you want, you can talk about geothermal until you're blue in the face, but I could care less because we've only got two real choices: one is nuclear and the other one is solar. And, speaking tongue-in-cheek, since we're all chemists—most of us are chemists—the one thing we have in common is that we *hate* physicists. I'm the captain of the team, and I'm telling you right now that we've got to beat the physicists. They've got nuclear—we've got solar. I'm going to tell you tonight how we can bring this solar baby home to victory!

First I have to give you a very short course on solar energy conversion. You can either buy giant textbooks and carry them around in wheelbarrows and study for years, or you can listen to me for 15 minutes—you can laugh any time and there's no charge. This is the quick course in energy conversion.



Hydrogen-fueled "Necars" (New Electric Cars) from Daimler-Chrysler: 1 (rear), 2 (left) and 3 (right). Necar 1's fuel cell took the van's entire cargo area; 2 had three rows of seats, but put fuel cell in ugly roof hump. Sleek Necar 3 includes both a small fuel cell and methanol-to-hydrogen reformer.

### Solar energy conversion

There are three types of conversion involving three fundamentals (figure 2). The first fundamental of solar conversion is light capturing—you have to capture the light that hits the Earth's surface. If you only absorbed this radiation with a blackbody, you would degrade light into heat. We've been heating swimming pools for years with this solar thermal energy; very inefficient, and it's not going to solve our problems. If we take one step forward and do light capture and then what we call "quantum conversion," using silicon semiconductors for example, we can make electricity, make electrons. Shine light on silicon solar panels and we have electricity we can use right here and now; you can now get very good ones with 10 percent overall efficiency in converting light to electricity.

As a matter of fact we could convert Los Angeles to a solar city today if we had the will to do so. We could put silicon on everybody's roof and make the whole city a collection area; feed electrons into the main power grid, store the electrons during the day from our homes and borrow them back at night. Instead of getting bills at the end of the month, we'd get checks. We could build a solar city right now if we had the will, but there's something even better.

If we just do light capture and electron transfer, we've converted solar energy into electricity, but we haven't stored it; we've got to use it or lose it. You can store it in batteries, sure, but the dream is to take the third and final step in solar conversion—do light capture, electron transfer and then some chemistry. Do chemical catalysis—take water and split it into hydrogen fuel and oxygen. If we can do that, if we can split water with sunlight and generate hydrogen fuel and oxygen, we'd have efficient storage.

The energy of sunlight would be stored in a hydrogen-hydrogen single bond; we can then use it anytime, to generate electricity day or night, by running the hydrogen and oxygen back through a fuel cell. Catalysis is the problem, ladies and gentlemen, that we've got to solve in the next ten years. A lot of smart people have been working on it and haven't quite solved it yet, but we're getting very, very close; we're knocking at the door.

Twenty-five years ago when I and many of my colleagues started working on this problem, we thought we were smarter than God. We thought we could cut corners and use mostly inorganic materials that would do this job better than nature. With green plants, nature has solved problems of optimizing light capture, electron transfer, and catalysis in converting carbon dioxide and water into carbohydrates and oxygen.

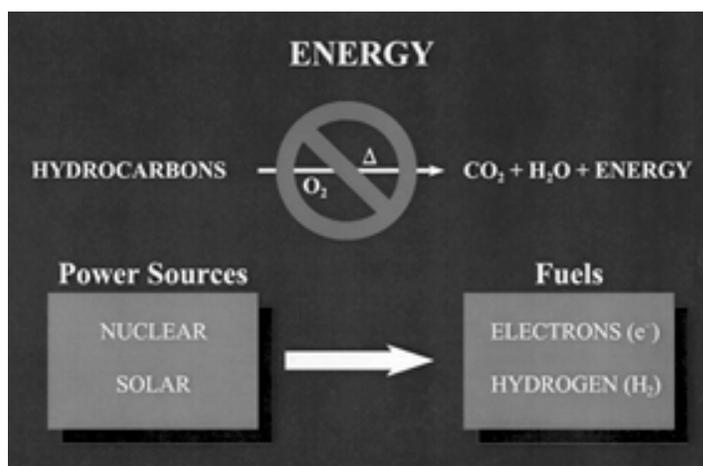


Figure 1. "We must stop burning hydrocarbons," argues Gray, "and go to clean fuels—electrons and hydrogen."

### Energy conversion in nature

Nature solved those problems, as you know, with a device that looks like figure 3, which you can get from green plants, called photosystem II. This is what we know about the structure of the oxygen-evolving complex in plants: light capture is done by P680, which is a collection of chlorophylls; there is a group of organic molecules optimizing electron flow. Light capture is optimized, electron transfer is optimized; and nature built the manganese-oxygen cluster at the bottom (figure 3) that does this marvelous four-electron oxidation of water to oxygen.

While I and several of my colleagues were trying to

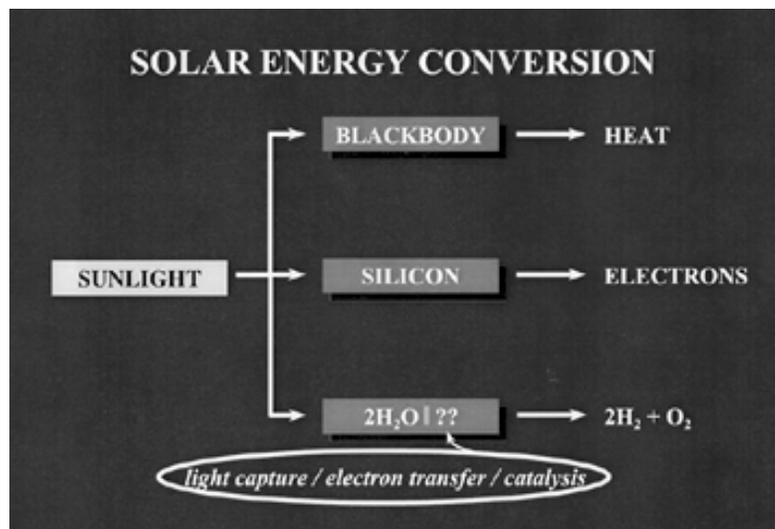


Figure 2. Of the three routes to converting solar energy, only the last two promise efficiency, with the last also offering storage and portability.

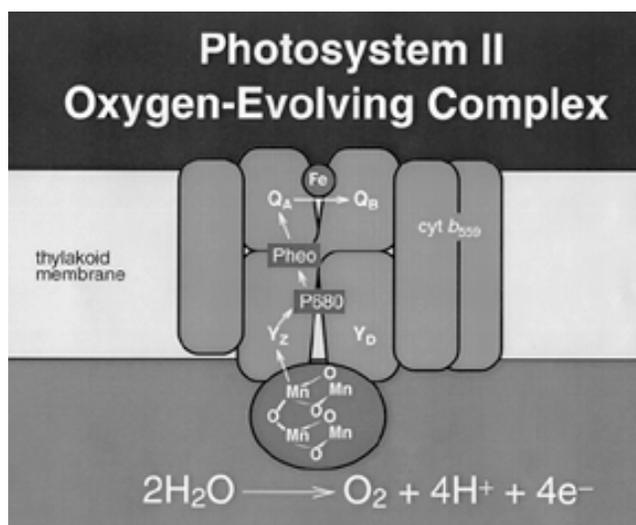


Figure 3. Nature's photosystem II provides some valuable hints on how to construct a practical water-splitter.

build something extremely simple that would do all of this, Eli Greenbaum at Oak Ridge, who was smarter than all of us, observed that nature had already solved most of this problem. Nature has optimized light capture, electron transfer, and built one of the two catalysts we need, the one for oxygen—all we've got to do is add the catalyst to make hydrogen.

So while all the rest of us were getting nowhere, Eli took a bunch of plants, extracted the chloroplasts, got a baggie with some platinum solution and then dropped the chloroplasts in. Then he pulled them out, dried them off, and put sunlight on them and split water, with 1 percent more efficiency than nature does. So here's the work (figure 4) of Eli "Green tree"—a wonderful name for a man who's working on artificial photosynthesis. Eli's platinized chloroplasts work beautifully if you want to split water. Hydrogen on the platinum side, oxygen on the manganese side, everything's optimized—and now you breathe a sigh of relief: the problem's solved and this talk is over; but it's not!

### Durability is vital

It's not over because I want to tell you something really important, and I don't want you ever to forget it. If you read about a high-efficiency water-splitting system and it's working fine, you say, OK, but how long will it last? Is it durable? Eli's isn't. Any time you generate oxygen with organic molecules around, you will make singlet oxygen, a very reactive molecule, that literally destroys organic matter. It oxidizes it completely. You can shine light on platinized chloroplasts and they split water just fine, then they start to poop out in a few hours. It's because they are full

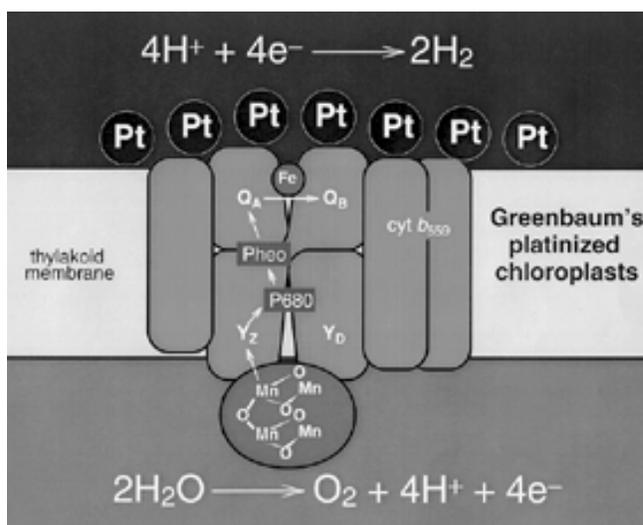


Figure 4. Simply platinizing chloroplasts can make an efficient, if short-lived splitting device.

of *wimpy* organic molecules. And so on to Lesson Number Two.

If Lesson Number One is "chemists *hate* physicists," Lesson Number Two is "inorganic chemists *hate* organic chemists." When Eli came in and showed us all up we were really set back because he was splitting water while we were still whistling Dixie. But when Eli's stuff pooped out, we said "oh, yeah!" It was my theme back then. We said, we inorganic chemists are now going to make something that's going to last *forever*. We're going to get rid of all those wimpy organic molecules, we're going to ban them, and we're going to make a tough, inorganic water-splitting system that's going to last forever, because—we are the Marines! And, sure enough, we did it! The cerium system shown in figure 5 has not a single organic molecule in it. You hit a cerium ion with deep UV light, and it reduces water to hydrogen making cerium+4, which all of you know is a powerful oxidant, and which, in the presence of a hard-core inorganic catalyst, ruthenium dioxide, will make oxygen. This thing will last forever.

### The question of efficiency

So it lasts forever! But what's the other question? You next ask, but Harry, is it efficient? Well, this one really isn't very efficient. The calculated efficiency of this system is well below  $10^{-10}$  percent! Now let me tell you first how low that is. If I filled the Pacific Ocean with this cerium system and irradiated it with sunlight for a billion years, it would not make enough hydrogen to drive my car around the parking lot. This system is really, really inefficient—but it lasts forever.

Now, I'll tell you why it's so inefficient: its light capture

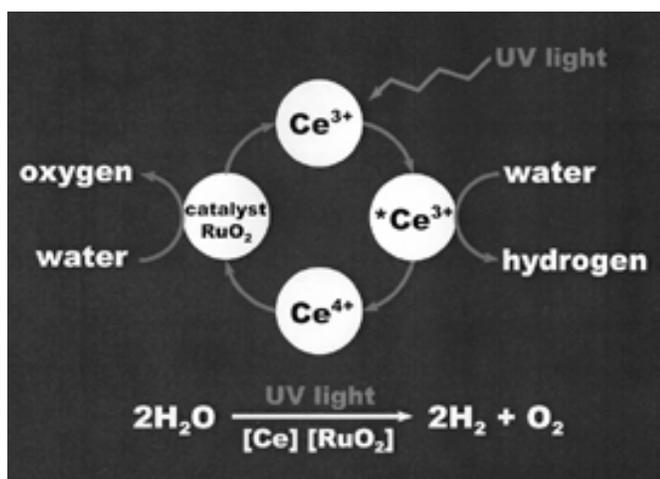


Figure 5. Cerium water-splitting system lasts forever, but captures only deep UV, limiting its efficiency to very low levels.

is no good. Deep UV is the only thing that will drive this cycle. And you know that pesky ozone layer up in the sky? Well, it won't let the deep UV through! It leaves us with two choices. The first is we get all those aerosol spray cans back and get rid of that pesky ozone layer so this baby will work. The other choice is that we've got to get smarter. We've got to make peace with our organic chemistry friends, and get some of those *wonderful* organic molecules back, combine them with metals and find some way to make a much more durable, efficient system.

### Alternate systems

Well, before we arrived at that conclusion, my group and I went through a number of other systems. We worked on an iridium system (figure 6) that uses near-UV light, and we ran it for a while—you can actually make enough hydrogen in this iridium system to drive to downtown Tucson and back. And then we made a breakthrough: Kent Mann, Nate Lewis and I discovered a rhodium system (figure 7) with isocyanide ligands that works with red light. It captures red light, makes hydrogen, then it can oxidize chloride to chlorine. Chlorine can oxidize water to oxygen, so you can complete a water-splitting cycle.

But there is a problem. To make a long story short, we analyzed the mechanism of this cycle, and it's very, very complicated, involving many species. About 1980, it was quite clear that there was no way we were ever going to get more than about 1 percent efficiency out of these rather simple systems, and that we'd better admit that we were not as smart as nature. We had to learn nature's lessons a little more carefully and reengineer, and that's what we did.

We went after the problem of how electrons tunnel long

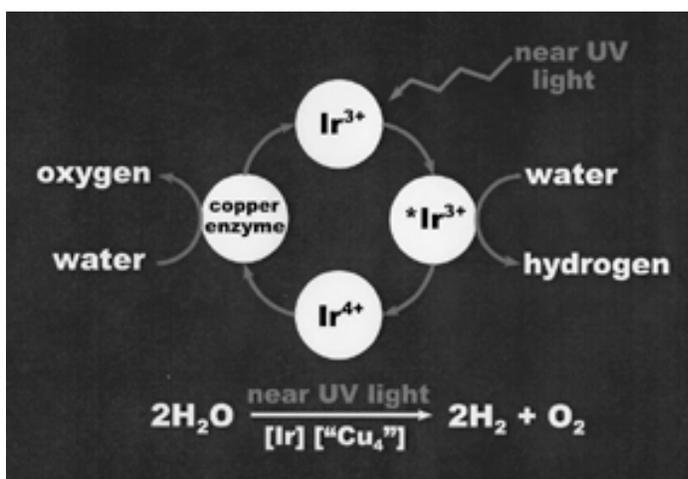


Figure 6. An iridium-copper enzyme system captures near UV light and boosts efficiency over the cerium system (left), but not to a practical level.

distances through photosystem II (figure 3, page 7). What controls this tunneling and therefore the efficiencies of these devices? We wanted to understand the green plant photosystem that nature has built for us, and then make it more durable by replacing as many of the organic components as possible with inorganic components. We wanted to replace all of those chlorophyll-like things with metals.

We were faced with the problem that all inorganic chemists know, and I will now share with you, that metals in water solution don't transfer electrons very rapidly (figure 8). There are enormous barriers; we call them reorganization barriers, reorganizing water around highly-charged metal ions—these are big barriers. And so it's actually dif-

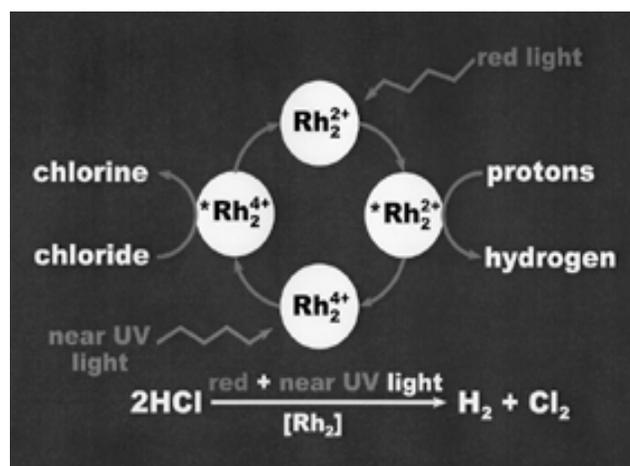


Figure 7. More complex rhodium system adds isocyanide ligands (not shown) that work with red light, and uses near UV to oxidize chloride to chlorine.

difficult for  $\text{Fe}^{2+}$  and  $\text{Fe}^{3+}$  to just transfer an electron in water, even when in contact with each other, much less in these big biological structures nature has built, where electrons are going 10, 15 and 20 angstroms—pretty fast.

### Fast electron transfer

I did a quick calculation for you just to dramatize this problem (figure 9). The barrier in water is so high that if you replace the organic components with iron it would take  $10^{16}$  years to get an electron to travel 20 Å. We don't have that much time!

However, we believed we could reduce  $10^{16}$  years to 1 millisecond by embedding iron ions in a protein, effectively making them the same size, reducing the reorganization barrier (figure 10). And that's when I got the team together again and said we've got to figure out whether it takes  $10^{16}$  years or 1 millisecond with iron embedded in a protein. And my group said, Harry, we certainly hope it's a millisecond, because if it isn't it's going to take a long time to get our degrees!

That's when we built the molecules that those of you who follow our work in the literature know about. We started taking proteins like cytochrome *c* (figure 11), which has a heme group with an iron, hanging rutheniums on it at distances of 18-20 Å from the iron, and then doing laser experiments to measure electron tunneling times (figure 12).

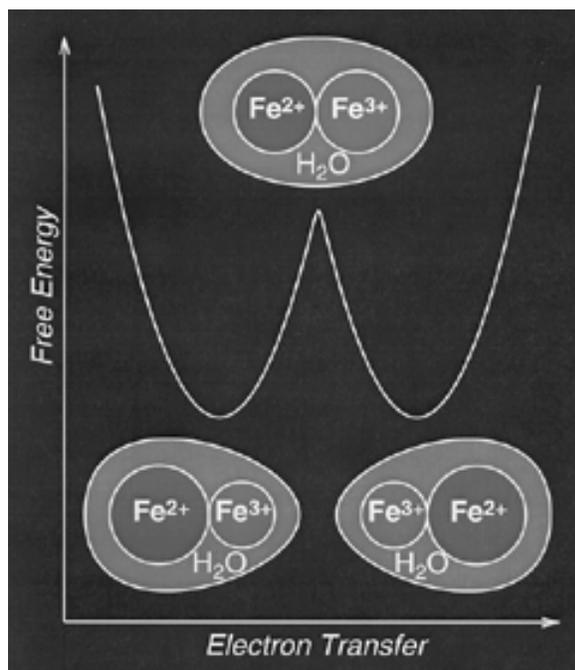


Figure 8. The reorganization of water molecules around highly-charged metal ions retards the transfer of electrons.

We measured the time it takes an electron to tunnel certain distances. And when we finished these experiments we proved that a protein with a metal ion embedded in it could undergo rapid electron tunneling over long molecular distances.

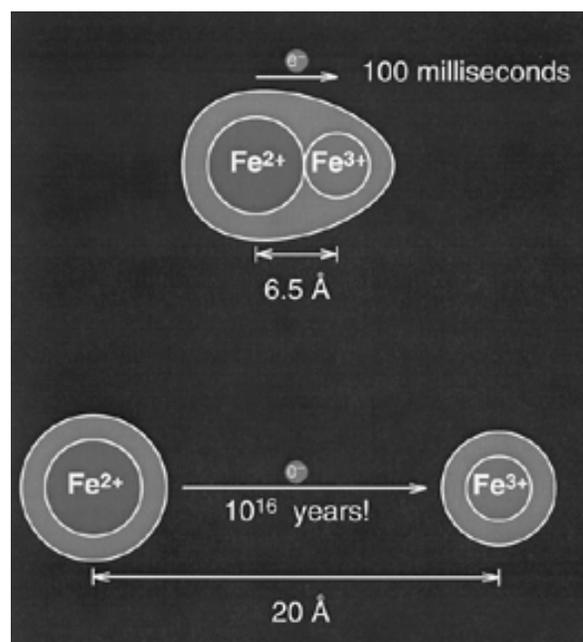
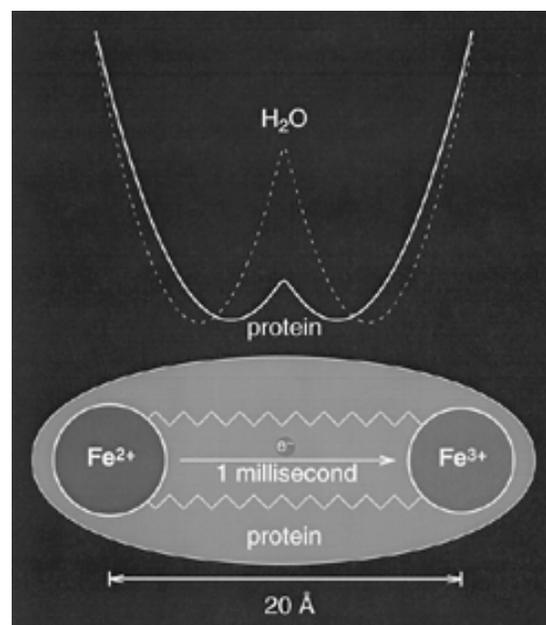


Figure 9. A quick calculation shows it would take an electron  $10^{16}$  years to travel just 20 angstroms. The answer to efficient tunneling or electron transfer? Bury the iron ions in a protein, reducing the barrier and thus the travel time to 1 millisecond, (figure 10, below).



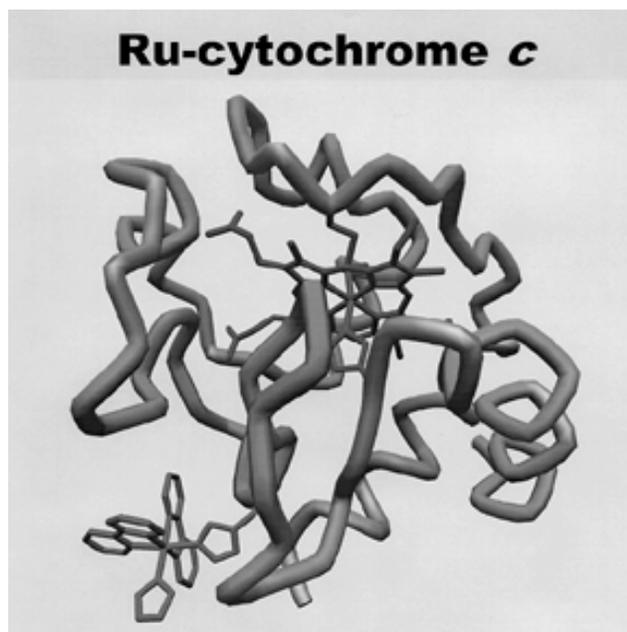


Figure 11. Step toward organic-inorganic water splitter: cytochrome c protein molecule with a heme group and a ruthenium complex.

### Tunneling timetables

Then we started generating what I call tunneling timetables, with "time"—from picoseconds, or trillionths of a second, to seconds, on one axis, and the distance of tunneling on the other axis (figure 13). Electron tunneling in a vacuum drops off very fast, that's the basis for STM. But we found an electron could tunnel through a protein in accord with the "organic" tunneling timetable. An exponential decay of 1 per angstrom means that at 20 angstroms the tunneling time is a microsecond rather than  $10^{16}$  years.

We noticed that there were two points in the timetable that were way off, and they were very important points, because they had to do with a complex that actually does solar conversion. Somehow that structure was built to make tunneling very slow. Why? When its structure was determined in the 1980s, we knew why. Because the reaction centers that harvest solar energy have the organic electron transfer components I told you about completely embedded in  $\alpha$  helical structures (figure 14). What is the secret of the alpha helix? Well, it's very poor at electron tunneling because of the long and winding chemical bond pathway the electrons have to travel (figure 15).

If you take the other kind of element in proteins, the beta strand, that's an expressway for an electron to tunnel. And we said "aha!" The tunneling time for an alpha helix is going to be much longer than the tunneling time for a beta strand; at any given distance, you're going to tunnel much faster in the beta strand than the alpha helix. But the story's much more complicated than that.

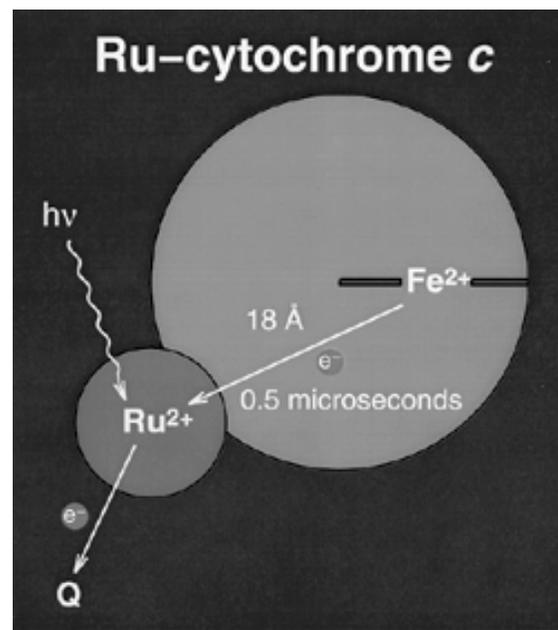
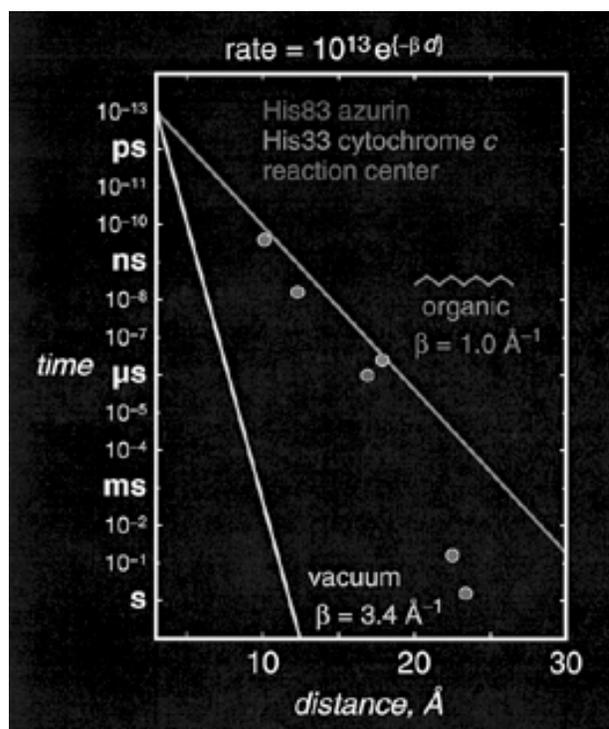


Figure 12. Laser experiments showed engineered protein could tunnel an electron in just .5 microseconds (500 nano-seconds). Figure 13 (below) is a tunneling timetable.



Ten years later, after mapping several proteins completely using protein engineering and an enormous number of laser experiments, we think we've solved this problem. The alpha helix is the most insulating in one direction,

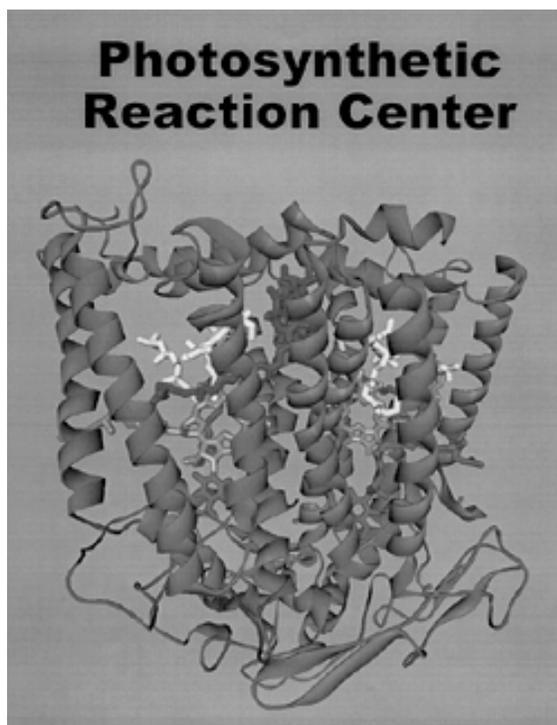


Figure 14. Photosynthetic reaction center is designed to efficiently harvest solar energy.

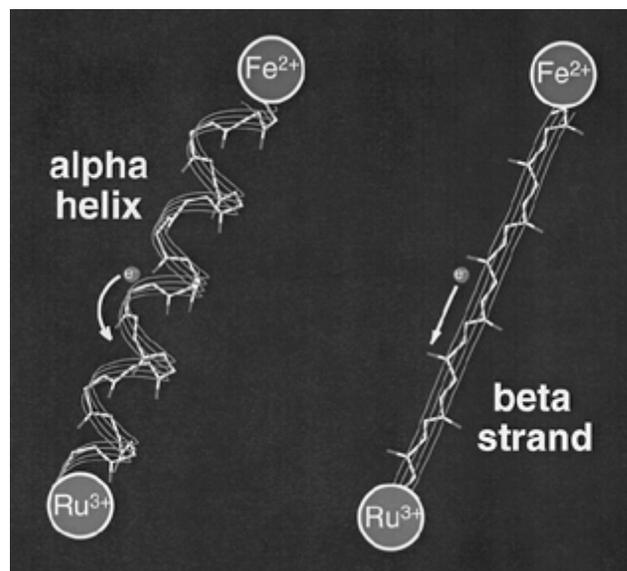


Figure 15. Alpha helix can be an insulator; the beta strand provides a better tunneling pathway.

Figure 16. Mapping various proteins reveals that the alpha helix is most insulating in one direction.

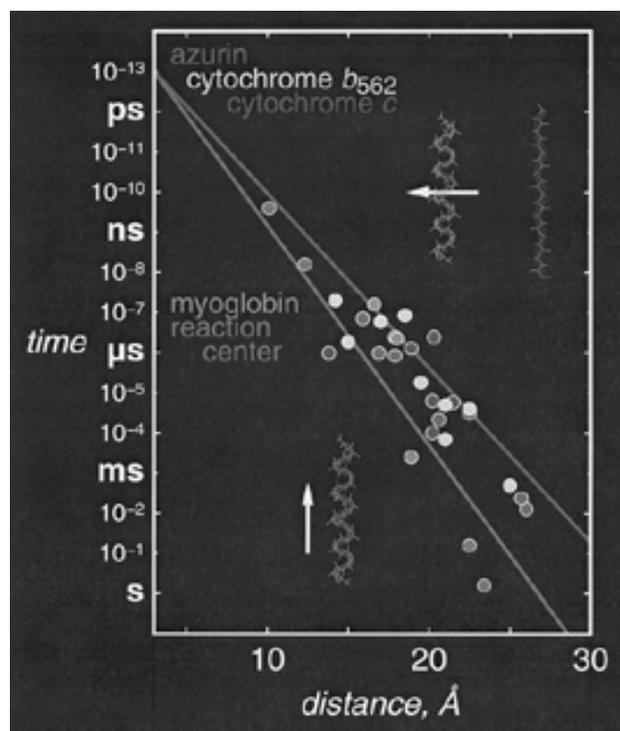
but interestingly enough—and surprising to us—the alpha helix in the other direction is very conductive. So if you want to run electrons fast, you go one way on an alpha helix or on a beta strand (figure 16). And if you want to stop electrons from tunneling, which you want to do in solar energy conversion, because the big problem is keeping electron and hole separated long enough to make oxygen from water, you can use an alpha helix as in figure 15.

**Back to photosynthesis**

We went back and looked at the reaction center in photosynthesis, and sure enough, all the initial fast acts from the special chlorophyll pair are along good pathways. But at the quinone, the electron should not recombine with the hole, otherwise you just get heat. There’s an alpha helix that prevents this electron to hole reaction.

The master tunneling timetable from our experiments shows that to tunnel really fast you have to use a series of conjugated double-bonds, or pi tunnels; those are really conductive (figure 17). From this tunneling timetable, you can now engineer a structure to put in conducting components for electron flow, and put in insulating components to keep the electrons and holes apart.

And now we can go back to nature’s complex (figure 18) and replace the chlorophyll with ruthenium complexes; and very importantly, on the oxygen side it’s going to be



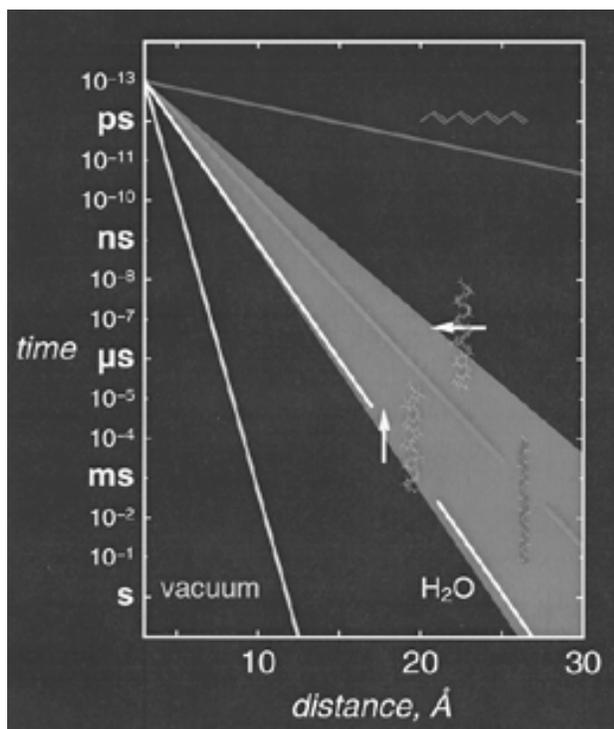


Figure 17. Master tunneling timetable contrasts fast electron tunneling through pi bonds (top line) with slow tunneling in vacuum, bracketing electron tunneling (gray area) in an alpha helix in various configurations; a practical water-splitter could thus use an alpha helix as a basic design element.

completely inorganic. We're going to give in a little bit by having organic stuff on the hydrogen side, but on the oxygen side, where we need the thing to be durable, it's going to be purely inorganic. We can put a hydrogenase enzyme on the hydrogen side, with molybdenum cyanide and ruthenium dioxide on the oxygen side. On the hydrogen side we can regulate electron flow by engineering the protein structure. We haven't put it together yet, but the electron tunneling rate on both sides suggests that we're getting close to an efficient, rugged molecular water-splitting device. That's good news.

**Semiconductor water-splitters**

The other good news is from NREL, the National Renewable Energy Laboratory. Not long ago the NREL people published in *Science* the world's record water-splitting solar cell (figure 19) in which they interfaced a gallium arsenide semiconductor with gallium indium phosphide. The gallium arsenide captures infrared light, the gallium indium phosphide captures visible and UV light on the hydrogen side, evolving hydrogen on the left; the holes produced by the gallium arsenide go over to platinum (right), evolving oxygen. This water splitter is 12.5 percent efficient, the world's record. So now, not only are the molecular devices coming fast, these cells with semiconductors to split water are coming very, very fast. I'd say in the last two or three years we've made an enormous leap towards the goal of an efficient, very durable, water-splitting device.

Daimler-Benz, Toyota, Honda, General Motors all know the inorganic chemists are going to come through for them.

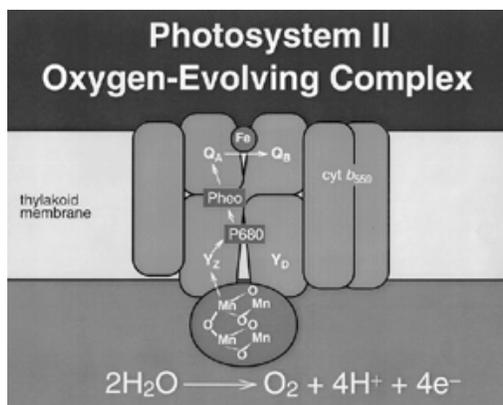
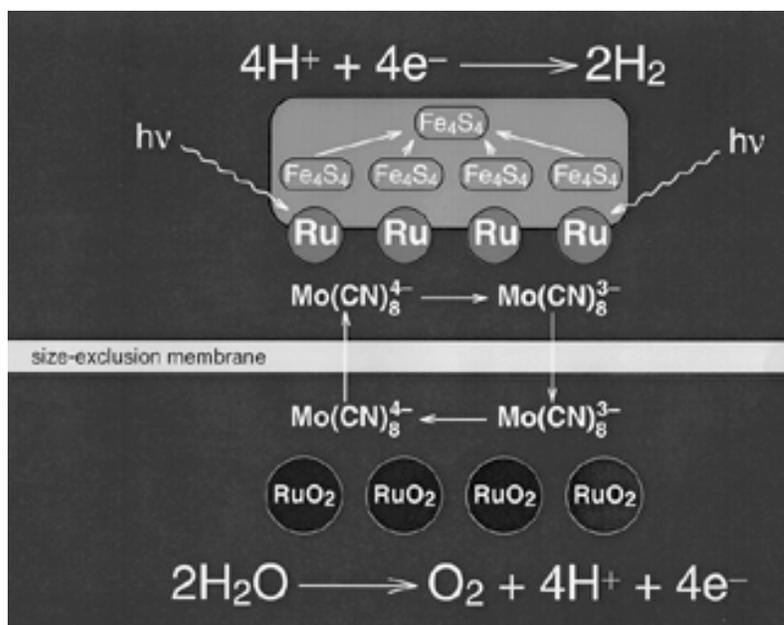


Figure 18: Going back to the drawing board, we will replace the chlorophyll in photosystem II with ruthenium complexes; replace some parts with a protein incorporating metals to evolve hydrogen, and make the oxygen side completely inorganic.



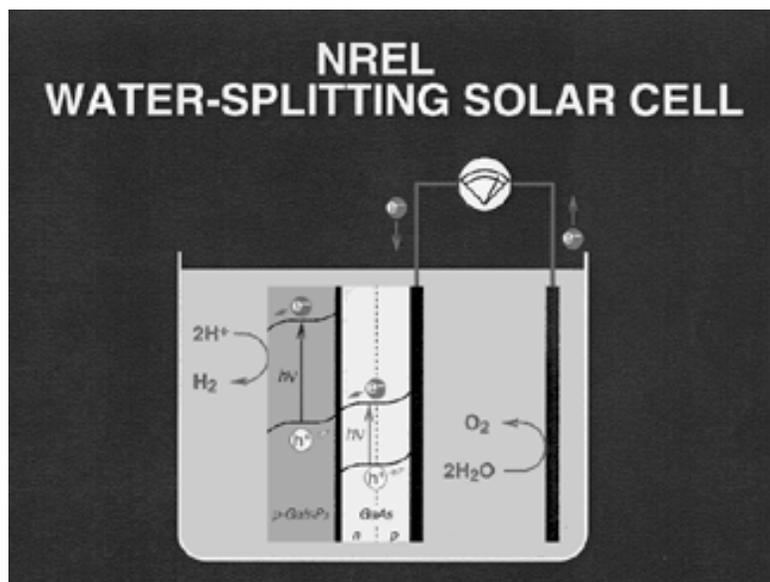


Figure 19: NREL water-splitting solar cell uses two semiconductors, (gallium arsenide and gallium indium phosphide) to split water with 12.5% efficiency. The sealed GaAs absorbs infrared and produces holes with the right amount of energy to produce oxygen when fed to a stable platinum electrode; the GaInP is used directly as an electrode, absorbs ultraviolet and visible light, and produces electrons with the right energy to evolve hydrogen.

So let me show you a couple of the cars. Nacar 2, by Daimler-Benz, now Daimler-Chrysler, has hydrogen tanks at the top(see minivan, page 5), it's running on hydrogen fuel cells, no emissions—this is a clean car. But you are not going to buy it because it has a big humpback, it looks ugly.

**Beautiful cars, 5,000 MPG**

The car you're going to see in the next ten years is going to have a solar panel charging the battery and an electric motor. It's going to have a fuel cell for long-range driving. In other words, it's going to be an electric vehicle that won't stop on the freeway after 40 miles because the battery just ran out. This is going to be a solar-charged battery car with fuel cells for long-range driving that will have a range of 500-700 miles and in Tucson and Los Angeles it could get 5,000 mpg. The fuel will either be hydrogen or more likely, in the first versions, methanol.

If you only drive around town, you'll get an infinite number of miles per gallon because you'll never turn on the fuel cell; the car will be run with a solar panel.

These cars are on the drawing board now. Someday your high school students will see these cars. They are going to be bought by people because they'll be better in absolute terms. They're going to be beautiful cars, they're going to burn rubber, and they're going to look like Jags—and they're going to get 5-10,000 mpg! People are going to buy these cars.

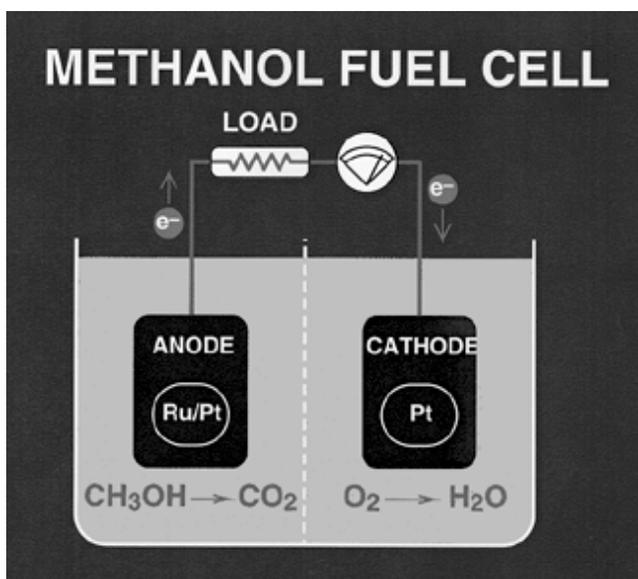


Figure 20 (above). Direct oxidation methanol fuel cell is an intermediate step between gasoline and hydrogen

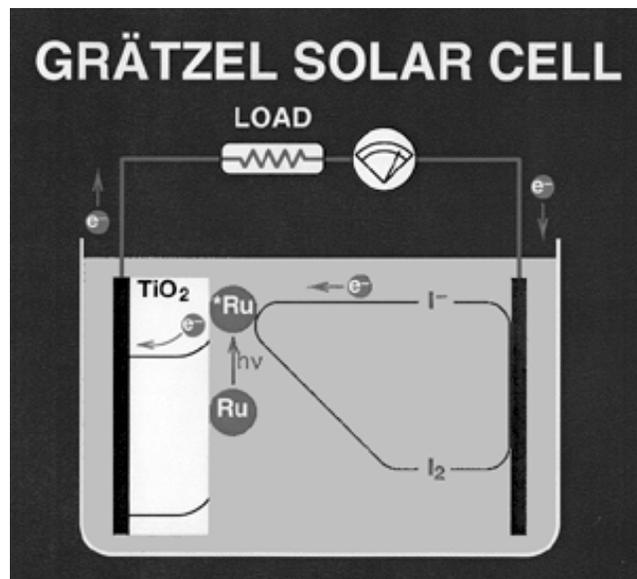


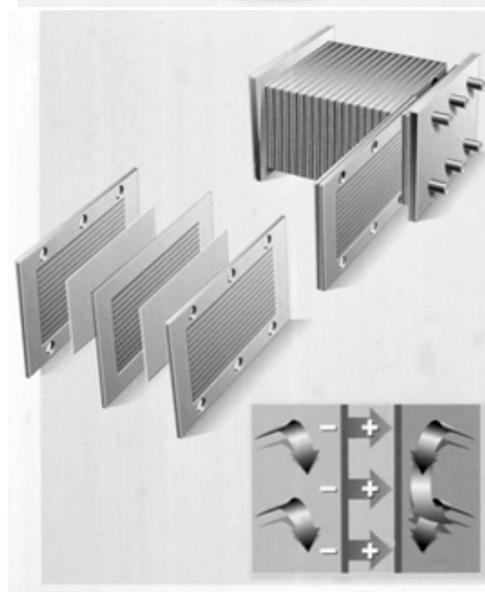
Figure 21. Grätzel solar cell is made by coating titanium dioxide with ruthenium. The solar-irradiated ruthenium injects electrons into the TiO<sub>2</sub>, making electricity.

Figure 22. Schematic drawing of fuel cell construction and placement of Daimler-Chrysler's Necar 4. Liquid hydrogen is stored in a cryogenic cylinder at the rear of the vehicle, and fed into a proton exchange membrane fuel cell. The car's top speed is 90 mph; it has a range of nearly 280 miles.

In order for this hybrid car to be possible, we're going to have to use a direct oxidation methanol fuel cell with a ruthenium-platinum anode (figure 20). And there is very good news from the semiconductor people. Michael Grätzel in Lausanne, and many other people around the world, are working on cheap semiconductors, sensitizing them with ruthenium complexes (figure 21). So we have some very good candidates now for the solar cells that could charge the batteries with sunlight.

Daimler-Benz has a car on the test track, Necar 4 (figure 22). It's a nice little car, it runs entirely on a fuel cell. The next version of this car may have a solar pack. Toyota has a hybrid car, gas-electric, Honda has one, they're starting to appear everywhere. We're moving down the track toward emission-free cars, but we still need help.

We need you to recruit your very best students to go into chemistry because every problem we've talked about tonight is a problem in chemistry that will have to be solved by chemists. And so we've got to keep talking it up, we've got to keep working on it. If we are successful, in 20 years, we will be driving cars that don't pollute; what is more, we also may have big power stations that don't pollute. We'll be running this planet with chemistry that's almost as good as the chemistry nature developed so many years ago.





*M. Bonner Denton is a professor of chemistry and of geosciences at the University of Arizona. His research interests include applying the latest technological advances in electronics, physics, optics, astronomy, acoustics, mechanical engineering, and computer science toward developing improved methods of chemical analysis. Research projects currently under way include: selective single photo ionization of complex mixtures; studies of new modes of radio-frequency-only quadrupole mass spectrometry; the use of charge injection array detectors in intelligent instrumentation; new hollow anode cathode discharges as spectroscopic sources; charge-coupled array detectors in rapid scan and low photon flux spectroscopy; studies of excitation phenomena in inductively coupled plasma discharges; element-selective detection in gas and liquid chromatography; and new modes of automated sample-handling methods.*

*Bonner Denton joined the faculty of the University of Arizona after receiving a Ph.D. from the University of Illinois at Urbana-Champaign in 1972. His numerous honors include an Alfred P. Sloan Research Fellowship, an ACS Division of Analytical Chemistry Award in Chemical Instrumentation, and the 1998 Pittsburgh Spectroscopy Award of the Pittsburgh Conference on Analytical Chemistry and Applied Spectroscopy.*

## Advances in the Application of Array Detectors for Improved Chemical Analysis

Adapted from a 2000 presentation by M. Bonner Denton

**M**y message to all of you today is something I know you already fully believe, deep down in your hearts—science and technology are fun. I am an analytical chemist, and analytical chemists often want to know what's there (qualitative analysis), how much is there (quantitative analysis), and what form it is (speciation). What we'd all like to have is what I refer to as the Mark I Magic Analyzer, where we take any sample—it might be an old boot—and put it in the hopper of the Mark I, and it would tell us everything about that sample (Figure 1). It gives us all the answers, and it even gives us the whole boot back, because we might want to wear it again. Unfortunately, today we don't have a Mark I Magic Analyzer. We're forced to perform most analyses using techniques that have been developed over many years. One of them is optical spectroscopy. The original detector for optical spectroscopy was simply the human eye, but technology has changed over the years. In the old days, chemists would run flame tests in which, for instance, a strontium solution would be exposed to a flame and would emit red light; sodium would emit yellow; and barium would emit green. Nowadays, instead of using a flame, which is only a few thousand degrees Kelvin, we use an inductively coupled plasma, which is a radio-frequency-excited argon discharge possessing temperatures approaching 10,000 degrees Kelvin or less. We still heat the samples and then look at the different colors produced, but we don't look at them with our eyes.

For many years, spectroscopic analysis was accomplished by collecting light from a source, whether it be flame or plasma, and transferring the light to a prism or grating where it is split into different colors. Either a

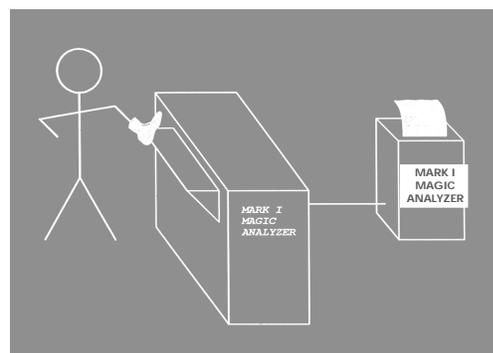


Figure 1. The hoped-for Mark I Magic Analyzer could provide analysis of any object without destroying it.

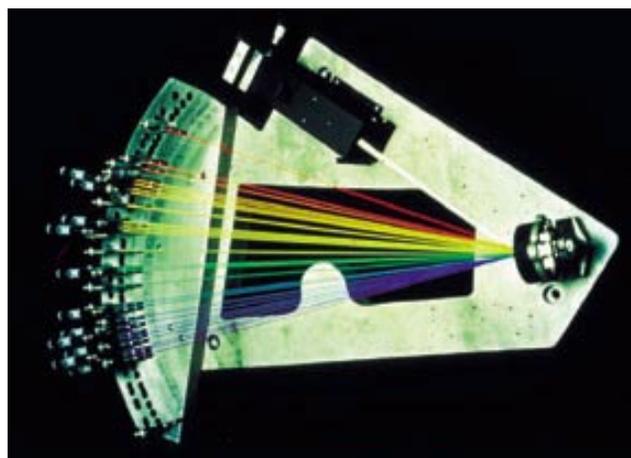


Figure 2. A multiwavelength polychromator system. Each detector must be aligned to the exact position required to observe an individual wavelength.

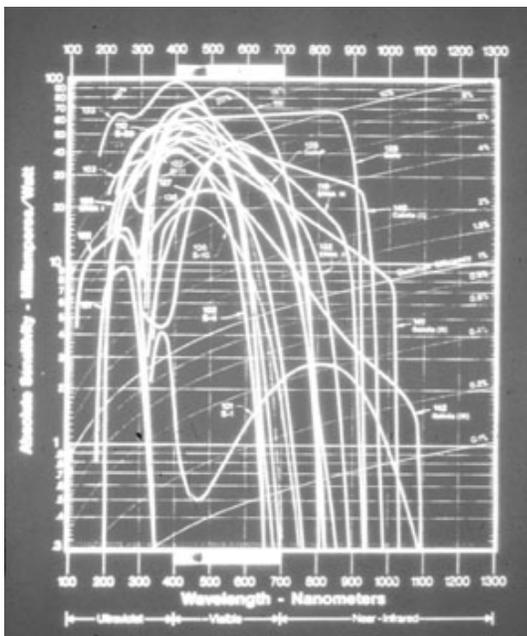
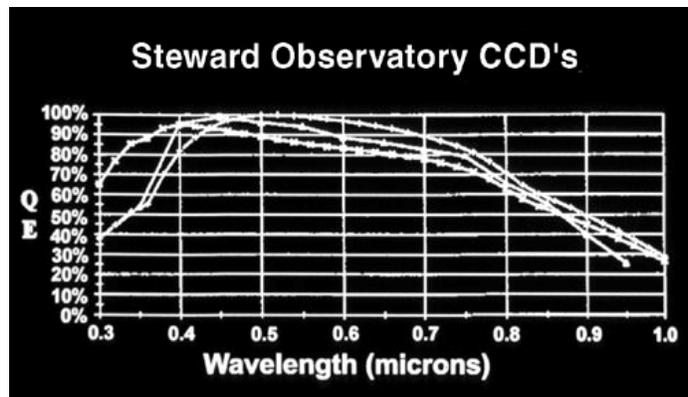
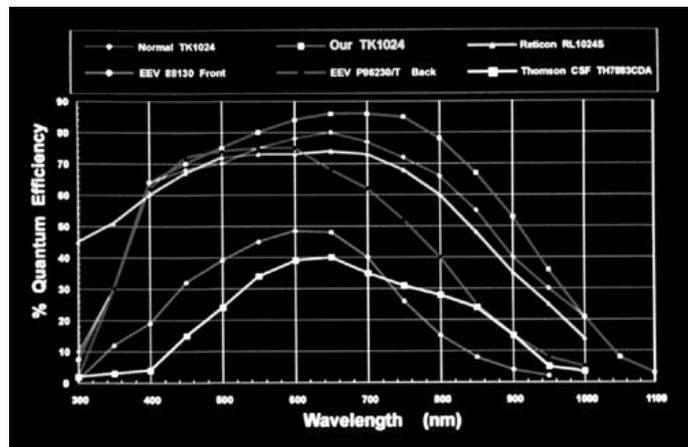


Figure 3. Photocathode spectrum (above) displayed quantum efficiencies (QEs) from a few percent to almost 40%. In contrast, CCDs often have QEs peaking at over 70% (Figure 4., top right), and several devices recently developed at Steward Observatory in Tucson have QEs approaching almost 100% (Figure 5., right).



single slit and photomultiplier detector or a series of individual slits and detectors were used to measure the intensity at each of the wavelengths. An example of a multi-wavelength polychromator system is shown in Figure 2. Each of the individual detectors must be laboriously aligned to the exact position required to observe an individual wavelength. It can take a competent engineer an entire day just to align one detector.

However, there are actually thousands of discrete wavelengths that provide meaningful information when complex samples are observed. For years, scientists had dreamed of some type of electronic readout that would allow us to observe all wavelengths simultaneously. In the early days of the '70s and '80s, a series of TV-camera-types of devices were evaluated. However, they were all found to be unsatisfactory, for one or more reasons. Not only was an extended wavelength region from 800 nm to 180 nm necessary, i.e., from the near-infrared through the ultraviolet, but extremely large dynamic ranges were required.

**Adapting CCD technology to spectroscopy**

In 1970, Willard Boyle and George Smith of Bell Laboratories described the first charge-coupled device (CCD), and

in 1972, Gerry Michon and Hugh Burke at General Electric developed the charge-injection device (CID). By the late 1970s, a number of astronomers started using the new CCD technology to explore the skies. These devices demonstrated promise that they could be the long-sought electronic equivalent of a photographic emulsion, possibly even exceeding its capabilities. My research group began collaborating with several astronomers, because it was clear to me that these solid-state devices had great promise for terrestrial spectroscopy. We received some early assistance from Dave Hunter, a General Electric field engineer, who agreed to provide us with a device and some preliminary engineering drawings. However, we were warned *never* to call the G.E. factory.

My group started incorporating these CIDs (as well as CCDs) into electronic systems capable of observing spectroscopic information. We developed new forms of readout, were the first to operate CIDs in the ultraviolet spectral region, and rapidly explored their potential in spectroscopy. The photomultiplier tube technology used previously provided quantum efficiencies from a few percent to less than 40 percent (Figure 3). In contrast, CCDs often have quantum efficiencies peaking at over 70 percent (Figure 4), and several devices recently developed by

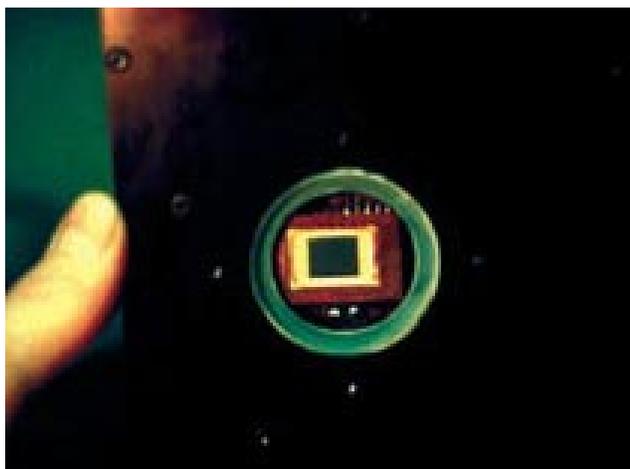


Figure 6. When first introduced, tiny devices, such as the one pictured above, with over 64,000 detector elements were greeted with scepticism.

Mike Lesser at Steward Observatory have QEs approaching almost 100 percent (Figure 5). These detectors, when properly operated, can exhibit extremely low dark current and read noise. One disadvantage, however, is that the detectors can require a significant amount of time to read the signal. As the number of detector elements increases, this becomes a larger limitation.

### Persistence does pay

In the early 1980s, I would attend national and international meetings and announce that I had discovered a wonderful new detector. Instead of having devices with only a single detector element, I described devices with over 64,000 detector elements. At that time, colleagues would say, "Bonner, you're crazy! Look how small your detector is (Figure 6). You'll never be able to use those tiny little detector elements. To be useful, the detector element must match the optical configuration of the spectrometer." Conventional instruments designed for photomultiplier tubes were all configured with slits and detectors on the order of 100 microns wide and 3 mm tall. "Bonner, your detector elements are only 20 microns by 20 microns. It's worthless. It will never work." So I have a message for teachers and students; you teachers already know it, but many students might not. When people say something won't work, don't necessarily believe them. Evaluate what *you* think the results will be and if you are convinced it *will* work, keep on trucking!

What we had to do was develop a custom set of optics, not an entirely trivial task. After trying several configurations, we decided that we needed some real expert help, so I made a call to the Optical Sciences Department. Luck-

ily, the University of Arizona has one of the most highly renowned optical sciences programs in the world. I recruited a graduate student, Russ Chipman, who had previously spent seven years in industry designing spectrometers. We turned him loose on a supercomputer and said, "Design a set of optics using a commercially available grating." Other than that, I said he could design anything he wanted.

Not being more specific was a mistake on my part, because when we tried to get someone to fabricate some of his custom components, we were told that it would not be possible to manufacture these elements. Persistence again paid off, because after searching the world over for a manufacturer, we found one right in our back yard—a man with world-class optical fabrication talents, Mr. Don Loomis, who before retirement had been in charge of Kitt Peak National Observatory's optical fabrication shop. He had a workshop behind his house that was amazing. When I showed him the plans for the components, he said he knew how to build them. He made the whole optical system, and it worked beautifully. Once again, we were rocking and rolling with science.

Instead of having a long, horizontal focal plane, these echelle optics took the equivalent of a 50-foot-long focal plane, broke it up into little stripes, and stacked the stripes one above the other, allowing the full X-Y array designed for television camera applications to "take a picture" of the spectroscopic information. At this point, a number of manufacturers of atomic spectroscopic instrumentation became interested in our approach, and we were able to

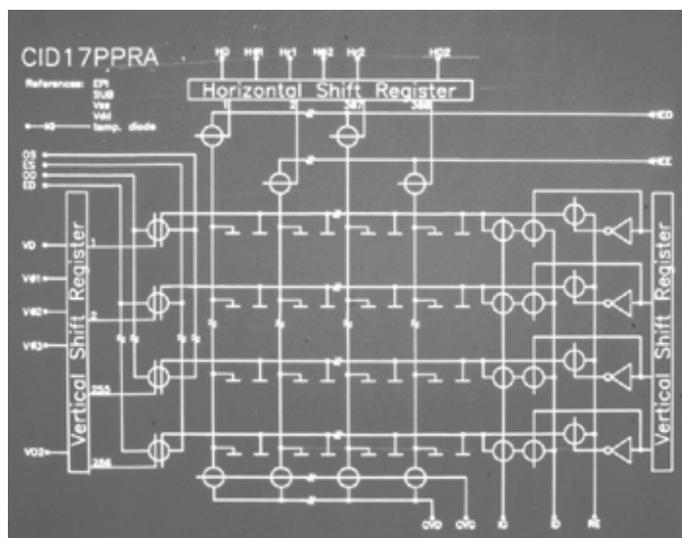


Figure 7. Diagram of optimized CCD detector with X-Y array based on configuration originally designed for television cameras.

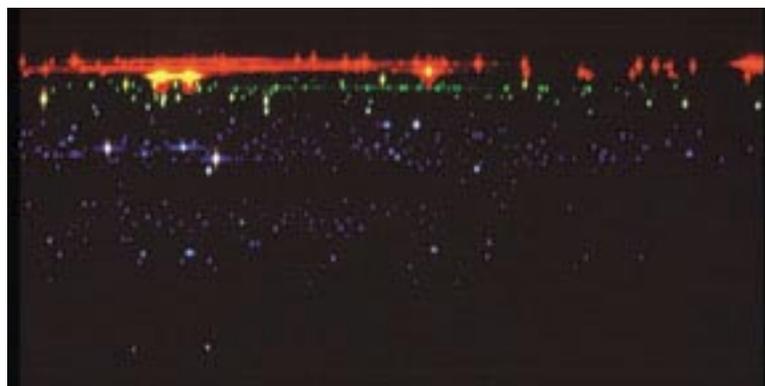
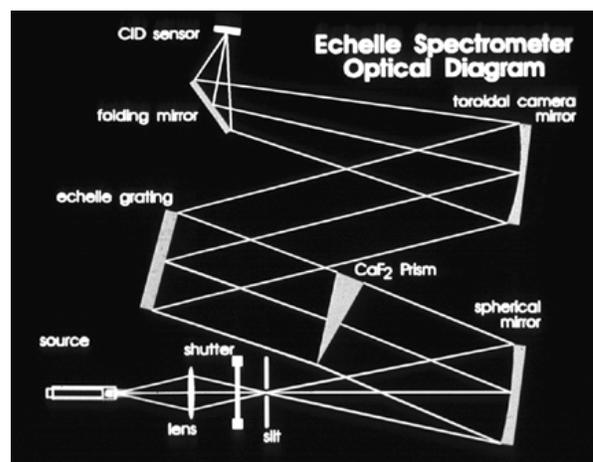
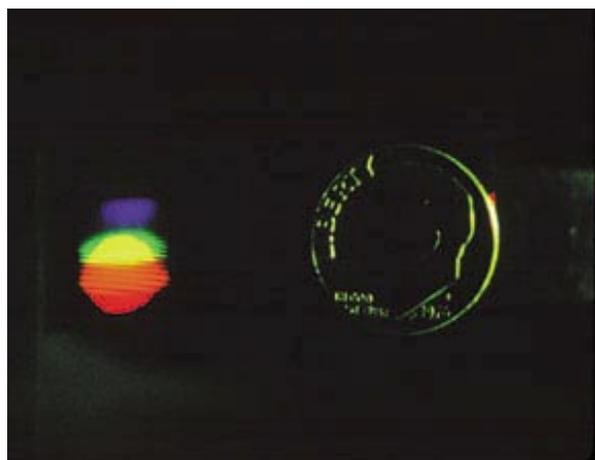


Figure 8. A continuum spectrum from an early echelle designed for spectroscopic application (top left). Echelle optics took the equivalent of a 50-foot-long focal plane, broke it up into little stripes, and stacked the stripes one above the other.

Figure 9. Top right, a schematic of a modern echelle system.

Figure 10. The atomic spectrum from a modern echelle system (left). The detection limits are better than those achieved with photomultiplier tubes, and these instruments are capable of observing all wavelengths simultaneously.

convince CIDTEC, the company that spun off from G.E. with the CID technology, of the potential financial promise of developing optimized CID detectors (Figure 7). An example of a continuum spectrum from one of the early echelles designed for spectroscopic application is shown in Figure 8. Today, similarly based array detector atomic emission instruments are being marketed by most major manufacturers. A modern echelle system is shown in Figure 9 and an atomic spectrum is shown in Figure 10. The detection limits are better than were ever achieved with photomultiplier tubes, and because these instruments are capable of observing all wavelengths simultaneously, they have complete flexibility for the analysis at hand.

### Real-world problems analyzed with new spectroscopic techniques

Several years ago, a major automotive manufacturer wanted to be able to accurately analyze catalytic converters for platinum, palladium and rhodium. They were buying large quantities of these metals in bulk from South Africa, Russia, and a few other suppliers. They would then ship the materials to outside contractors who would incorporate the platinoids into proprietary wash coats for

application on a ceramic brick to serve as the active element in a catalytic converter. Each of the individual wash coat manufacturers had their own proprietary formulation, which was not known to the automotive manufacturer. The choice of wash coat and hence vendor was dependent on empirical tests performed with each different type of engine. Unfortunately, many of the wash coat matrices were so complicated that they were found to be very difficult to analyze for the platinoids. The best the automotive manufacturer could achieve was plus or minus 7–10% for the concentrations of platinum, palladium and rhodium. I am sure everyone here is familiar with the expense of platinum, but in reality, platinum is relatively cheap compared to rhodium, which varies widely but often falls within the range of \$3,000 to \$6,000 per ounce. If the contractors were shaving off 5–7% of the rhodium and holding it back, it would certainly be to their own financial advantage. Obviously, the bean counters at the automotive company did not want to see this happen, so they came to us and asked if we could devise new techniques capable of accurately analyzing catalytic converters.

We first developed new techniques for solublizing the entire brick and discovered that the resulting solution pro-

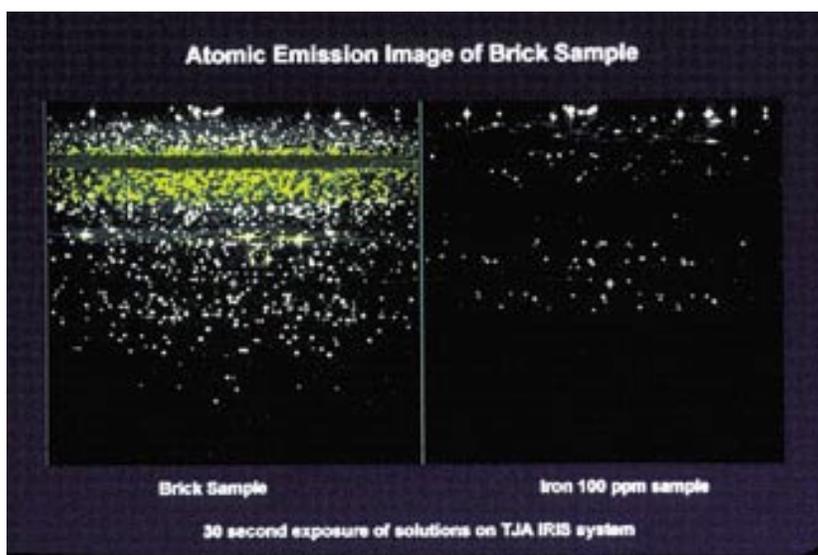


Figure 11. Atomic emission images of a solubilized brick sample from a catalytic converter (left) and the relatively complex but much simpler spectrum produced by a 100 ppm iron sample (right).

duced a very complex spectrum. Figure 11 contrasts a solubilized brick sample with the relatively complex but much simpler spectrum produced by a 100 ppm iron sample. We laboriously attacked the many potential spectral interferences by first looking up coincident wavelengths and nearby wavelengths for all the elements found in the sample. We then conducted numerous spectroscopic analyses on the elements present in the brick, as well as on solutions containing combinations of the elements present. We eventually were able to determine appropriate wavelengths that were either interference-free or that exhibited a minimal interference. In the end, we developed a protocol that allowed for an effective and accurate analysis of catalytic converters.

### Artificial intelligence enhances new spectrometer technology

After we were through with the method development,

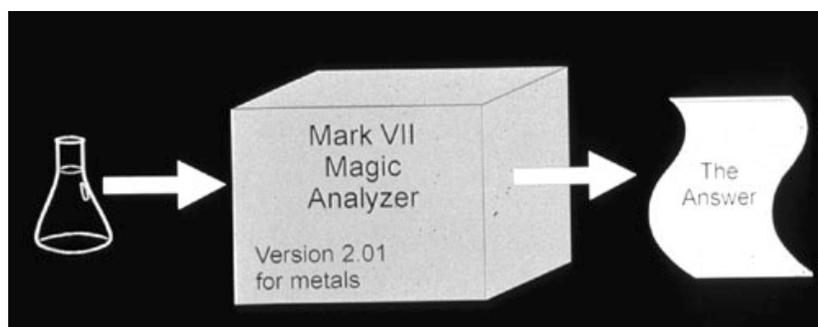


Figure 12. A better way to develop analytic protocol by employing artificial intelligence—an instrument with a “built-in brain”—is close to being a reality.

we asked ourselves: “Wasn’t this the brute force approach? Wasn’t there a better way to develop this analytic protocol?” Then we asked ourselves: “Why not employ some form of artificial intelligence?” How could we make the instrument develop the method automatically? Why should we spend months measuring and cross-referencing all of these interferences to pick out appropriate wavelengths? What we really want is the Mark VII Magic Metals Analyzer (Figure 12). We have been working on this concept ever since and are almost there—the equivalent of an instrument with a built-in brain.

This capability is achieved by employing mathematical and statistical algorithms from a field of chemistry known as chemometrics. After running a suitable “training” set, we can now analyze simultaneously for 31 elements, but instead of just observing 31 wavelengths, we actually measure 1,632 wavelengths. The instrument immediately reports semi-quantitative analyses without having to run any new standards. Often these numbers are within a few percent to 10% of the actual value, satisfying the requirements for most routine analyses. However, if more accuracy is required, as was the case with the catalytic converters, the instrument also provides a list of interference-free wavelengths optimized for whatever elements we wish to analyze. The instrument also provides appropriate concentrations for standards to provide the desired analysis. Instead of the months of time it took us to develop the technique for catalytic converters, we now immediately know all the elements present in the unknown sample and have a semi-quantitative determination of their concentrations, as well as of the wavelengths and standard concentrations necessary for high accuracy quantitative analysis.

We have also been using the same technology for an element-selective chromatographic detector. We can take the output of a high performance liquid chromatograph (HPLC) (Figure 13) or a gas chromatograph (GC) and observe the atomic emission from all the elements in the effluent. If one observes not only the UV and visible but also the vacuum UV spectral regions, one can analyze for all the elements in the periodic table, producing an element-selective detector capable of providing qualitative and quantitative data for all elements.

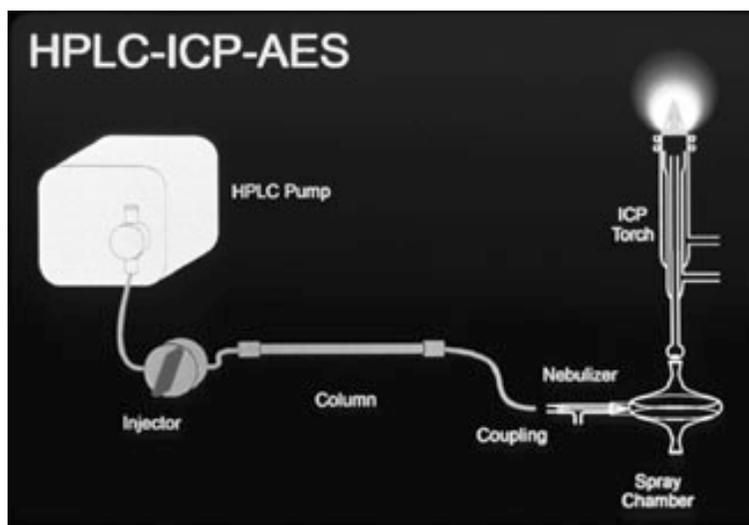


Figure 13. With the output of a high performance liquid chromatograph (HPLC) or a gas chromatograph the atomic emission from all the elements in the effluent being analyzed could be observed.

#### Rapid solids analysis made possible by DC-plasma-based system

A major challenge for these new spectroscopic systems is the analysis of real-world samples, such as the catalytic converter. Many academics are great at developing their favorite techniques for analyzing artificial samples in triply distilled, deionized water, but seldom test their method's capability to analyze complex real-world samples. Recently, we were challenged to develop a method for analyzing raw sewage at a Pima County Wastewater treatment plant for toxic metals. This is a real-world, hard-to-analyze sample that contains approximately 9 percent suspended

solids. The inductively coupled plasma I mentioned earlier will not tolerate such high levels of suspended solids. If you put such a sample into an ICP, it will either plug the nebulizer or quench the plasma. We therefore were forced to explore other approaches. The problem at the treatment plant is that individuals arrive with both large and small pumper trucks and, of course, always say that they have just pumped out a few residential septic tanks. However, the County must make certain, because if in fact one of those pumper trucks is full of high valency chromium from XYZ plating company, and this is fed into the wastewater process system, all the little beasts that digest the sewage may be killed. You have heard about being up to your neck in something before. Well, this is really it! When the process in the plant stops running and sewage is no longer properly digested, you have an unacceptable disaster on your hands.

To prevent this, Pima County must analyze every single tanker truck that arrives at their facilities to ensure that no toxic materials are introduced into their process. Conventional techniques required dissolving the suspended solids with a complex chemical digestion process, which can take several hours. Such a delay, however, is extremely unpopular with tanker-truck drivers. We therefore developed a DC plasma array detector technique that will accept these very high levels of suspended and dissolved solids. We employed a conventional DC plasma in a new optical geometry with an improved nebulizer spray chamber (Figure 14). With this improved technology, the system was able to meet all detection limit and accuracy criteria necessary for the wastewater analysis.

#### New detectors applied to arc and spark spectrometers

Two other techniques utilized for directly analyzing solids—arc and spark spectroscopies—have been around for many years. These spectrometers, commonly used for quality control in the metallurgical industry, excite the sample with either a high voltage spark or arc. The resulting atomic emission was observed with the direct reader, shown earlier in Figure 2. Recently, Baird Corporation and Thermo Jarrell Ash Corporation have introduced arc and spark spectrometers using a charge-

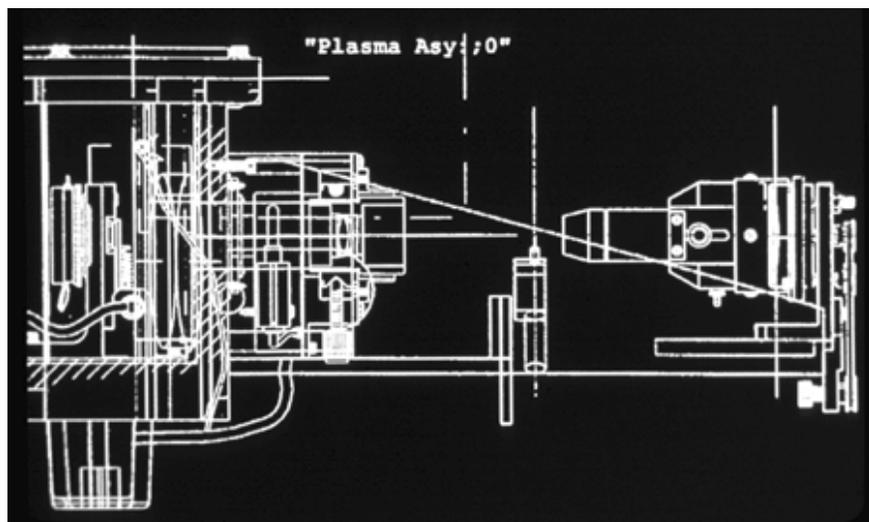


Figure 14. A conventional DC plasma in a new optical geometry with an improved nebulizer spray chamber. With this improved technology, the system was able to meet all detection limit and accuracy criteria necessary for the wastewater analysis.

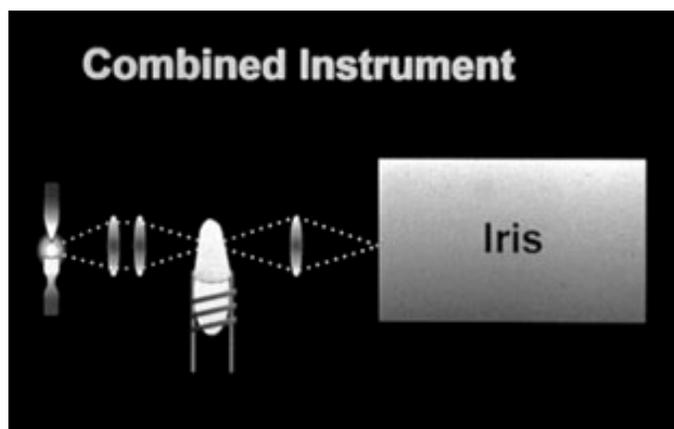
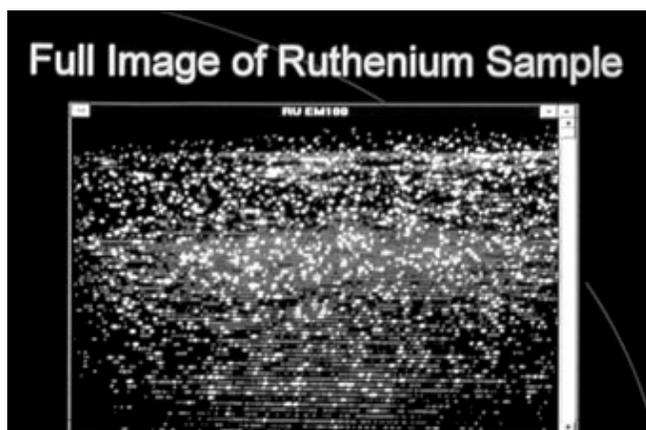


Figure 15. The Denton research group's version of a combined ICP/Arc/Spark instrument (above). An illustration of the complexity of such a spectrum is shown in Figure 16, below.



injection device to simultaneously observe all wavelengths. Our research group's version of a combined ICP/Arc/Spark instrument is shown in Figure 15. An illustration of the complexity of such a spectrum is shown in Figure 16. Variations in background between samples, if not properly corrected, can greatly degrade quantitation. Figure 17 shows a plot of concentration versus response without background correction. The results are poor. Because the array detector simultaneously monitors both the analytical wavelength and backgrounds on both sides of that wavelength, accurate background correction is readily accomplished (see Figure 18). We now have observed a reasonable correlation between increasing signal with increasing concentration and a reduction in the size of error bars. Atomic emission spectroscopy with array detectors demonstrates high reliability, high sensitivity, and a large linear dynamic range. Additionally, our ongoing research shows great promise for the development of practical, expert systems capable of rapidly developing analytical protocols for new classes of samples.

### The revolution in Raman spectroscopy

The next topic I wish to discuss is Raman spectroscopy. Raman spectroscopy was first employed by physicists and physical chemists studying fundamental phenomena. In the early days of Raman, it could take a whole day of exposure to a mercury arc to be able to observe a Raman spectrum of a pure compound. With the introduction of laser excitation, many spectroscopists felt that Raman spectroscopy would soon become a routine technique. It didn't happen. Today, the combination of array detectors, new solid-state diode lasers, and a variety of other new optical components are having a major impact on Raman. In fact, I believe these advances are transforming Raman into the

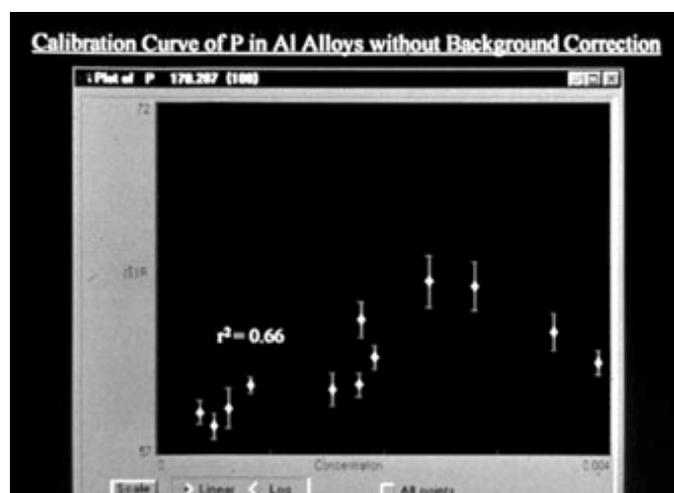
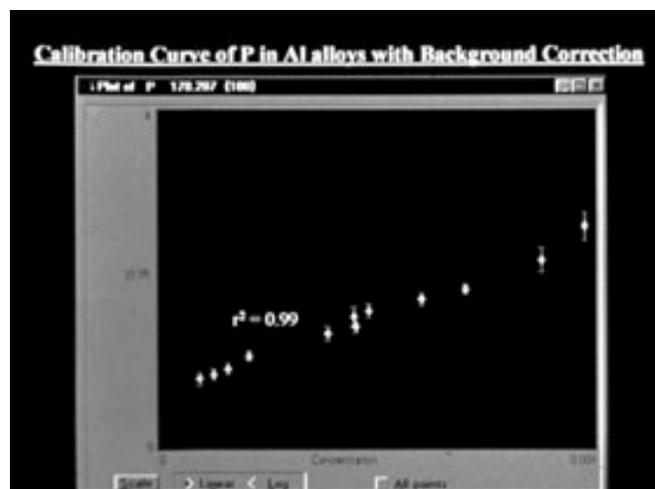


Figure 17. In a plot of concentration versus response without background correction the results are poor.

Figure 18. Accurate background correction is readily accomplished when the array detector simultaneously monitors both the analytical wavelength and backgrounds on both sides of that wavelength



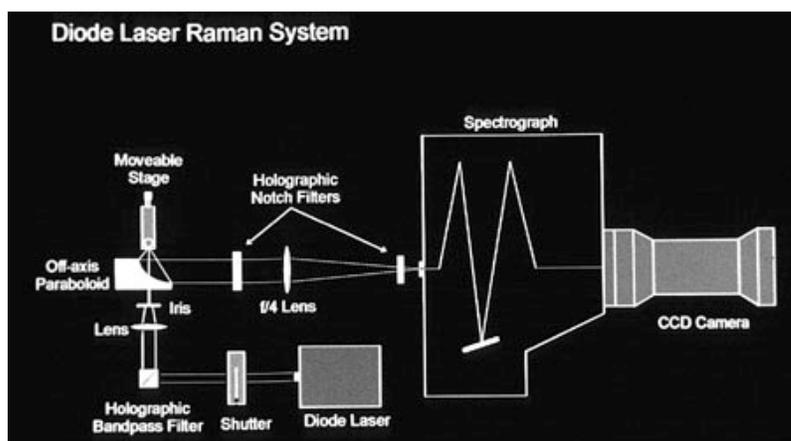


Figure 19. A schematic of Raman technology that could be developed into a practical commercial instrument.

awakening giant of analytical spectroscopy.

Several years ago, the Environmental Protection Agency approached my research group to explore Raman's capability in environmental analysis. While they already possessed mass spectrometers capable of performing most environmental analyses, these were large, expensive instruments requiring specialized laboratory space. The EPA wanted a technique like FTIR (Fourier transform infrared spectroscopy) that could be readily transported to toxic waste sites. Additionally, they were interested in coupling Raman spectroscopy to fiber optic probes, which would allow monitoring of organic pollutant plumes by lowering the fiber down water wells. The instrument configured by my group employed a new generation of diode laser, high quantum efficiency CCD array detector technology, and a new type of optical component, the volume phase holographic element. This technology, originally developed for the military, provides such high rejection of the Rayleigh scattered light from the laser that much simpler and higher throughput polychromators can be utilized.

The EPA asked us to examine a number of environmentally important dyes often found in municipal wastewater discharges. Our studies into detection limits of these dyes demonstrated that we could detect many of the compounds at levels as low as parts per million with integration times not in days or even hours, but in seconds to a couple of minutes. The detection limits by diode laser fiber optic Raman spectroscopy that had appeared in the literature at that time were on the order of 0.05 molar. Our ppm detection limits equated to  $10^{-7}$  molar, a decrease of 5 orders of magnitude—a significant improvement.

#### New databases for Raman spectroscopy

At about this time, a major instrument manufacturer

came to my research group to explore whether some of the Raman technology we were evaluating (Figure 19) could be developed into a practical commercial instrument. While instrumentation was rapidly approaching a level of sophistication suitable for routine analysis, one key component for success was still missing: a database containing spectra of a large number of compounds. However, Nicolet Instruments had already teamed with Aldrich Chemical Company, a major supplier of fine and specialty chemicals, to develop a database of almost every compound Aldrich offered for sale. We were able to acquire one of the first beta test versions of this database.

Our initial experiment was to search the database

#### Mixture of Hexane, Heptane and Octane Raman Search Results

1. Hexanes
2. Hexane
3. Heptane
4. Octane
5. Nonane
6. Undecane
7. Dodecane
8. Tetradecane
9. Decane
10. Hexadecane

Figure 20. Raman spectroscopy is capable of qualitatively identifying components in mixtures not easily determined with IR. A mixture of hexane, heptane and octane yielded a first hit of hexanes—not quite right but not really wrong, either: The second, third and fourth results were all correct. In Figure 21, below, IR recognized that it was a hydrocarbon mixture but was not as effective in picking out specific components.

#### Mixture of Hexane, Heptane, and Octane IR Search Results Using Multiple Libraries

1. Decane
2. Hexadecane
3. Octane
4. Dodecane
5. Ethylene, propylene, diene terpolymer
6. 2-Methylbutane
7. Pentane
8. Heptane
9. Poly(1-butene), isotactic, low m.w.
10. Petroleum ether

using a spectrum obtained from an Excedrin tablet (clearly a multicomponent, real-world mixture). The first compound identified was acetaminophen. The second compound found was acetylsalicylic acid. The third compound was caffeine. Amazingly, these are the active components of Excedrin. The full extent of sample preparation was simply to break the Excedrin tablet in half and place it on the spectrometer stage. This was the first sample we had ever searched in the database, and the ease and accuracy of the search astonished me.

By comparison, the results using IR spectroscopy were less accurate and required much more elaborate sample preparation than those using Raman. Just preparing the KBr pellet took one of my graduate students half a day, because he was not skilled in the technique. The first pellet he made had too much Excedrin. The next one didn't have enough. The next few were cloudy. Even after he obtained a good pellet, identification of the compounds left something to be desired. The only component correctly recognized was acetylsalicylic acid in KBr, and this was hit number seven, after six preceding incorrect hits.

Raman is capable of qualitatively identifying components in other mixtures not easy to determine with IR. For instance, a mixture of hexane, heptane and octane yielded a first hit of hexanes—not quite right, but not really wrong either: the search algorithm recognized that there was hexane in the mixture, but it wasn't pure hexane. The second hit was hexane. The third hit was heptane. The fourth hit was octane (Figure 20). That was exactly what was in our mixture. Using IR spectroscopy on the same mixture, octane was the third hit and heptane was eighth. The other hits demonstrated that IR clearly recognized that it was a hydrocarbon mixture but was not as effective in picking out specific components (Figure 21).

### Students begin using Raman to attack laboratory course challenges

The University of Arizona has a program, begun in one of our advanced graduate courses and since extended to our undergraduate course in chemical instrumentation, that involves giving the students real-world samples and posing a series of chemical problems for them to solve by analyzing specified components. In general, each student will be faced with a variety of different qualitative and quantitative analyses for organic and inorganic species. To explore the robustness of our Raman technology, we invited the students of these courses into our research laboratory and allowed them to use our research system to attack the various types of problems with which they were confronted.

One student was given a can of silver spray paint and

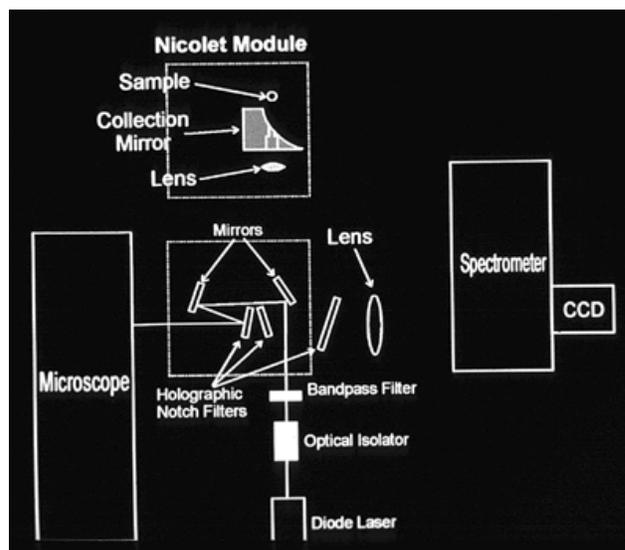


Figure 22. A schematic of a microscope system, using interchangeable modules, that is capable of both macro and micro sample analysis.

was asked to determine the carrier solvent, the propellant gas, the metals giving the paint its silver metallic color, and the composition of the nozzle and the plastic cap. The student dropped the plastic cap into the Raman instrument and quickly correctly identified the cap's composition as low-density polyethylene. He subsequently ran the spray button and quickly identified it as being made of nylon 6,6. In general, the students found Raman spectroscopy very effective for identifying a wide range of organic components with a minimum of sample preparation.

Nicolet Instruments expressed a wish to have my research group develop an optimized Raman microscope. We decided that it would be better and more flexible to build an instrument capable of both macro and micro sample analysis, so we developed a system using interchangeable modules (Figure 22). In one experiment using the microscope system, a student rapidly analyzed the components in a timed-release cold capsule. Each of the colored beads could be individually analyzed for content and concentration. In another case, a student was required to identify both the polymer and the active ingredient in an anti-flea dog collar. Under the microscope, the crystals of the active ingredient were seen to be embedded in the polymer. Spectra of the polymer portion were quickly identified as polyvinyl chloride. The active ingredient, Stirifos, was unfortunately not in the database, so the student had to interpret each of the observed Raman peaks as to their probable origin and use this in combination with other data to deduce the correct answer.

Raman spectroscopy is capable of providing high accu-

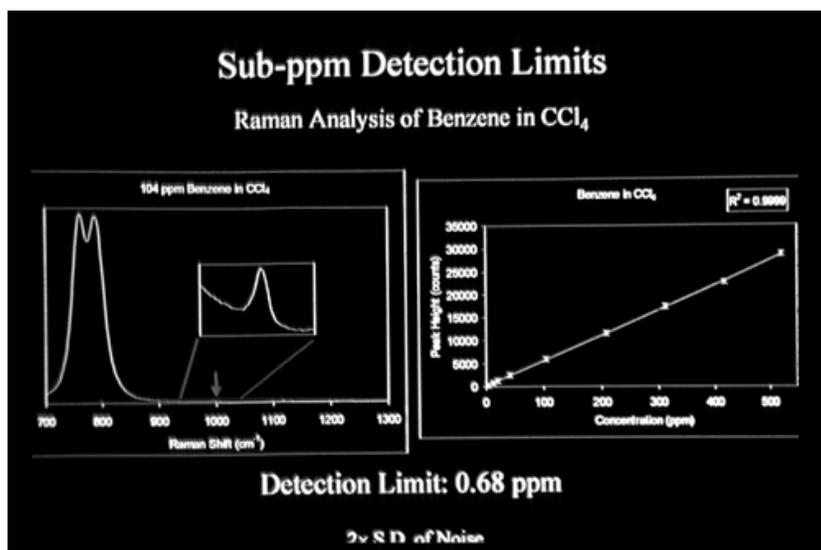


Figure 23. On the left, the Raman spectral region for the quantitation of a benzene in  $\text{CCl}_4$  at 0.68 ppm. The calibration plot on the right demonstrates the tremendous dynamic range of the CCD detector.

racy and sensitive quantitative analysis. The left side of Figure 23 depicts the spectral region for the quantitation of benzene in  $\text{CCl}_4$ , with a detection limit of 0.68 ppm. The right side of Figure 23 demonstrates the tremendous dynamic range of the CCD detector. When one observes the spectrum, one initially sees a huge peak for  $\text{CCl}_4$ , with an arrow pointing to the peak used for quantitating benzene. Magnifying the benzene wavelength, as shown in the box, demonstrates that while the benzene peak is small compared to the  $\text{CCl}_4$  peak, it is still well out of the noise level.

### The future of Raman spectroscopy in medical applications holds promise

Fast and accurate cancer detection is of great interest and the focus of major research efforts. Figure 24 shows spectra of two breast biopsy samples observed by Richard McCreery's group at Ohio State University. The upper spectrum is of normal tissue, while the lower spectrum is of tissue with infiltrating ductal carcinoma. These observations point to the possibility that Raman spectroscopy using a fiber optic probe might be less invasive than today's

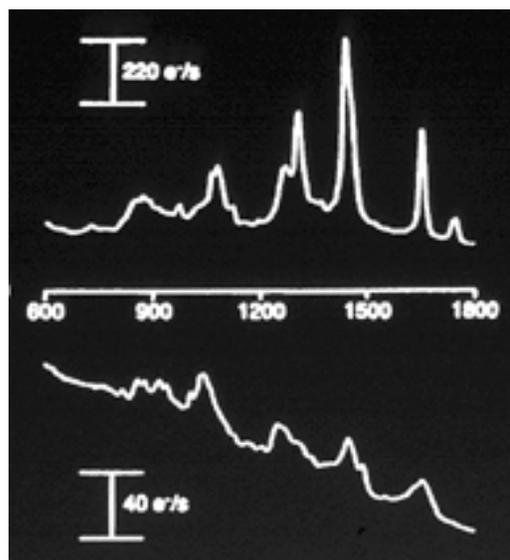
Figure 24. It is possible that Raman spectroscopy using a fiber optic probe might be as effective and less invasive for tissue analysis. The upper spectrum is of normal tissue, while the lower spectrum is of tissue with infiltrating ductal carcinoma. (C.J. Frank, R.L. McCreery, D.C.B. Redd, "Raman Spectroscopy of Normal and Diseased Human Breast Tissue," *Analytical Chemistry* 67, 1995: 777.)

approaches. If the technique can discriminate between normal tissue and a number of different types of malignant tissue, Raman could find itself a major player in cancer detection.

### New methods for high sample throughput analysis

Another area I want to mention briefly is the use of chromatographic separation methods coupled with parallel detection techniques to achieve high sample throughput. Methods capable of analyzing a large number of samples in a relatively short period of time are in increasing demand in many areas of science, nutrition, and environmental monitoring. We are currently investigating new approaches to an old technique: thin-layer chromatography (TLC). Many research scientists today regard TLC as an unsophisticated technique best suited for demonstrating chromatographic methods to high school students by separating different colored components from a single initial sample spot of food coloring or ink. However, TLC offers tremendous capabilities for much more sophisticated analysis. It provides the ability to analyze a number of samples simultaneously on the same chromatographic plate, that is, TLC is a parallel processing technique.

Using a precision sample application, I can place anywhere from 10 to 50 sample spots on a single plate and while that plate is developing I can be placing sample spots on many more plates (Figure 25). If I can subsequently in a very short time read simultaneously all of the components that have separated out during the chromatographic pro-



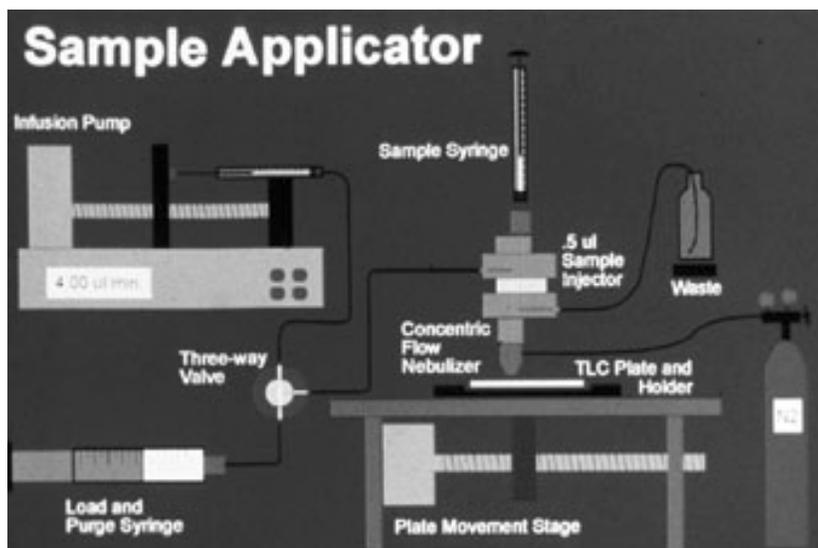


Figure 25. Using a precision sample application, from 10 to 50 sample spots can be placed on a single plate and while that plate is developing, sample spots can be placed on many other plates.

cess from each of the original sample spots, then I will have achieved a very high throughput analytical system. This readout is easily accomplished using one of our high dynamic range CCD cameras by merely taking a "picture" of the chromatographic plate.

Compared to standard TV cameras, which provide a grayscale resolution of 1 part in 256, scientifically operated CCD detectors are capable of grayscale resolution of 1 part in 64,000 to 1 part in 260,000. To give you an idea of how incremental this grayscale is, I can take one of our scientific CCD cameras outside of this building at high noon in Tucson, Arizona, point it at the sky, focus at infinity, and see stars. I can't see the stars with my human eyes—all I see is the blue-scattered light from our atmosphere. But the stars don't go away during the day; they are still there. The difference is that the scientific CCD technology, with its tremendous grayscale resolution, can actually resolve the slightly brighter spot that is the star from the blue-scattered background. The high dynamic range of the CCD system, coupled with its intrinsic low light detection capabilities, offers superb detection limits and a large linear dynamic range. When we carry this technology over to the observation of thin-layer chromatography plates, we are able to realize astounding sensitivities.

We became interested in applying this new approach for rapid sample analysis to the determination of aflatoxins in foods. Aflatoxins occur naturally, produced by fungi that grow on grains such as peanuts and corn. They are highly carcinogenic and are known to cause liver cancer. If contaminated grain is fed to chickens or cattle, and the aflatoxins are at a relatively low level (below toxic levels), the aflatoxins can be passed along in eggs, milk and other animal products and can move into the human food chain. Because humans obviously do not want to be exposed to

these carcinogens, we would like to have a rapid and sensitive screening test to ensure that these foodstuffs are safe for human consumption.

The most sensitive test results previously published utilized high-performance liquid chromatography mass spectrometry (HPLC-MS), which not only requires very expensive instrumentation but also a significant amount of time for the analysis of each sample. A system using high-performance thin-layer chromatography (HPTLC) using an autospotter and automated plate development (Figure 26) could

easily analyze 50 samples per minute, instead of the one sample every 15 to 30 minutes with HPLC mass spectrometer. One real-world test that demonstrated the sensitivity and applicability of the HPTLC method was the analysis of a popular brand of peanut butter, in which we were able to detect 3 to 5 picograms of each of the aflatoxins—just below the EPA level allowed for aflatoxins in a product for human consumption.

Unfortunately, HPTLC does not provide the chromatographic separation powers of HPLC. Sometimes the spots occupied by two different compounds will overlap. Different chromatographic conditions can be explored to try to achieve a full spot separation. Unfortunately, this is not always easy, or even possible, to accomplish. Recently, we

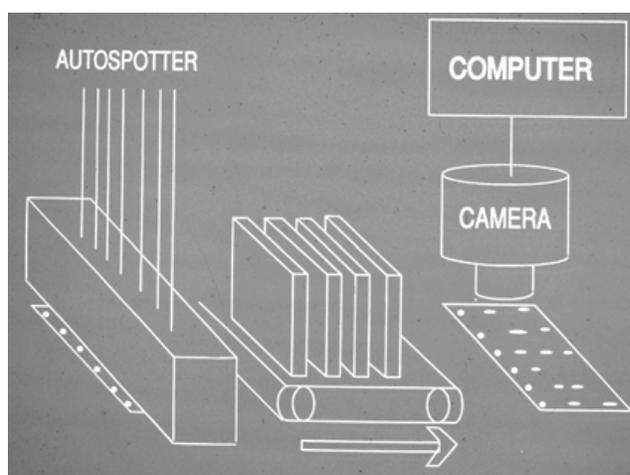


Figure 26. A high-performance thin-layer chromatography system using an autospotter and automated plate development. This system could easily analyze 50 samples per minute, instead of the one sample every 15 to 30 minutes with a HPLC-mass spectrometer.

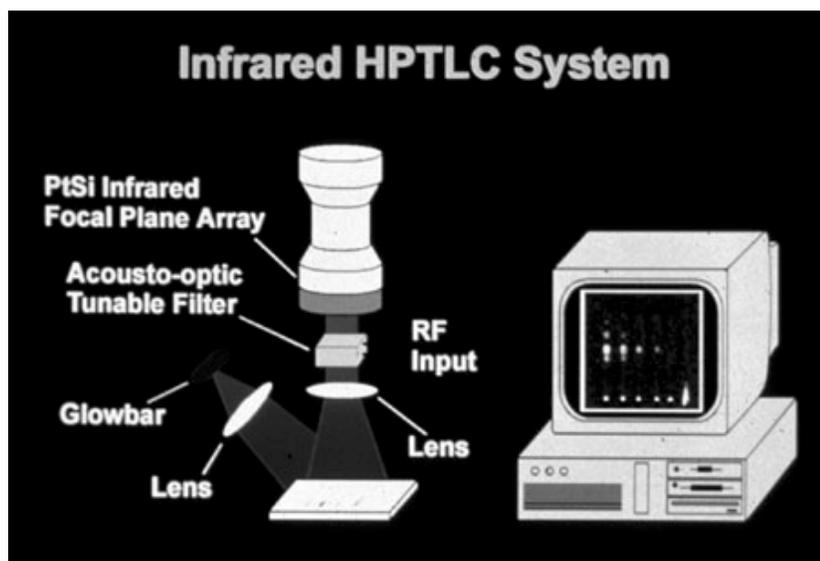


Figure 27. A thin-layer plate reader using a PtSi infrared focal plane array in conjunction with an acousto-optical tunable filter allows imaging of a plate at a computer-selected infrared wavelength characteristic of a functional group in one molecule which is absent in the overlapping species.

have been applying technology originally developed by the military for surveillance and night vision applications to enable observation of infrared and near-infrared functional group signatures. We have been developing a thin-layer plate reader using a PtSi infrared focal plane array in conjunction with an acousto-optical tunable filter to allow imaging of the plate at a computer-selected infrared wave-

length characteristic of a functional group in one molecule but absent in the overlapping species (Figure 27). For example, if one overlapped component has a carbonyl group and another one has an alcohol group, we should be able to differentiate between them by looking at the characteristic wavelengths associated with each of the groups.

Another technique widely used in biomedical research is combinatorial chemistry, where two linear arrays of chemicals are combined to create a two-dimensional array of reaction products. This technique commonly uses microtitre plates, which range in size from 24 wells to thousands of wells, with the 96-well plate currently the most common. Such approaches generate a tremendous number of individual mixtures requiring analysis. Today, the technology consists of reading out each individual well or, at best, reading out a row of wells. Why not read all the wells on the plate simultaneously? (Figure 28) And, as the trend continues to an even larger number of wells on a single plate, the simultaneous readout technology could be easily expanded. Recently we studied a variety of different samples in multiwell plates, using both fluorescence and UV absorption detection. Results indicate that the simultaneous approach provides high accuracy with low detection limits.

#### New array detectors for mass spectrometry

Last but not least, I want to mention another project that we are currently gearing up—a new mass spectrometer detector capable of observing all the ions over a large mass region simultaneously. We are adapting technology originally developed for infrared focal plane detectors used in night vision systems. Infrared focal plane detectors are often fabricated using exotic materials such as HgCdTe to convert the infrared photons into electric charge. The exotic materials are placed on top of a silicon readout multiplexer that amplifies the resulting charge and transfers it to the support

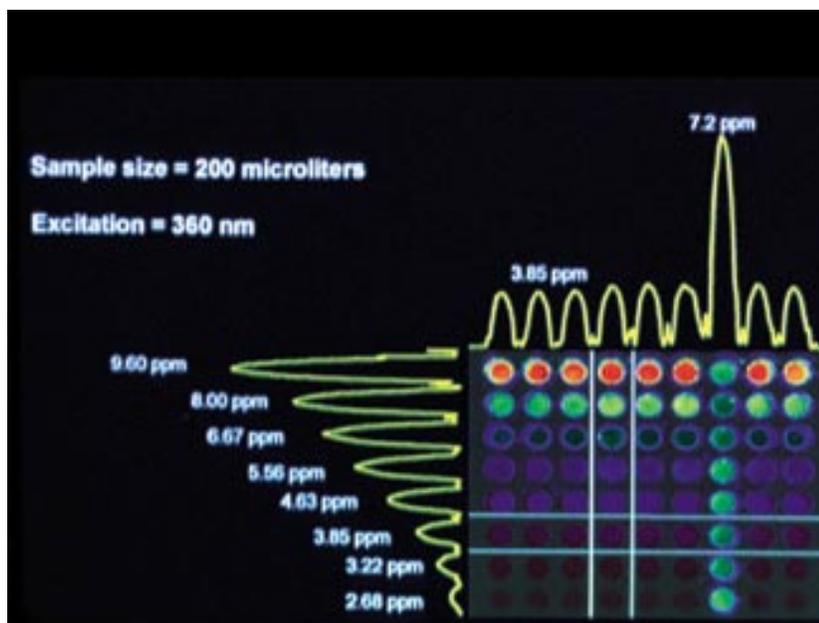


Figure 28. A variety of different samples in multiwell plates can be read simultaneously rather than individually, using either fluorescence or UV absorption detection. The results indicate that this approach provides high accuracy with low detection limits.

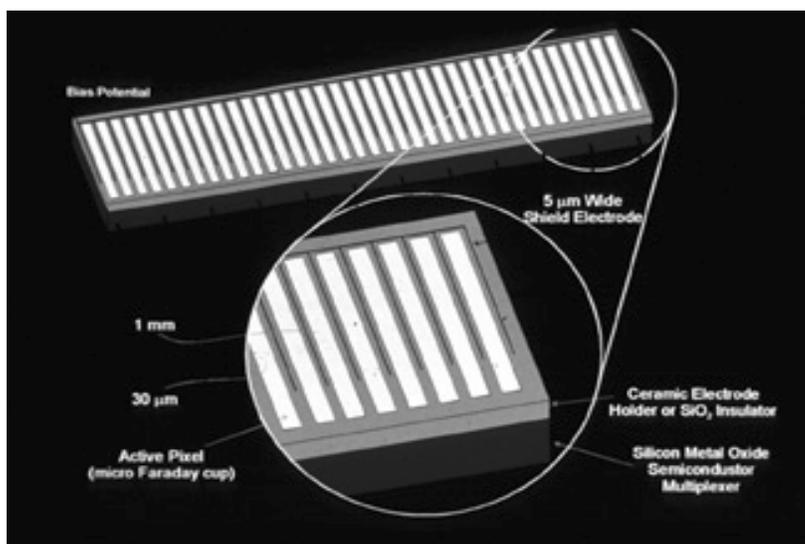


Figure 29. If a multiplexer were configured such that the amplifier array received charge from a micro electrode, and the electrode arrays were placed on the focal plane of a mass spectrometer, it should be possible to detect ions with less than 1 electron of read noise.

electronic system. These silicon multiplexers are capable of very high performance. However, by drawing on a technology developed by Jim Janesick while he was at the Jet Propulsion Laboratory, even higher performance should be possible. Jim's approach for reading out a charge-coupled device demonstrated subelectron read noise. If a multiplexer were fabricated using these approaches and, instead of receiving charge from the infrared material, were configured such that the amplifier array received charge from a micro electrode, as shown in Figure 29, and these electrode arrays were placed on the focal plane of a mass spectrometer, it should be possible to detect ions with less than one electron of read noise. We are in the process of applying this technology to implement an array mass spectrom-

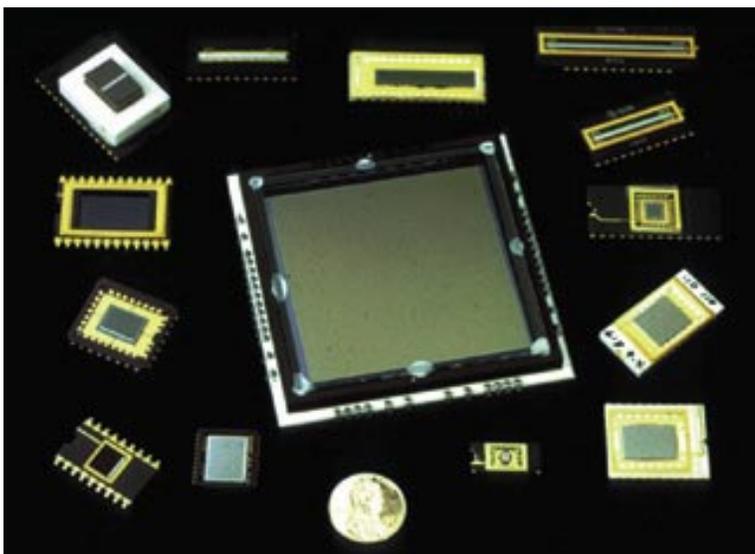
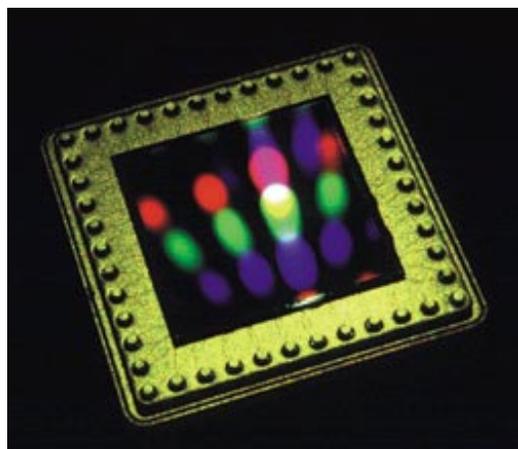
eter detector that will allow us to achieve many of the advances produced by focal plane array detectors in optical spectroscopy. Our operational phrase describing this new mass spec detector technology is the measurement of "all the ions all the time." We intend to be able to analyze a large mass region by simultaneously observing all masses occurring in that region while vastly increasing our sensitivity.

**Innovations require multidisciplinary solutions**

The array detector technology that I've discussed with you today has already revolutionized many areas of optical spectroscopy and promises to revolutionize others. It is also dramatically changing the speed of X-ray crystallography and holds great potential in the world of mass spectrometry. My research group has worked with a large variety of devices, many of which were custom fabricated.

They range from single-element devices to very large arrays (Figures 30, 31). At this time, I would like to reemphasize the concept of using technology from other disciplines to solve problems encountered in your area of science. Chemistry today is increasingly multidisciplinary, and innovative solutions often require optimal utilization of technologies from many other fields.

Figures 30, 31. New array detector technology uses a variety of devices, many of which are custom fabricated. They range from single-element devices to very large arrays.





## Proteins: Folding, Form and Function

Adapted from a 1994 presentation by Brian W. Matthews

Widely known for his X-ray crystallography studies of protein structure and function, Brian Matthews was born in Australia and took his undergraduate and graduate training at the University of Adelaide.

In 1963 he moved to the Medical Research Council Laboratory of Molecular Biology, Cambridge, England, and was a member of the group led by David Blow that determined the three-dimensional structure of  $\alpha$ -chymotrypsin.

After spending two years at the U.S. National Institutes of Health in the laboratory of David Davies, he moved in 1969 to the University of Oregon to establish a new protein crystallographic laboratory. At the University of Oregon he has served as director of the Institute of Molecular Biology and as chairman of the department of physics.

He is a member of the U.S. National Academy of Sciences and the American Academy of Arts and Sciences and is currently an investigator of the Howard Hughes Medical Institute and member of the Institute of Molecular Biology at the University of Oregon.

**I'd like to give you** you a feeling for the form and function of proteins. You come from a number of different disciplines, and I'm worried that the biologists among you are going to be bored and say, look, I know all this, why don't you tell us something new? On the other hand, I'm concerned that, even though it's morning, the physicists and others will go to sleep when I show the first slide. So I'm going to follow a middle course and hopefully give a little bit of something for everyone.

As you know, I come from Australia. What sort of credentials does one have to establish himself in the sophisticated United States? I usually quote the following so you can calibrate my background. "In America only the successful writer is important. In France all writers are important. In England no writer is important. In Australia you have to explain what a writer is."

Figure 1 is a Chinese postage stamp. What you see is a photograph of a wire model of a protein. The model is about three feet in each direction, but of course the actual protein itself is much, much smaller. The scale is about 800 million to one. The caption says, "1.8 Angstrom structure of insulin determined by the Peking insulin structure group." This structure was worked out during the Cultural Revolution. No one in the West was aware of this work, although there were parallel studies going on at Oxford. This figure will give you an impression of the complexity of the protein structures I'm going to talk about. They're made up of thousands of atoms, folded together in complicated ways, yet they're very, very tiny.

The technique of X-ray crystallography is a way to get information at the level of individual atoms. I don't have time to tell you about the technique, but I'll run through some of the results.

Figure 2 is a protein, called thermolysin, which was an early subject of the work we did in Eugene. We chose this substance because it is isolated from a bug that lives in hot springs and we were interested in how this biological material could survive in a hot environment. It turned out that the knowledge of this structure could be transferred to an understanding of the proteins involved in human blood pressure regulation.

In figure 3 you'll see a protein that does pretty much the same thing as the one in figure 2, but in a very different way. It has a completely different fold, but the activity is more or less the same.



Figure 1. Chinese postage stamp

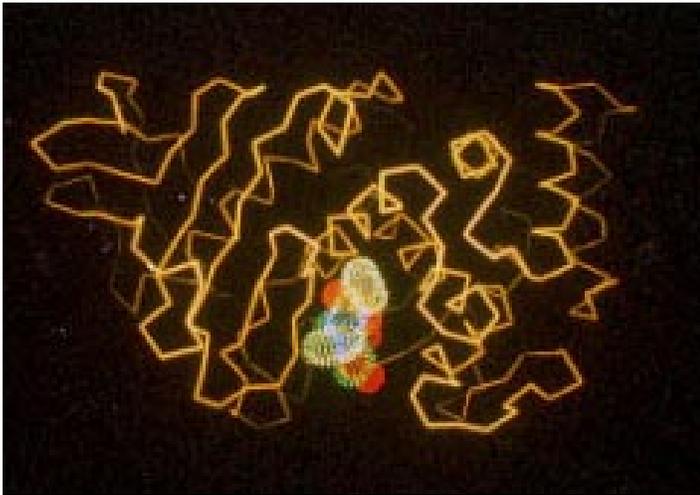


Figure 2. Thermolysin

Figure 3. A protein similar to thermolysin that folds in a different way.

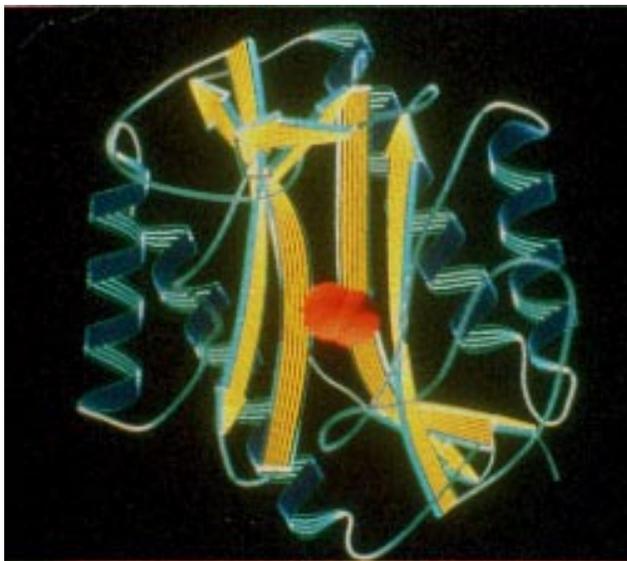


Figure 4 also early work from my lab, provides another example. This is a protein known to bind at a specific recognition site on DNA, shown here in a stylized way (if these models were to scale the DNA would be some miles long). Note that the DNA has grooves. When we first looked at the protein we wanted to infer how it might bind to DNA. We noticed that there were protrusions ( $\infty$ -helices) on the protein that would fit into the grooves on the DNA. We've subsequently worked out the structure of the complex and it is, indeed, very much like the model shown in figure 5.

I'd like to focus on how these proteins fold. They are made up by linking together a series of units called amino acids. You can think of them as beads on a string. The beads

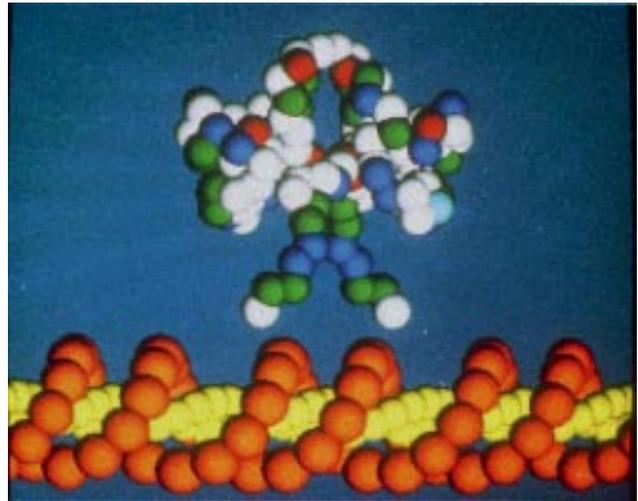
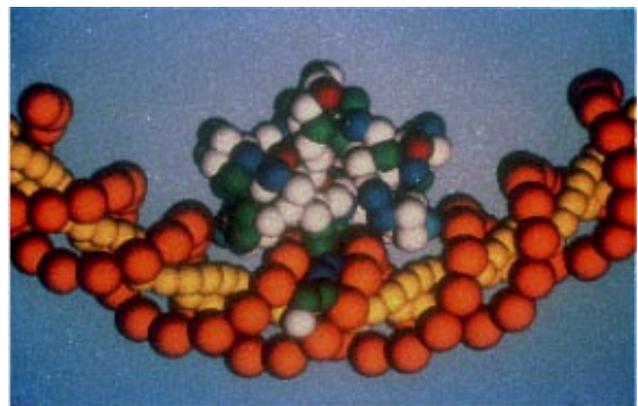


Figure 4. Protrusions on this protein (top) fit into the grooves on DNA.

Figure 5. Protein-DNA complex.



come in 20 different colors. It's the sequence and thenumber of beads that de-termines the folded structure. Similarly, a protein is characterized by the number and sequence of amino acids. There are three key points to remember:

1. The amino acid sequence determines the folded 3-D structure.
2. The protein folds spontaneously to give the 3-D structure.
3. The active folded 3-D structure is the one with minimum free energy.

The protein I'm going to focus on is phage lysozyme (Figure 6). There are 164 amino acids in this protein and we would like to know how these 164 amino acids determine the shape. We don't know the answer, but we have made progress.



Figure 6. Phage lysozyme

**Homegrown mutants**

Until about ten years ago, it was possible to look only at proteins that exist in nature. You *could* make mutants, but they were random and you could not control where they would be. More recently, thanks to what's called directed mutagenesis, you can, in the test tube, change one or more of the amino acids to something else.

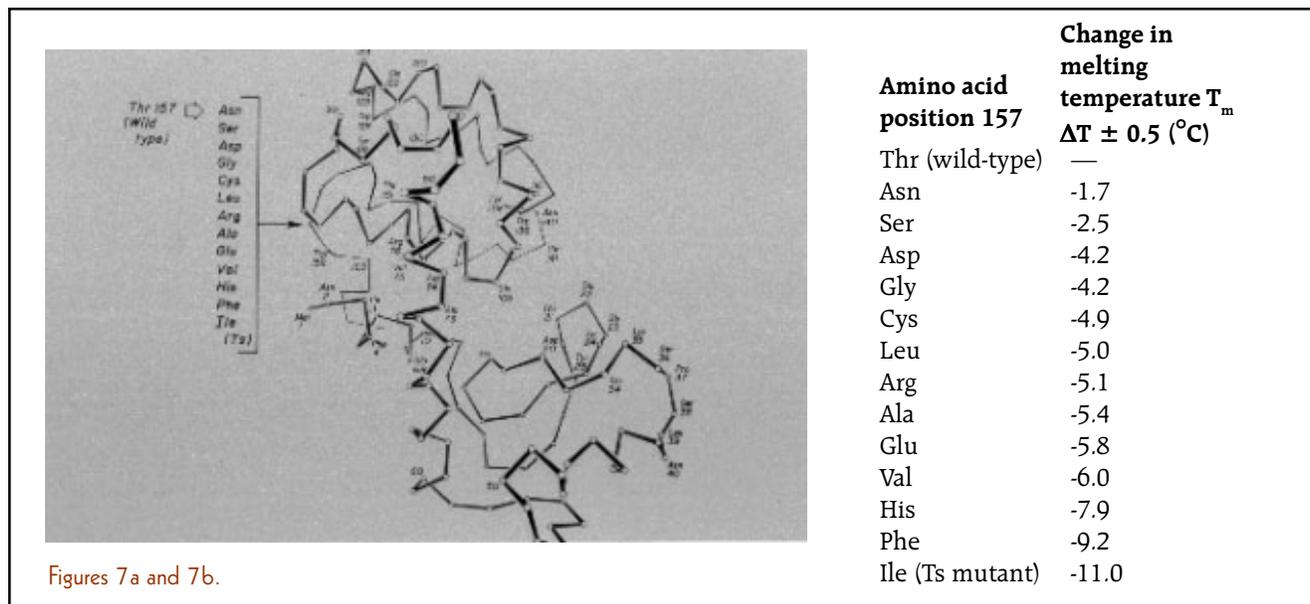
Suppose we look at the lysozyme structure in figure 6 and focus on amino acid number 157. We can determine whether position 157 is important by changing it to something else and, better than that, we can actually change it to a whole series of different amino acids. This is shown in figure 7a. Figure 7b is a list of the relative stabilities (at

pH 2) of lysozymes with different substitutions at position 157.

You can take lysozyme and heat it, and at a certain temperature it unravels. If you cool it down again, it very quickly goes back to native form. The temperature at which it unravels is a measure of its stability. In this particular case, we never found a variant that made lysozyme more stable than the native enzyme. Most of the time, the protein didn't much care what amino acid was put at position 157; *the protein was tolerant of change*.

Now let's think about the question of how many of the 164 amino acids are important for folding? How many are really necessary and how many participate? We have found that the amino acids on the surface of the protein can often be replaced with little effect on stability or folding. On the other hand, the protein is affected more by substitutions on the inside. So we haven't solved the whole problem, but we've established that while the 164 amino acids determine the folded structure, about half of those amino acids are on the outside and don't seem to be directly involved in the folding process. It's the inside parts that are the key, and that we need to focus on.

I want to talk about the inside of the protein and the sort of experiment one can do there. A vital question is what happens if we take a bulky buried residue in the middle of the protein and replace it with something small? The protein could respond in different ways. One way would be to repack to avoid a cavity. Another scenario would be that the protein structure in fact remains the same, in which case there is a cavity. If there is a cavity, is it truly empty?



Figures 7a and 7b.

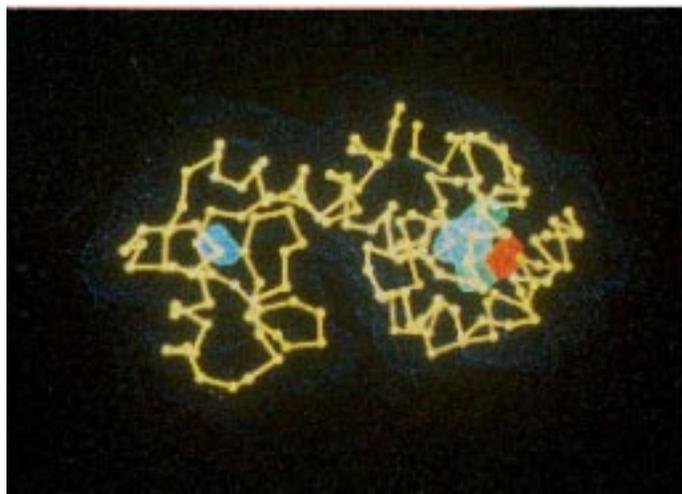


Figure 8. Protein with cavities.

### Cavity-creating mutants

We've done these experiments and the answer is that every time you do them you get a different result. Sometimes you *do* get a cavity, and sometimes the protein re-packs, and sometimes it's half way in between. In "cavity-creating mutants" above, two extreme examples are illustrated.

These examples illustrate that if you change the residue at position number 99, the protein structure hardly changes at all. You get a cavity big enough to accommodate five water molecules. But as best we can tell, it's completely empty. It's very destabilizing. The protein does not like this at all and its melting temperature goes down by about 16 degrees. This is one of the most destabilizing mutants we have. The second replacement is chemically the same, a leucine to an alanine. Here the protein structure adjusts and the cavity is much smaller.

We're using these experiments to work out the energetics of the way in which different amino acids in the inside of the protein contribute to stability. We have proteins with different cavities in the inside. Figure 8 is a composite showing different examples. The cavities appear empty, but you could ask is it possible to get something to go into these cavities? This is a question which has to do with the dynamics and the flexibility of proteins. The cavity is surrounded by solid protein and it's about 7 Angstroms from the edge of the cavity out to the surface. There's no connection from the cavity to the outside of the protein so if anything is going to get in and out of this cavity, there is no way it can do that with the protein structure remaining rigid.

When she was in my lab, Liz Eriksson decided to see if she could get a substance to go into the cavity. She chose benzene because it is a very well-known chemical and it's

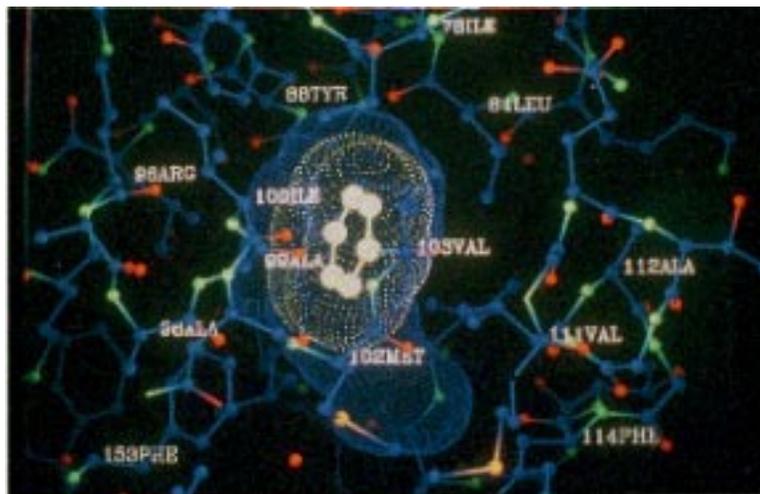


Figure 9. Benzene inserted into a protein cavity.

the sort of molecule that chemically you would expect might like to be in the cavity rather than out in the solvent. Actually, it turned out to be very easy to get benzene into the middle of the protein, no trouble at all. It can be seen in figure 9.

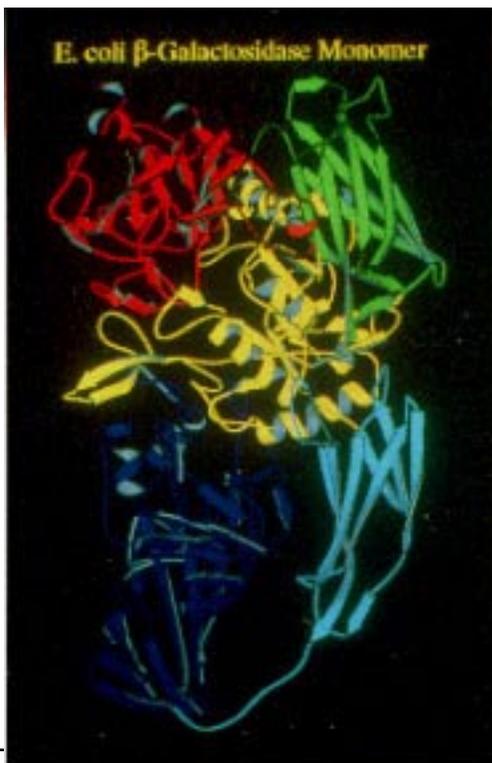
This tells us that these protein molecules are not fixed, rigid and stationary. They're clearly very flexible, they're dynamic. In some way or another, they can open up to allow the benzene to penetrate to the inside.

So far I've talked about a simple little protein called T4-lysozyme. I now want to tell you about something that's much bigger. This work is brand-new, not yet published.

Shown in figure 10 are crystals of a protein called  $\beta$ -galactosidase, probably the most widely used protein in molecular biology and biotechnology. It has a very long history and was key in discovering how genes are turned on and off. The protein has a number of very desirable features, and it has substrates that dramatically change color, making it a very sensitive assay in various contexts.

The problem with this protein is that it's big, and part of the reason I want to tell you about it is to contrast small proteins and big proteins. On the bottom of figure 11 at left is T4 lysozyme and on top is  $\beta$ -galactosidase, drawn to scale. Lysozyme T4 has 164 amino acids from one end to the other;  $\beta$ -galactosidase has 1,023. Furthermore,  $\beta$ -galactosidase is not just a monomer; the active protein is a tetramer made up of four subunits, each with 1,023 amino acids.

A very talented graduate student, Ray Jacobson, attempted as a thesis project to determine the structure of this protein. I'm going to tell you about the result to give you a feeling of the relationship between small and large proteins. The question is whether knowledge from small proteins will be helpful in thinking about larger proteins.



main (figure 14) looks like a part of an immunoglobulin, the proteins that we use to defend against various diseases. This shape has also been seen before.

Now we go to the third domain (figure 15) which is called an alpha beta barrel. It looks complicated in the illustration, but it has strands in the middle and helices on the outside arranged in a cyclic pattern. There are many proteins that have this shape and, when they do, the active site is always at the end of the barrel. That's exactly the case here except there is a difference, there's a *twist*. In our case there's a piece from another domain that comes across and joins to make the whole of the active site. So you can't just have activity with this domain by itself. The need for this extra piece explains a number of the funny properties of this protein that make it very interesting and useful in biotechnology.

Going to the fourth domain (figure 16), this, again, is like an immunoglobulin domain.

The fifth domain (figure 17) is a complicated, twisted shape that had not been seen before so there was something that was completely new and different in this structure.

**Data from different systems sheds light on proteins**

These figures suggest that understanding large proteins may not be so formidable as you might at first think. Large proteins tend to be modular. They tend to be made up of smaller domains that often *have* been encountered before. Information from one source—from working on a bacteria, for example—can be transferred to another field. Often knowledge developed in one system, quite unexpectedly in many cases, can be transferred and can be useful in other applications.

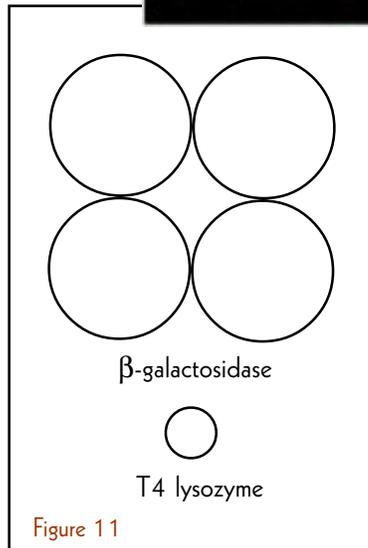


Figure 11

Figure 10.  
beta-galactosidase

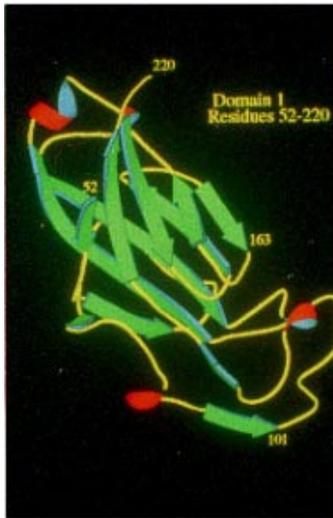
As I've already told you, this protein is very large. To this point, there is no protein longer than 1,023 amino acids for which the structure is known. Nevertheless, Ray did manage to work out the structure. It is shown in figure 12. This figure shows the shape of one monomer, drawn according to the spectrum so the first part is in red and then it goes

to the yellow and then to the green and to the light blue and the dark blue. The first point is that the protein folds up into five units (see figures 13–17, page 121) or domains, 1-2-3-4-5.

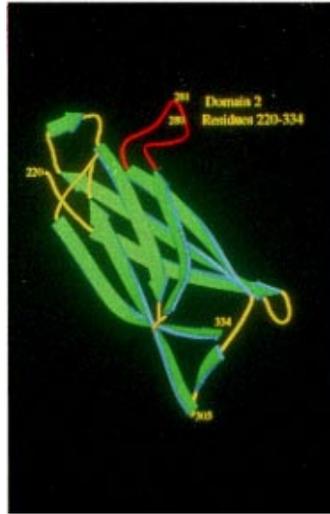
I'll quickly run through these domains just to see how they look. The first domain (figure 13) has these "strands" that go back and forth. We've seen this shape before, it's called a "jelly roll." It's a protein fold that occurs in many many proteins that are parts of viruses. The second do-



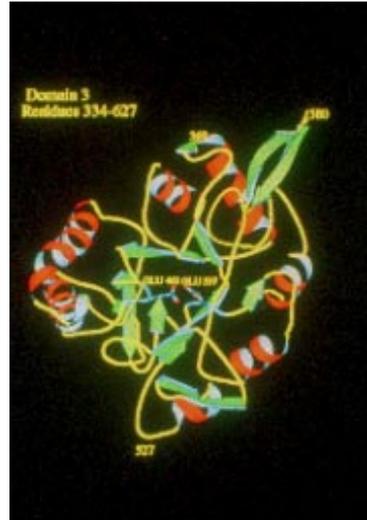
Figure 12.



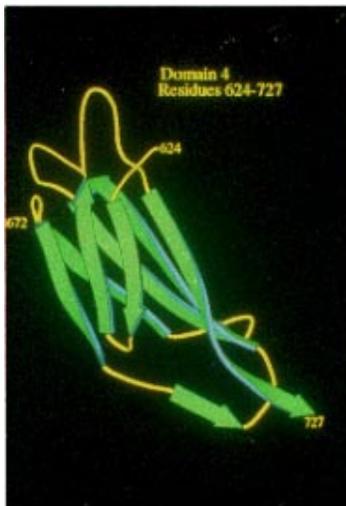
13



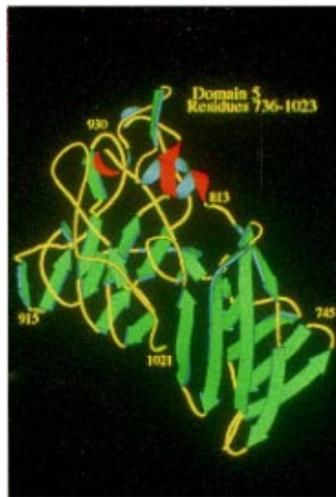
14



15



16



17

In figures 13 to 17,  $\beta$ -galactosidase folds up into five units or domains.

I wanted to give you a final thought, realizing that you *are* teachers. I have a quote to share that I think you will identify with.

A young man was interviewed by the *Register Guard* which is the Eugene, Oregon paper. They were discussing how this young man had decided to come and play on the University of Oregon football team. He said he finally chose Oregon over Montana and admitted he had few scholarship offers from larger schools because of his grades. "I've had trouble in school," he admitted, "but it's not because I can't learn, it's because I didn't try. I'm making up for that now. Since I *started going to class*, everything has gotten better."



## The Discovery of RNA Catalysis

Adapted from a 1994 presentation by Thomas R. Cech

*"Tom Cech is one of America's most distinguished scientists," commented Research Corporation President John P. Schaefer in his introduction to Dr. Cech's presentation. His breakthrough discovery of catalytic RNA established a link between RNA and the evolutionary chain and brought him the 1989 Nobel Prize in chemistry at the age of 41.*

*Thomas R. Cech is president of the Howard Hughes Medical Institute where he has been an investigator since 1988. He is also American Cancer Society Professor at the University of Colorado at Boulder and professor of biochemistry, biophysics and genetics at the University of Colorado Health Sciences Center.*

*In addition to the Nobel Prize, Dr. Cech's honors have included the Pfizer Award in Enzyme Chemistry, the Heineken Prize, the Lasker Award and the National Medal of Science.*

**I'd like to talk about** a discovery, one in which I was fortunate enough to be involved. For a number of years I felt my seminars should be restricted to what we'd done in the last six months, the unpublished, hot new stuff from the laboratory. People might think we were resting on our laurels if I mentioned anything older. Then I realized that the discovery story was fun to tell and, when I didn't, people would often raise their hands after I spoke about recent work and ask, how did you ever get started on this in the first place?

There's another reason for talking about the discovery of RNA catalysis. When you read that a scientific truth has been established, you get the impression someone had an idea and did a series of linear experiments one after another to confirm it. And then it was printed in the textbook.

I suppose there may have been things that were discovered in such a manner, but usually, if it were that easy, somebody else would have done it long ago. So I think our story is really much more typical: people were trying to do "A" when they tripped over "B." As Louis Pasteur put it, "chance favors only the prepared mind." There's some luck involved, but you have to have the training to know if it is worth looking closer when you stub your toe on something. I don't know how to teach people to make those decisions; I don't know how often *any* of us makes those decisions properly. But somehow, after students come into our labs and leave a few years later, they have picked up some intuition about which fork in the road to follow.

### A short course in molecular biology

As you know, information is stored in all living cells in the form of deoxyribonucleic acid, the famous double helix of DNA. By information we mean coding which tells the cell how to make a particular protein. A protein could be an enzyme that facilitates a particular chemical reaction inside of a cell, or it could be a DNA-binding protein involved in gene regulation, or it could be hemoglobin, the oxygen-carrying protein. For each protein the cell needs, it has a piece of a chromosome, a DNA molecule that encodes it.

How does the transfer of this information take place? The DNA is copied into the chemically similar molecule RNA (ribonucleic acid), the main difference being the presence of one extra oxygen atom on each of the repeating units. The individual monomer units on this polymer (the four-letter monomer code for RNA is A, G, C and U) are copied in order from the A, G, C and T on the DNA (T and U are chemically similar units that have the same coding capacity). The order of these monomer units, or nucleotides, along the RNA determines the order of amino acids that are laid down in the protein.

The code is read during the process of protein synthesis, three of these building blocks at a time, so it's every triplet of nucleotides along the RNA that specifies a particular amino acid. For example, three U's in a row is phenylalanine and A-U-G is methionine. The order of amino acids determines not only the sequence of the protein, but also how it folds up. This in turn determines what function the protein has in the cell. The

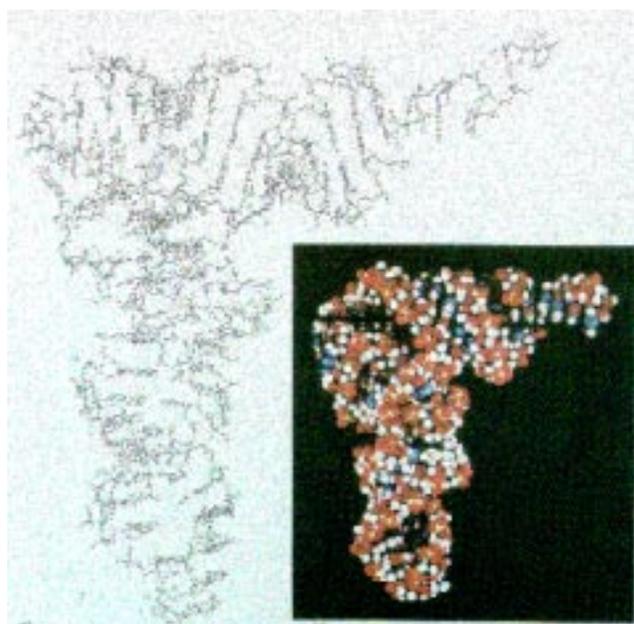


Figure 1. Transfer RNA

function can be as different as being a digestive enzyme (pepsin) which helps break down the breakfast you just ingested, or a protein like myosin which is responsible for muscle movement. Each of those, of course, has a different amino acid sequence specified by a different gene.

A simple, everyday analogy is that the DNA is like an archival copy of your favorite video. From this master you can make many informationally identical copies by duplicating the videotape. Those copies are analogous to RNA. But just having a lot of information is of no value without a machine to plug it into. A VCR is analogous to the cell's protein synthetic machinery, the ribosome; you plug in the information and out comes the final "image" or polypeptide, the purpose for storing all this information in the first place.

All of this has been known since the 1960s and is often called the *central dogma of molecular biology*. However, the view of these three molecules, DNA, RNA and protein—at least in certain parts of Boulder, Colorado—is now a little different. We put RNA in big letters in the middle, noting that of these three essential macromolecules, RNA shares with DNA the ability to be an information carrier. (In fact, there are some viruses that don't even have DNA; their genomes are in the form of RNA to start with.) But RNA also shares with protein the ability to catalyze biochemical reactions. The fact that RNA embodies both information transfer and biocatalysis has led to new ideas about the origin of life, which I'll talk about later.

Why did it come as such a surprise that RNA, a nucleic acid, could be a biocatalyst? When you think of biocatalysts

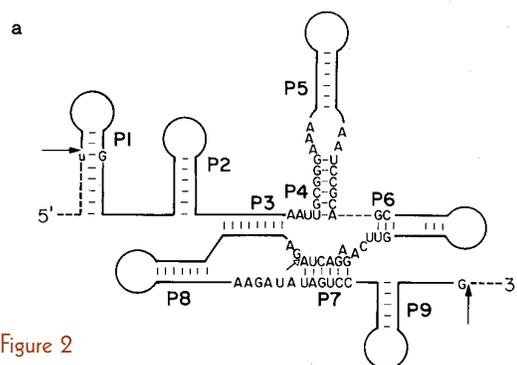


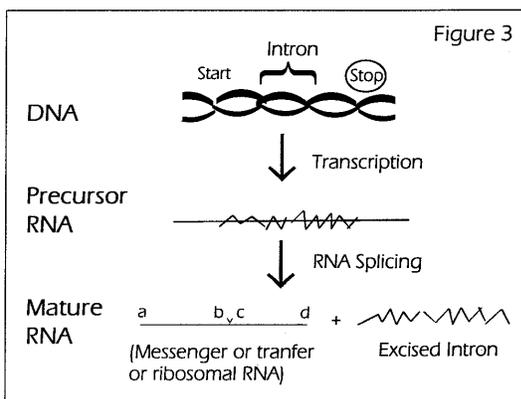
Figure 2

in living cells, you normally think of proteins—globular macromolecules with an active site cleft with a shape and a chemical structure complementary to those of the substrate. The substrate, perhaps a sugar molecule, is bound in that cleft. Everything is then set up just right so that the amino acids of the protein encourage a particular reaction. In technical terms, they lower the activation energy by stabilizing the transition state for this particular occurrence.

When you turn to nucleic acids, you envision a very different structure. The nucleic acid with which people are most familiar, appearing on the cover of *Newsweek* or *Time* magazine, is, of course, the double helix of DNA, a molecule that doesn't have a globular structure with a specific active site cleft. Nor does it have a variety of chemical functional grooves poised in such a way that you would predict it to chemically interact with the transition state and encourage a reaction to take place. (DNA does have grooves, but their structure is largely repetitive along the helix.) DNA has a rather boringly repetitive structure, well suited to its main job of storing information. After all, you don't expect the tape on which your videos are recorded to have a lot of character.

When you consider DNA's sister nucleic acid molecule, ribonucleic acid or RNA, you see a single chain of nucleic acid has folded back on itself (figure 1). Even in one of nature's simplest RNA molecules, called transfer RNA, there are hairpin stem-loops, a complex tertiary fold, a single-stranded extension at one end, a dip in the molecule at another point, and a small hole. I think you can appreciate that there's more richness of structural texture with this molecule than with DNA, that there's an ability to fold into something that looks a bit more like a globular protein.

The transfer RNA molecule in figure 1 is involved in protein synthesis, so it contributes to a catalytic process. But it, by itself, has not been shown to catalyze any biologically relevant reaction. The RNA molecule shown in figure 2, on the other hand, represents one of a growing list of RNAs that act as biocatalysts. You'll notice immediately



that the picture isn't nearly as pretty. We don't have a gorgeous three-dimensional representation, but instead a two-dimensional road map to the base-pairing arrangement. This represents the first order of folding of this catalytic RNA molecule. One of the current goals of research in our laboratory is to obtain an accurate picture of such a molecule at atomic resolution, which would allow us to understand the folding and the disposition of atoms in three-dimensional states that allow a catalytic RNA to actively participate in cellular biochemistry.

### Our story

How did we come to the realization that RNA could have such activity? As I already hinted, it involved a large dose of serendipity. We were studying *Tetrahymena thermophila*, a ciliated protozoan related to *Paramecium*, which I'm sure all of you have looked at through the microscope. (*Tetrahymena* is a huge cell; *Paramecium* is about twice as big and therefore you don't need a very expensive microscope to watch these organisms beating their cilia and moving gracefully through the water.) This organism is found in freshwater ponds throughout the world.

Why did we choose *Tetrahymena* as an experimental organism? Like a human cell, it has a nucleus which contains the genetic information separate from the cytoplasm where protein synthesis takes place. It's a eukaryote; it belongs to the same biological kingdom as higher plants and animals, but is as easy to grow and to work with as bacteria. In fact, you can harvest a million cells to do an experiment, and no animal rights activists will pound on the door.

Within the large, so-called macronucleus of *Tetrahymena*, there are 10,000 identical copies of a particular gene, all doing the same thing at the same time. That provided a nice opportunity to look at a specific gene, a specific piece of DNA, and see how it is expressed at a particular time. We wanted to study how an RNA copy is made from DNA,

the process called *transcription*. We wanted to identify the signals that cause the copying enzyme to initiate at the so-called promoter, continue making RNA, and then terminate at a termination site (figure 3).

In 1979 in the process of studying DNA-to-RNA synthesis, we stumbled upon the fact that the gene we were looking at, like many other genes in eukaryotes, was interrupted. Its normal flow of information was interrupted by a stretch of noncoding DNA, called an intervening sequence or intron. This phenomenon of split genes had been discovered in higher organisms two years earlier by Phil Sharp at MIT and a group led by Rich Roberts at Cold Spring Harbor. You may recall that a 1993 Nobel Prize was given to Sharp and Roberts for their discovery: in higher organisms, genetic information is not always continuous, but often interrupted.

By the time we found this new example two years later, it was, perhaps, the 100th split gene found in a eukaryotic organism. So this alone was not cause for celebration. It was already known that the copying process did not distinguish between coding and intervening sequences. The whole thing is copied into a precursor RNA, and the interrupting information is removed in a subsequent step called RNA splicing, so named for the way sailors used to repair a bad piece of rope (see figure 3). The rope was cut on either side of the frayed region (points b and c), point b reattached to point c to make a good piece of rope, and the middle discarded.

Although we had set up a system to study transcription, the system seemed to want to tell us about RNA splicing. When we tried to look at transcription, we observed splicing in the same test tube. Well, it was a little hard to remain detached, and our curiosity started getting the better of us. At this point there was only one other lab in the world that had seen RNA splicing take place outside of the cell, a reaction *in vitro* (in glass, despite the fact that we now use plastic test tubes). Getting a reaction to work *in vitro* is the key to being able to dissect, at a chemical level, the requirements for a reaction, its components and the mechanism by which it works. With only one other lab that had seen splicing taking place, we thought we might make more of a contribution by studying it rather than the transcription of RNA. There were over a hundred laboratories trying to understand the latter.

### World's strangest enzyme?

If you're a biochemist, how do you go about trying to clarify splicing? We wrote a grant proposal saying we were going to purify the enzyme that's responsible for converting the precursor RNA to spliced products. We took it for granted that the enzyme would, of course, be a polypeptide, because you can read in any biology book that any

reaction which takes place with such exquisite specificity and efficiency must be catalyzed by an enzyme, and we all know that enzymes are proteins. How much specificity? The RNA chain contains 7,000 nucleotides. Of those 7,000 it's only one position, let's say 6,256, which is always joined to position 6,669—two specific points along this large RNA chain which are chosen for the breakage and rejoining reaction. Getting back to the grant, we told the National Institutes of Health that we were going to purify this protein, and they said, wonderful, do it. Of course, our curiosity was such that we would have done it regardless of what they said.

First we isolated the precursor RNA and incubated it with an extract made from *Tetrahymena* nuclei, because we knew the nucleus must contain the protein enzyme that would catalyze the reaction. The first time we did that experiment, the precursor was converted to the spliced products. It's unusual in science for anything to work the first time, so we were very gratified.

But then, next-door to that sample was our control experiment in which the precursor was incubated with the small molecules, salts and nucleotides that are present in any cell, but not the one critical component, the extract from the *Tetrahymena* nuclei. Much to our surprise, the second test tube showed just as much splicing as the first one! Well, that wasn't one of the possible outcomes of this reaction. My initial response was, "You must have mixed up the tubes. Just repeat that control experiment; if we can get it right the next time, we'll be okay."

However, the unexpected result proved to be reproducible. I had to come up with some explanation. As you'll see, my first attempt was wrong, but a hypothesis has value in science if it flows naturally from what you know and you can test it. If you can get a clean yes or no answer, you can proceed. My first hypothesis was that what we were calling pure RNA had the splicing enzyme already associated with it. (Of course, the unknown activity *had* to be a protein.) When we incubated our RNA in the presence of the small molecules found in all cells, it finished off a reaction that it had started in the living cell. But if we could completely deproteinize the RNA, the splicing activity would be inactivated, and we could then find the cellular components that would reconstitute the active ensemble of macromolecules.

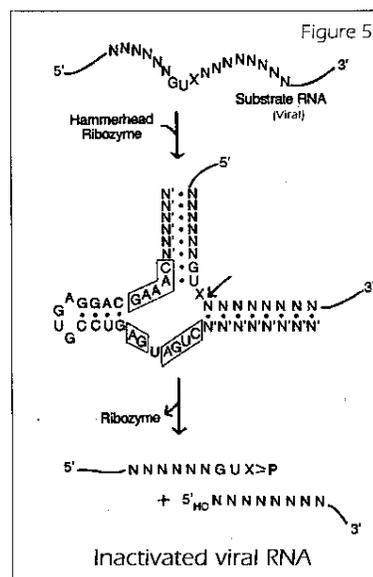
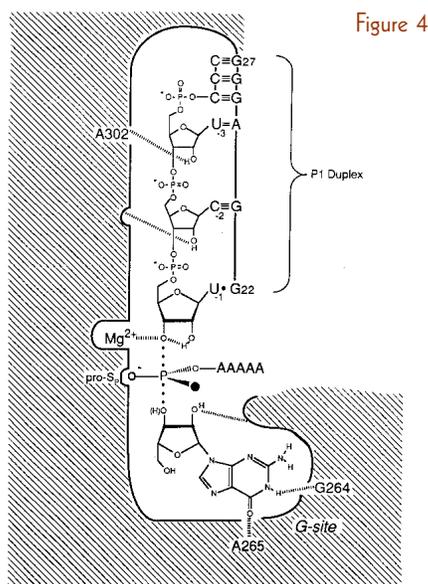
There are well-known ways to destroy protein. Most of them don't like to be boiled, for example. But RNA doesn't mind being boiled, so we boiled our RNA to eliminate any associated protein, and then cooled it and added the small molecules. The splicing continued. Well, now we had to revise the hypothesis: perhaps this was an interesting, very thermoresistant protein that was causing the RNA reaction.

Next we turned to detergents. When you wash your clothes you use high temperature and detergent, conditions that denature protein chains quite effectively. Lo and behold, our splicing reaction was insensitive to detergent as well as boiling. "Well," I said, "we'll get it next time. We'll boil it in detergent." Proteins really dislike that sort of treatment, but when we lowered the temperature and added the small molecules, the splicing once again continued. Next we purchased large quantities of nonspecific proteases, enzymes that are, in fact, used in enzyme-active laundry detergents. These degrade polypeptides down to amino acids. "Protein can't possibly be resistant to protease in the presence of detergent *and* boiling," we said. But, as you might guess by this time, none of these treatments prevented splicing.

This brings us to the Christmas of 1981 when one of my graduate students gave me a daisy with petals labeled, "It's a protein, it's not a protein." Clearly, the hypothesis was in need of revision. But we couldn't make the novel announcement—"RNA, by itself, can fold up and form an active site for a very specific, efficient reaction"—based solely on negative evidence, on the fact that we could not stop the activity with protein inhibitors. We needed some kind of positive evidence before I was going to stick my neck out and make such a pronouncement.

This evidence came from genetic engineering. The idea was to make RNA in as artificial a way as possible, RNA that had never seen a *Tetrahymena* cell. If we could demonstrate that RNA made by a completely independent pathway still had splicing activity, then it would be time to believe that the activity was intrinsic to the RNA. Back in 1981 it was much less common to make bacterial plasmids to do cloning than it is now, so it took us a while to learn the technology. The *Tetrahymena* DNA that included the intron was put into a bacterial plasmid, thereby allowing it to be grown in *E. coli*, and then purified. Carol Cech's lab was working on *E. coli* RNA polymerase, and they provided a purified sample of the one enzyme we would need to convert the DNA into a complementary copy of RNA. We made transcripts that contained the entire intervening sequence and some of the flanking coding sequences. We removed that one enzyme we had added, the RNA polymerase. When we then added the small molecules that are present in all living cells, the RNA once again underwent splicing.

Importantly, the sites along this synthetic transcript where the cutting and rejoining reactions took place were identical to the splicing sites that were utilized in the living cell. This gave us confidence that this very artificial reaction had to be relevant to biology.



### Implications of RNA catalysis

Where has this work taken us? Most of our current efforts are occupied with trying to understand, at the level of individual atoms, what constitutes an active site of an RNA enzyme. What allows reactions to occur with such high efficiency and specificity. We have mapped numerous interactions. Last year we located a specific metal ion in the active site and were able to determine its contribution to making this reaction proceed with such high velocity (figure 4).

By the way, the amount of rate acceleration that RNA provides to the reaction is 10 to the 11th power. That's the difference between the reaction taking place in a minute in this catalytic center, as compared to 16,000 years if it were uncatalyzed. This is a huge amount of rate acceleration and puts this RNA-catalyzed reaction in the mid-range of what protein enzymes can do.

Another sort of information that's coming out of this work is much more biological: the recent understanding that some RNA mechanisms are fundamental to all living cells. The ribosome is the protein synthetic machine that's the equivalent of the cell's VCR unit: you plug in the RNA and get the protein as the readout. The ribosome is an unusual macromolecular catalyst in that it has, by weight, about as much RNA as it has protein. Why should there be RNA in this enzymatic machine? Now it's thought that the RNA of the ribosome is not just a *framework* to hold everything together, but another RNA catalyst. Perhaps, at its heart, it's really a "ribozyme," the word we coined to refer to an enzyme composed of ribonucleic acid.

Ongoing research in other laboratories, including that of Harry Noller at the University of California at Santa Cruz,

is aimed at answering this question: is the ribosome, responsible for protein synthesis in all living organisms, also an example of an RNA catalyst?

Campbell's Primordial Soup hasn't been seen on your grocer's shelf; you just heat it to very high temperature and stir it for millions of years. There used to be a conceptual chasm to jump across: it's hard to envision life without some kind of an informational molecule that's copied to give rise to the next generation. It's also hard to envision an informational molecule reproducing itself without a catalyst.

Informational molecules, as far as we know today, mean nucleic acid. It's a rather sophisticated catalytic operation to copy the information from one nucleic acid molecule into another. And if biological catalysis means protein, then how are we to envision the origin of life? Did random chemical processes have to bring just the right polypeptide and the right nucleic acid together in the same droplet of water at the same time? How likely would that be?

Now we know that one of these molecules, RNA, can do both jobs: it alone can both carry information and engage in catalysis. Perhaps at the beginning there was RNA which copied itself, and the proteins came later. The chemical plausibility of that scenario is becoming more and more certain. RNA has been found to be able to catalyze the stitching together of nucleotides into a polymer of RNA, exactly the sort of reaction that would be needed for RNA to copy itself. Even if we prove this is chemically feasible, is it what occurred four billion years ago? This question will provide good panel discussions for years to come.

### Agents that attack viral RNA?

As often happens in cases where scientists are simply following their curiosity and not trying to do anything practical, something pops up that may have useful spin-offs. A few years after discovering self-splicing, we were able to reengineer the *Tetrahymena* intron RNA by removing its internal reaction sites. This made it a stable entity which, when it encountered another RNA molecule, would form a complex and would cut and splice the different RNA molecule—a real enzyme, acting on other RNA molecules rather than working on part of itself. This concept has been applied by Olke Uhlenbeck (University of Colorado) and by Jim Haseloff and Wayne Gerlach (CSIRO, Australia) to another RNA catalyst, one which is extremely small. This is called the hammerhead ribozyme, because its structure resembles the shape of a carpenter's hammer (figure 5).

When the hammerhead ribozyme's structure forms, either in a test tube or in a cell, it binds a metal ion (magnesium ions are ubiquitous in life) in such a way that a particular bond is severed. This destroys the covalent continuity of the chain at the point shown with an arrow in figure 5.

Why is that useful? Imagine that this RNA, the so-called substrate RNA, is the HIV RNA, the causative agent of AIDS; or imagine that it's a flu virus RNA or a polio virus RNA or an RNA that's responsible for cancer. To a large extent, these RNAs have been identified, and we'd like to get rid of them in an effective way. It turns out that, given any RNA sequence that you want to eliminate, you can design a hammerhead ribozyme to bind and inactivate that RNA.

You simply look for the sequence of G followed by U along the RNA, which is required for cleavage by this catalytic apparatus. That sequence occurs all the time; because there are only four different bases, G followed by U is a very common occurrence. Then you examine the flanking sequences (which are generically represented as Ns in figure 5) and if a particular nucleotide in the target is an A, you put a U opposite it in the "arms" of the hammerhead ribozyme. If the next one is a G, put a C opposite. If it's a C, put a G opposite. By following the rules of Watson-Crick complementary base-pairing, one can make a designer ribozyme to destroy any specific RNA.

By the time each of the arms in our designer ribozyme is about seven or eight nucleotides long, the statistics are such that you're unlikely to find any other RNA (even in a human cell, which is about as complicated as you get) that would have that same sequence. So you have a potential therapeutic agent with specificity far beyond that of agents which comprise current chemotherapy.

This works very effectively and very efficiently in test tubes. Now the question is, can we use the same techniques to make therapeutic agents useful in living organisms? Experiments are now being done at the cellular stage. In quite a number of cases, ribozymes have been shown to destroy HIV RNA, among other RNA targets in cultures of human cells. There's a long road between success at the level of a tissue culture cell growing on a petri dish, and success in an animal or human. Thus, this remains one of the most challenging areas for RNA research in the future.



## The Human Genome Project: Implications for Biology and Medicine in the 21st Century

Adapted from a 1996 presentation by Leroy Hood

*Leroy Hood's laboratory has played a major role in developing automated microchemical instrumentation that permits the sequence analysis of proteins and DNA, and the synthesis of peptides and gene fragments. More recently, Dr. Hood has applied his laboratory's expertise in large-scale DNA mapping and sequencing to the analysis of the human and mouse T-cell receptor loci—an important effort of the Human Genome Project and in the genomic analysis of the AIDS virus. His laboratory is also interested in the study of autoimmune diseases and new approaches to cancer biology.*

*Dr. Hood, the William Gates III Professor of Molecular Biotechnology at the University of Washington, has an M.D. from the Johns Hopkins Medical School and a Ph.D. in biochemistry from the California Institute of Technology.*

**O**f all the scientific endeavors humankind has ever undertaken, the Human Genome Project is one of the most exciting and profound. The object of study is humans themselves, what they are and how they function. To what extent can we learn what humans are all about? Much of this new knowledge will bear on problems in biology and in medicine; the ethical and social dimensions will be the subjects of a separate presentation.

I joined the University of Washington about three years ago, because of the exciting future of biology and our new Department of Molecular Biotechnology. Bill Gates of Microsoft was also enthusiastic, and when I met him he made an observation with which I agree: as we look to the next century there will be two technologies that are going to dominate in both the intellectual and scientific sense and in the impact they'll have on our lives: the information sciences and biotechnology. The Human Genome Project is a big part of the second.

Both of these technologies deal with information, with the storage and manipulation of digitized information, and with biological information. The collision between these two disciplines is going to create a third: computational biology. In a sense it will become the new theoretical biology and will be profoundly essential for us to understand biotechnology, the Human Genome Project, and many other biological endeavors.

In fact, if you ask what has really changed in biology, biotechnology, and medicine in the last five years, it is our ability to decipher biological information. We have enormously powerful tools that can, at unprecedented rates, look at the basic plans from which we human beings are constructed. As we move into the 21st century we are going to learn how to manipulate that information so as to understand the profoundly complex biology of the human organism. Equally important, we will discover the root causes of mistakes in the basic plan that lead to diseases, and how to correct them.

Before we talk about the genome project in any detail, we should understand what we mean by biological information. For you teachers, more and more of the biology curriculum will be oriented toward this concept of information. There are three types of biological information, one of which is expressed in our genes and chromosomes. We all know that this is a four-letter language, a digital type of information. It's not the same language as computers, which use a base-two, but information expressed by the four letters of the DNA alphabet, with genes and chromosomes carrying the data that is ultimately expressed as a human being.

This linear digital information has a number of interesting properties. The essence of DNA is one-dimensional molecular complementarity, the complementary pairing of pairs of bases—the "A's" and the "T's" and the "G's" and the "C's"—with one another. The reason this molecular complementarity is so important is that double-stranded DNA can be copied in a tiered, template-like fashion. That is, either strand specifies the information needed to reconstruct the other strand. This is the essence of DNA replication, of

how you express information for genes (figure 1). It is also the essence of mutations and disease, and finally, it is the essence of powerful laboratory technology.

You can, for example, take a chromosome which is perhaps 130 million letters long in DNA language, shear it into smaller fragments, and then tear the two complementary DNA strands apart. Because of molecular complementarity, these strands, under appropriate conditions, have the ability to recognize and rejoin themselves to one another. This is the process of molecular hybridization, and it is one of the fundamental techniques in modern molecular biotechnology that allows us to study and interpolate information.

**Tools for biotechnology**

Other fundamental tools for exploring our chromosomes arise from three kinds of enzymes that operate on DNA molecules. We have the so-called *restriction enzymes* that cut DNA by recognizing particular sets of nucleotide sequences and chopping at those particular sites. Secondly, there are enzymes called *ligases* that have the ability to join fragments of DNA together (figure 2). Then there are very important enzymes that can make an exact comple-

mentary, second strand from one strand of DNA. These are called the *DNA polymerases*, and they are the essence of the recombinant DNA technology. Almost all of the major techniques in biotechnology rest on the ability of these enzymes to manipulate DNA in one way or another.

Some work starts with enzymes that can cut human DNA into small pieces. These small pieces of DNA can then be sliced again with a second kind of enzyme and placed in an appropriate carrier for insertion into bacteria. You can also have a recombinant DNA molecule with a human gene, and make many copies of it using a third enzyme, a polymerase. The digital information of human chromosomes is manifest in perhaps 100,000 different units, the so-called genes. Each gene can be expressed independently as a molecular copy of the DNA of the corresponding region of the chromosome; the copy is a related nucleic acid—messenger RNA. This template then employs specialized machinery to make proteins; proteins, in a sense, are the final output of our genes. These proteins are based on a

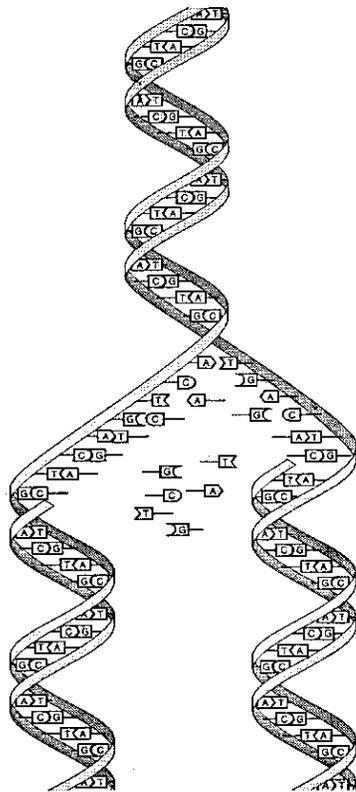


Figure 1. Replication of DNA  
(Source: Office of Technology Assessment)

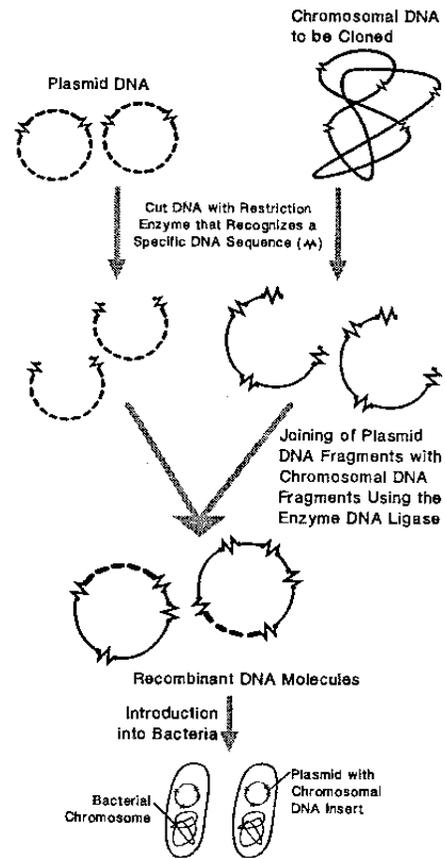


Figure 2. DNA cloning in plasmids  
(Source: Office of Technology Assessment)

20-letter alphabet, 20 different amino acids. It is the order of these amino acids that confer on proteins their fundamental property of being able to go from a one-dimensional string to a three-dimensional molecular machine.

A common enzyme that operates in our digestive tracts for predigesting carbohydrates includes an active site with just the particular letters of the amino acid alphabet that add up to the right chemistry for carbohydrate cleavage. They are juxtaposed in the appropriate three-dimensional space so that cleavage can occur. This, then, is the second type of information: three-dimensional proteins in these molecular machines that catalyze the chemistry of life. They give the body shape and form; when you look at another person everything you see is protein—hair, skin, eyes. These are all products of the fundamental software of life that is embedded in your chromosomes.

There are two problems with proteins, both intriguing, that raise fundamental questions in contemporary biology. One is called the protein folding problem: what are the rules which specify a particular three-dimensional shape for a specific chain of amino acids? When we solve the protein folding problem we can use the knowledge to design drugs and foods. The three dimensional shapes will have the appropriate complementarity to carry out whatever task we wish. And that leads us to the second major challenge: given the particular shape of a three-dimensional protein, how do we figure out what it does, how this machine really operates? In some sense that is a deeper and more difficult problem than the first one. The answers to these fundamental questions are probably going to come from theoretical biology and computational techniques that we're just beginning to devise.

### The challenge of complex systems

If we have one-dimensional information in chromosomes, and three-dimensional information from the genes expressed as proteins, what is the third kind of biological information? From the medical point of view it is the most important of all: the information that arises from complex systems and networks. Your brain has  $10^{12}$  neurons which construct  $10^{15}$  connections, and of networks and complex systems have emerging properties which arise as a consequence of the system as a whole. The emerging properties of the brain are memory, consciousness and the ability to learn. The study of one neuron for 20 years wouldn't tell us one iota about these emerging properties, because they're system properties.

The future of biology is very much bound up in understanding how emerging properties can arise from complex systems, and this is the challenge of both biology and medicine. A simple disease like prostate cancer is a very complex system which requires the study of many different

components. In fact, prostate cancer may be five or ten diseases rather than one. One of the first things we must do with a complex system is to break it down into subsets, subsystems that can be studied. We have to define the elements in the system, establish the connections they make, and understand the physiology arising from those connections. We'll have to develop complex computational models because the complex systems won't be simple and intuitive, or susceptible to simple logic.

To decipher one-dimensional chromosomal information, we determine the order of letters all the way across the 130 million letters of the DNA alphabet. This is the sequence of each of the 24 types of human chromosomes and it's what the Human Genome Project is all about. A second type of deciphering is to extract the meaning that 3.7 billion years of evolution has embedded in those digital strands, an enormously complex computational problem.

This two-pronged deciphering process also applies to proteins: it's one thing to determine structure and another to understand how that structure correlates with function. And the same is true of complex systems. To identify the elements and their connections is one thing; to understand how those interactions lead to emerging properties of the brain and the immune system, to the development of human organisms, is quite another. Again, these are going to be fundamentally challenging computational problems.

We need to look at entire systems to study the brain, the immune system, cancer, or autoimmune disease, and new tools will be needed. It is interesting to observe that many developed thus far have come from the Human Genome Project. This endeavor was first discussed about 1985, and officially got started in about 1990. Enormous progress over this five years has given us an appreciation of what the project will ultimately do for us.

Humans have 24 different chromosomes. There are 23 pairs, and an extra in the pairs because of the X and Y chromosomes. Chromosomes, all contained in the cell nucleus, are fascinating: you have in each of your  $10^{14}$  cells roughly six feet of DNA, and yet that DNA folds down into an incredibly compact chromosome much smaller than the point of a pencil. The engineering process by which this folding occurs can't be duplicated by contemporary techniques.

What is the genome all about? It is the software which governs the most fascinating of all biological processes, that of human development. We all started as a single cell; went through a series of highly programmed and reproducible cell divisions to make spheres and cylinders that flatten into the three different germ layers; and ultimately we grew into adult human organisms with  $10^{14}$  cells. The essence of development is the chromosome choreography

that expresses itself in different cell types, different subsets of the genes. It is what confers on muscle cells their distinct properties, and a different set of properties on brain cells. Understanding the molecular choreography by which your genes are differentially expressed in different cells is the deepest and most fundamental problem of developmental biology, and it's one that can be attacked with the digital information that's present in our chromosomes.

The amount of information is enormous. In one set of chromosomes from either parent there are three billion nucleotides. If that was translated into an encyclopedia for constructing humans, one would have 500 volumes of 1000 pages, each page containing 1000 words averaging six letters of the DNA language. When the genome project is finished, we'll be able to go to volume 240 and open it up to chapter three and begin reading. This encyclopedia will be highly abstract; you will understand a few words here and there, but the essential meaning will be lacking. In short, it will contain a variety of Rosetta Stones to be deciphered over hundreds of years.

The Human Genome Project is also about sequencing the genomes of simple organisms like bacteria, yeasts, a small worm, and a fly. These may be a hundredth or a thousandth the size of the human genome and are much easier to analyze. These genomes contain many genes that are very similar to human genes. We can traverse genomes from humans to yeasts, and come back to humans with an enhanced understanding and more precise questions on how similar kinds of genes function. Another model organism is the mouse, which has a complement of DNA similar to ours. We sequence that genome because it will serve as a wonderful model system for studying human disease. If we discover a lesioned gene in the form of a mutation in humans, we can create the same mutation in a mouse to help to understand its origin and what types of therapies might work.

### Creating genetic maps

The Human Genome Project will deal with chromosomal digital information at three levels. Genetic maps can be made by following with genetic markers certain traits in human organisms that vary—eye color, hair color, susceptibility to certain diseases. We can then assign these traits to certain chromosomes and, in fact, order the traits with respect to one another. A genetic map is kind of like a road map of the U.S.; you can say the disease gene you're interested in is in Tucson, but what about the second level—the street address, the house at which that gene resides?

Creating a physical map involves simply treating a chromosome with DNA cutting enzymes and making small pieces that we can manipulate with recombinant DNA tech-

nology. Because everything fits together like a jigsaw puzzle, different DNA fragments in a test tube will give us their relative order. This information is useful in saying, for example, that a particular fragment has this interesting gene that predisposes to schizophrenia, but it's also useful as the substrate for the third type of map. This is the sequence map of the DNA letters all the way across a string; as many as 130 million. Once we go from fragment to DNA sequence information, we can identify the exact molecular address of any particular gene and work to understand how it carries out its functions.

For each of these major types of maps we have had to develop technologies for analyzing information very rapidly and on a very large scale. This is the challenge taken up by our Department of Molecular Biotechnology at the University of Washington. It is creating instrumentation and tools to be able to decipher one-dimensional, three-dimensional, and in a sense, the four-dimensional information presented by complex systems.

Advancing biology is intimately tied to new technologies that will let you breach the barriers. In sequencing human chromosomes, for example, a real barrier was manual keypunch methodologies that were so slow and agonizingly difficult that it would have taken forever to tackle the entire genome. New machines open new vistas for exploration—a machine that automates sequencing, for example—but you come to new barriers requiring new technology to overcome. Biologists of the future should be trained to think not only about biology, but about developing tools, especially computational. Students must learn about computers, a vital tool in biology as well as in other fields.

Three kinds of tools that have been developed to help define the human genome are large-scale DNA sequencing tools, large-scale genetic mapping tools, and a very powerful new approach to explore fundamental questions like what causes cancer. Large-scale DNA sequencing began about ten years ago when we developed an instrument with relatively simple logic—it used fluorescent groups to color-code the four different letters of the DNA language (figure 3). Someone else discovered a way of taking a piece of DNA and creating from it an array of fragments that were successively one shorter all the way across the unknown sequence.

### Color-coding and sequencing DNA

The end result was that we could color-code all of the "C" fragments in one color, all the "T" fragments in another color, and the "A" and "G" fragments in traditional colors. Then we could mix this array of fragments together—each terminating with the four letters of the DNA

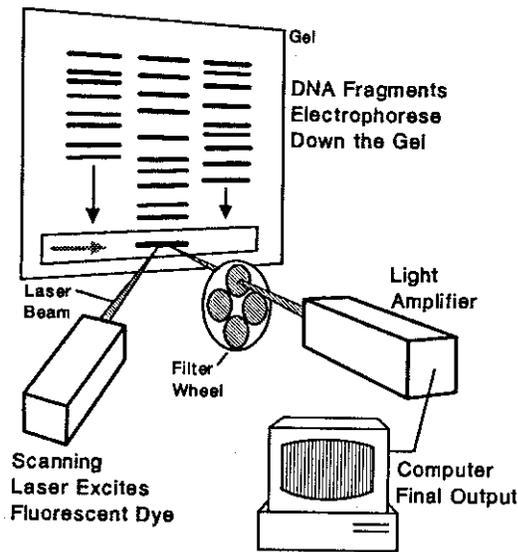


Figure 3. DNA sequencing using fluorescently labeled DNA

language—and separate them by size using gel electrophoresis, with the shortest migrating the most rapidly. The different sized bands had letters that successively read out the DNA sequence. We used a laser beam to interrogate the colors of the bands, fed the color information into a computer, and programmed it to translate colors into DNA sequence.

The first machine we developed could run the 24 samples of human DNA simultaneously, reading several hundred letters of the DNA language at a time. The instruments we have now can read 700 letters at a time, and run about 36 fragments simultaneously. Many centers have 20 of these instruments, and so can do DNA sequencing very rapidly. We have used these machines to decipher a family of genes that plays a critical role in immunity: the T-cell receptor family. It's the largest human sequence ever done, 700,000 letters of DNA language. We used computational techniques to convert this digital sequence into an analog readout. Just one part of the sequence contains more than 90 different genes, and we can identify all of them with straightforward computational techniques and display them in analog fashion so humans can understand them.

Chromosomes have many different languages. One is the language of genes, and another is the language of how the genes are regulated and differentially expressed in different cells. Then, how do you coordinate the expression of many different genes? But there is a language that has to do with how chromosomes replicate; one dealing with how our chromosomes evolve, and a language that deals

with how chromosomes move from one end of the cell to the other during duplication. The point is this: all of these languages embed themselves as information represented by arrangements of DNA letters. We don't even begin to understand these languages, and only very powerful new computational techniques will make them accessible. Biology and mathematics will be very closely tied together.

### Large-scale mapping and markers

The instrumentation for doing large-scale genetic mapping is the same as that used for DNA sequencing. Genetic mapping is the ability to localize sequences at the DNA level, little bits of DNA that vary from one individual to the next. Each of these sites is a genetic marker. The genome project has created 7,000 genetic markers for both humans and mice, scattered across all the different chromosomes; the map of human chromosomes looks the same as that of the mouse.

The importance of these markers—as one example—is that you can study human families with prostate cancer by following the segregation of the trait for prostate cancer among different members. If you study large enough families, you'll find that two markers are always found in adults who can get prostate cancer. This localizes, between these two markers, the gene responsible for prostate cancer. With these genome-wide genetic markers, we are beginning to pinpoint the genes associated with normal physiology and those that cause disease.

The genetic markers that we analyze on the DNA sequencer are, for the most part, stretches of common, simple nucleotide sequences, a "CACACA" repeat, for example. The reason they are polymorphic from one individual to the next is differences in the process of DNA replication in these regions. Very often mistakes are made which copy extra "CA's" or delete "CA's;" maybe you'll have ten of these "CA" pairs whereas I might have 20 in my homologous chromosomes.

We can actually synthesize an oligonucleotide (primer) that is complementary to a region just outside the "CA" repeat on one side, and another primer on the other side. We can then carry out the polymerase chain reaction procedure, the capacity to amplify the DNA between these two primers about a millionfold. The DNA can then be sized on a DNA sequencer, so bands of one size are one allele, and bands of another size are another; if you have both alleles, you see both of the bands. Polymerase chain reaction is a powerful adjunct to genetic mapping.

To genetically map interesting "CA" strings, we start with three or four good genetic markers scattered across a family's genomes, look at 100 or 200 families, then pinpoint six or seven locations that may have genes. We next



Students analyze molecular data with help from Leroy Hood.

get very dense markers across those regions, specify them as sites, then move to DNA sequencing that takes us precisely to the individual gene. The ability to do this sort of genetic mapping has increased 10,000-fold in just the last five years, and it is a revolution.

The third kind of technology we're going to talk about involves large-scale DNA arrays and DNA hybridization. We can do spectacular things with machines that can take a particular cell, a tumor cell for example, and convert the five or ten thousand messenger RNAs that are present into DNA copies. These are called cDNA clones, and this machine will array 25,000 different cDNA clones on a single filter the size of a normal sheet of paper. We now have 25,000 units of DNA information that can be used to interrogate a whole array of prostate tumors to find out which genes are expressed in them. The size of the reaction will be proportional to the amount of the messenger RNA that's present in a tumor; some things will be present in many copies, and some expressed in relatively few. The important point is, if we look at a whole series of prostate tumors, we can divide them into basic types or categories. These will be characterized by the nature of the patterns we see on their very complex DNA arrays: tumors of certain types that share one diagnostic feature and a particular array of patterns.

In fact, in collaboration with another organization, we are already using these technologies to interrogate cancers of the ovary and the prostate, and are beginning to discover the genes that are diagnostic for the different types.

We will ultimately discover genes that contribute to the diagnostic process itself. This is a large-scale procedure: in the past biologists studied one protein, one gene at a time. We can now study 25,000 or 100,000, and that is the only way we will come to understand the biology of complex systems.

We're also working on the ability to synthesize fragments of DNA on chips. In the future we'll be able to make 100,000 different little DNA fragments, and we'll be able to synthesize on a chip the fragments that can diagnostically pinpoint the 100 most common human tumors. When you go in for a routine physical they'll run a simple blood test and say "no problem" or "we'd better check out your prostate." These biological computer chips of the future are going to play a critical role in diagnosis.

The role of computers in biology is dramatized by a computer chip we constructed with the Jet Propulsion Laboratory that is the size of your thumbnail and has 400,000 coprocessors on it. To give you an idea of its power, a board of these chips can complete a task in 3.5 seconds that takes supercomputers up to 12 minutes. This board costs about 1/10,000 as much as a supercomputer, and it's something most biologists could actually run. These specialized coprocessors, as they are called, will be the mainstay for biological computation.

### Ten years, 100,000 genes

What about the implications of the genome project for biology and medicine? In perhaps ten years we'll have the precise identification of all 100,000 human genes; the enormous task remaining will be to determine what they all do. In the past biologists started with function, and then got to the protein and then the gene; how can we go from gene to protein, and from protein to an understanding of its function? This is the fundamental question facing genomic biology today.

More than half of an individual's genes are expressed in the brain; many of those in just a few cells of the brain, perhaps during a narrow window during development. As a consequence, the only way we'll ever identify many of those genes is through a genome-wide analysis. Then the next questions will be, how do you determine from a gene *where* it's expressed, and how do you go from where it's expressed to what that gene actually *does*? One thing you'd like to be able to do, is to decipher the regulatory elements that govern gene expression.

These regulatory elements do three things: they govern when during the life of the organism the gene is expressed; they govern the tissue in which gene is expressed, and they govern how much of the gene is expressed. With these three specifications we can get to particular locations

and guess about gene function. In the next ten years we'll decipher this regulatory language, and we'll use computers to order up representations of all the genes from the liver, or all the genes present in a particular cell. A subset of those 10,000 genes will then light up.

After you've localized the gene and identified where it's expressed, the problem—as we've noted—is to understand function. The real key is that the genes and the corresponding proteins are made up of functional entities called *domains*, and the domains are made up of subgroups of entities called *motifs*. Think of this as an analogy to a train: domains are the different cars of the train and the motifs are the different components of each car. For a caboose, a motif might be a window, a stove, a chimney and a wheel. We can use the motif to get an idea of the structure of a gene. Secondly, the motif allows us to begin guessing as to function, because motifs fold in particular ways that are individually unique.

Once we have the entire human genome sequence we can begin to use very powerful computational techniques to identify the motifs. The motifs are difficult because they are degenerate; that is many different combinations of the protein alphabet, building blocks and letters, can share particular motifs. Given the lexicon of known motifs, however, we'll actually be able to solve the protein folding problem in the not too far distant future. We will take the particular order of letters, and deconstruct them into a motif, and then computationally build a three-dimensional structure. Going from the three-dimensional structure to function is another challenging and complicated task.

The Human Genome Project will tell us much about evolution. There are families of genes scattered across the human genome, for example, that constitute all of the specific immune receptor families. They are called the immunoglobulin gene superfamily, and more than 200 different members have been serendipitously discovered. Given the whole genome, we can write out the entire list of members, interesting because all are involved with molecular recognition at the cell surface, and the activation of differentiation in genes.

Medicine is interested in individual variation. Look at a picture of Willie Shoemaker on the one hand, and Wilt Chamberlain on the other. These differences are encoded in the genes and the genome project will give us an understanding of the differences in normal physiology and—even more importantly, the differences in disease physiology that are a consequence of our genes.

### Genes that cause disease

Genes can cause, or predispose, to disease. If you have a certain mutation in a gene that encodes hemoglobin, it

will inevitably lead to sickle cell disease. But rather than leading to disease, the lesion in the gene usually just confers on individuals a probability that they'll get it. Another example is the gene that's recently been isolated for breast cancer. But when we say that we've identified the gene that predisposes to breast cancer or a gene that predisposes to cystic fibrosis, you have to remember that it is a long way from being able to diagnose genetic disease to figuring out a therapy. The gene implicated in cystic fibrosis was discovered in 1987; we're no closer to be able to treat cystic fibrosis now than we were then.

The difference between the ability to diagnose a disease and to treat it is one of several ethical dilemmas, the topic of my second presentation. Almost any disease you can think of is caused not by one gene, but by a set of genes. In fact, at least four different genes predispose to Alzheimer's disease. As I pointed out earlier, we know almost certainly that all diseases are complex systems; they're not one, but a number of different diseases. Genetics testifies to the nature and type of a disease, and you can't cure it until you've broken it down into its component parts.

In 20 to 25 years we'll discover a number of genes that predispose to diseases. We'll be able to take DNA from each of us and analyze it with computer techniques. We can then print out our disease susceptibilities and probability coefficients. If we do this type of DNA fingerprinting, we'll need therapies to circumvent genetic predispositions for disease. These will be nothing more than our ability to manipulate the three different types of biological information that we talked about before: gene therapies will manipulate DNA; the protein engineering we'll be able to do will provide ways of manipulating proteins; and there are fascinating ways that we can manipulate the immune system. With these preventive measures and the ability to predict individual health prospects, we can move into preventive medicine, and focus on keeping you well rather than on curing you when you get sick.

This is going to cause fundamental revolutions in medicine, in how we train doctors, and in how we educate the public. How do you tell a patient that she has 70 percent chance of having breast cancer by the time she's 65? What does that mean to her, and what kind of advice do you give her?

### Genome project is interdisciplinary

I joined the University of Washington over three years ago to help start the first department ever in molecular biotechnology. It had two objectives: to do interdisciplinary science, and to train students to respond to a fundamental change in the social contract between scientists

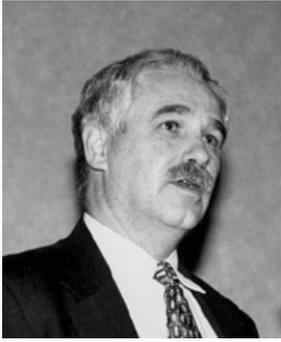
and society. As to interdisciplinary science, the department wants to create large-scale tools for analyzing the three types of biological information we've mentioned. This will require leading-edge technologies from all of the other disciplines—math, physics, chemistry, computer science and engineering.

One challenge is the language barriers that separate different disciplines. A physicist speaks a different language than a biologist, and looks at life science in a different and interesting way. One way to bridge the language barrier is to spend a lot of time with different people, exactly what you would do if you went to Russia and didn't speak Russian. But we have a second approach, and that is to give exceptional students a mentor in biology, and a second mentor in another discipline such as computer science. We will provide the tools they'll need to transcend the challenges of their own disciplines. The student can serve as an intermediary between people who speak different

scientific languages.

There is an emerging discipline—theoretical biology—that within five years will be very well established. We will bring into the field applied mathematicians and computer scientists who have the skill to decipher the three levels of information that we mentioned previously.

We will finally have the tools to attack this complicated network that we call human development. We're beginning to nibble at the edges because of our growing ability to decipher enormous amounts of biological information and thus analyze biological complexity. The ability to train people who can tie together both biological discovery and technical development, who can bring to biology the powerful tools and techniques of the other disciplines, will foster the emergence of theoretical biology. Most fundamental of all, scientists must share their knowledge with society.



## Peeking at the Blueprints: Gene Discovery and Modern Biology

Adapted from a 1996 presentation by David J. Galas

*David J. Galas is the executive vice president and chief scientific officer of the Darwin Molecular Corporation, an innovative biotechnology firm. His research has included the transposition of genetic elements and its consequences, and DNA-protein interactions. Among his own innovations in molecular biology is the widely-used DNA "footprinting" technique, a method for defining DNA sequence-specific sites for DNA-binding proteins controlling gene expression for DNA replication and recombination.*

*Formerly director for Health and Environmental Research at the U.S. Department of Energy, Dr. Galas had responsibility for the Human Genome Project as well as biological and environmental research. Before serving at the Department of Energy, he spent nine years as professor of molecular biology at the University of Southern California.*

*David Galas received his B.A. in physics at the University of California, Berkeley, followed by a Master's and a Ph.D. in physics from the University of California, Davis.*

**I**t's become something of a cliché to say that there's a revolution going on in the life sciences, but in this case it's absolutely true. The changes taking place are legion, but one change everybody is aware of is the ease with which we can now obtain information about the genetic sequence encoded in our DNA. Fundamental genetic information, in the form of DNA sequences and gene locations, is becoming available rapidly and in tremendous quantities. Most working biologists would never have suspected 20 years ago that we'd be able to do the kinds of things that we can do today. The idea that you'd be able to sequence DNA as a routine laboratory exercise was unheard of, and most people thought we would have to infer by indirect means how biology worked at the molecular level. This was the state of the science not so very long ago.

I am in a familiar trap in that I'm forced to make predictions about what's going to happen tomorrow. Just a few years ago we all felt unsure about some of the goals of the genome project, in the sense that there were real questions as to whether we could actually reach them. But it's turned out, in fact, that we were far too conservative about what actually could be done. We have realized, therefore, that Yogi Berra's observation that, "it's very difficult to make predictions, particularly about the future," is quite appropriate. With that important qualification, I'll forge ahead and try to make some predictions where I must. I'd like to leave you with the message, however, that it's absolutely essential that we try to see ahead. Almost no one realizes that many of those things I'll talk about here, things that you might think are wild speculations, are probably just a little speculative. They will be here much faster than you think, so we better start preparing for them.

### **The genome: surface still unscratched**

The questions we need to first consider are these: what is biological information and how does it determine structure and organization? What are genes? We really need to discuss them, particularly human disease-related genes. I consider these the key to understanding the biological mechanism and therefore the key to the future of medicine. The title I originally gave for this talk, "Peeking at the Blueprints," is interesting in several respects and I think I'll keep it. There are two reasons: first, we need to talk about what we mean by "blueprints." "Blueprints" is a metaphor that's been used before to describe the genome, but genes are not really blueprints at all; they're altogether different. The other interesting aspect of the title is "peeking at." You might get the idea that we know a great deal about the genome from what's been reported about the rapid pace of gene discovery. Indeed we do know quite a bit about it, but not in a relative sense. We haven't really even scratched the surface of the richness of genetic biological information in the genome. There are enormous amounts of information yet to be learned, and much more to be understood.

A famous electron micrograph made some years ago shows a chromosome which has been isolated. What you see is the protein skeleton of the chromosome which has been treated in such a way that the DNA released from the chromosome is unraveled. In a gray

area of the electron micrograph are seen the individual DNA strands. There is essentially one molecule on each of the chromatids so two molecules of DNA represent one chromosome. There's an enormous amount of information in the sense that it's a long length of letters compacted so dramatically in the cell; but in another sense, it's not a great deal of information. Scale is really an interesting aspect of biology, particularly to a former physicist. One of my favorite little organisms is the diatom, a small phytoplankton that lives in the oceans. The dimensions of each exquisite little calcite structure is about 1 micron. If you look at diatoms from space, the spring bloom of phytoplankton in the north Atlantic, you see very large "biological" structures, tens of kilometers long. They're not really biological, but are made up entirely of organisms. The ratio of the size of these oceanic swirls to the size of an individual is the same as the size of the genome to the size of an individual gene.

Another analogy is that if you consider the whole genome as equivalent to the surface of the earth (remember those beautiful photographs of the blue-white earth surface photographed from the moon), the size of a single base pair would perhaps be the width of my pen. We're talking about an enormous difference in scale; for some reason, nine orders of magnitude seems to characterize the biggest and smallest objects we're talking about.

Scale, however, is really not the key to understanding the human genome; the key is the way in which the information is interrelated. The human genome is not a lot of information in the sense that there are 24 chromosomes, 100,000 genes, three billion base pairs—even if every one of them were a meaningful fit we could put all the information on a relatively large hard disk on a home computer. It's not a lot of information at that level. At another level, however, it's an exquisitely large amount of information because of the four-dimensional structures Lee Hood has talked about: the way in which the information is related to itself by having genes that control genes and make products that interact with each other and signal the expression of still other genes. They form an intricate and sophisticated network that we do not yet understand. So then, what is the information that will come from the genome project? It's access to 100,000 genes, which is very important for medicine and biology; it's protein sequences, and the clues to understanding biological structures; and it's the statistical and probabilistic differences between organisms, as defined by the way their genomes are described. The last is important, indeed central, in understanding biology and evolution, and also important in medicine.

Let's briefly address the question of what "blueprints" means in this context. Well, once again, "blueprints" re-

fers to proteins, nucleic acids, networks, cells and organs, organisms and species. And the blueprints of what? As you go to scale, you see that there are very different types of information. We don't understand much about the way this information is recorded. What's the problem, you say, don't we know what the genetic code is? If we have a piece of DNA, we know how to read out the protein, and we know how to read out the sequence of an RNA that is encoded by a gene, but networks we understand extremely little about, and we don't even understand how to fold a protein from the one-dimensional information. We certainly don't understand much about what makes a full cell function, how all the genes are interrelated, how to make an organ from an individual cell, and therefore how an organism functions. In short, still well beyond our grasp is an understanding in complete detail of the way in which evolution works on a molecular level.

### **Biology's golden age**

Nonetheless we have all the tools in hand, at least in principle, to solve these problems. The intellectual attractiveness of biology as compared to physics is that you can ask relatively simple (deceptively simple) questions, even as a layman, and go two or three layers deeper into the question and quickly reach the unknown. Yet in principle we have all the tools today—chemistry, molecular biology, and biophysics—to solve those problems. This has to be the essence of a golden age in science—having an abundance of good questions and having the tools to actually answer them all. It may take you 10 or 20 years to do the work, and it is difficult to understand the profundity of some of the questions, but we are indeed in the golden age of biology.

Coming back to the concept of blueprints, you must realize that the macromolecular structures encoded by the genome include not only those which handle the genome's requirements for chromosomes and cell division and so on, but also the macromolecular structures that encode all the apparatus to interpret that information. The transcription apparatus of the cells are all encoded by the information that is read by those same things. So there is a recursive sense in which this information is integrated, and you can make, in principle, changes in a tRNA which would cause you to misread that information or to read it in a different way, which would have repercussions. In the simple sense, then, the metaphor of blueprints is not really correct. And there are other issues that we simply don't have time to discuss.

Let me turn for a moment to the rate of change in this science. One way of focusing on the rate of change in biology is to talk about the goals for the Human Genome Project. An October 1993 *Science* article I published with

Francis Collins describes a committee decision on the goals for 1998: to get a genetic map at a particular resolution; a physical map at 100 kilobases resolution; and the ability to sequence human DNA at about 50 million base pairs per year (see figure 1). Gene identification was a specific goal. In 1996 the genetic map is completed to a higher resolution than called for, and the physical map is not quite there yet, but will be done well before 1998. Although the DNA sequencing is not at 50 million base pairs per year, it will certainly be there, probably before 1998.

The reasons for this progress are not technological breakthroughs: it's simply that a lot of smart people have been using applied chemistry, physics, computational approaches, and engineering skills to make incremental improvements in the instrumentation. The things that Lee Hood has invented are examples. Instruments are more powerful and more efficient, and the DNA sequence information, as measured by how much has gone into the public databases for the last decade, has been increasing by a factor of two every 18 months. The big question is: are technological changes going to actually increase that rate of decrease in the doubling time of DNA sequence information? I think this will be the case, but even if the rate remains the same, we'll have a hundredfold more information by 2005, and more than a thousandfold if even a modest increase is experienced. In fact, I expect we'll finish the first full sequence of the human genome by 2004.

### Decoding the DNA "tape"

What about the structure and organization of this information? A scanning electron micrograph of the "Y" chromosome and the "X" chromosome as they appear in the most compacted form in the cell cycle is revealing as are other views of the chromosome, including an scanning electron microscope view of a not-so-condensed chromosome. From some views the DNA looks like a tape of information which is being held and handled by a very sophisticated set of molecular machinery. This machinery is designed to handle the DNA, to compact it during cell division, and to let it out so that the information can be read. A geneticist's view of the chromosome is linear and fearfully dull. But the average chromosome is about 100 megabases in length, and you can usually isolate a specific gene with genetic mapping techniques. We're achieving excellent resolution with genetic maps, in many cases down to the 1-3 megabase range.

It is also possible to obtain the physical pieces of DNA, as Lee Hood described, that constitute a physical map. An important point is that DNA sequencing has now reached such a level of efficiency that you can actually do DNA sequencing at about the same scale that you work at when you come down from the top end and make genetic maps.

In other words, it is now possible to do genetic mapping, then pull out the pieces of DNA identified by genetic mapping and simply sequence them and define all the genes in the mapped region. This represents something of a technical shift, and it makes a big difference. It will dramatically increase the speed at which genetic information is obtained.

Let me just step back for a moment here and tell you what a gene actually looks like, encoded as it is on a one-dimensional strand of DNA. Most genes have an interspersing of protein-coding regions, the so-called *exons*. These are scattered throughout the DNA, but in a linear order, such that an amino acid is encoded by one exon and the next piece of that protein is encoded by the next exon, and so on. The regulatory information, however, is not so linearly arrayed. The regulatory information tells specific cells and their apparatus when and how to turn the gene on or off, and at what level. Much of this information is arrayed at the upper end of the gene where the transcription begins, but pieces of it are actually scattered throughout the genome. The spacing between the genes is something that is still not understood very well because we don't have enough of the genome sequence to put that structure together. Sequences of the T-cell receptor region, for example, are beginning to elucidate some of those things, but we need to know much more before we can draw solid conclusions.

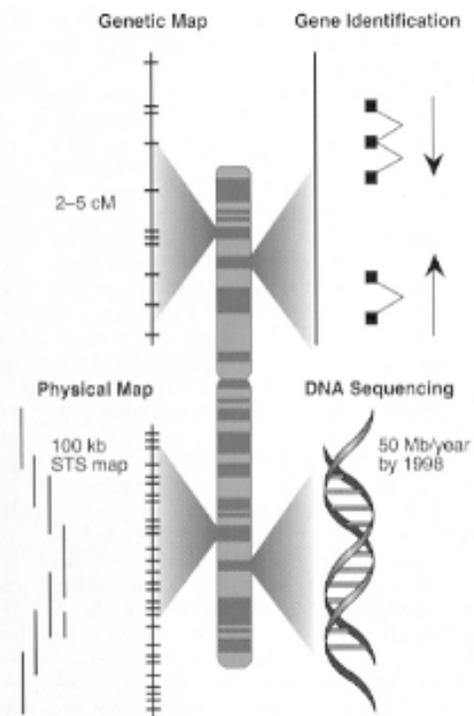


Figure 1

I want to make another point about the way genetic information is found in the human genome—probably in all genomes, but certainly in mammalian genomes. It appears that speculations a number of years ago, particularly by Kimura, are correct: gene duplication and drift is responsible for a lot of evolution. If you simply look at the genome now, you realize he had it right, more so than we thought (figure 2). You can see in the genome, in fact, many cases in which genes like the ones that Lee Hood discussed, the immunoglobulin superfamilies, have duplicated and changed in their sequence. And in fact, you find in the genome many places in which there are multiple genes that are very similar in sequence, and that are obviously derived from one another.

A very simplified map of one region on chromosome 19, for example, shows the carcinoembryonic-antigen genes plus some other genes that are very closely related. Chromosome 19 is where all of the carcinoembryonic-antigen genes in the human genome reside. They're in a cluster of 30 genes and, by comparing the sequences of these genes, you can see that they are derived from one another, and which ones are the most closely related. In fact, if you compare them with the same set of genes in mice, you can see where the divergence probably took place; this is the case for many different kinds of genes.

Let me give you one other example, a particularly interesting one because it refers to a set of gene encoding proteins called "zinc-finger" proteins. We don't need to discuss why they're zinc-finger proteins, but their characteristic is that they have zinc atoms held in the molecular structure in a specific way. The reason they are interesting is that this set of genes is responsible for binding DNA in such a way that it's clear that it's involved in changing the expression levels of other genes—in regulation of the cells. They turn genes on and off, and it is estimated that there are about 600 of them in the human genome. There appears to be at least 200 of them on chromosome 19, and they appear to be in clusters. In this case, however, these genes have a sequence which is so similar, in the segment they use to hold the zinc atom, that you can tell by hybridization that there is a zinc-finger gene present, even if you can't see it. It has been possible to map out a set, and mark on a map of chromosome 19 all the places where zinc-fingered genes have been identified. So the structure of the genome is beginning to be put together. We're also beginning to glimpse the organizational structure of the genome, how it functions, and how it got to be the way it is through evolution. This brings up a fascinating point about what can actually be seen in the structure of the genome—the remnants of its history.

**Genes—of mice and men**

A comparison of the "X" chromosomes of mouse and human shows why the mouse is such a terrific model for doing human genetics. We have a map, crude at the moment, but a map, nonetheless, that shows how the pieces of the mouse chromosomes are related to corresponding pieces of the human chromosomes. A list of a set of genes in humans and a set of genes in mouse—although they have a few different letters throughout—shows they are the same genes. And they appear in pieces that enable you to see that something like 150 to 170 major breaks and rejoins have occurred since these two organisms diverged some 100 million years ago. This not only enables you to get at really interesting evolutionary questions, but it lets you explore mouse genes and in many cases to go directly to the right spot on the human chromosomes and begin to look for the corresponding human genes. The use of this kind of mapping tool, that is, doing mouse genetics and then using it as an aid to understand human genetics, is a major change in the way we do things.

A different kind of gene, the so-called homeotic genes, were originally discovered in the fly. And in fact the Nobel Prize was awarded this year to Ed Lewis and several others for their discovery and elucidation of these genes. These genes are found in all organisms from mammals to flies. This kind of comparative evolutionary information can be very useful. If genes in different species are the same or very similar, you can get into biological function, and even use the information for medical research. All of a sudden evolution has intruded into medical research. Many of the

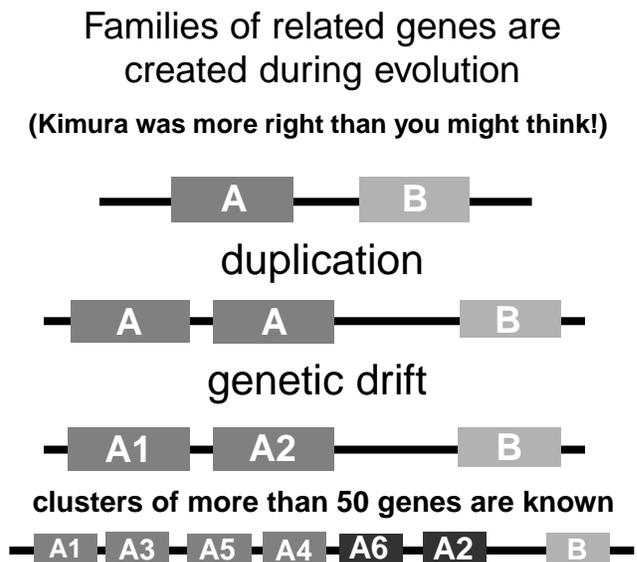


Figure 2

model organisms that were mentioned by Lee Hood are under study right now, and it looks like the yeast genome sequence of the most widely-used yeast model will probably be done in about five months. Likewise, it looks like the genome sequence for a small worm should be completed within the next two to three years. When those things are done we'll begin to see a tremendous amount of work in those model organisms, not for their own sake, but to understand the way in which human genes function. And increasingly it will be useful to do this kind of research to understand how the whole genome works together.

One of the major techniques for finding genes now, is to sequence something in the region of interest and then look in all the databases for all the genes in any organism to see if there's enough of a match to hint at what the function of the sequenced region is. You may find the answer in the worm, you may find it in the fly, you may find it in the mouse, you may find it in yeast, and it can be tremendously useful.

### Genetic bases for Alzheimer's

Let's turn for a moment to some of these disease mechanisms, which are key to understanding the way in which the new genomics is impacting medicine. Before I talk about Alzheimer's disease, let me just say that there are significant ethical, legal and social issues associated with an understanding of human genetics at a deep level. My strong feeling is that those issues should be brought to the forefront and dealt with immediately. It is urgent that we do this, for things are moving along at a fast pace. I think most people who have some responsibility for dealing with these issues have absolutely no idea how quickly things are moving. Change is going to be upon us, and there are many examples of things that need to be dealt with including setting standards and specifying legal rights.

We've been particularly lucky, my colleagues and myself, and we've had a very fruitful summer. One of the things we did find was a new gene that predisposes for Alzheimer's disease. It's been known for many years that there's a definite genetic component in Alzheimer's disease, even though it's difficult to study genetically because the onset is so late. It was shown quite clearly a number of years ago, by looking at siblings and twins and by other studies, that there is at least a significant genetic component—it is heritable in some part. Two genes were implicated some time ago, one a gene on chromosome 21, which is called the amyloid precursor protein gene. One of the more interesting things about it is that we still don't know what it does. Nonetheless it clearly is involved in Alzheimer's disease, and one of the substances that's cleaved from the protein that is encoded by this gene ap-

pears to be involved in the amyloid plaques that appear in Alzheimer's brains. Whether or not it's the causative agent is still not clear. Another gene which is in somewhat the same category is the apolipoprotein E, or ApoE, gene. It appears to be a modulator to a susceptibility to Alzheimer's, and interacts with these other genes in some way.

People have been trying to study the biochemistry of Alzheimer's, and about two years ago at the University of Washington Gerry Schellenberg looked at very early onset families. Individuals in these families appear to get Alzheimer's disease at a very young age. By doing a classic genetic linkage map and by looking at a large number of families, he mapped the gene to a locus on chromosome 14. The gene was cloned early this summer by a group from the University of Toronto, and that led to a whole series of interesting events. When we saw the sequence of the chromosome 14 gene—a component of our research project that we were worrying about—it led us very quickly to another gene as well.

Let me tell you a little about this new gene and its discovery. It's an interesting detective story. There is a scientist at the University of Washington's Veteran's Administration Hospital named Tom Bird, who had been following for many years families with early-onset Alzheimer's disease. The problem he was seeing for some patients clearly was not in the genetic maps where chromosome 14 fits. And Tom, an ardent historian and a very careful man, discovered by chance that members of this second category of Alzheimer's families came originally from villages on the Volga River. It's an interesting story because he picked it up only because people would tell him, "Well, we're from Russia, but we're not Russian, we're German," and he followed up on the lead.

The history of these people is that in the 18th century Catherine the Great, who was German, decided that there weren't nearly enough Germans in that region of Russia. There was a lot of free land, so she recruited people from Germany and they founded a series of villages on the Volga River. Someone who came from Germany in the 18th century had a mutation that enabled us to identify this second Alzheimer's gene. By carefully piecing this together, Tom figured out not only that everybody was coming from these villages, but that there were two specific villages that had families with early-onset Alzheimer's. Many of these people migrated after World War I when they were persecuted by the Russians, and finally in World War II, when the villages were completely leveled, more migrated to the United States. In fact, eastern Washington has many of the so-called "Volga Germans." Tom became well acquainted with these people and managed to collect a number of interesting histories, family pedigrees and DNA samples.

It turns out that, like the chromosome 14 gene, this

second gene has an Alzheimer's susceptibility trait that is dominant, or appears to be: you only have to have one copy of the mutated gene in order to express the phenotype. This makes it a very intrusive gene, tragic for the affected families. In the Volga Germans, the trait was mapped to chromosome 1. What happened next is a bit technical, but it turned out that when the chromosome 14 gene was mapped, we still had a long way to go to isolate the other Alzheimer's gene which was on chromosome 1. How were we going to identify it? Fortunately, the chromosome 14 gene actually turned out to be homologous to a piece of random DNA in the public databases. This, in turn, was sufficiently similar to a cloned piece of DNA from the region we were interested in, that we were able to identify the gene that it corresponded to. And in fact, it turns out that the chromosome 14 gene and the chromosome 1 gene are very similar—we have found the right one.

The way the Alzheimer's story now appears is that there are four genes that affect the frequency of Alzheimer's disease. So far the mutations that we found in only these two genes seem to affect early-onset Alzheimer's. But the disease looks identical, in early or late onset, so these two are probably very important in the pathogenic process that produces Alzheimer's disease. If you compare the amino acid sequence of the chromosome 1 gene to the chromosome 14 gene, you see that these are very similar proteins. We can infer something about the structure from the sequence of these genes: they have many regions that are very similar to those found in many other proteins, and that indicate internal membrane components. We don't understand the structure of a segment on the inside of the membrane, a segment on the outside, and a bunch of membrane components. The amino acid changes are fascinating; there is one amino acid change, and everybody who had it had early-onset Alzheimer's disease.

This is a very unusual case, both for genetics and for the position of the amino acid change. It is possible to draw a cartoon of what the structure is like given the position of the Volga German mutation; a similar drawing can be made for the other mutation that predisposes to early-onset Alzheimer's, that in the chromosome 14 gene. These represent places where mutations are now known to predispose for early-onset Alzheimer's disease. What this means is that we need to find out what these mutations are doing. They're in the membrane, but we've never seen any proteins or any genes like this before, and now we have two of them that are clearly linked to Alzheimer's disease. In appearance, the structure of the gene in the genome doesn't tell you anything—it looks like a normal gene. It includes introns (spacers) and the exons with the protein coding divisions and some leader sequences that look pretty normal.

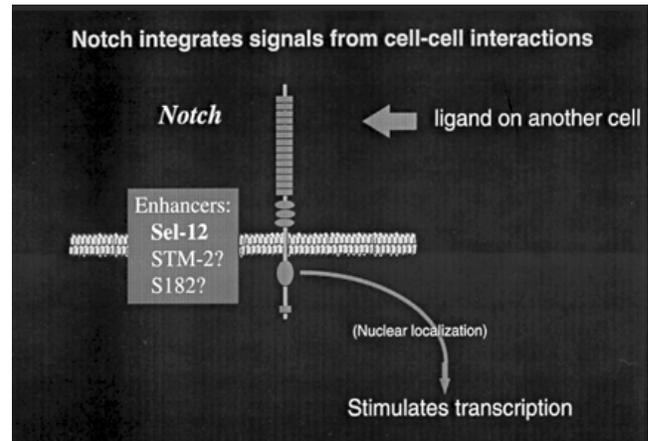


Figure 3

### Clues found in nematode genome

Something else interesting and unexpected in this story happened later in the summer, perhaps an indication of how genetics will affect medicine in the future. Developmental genetics on the *Caenorhabditis elegans* (nematode) genome provided an important clue. *Caenorhabditis* is a very unusual organism, at least as far as we know now, in that it goes in a "hard wired" fashion in growing from a single egg to a full organism. The lineage map of every cell that ends up in the adult organism has now been worked out, which makes it a great model for studying developmental genetics.

One nematode gene, called *sel12*, was found to be homologous to both Alzheimer's disease genes, and affects late development in *Caenorhabditis*. That fact is only a curiosity, but it provides a rare and powerful insight to enable you to understand the mechanism. We do know something about the way that gene works in the nematode, even though we don't understand anything about the way the homologous human genes work. The nematode gene appears to modulate a certain signaling pathway involved in cell-cell interactions (figure 3). One of the components of that signaling apparatus, something that's called a "notch gene," is also present in fruit flies and man. (It's called "notch" because of the way the fruit flies looked when they had a mutation in what turned out to be this gene.) What model organisms have told us then, is that there is a biological pathway, a molecular pathway, that has been implicated in the functioning of a mutated gene that causes Alzheimer's disease. We don't know if this signaling is causing Alzheimer's, and we have no idea how that may be connected to the pathogenesis of the disease. But, by finding those genes in humans and comparing them to genes in the nematode, we have a hot lead about how the disease mechanisms might work, because we've got a

biological pathway.

The important goal now is to establish a connection, any connection, to Alzheimer's pathogenesis. We will look at all mammalian notch genes—there are more than three or four of them—and figure out what functions are involved. I don't want to go into the details of what is and is not known; suffice it to say that it is still complicated, but there are at least three of these kinds of genes in mouse and man. The important thing is that we know that if the signaling pathway is involved, it's suppressing myogenesis (the development of muscle cells) and neurogenesis (the development of neurons from their precursor cells). We don't know whether or not that has anything to do with Alzheimer's, but things have really changed. A model for the nematode shows how the signal is transmitted from cell to cell. It turns genes on internally, and the affected genes in this signaling include the two Alzheimer's disease genes. It has to be shown yet that these really work this way, but it all fits together in a way that at least gives us a path for investigation. This is really tremendously exciting.

As another example, there's a syndrome that was discovered many years ago, Werner's syndrome, which is extremely rare and is genetically predetermined. People who have this syndrome die of premature aging. It's like a progeria, but it's more general and perhaps more like a natural aging; victims die in their 40s and they look like they're about 80. They have a significant increase in the major diseases that will afflict many of us as we age: cancer, cardiovascular disease, osteoporosis, diabetes, cataracts. Other interesting characteristics are that Werner's syndrome causes an instability of chromosomes in certain cell cultures, and it appears to be determined by a single gene on chromosome 8. In this case, unlike Alzheimer's genes, it's a recessive gene, so that means you have to have two mutated copies to get the disease. Werner's syndrome is extremely rare, but we were interested in it not only per se, but to find the gene responsible, determine if there are mechanisms that connect all these diseases, and perhaps get some insight into the process of aging. That, of course, is an area in which all of us are increasingly interested.

### **Werner's syndrome gene identified**

We've actually found this gene, and just recently. We were collaborating with Gerry Schellenberg of the University of Washington, Seattle, Veteran's Administration, who we also worked with on Alzheimer's. The way we found the gene was unusual because it was the first, or one of the first, cases in which we identified the region by genetic mapping and then found the gene by sequencing the whole region of the chromosome. By building a physical map of the region, we could lay our hands on the DNA,

simply infer from the map where the most likely place for that gene was, and then just start sequencing. We sequenced about 650,000 base pairs of DNA and found the gene—identified by the mutations in the patients with Werner's syndrome. We wish we'd found it a little earlier than that, but that's the way things go and it's still a wonderful discovery.

The discovery of that gene is another demonstration of the way things are changing. As we sequenced this whole region, we were looking for genes. This is one of the principle difficulties in the present state of the science. Even if you've got the book that is the genome, it's not always obvious where the genes are and what letters spell "gene," much less what each of them do. In this case were many pieces of DNA sequence from cDNAs and other sources present in the public databases. These cDNAs are pieces of DNA that are expressed as genes by being copied from the DNA genome into messenger RNA, which tells you only that they are coding regions of some kind, nothing about what they do. These things have been put into the databases at tremendous rates over the last few years, and are extremely useful. We simply sequence the genomic DNA, and then search via computer, retrieving these databases from the Internet back to our lab to make a comparison. And as soon as you find a hit in cDNA data, which is expressed somewhere in some cell, you know there's a gene there, and then you examine that gene for mutations. One can find genes that way, and that's the way this work was done.

The genetics of Werner's syndrome is interesting. It's rare, as I said, and we've found five mutations so far that cause it. Two of them are not changes in the protein structure at all, not amino acid substitutions, but changes in the gene sequence right at the boundaries of the exons, right at the boundaries of the coding region. This actually causes the apparatus not to recognize it as the coding region, so it skips that one and goes on to the next, leaving out a piece of the DNA. There are two splicing defects, two stop codons instead of amino acid sequence, and one change in amino acid sequence. And all of these mutations cause this particular disease when they're present in two copies. Although this is just a beginning, just a peek at the blueprints, much of this kind of information will be available in the future and understood as to biological function. This kind of work is going on in many laboratories, and by groups concerned with various medical and fundamental biological problems.

### **Dramatic changes in the offing**

Where will the future take us? I don't know how fast things will change, but it will be faster than we think. I would like to draw a few conclusions by referring to a

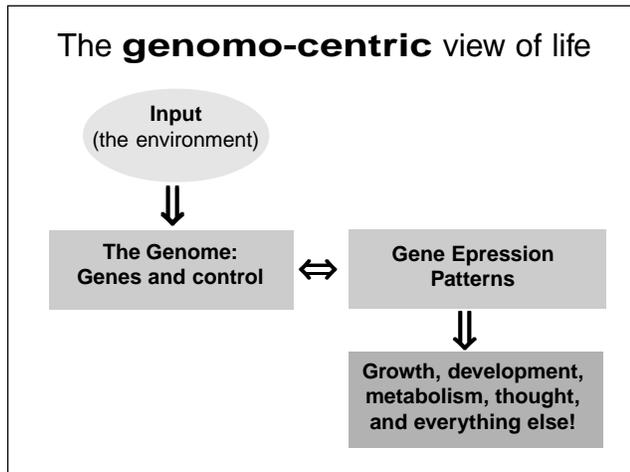


Figure 4

"genomocentric" view of life which is outrageously oversimplified (figure 4). The genome is the repository of everything that turns on or off the genes in response to input from the environment. These are the expression patterns in the various tissues of the body, the cDNA message patterns that Lee Hood talked about (Chapter XX), and they determine ultimately what happens in every cell, what type of cell it is, and how it responds to other cells or to the environment. These patterns determine growth, development, metabolism, and everything else—this is the genomocentric view of life. We are working very hard on deciphering these components, and we're just beginning to see how things are organized. In some few instances we

know how things work, but those are extremely rare; mostly we know a few pieces of the puzzle or nothing. We probably understand with certainty at any level, less than one percent of the genes in the human genome.

What are the implications for medicine? Assume for the moment that the pharmaceuticals we have available reflect a large fraction of what we're able to do medically. Then ask which protein targets in the human body the drugs at hand are targeted against. There's a certain number; they are a very small fraction of the one percent that we vaguely understand. My conclusion would be that in the 21st century most of the pharmaceuticals are going to be directed against gene products that come from genes that we either don't understand now, or more likely, that we don't even know about.

The message I'd like you to take away is that things will change dramatically in medicine, because we know so little now—and there will be other changes. If you can look carefully at the genetic complement that is present in each individual, you eventually (when you know more) can begin to determine just what you need to do to prevent undesirable medical outcomes. And so a migration from acute care to preventative medicine will take place, and it will be based on human genetics.

This is a quick overview of some of the exciting things going on in this area of science, and my attempt to beat Yogi Berra's dictum by trying to say something about the future. I may be all wrong, but I'm absolutely sure it will be exciting and also productive for society. I hope you will become part of the excitement.

**PHOTOGRAPHY AND ILLUSTRATION CREDITS**

Page 19: From *The God Particle*, by Leon Lederman with Dick Teresi, Dell Publishing, New York, 1994.

Pages 22–24: Fermi National Accelerator Laboratory

Page 26: adapted from NCSA/University of Illinois at Urbana-Champaign web site

Pages 32, 38, 39: NASA, COBE Satellite images

Page 33: Harvard-Smithsonian Center for Astrophysics, Geller, da Costa, Huchra and Falco

Pages 34, 35: From *Wrinkles in Time*, by George Smoot with Roger Lewin, W. Morrow, New York, 1993.

Page 37: NASA Goddard Space Flight Center

Pages 56, 59, 60: NASA

Pages 57, 58: Eugene Shoemaker

Pages 61–68: Christopher Impey

Pages 69, 73, 74, 76–78: Thomas V. O'Halloran

Page 70: Brookhaven Protein Data Bank

Page 72: A. Rosenzweig, J. Penner-Hahn, V. Culotta

Pages 80–92: Thomas Eisner and Jerrold Meinwald

Pages 93, 102: Copyright 1998, DaimlerChrysler

Pages 94–101: Harry B. Gray

Pages 103–115: M. Bonner Denton

Page 112: Figure 24 courtesy of R.L. McCreery

Pages 115–120: Brian W. Matthews

Pages 122–126: Thomas R. Cech

Page 133: Mary Levin, University Photography, University of Washington, Seattle

Pages 136–143: Dr. David Galas