The technology, however, is ready. While NRAO astronomers waited for the green light, they linked up with teams of researchers at the University of California, Berkeley, Caltech, and elsewhere to reduce the background noise of millimeter-wave antennas and receivers to the point of approaching limits set by quantum mechanics. Thanks to that work, the biggest remaining challenge is to find a site that could take advantage of such precise technology. It must be high and dry enough to minimize the Achilles' heel of millimeter astronomy—water vapor, which absorbs the incoming signals.

After an extensive search, NRAO officials have a promising candidate: a 20-kilometer-square site in the Atacama desert of northern Chile, at an elevation slightly above 5000 meters in the Andes. Owen calls the site a "pleasant surprise" because of its proximity to a major highway over the Andes into Argentina. In addition, it's only an hour away from the region's major village, San Pedro d'Atacama, which sits at a much more hospitable 2600-meter elevation.

NRAO is also weighing a site on Mauna Kea, a 4200-meter extinct volcano in Hawaii that is already bristling with telescopes. Although Owen, who is head of the site-selection committee, emphasizes that no decision has been made, South America seems to have the votes of most of his colleagues. The Chilean site offers more room—the greater the distance between dishes, the better the resolution—and better atmospheric conditions, which means more observing days, says Robert Brown, NRAO's associate director: "It allows us to get the ultimate out of the instrument." The site's biggest disadvantage is that it is on foreign soil, but the Chilean parliament is working on legislation that would create scientific preserves and delineate a process by which other countries could gain long-term rights to the property.

Politics is also a major concern at home, where astronomers realize it could be tough to steer a big-science project through a tight-fisted Congress. "I hope they see the project as an asset, not just to science but to the communications industry, which is moving toward millimeter wavelengths," says Brown. But Owen acknowledges the obvious peril of a non-U.S. site: "There's no senator from Chile in Congress."

Right now, however, radio astronomers are encouraged that their plans may finally be getting off the drawing board. Jack Welch of the University of California, Berkeley, a pioneer in the field who served on the original NRAO panel that proposed a millimeter array, says the arguments in its favor are overwhelming. "The MMA will be superior to what anybody is now doing," he says. "And the science will be spectacular."

–**Jeffrey Mervis**

BIOINFORMATICS

How to Get Databases Talking The Same Language

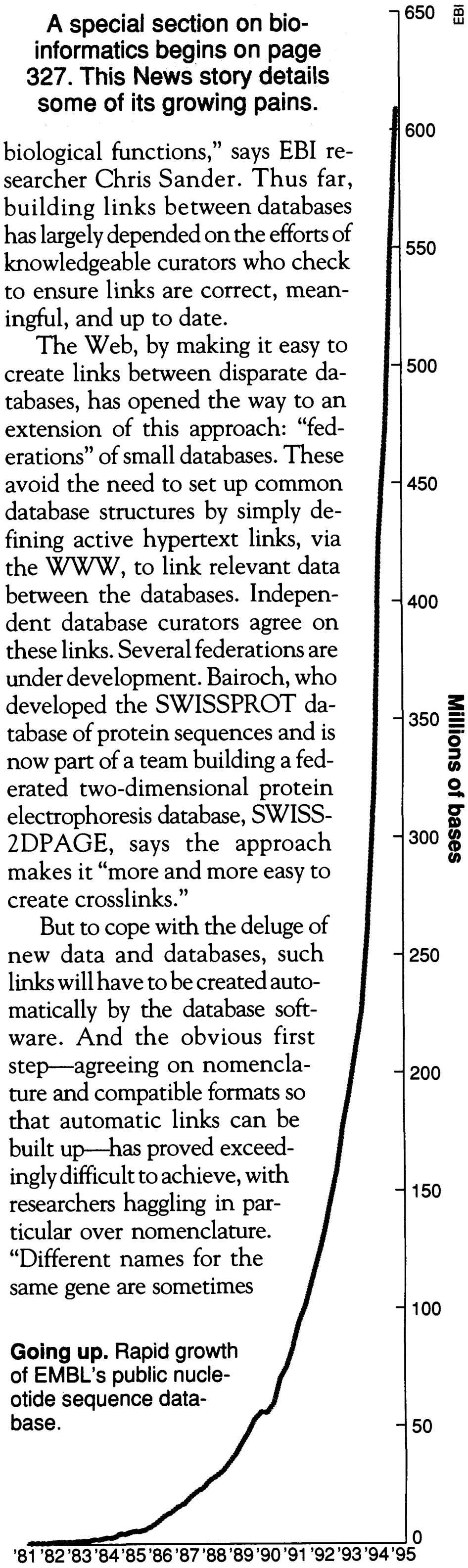
A special section on bioinformatics begins on page 327. This News story details some of its growing pains.

CAMBRIDGE, U.K.—Biologists have been full participants in the Web mania that has swept the Internet, transforming it from a talking shop for researchers into a medium of mass communication that encompasses science, business, and leisure. Indeed, the easy access to data provided by the World Wide Web (WWW) has fed the exponential growth of biological databases housing information on everything from genome sequences to natural history collections. But this new world is increasingly balkanized. Differences in the structure of databases and in their nomenclature mean that a researcher working on a gene or protein in one species may find it exceedingly difficult to find data on the same gene in databases for other species. "There's a significant lack of interoperability. There are now hundreds of databases. There's stuff coming out of the woodwork, but we don't know what to do with it," says Graham Cameron, head of services at the European Bioinformatics Institute (EBI) near Cambridge.

Cameron and others hope to turn this fractured landscape into something more coherent. These bioinformatics experts are trying to develop common standards and names for databases and to establish links between them so that when researchers seek information from one database, they are automatically linked to other databases with additional data. "Anyone interested in the yeast gene *TUP1* faces seven different names for the same gene in other species," says biochemist Amos Bairoch of the University of Geneva.

Efforts to standardize nomenclature have met with resistance from some database curators and specialists in some research fields, who hesitate to change preferred terminology. But researchers are optimistic that Web tools developed for other uses—such as the portable programming language Java—may come to the rescue by creating bridges between databases even if they differ radically in structure and nomenclature. All these efforts, however, are jeopardized by the uncertain funding for many smaller databases, says biological software designer Theresa Attwood of University College London.

The problem of nomenclature stems from the diversity of biological research communities studying gene and protein functions and the lack of common vocabulary among them. "Biochemistry is not necessarily a one-to-one mapping with genetics, but there is a clear relationship. We need to formalize a description of biological functions," says EBI researcher Chris Sander. Thus far, building links between databases has largely depended on the efforts of knowledgeable curators who check to ensure links are correct, meaningful, and up to date.



Going up. Rapid growth of EMBL's public nucleotide sequence database.

The Web, by making it easy to create links between disparate databases, has opened the way to an extension of this approach: "federations" of small databases. These avoid the need to set up common database structures by simply defining active hypertext links, via the WWW, to link relevant data between the databases. Independent database curators agree on these links. Several federations are under development. Bairoch, who developed the SWISSPROT database of protein sequences and is now part of a team building a federated two-dimensional protein electrophoresis database, SWISS-2DPAGE, says the approach makes it "more and more easy to create crosslinks."

But to cope with the deluge of new data and databases, such links will have to be created automatically by the database software. And the obvious first step—agreeing on nomenclature and compatible formats so that automatic links can be built up—has proved exceedingly difficult to achieve, with researchers haggling in particular over nomenclature. "Different names for the same gene are sometimes strongly held—it's really a mess. Efforts to assign a neutral name often lead to angry letters and disputes," says Bairoch. "Last June, SWISSPROT was fully integrated with the EMBL [European Molecular Biology Laboratory] nucleotide database. It has taken years to get things working," he says.

Similarly, the complete genome sequences for the bacteria *Bacillus subtilis* and *Escherichia coli* are both due to be completed this year. They have many genes in common—but under different names, creating a terminological impasse for researchers trying to find genes in common. The impasse was resolved "in a semidictatorial way," says Bairoch. "The *B. subtilis* names will switch to *E. coli* nomenclature with the old names kept as synonyms."

Despite the benefits of standardization, "the lesson has been that it can be incredibly difficult to gain consensus and to agree who has the right to change schemes. Standards are too slow and expensive," says informatics researcher Tomás Flores at the EBI. He and others are pinning their hopes on a different approach: standardizing only the messages between databases rather than their internal detail. Such an approach, referred to by programmers as "object-oriented," would involve taking chunks of data and "wrapping" them in a format that is standard across all databases—hence, the only agreement needed is on the interfaces. "The goal is for a process that's invisible. Researchers interested in a particular gene or protein would get all the information available without having to know where it was stored," says EBI's Cameron.

A group of European bioinformatics centers last year won funding from the European Union to study an interface being developed by the world's largest software consortium, the Object Management Group, which comprises most major software and information technology companies. The approach, dubbed the Common Object Request Broker Architecture (CORBA), has already been embraced by companies from aircraft manufacturers to banks. "The idea behind CORBA is that biologists will never entirely agree on common formats for data entry in databases," says Flores at the EBI, one of the centers involved. So, rather than imposing external rules, the CORBA approach tries to separate data access from data management. The project will cover several of the larger nucleotide and protein sequence databases, including EMBL and SWISSPROT, and several new and emerging ones.

The CORBA approach has also been boosted by researchers' growing interest in Java, a computer language that allows software to be sent through Internet links to carry out small applications, called "applets," remotely (see *Science*, 2 August 1996, p. 591). Using Java, researchers not only can access a database, but can interrogate it in an intelligent way through applets, regardless of database format. Java will use CORBA standards as a tool allowing researchers to range freely among databases. "The success of Java makes a CORBA-based approach much more likely to become mainstream," says Flores.

But funding shortages may interfere with these ambitious plans. Some funding agencies do not see some of the new specialized databases as a core, knowledge-generating activity, so grant decisions can be capricious. Even SWISSPROT, one of the oldest protein sequence databases, was endangered last year when Swiss funding agencies threatened to withdraw funding for expansion of what they saw as an increasingly international resource. "Many funding bodies will support the creation of a new database but are less willing to fund development and continuing support for them. It's crazy, and I hope the tide of things will change." says University College London's Attwood.

But in spite of the hurdles, bioinformatics researchers are mostly confident that they will be able to lower the barriers between databases. "Over 10 years, I'm optimistic because the need for bioinformatics is so clear. An increasing fraction of biology is devoted to information handling—it's intrinsically an information science," says EBI's Sander. "The needs eventually will be met."

–Nigel Williams

NEUTRON SOURCES

SINQ Quickens Pulse to Cool Its Beams

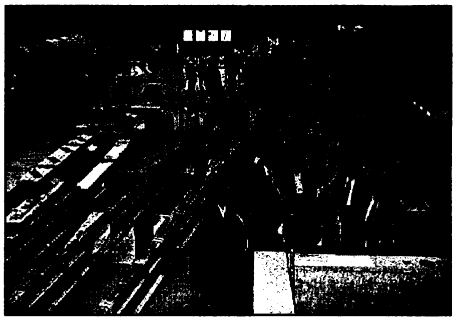
Today, the Paul Scherrer Institute (PSI) in Villach, Switzerland, formally inaugurates a $65 million neutron source that will help close one gap and bridge another. The first is the gap between supply and demand, as researchers clamor for access to neutron beams for research ranging from materials science to probing biological compounds. The new facility, called SINQ, will make a modest contribution toward solving that problem. The second gap is technological.

Neutron sources come in two distinct types: continuous sources, with neutrons provided by fission in a nuclear reactor; and pulsed sources, with neutrons produced by bombarding a neutron-rich target with proton pulses from an accelerator. Both sides have their vociferous advocates: Continuous sources produce very slow-moving, or "cold," neutrons that are ideal for probing delicate biological materials, while pulsed sources excel in time-of-flight spectrometry, where the momentum transferred from the neutrons to atoms in the sample is measured. SINQ, a new accelerator-based, or "spallation," source at PSI, has hallmarks of both: It produces a cold, continuous beam previously only seen at reactor sources. "It is quite a brave project to combine the continuous nature of a reactor and the generation of neutrons by spallation. We will all be extremely interested to see the kind of science that comes from such a source," says Colin Carlile of Britain's Rutherford Appleton Laboratory, home of ISIS, the world's most powerful spallation source.

Both types of source produce energetic neutrons that researchers must first "cool" by passing them through a "moderator" such as water that absorbs their excess energy and reduces their velocity. But for pulsed sources, too much moderation spreads out the pulses of neutrons, and sharp pulses are essential for time-of-flight measurements.

SINQ takes a slightly different approach from other spallation sources. It uses a cyclotron to produce proton pulses that bombard the target at a very high rate, in the megahertz range. When the resulting rapid-fire neutron pulses pass through a moderator, they smear out to create a continuous neutron beam. As a result, SINQ loses the benefit of pulses but gains high-quality cold beams. "[SINQ] will be comparable to the world's best source of cold neutrons at ILL [the Institut Laue-Langevin] at Grenoble," says Albert Furrer, scientific program head at PSI.

One of SINQ's key missions will be the development of targets for the proposed European Spallation Source, a pulsed source that would produce neutron beams about 30 times as bright as those at ISIS. But the main beneficiary will be the neutron community: SINQ will welcome experimenters from universities and industry, including "materials scientists, molecular biologists, crystallographers. … We hope to have an international group of researchers," says Walter Fischer, SINQ's main project leader.



PSI

Particle pathways. SINQ's neutron guides direct neutrons from source to instrument.

–Alexander Hellemans

Alexander Hellemans is a writer in Paris.