Development and Use of a Virtual NMR Facility

Kelly A. Keating,* James D. Myers,*,1 Jeffrey G. Pelton,† Raymond A. Bair,* David E. Wemmer,† and Paul D. Ellis*

*Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352; and †Department of Chemistry and Physical Biosciences Division, Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720

We have developed a "virtual NMR facility" (VNMRF) to enhance access to the NMR spectrometers in Pacific Northwest National Laboratory's Environmental Molecular Sciences Laboratory (EMSL). We use the term virtual facility to describe a real NMR facility made accessible via the Internet. The VNMRF combines secure remote operation of the EMSL's NMR spectrometers over the Internet with real-time videoconferencing, remotely controlled laboratory cameras, real-time computer display sharing, a Web-based electronic laboratory notebook, and other capabilities. Remote VNMRF users can see and converse with EMSL researchers, directly and securely control the EMSL spectrometers, and collaboratively analyze results. A customized Electronic Laboratory Notebook allows interactive Web-based access to group notes, experimental parameters, proposed molecular structures, and other aspects of a research project. This paper describes our experience developing a VNMRF and details the specific capabilities available through the EMSL VNMRF. We show how the VNMRF has evolved during a test project and present an evaluation of its impact in the EMSL and its potential as a model for other scientific facilities. All Collaboratory software used in the VNMRF is freely available from www.emsl.pnl.gov:2080/docs/collab. © 2000 Academic Press

Key Words: collaboratory; collaboratorium; virtual NMR facility; electronic notebook; TeleViewer.

INTRODUCTION

Modern wire and magnet technology continues to evolve and as a result brings to the NMR community ever-larger magnets (>18 T) for research. Concomitant with the larger magnets are increased purchase costs, increased facility space demands, and support requirements. Recently, a national committee (*1*) addressing the next generation of NMR facilities proposed the creation of 10 U.S. NMR sites from existing, leading NMR facilities to become "sectors" which will each eventually house 900- and $\geq$1000-MHz instruments capable of imaging, solid-state, and liquid NMR studies. These high field magnets afford better resolution and higher sensitivity enabling researchers to work on more complex systems at lower sample

concentrations. Within the context of structural biology, high field systems enhance our ability to study larger biomolecules ($\geq$30 kDa) through the use of novel pulse sequences that take advantage of the higher magnetic field, e.g., the transverse relaxation optimized spectroscopy (*2*).

Given their myriad advantages, the demand for high field systems will continue to grow, but an individual institution's ability to purchase them will probably remain limited. It has become more common to create multi-institution NMR facilities where several universities participate in submitting a grant application for an NMR instrument that is then housed at one of the campuses but "belongs" to all. Regional and national NMR user facilities have also been built, providing open, peer-reviewed access to a wide variety of state-of-the-art instruments (*3*). Using such facilities today is not without its drawbacks. The researcher's investment, including time away from one's home institution and travel costs, is significant and scheduling flexibility is limited. Further, the amount of time for in-depth discussion and consultation with facility staff is limited to the trip's duration. It is tempting to ask whether modern software and communication technologies could essentially eliminate the need to travel to a facility yet enable researchers to utilize its full range of instruments and capabilities as well as consult with its resident scientists in a much more flexible manner. This is what we will define here as a collaboratory, i.e., a research team working without regard to geographic location.

As a result of this possibility, we pursued the development of a "virtual NMR facility" (VNMRF). We define the term virtual facility here as an instrument-centered collaboratory. The Environmental Molecular Science Laboratory (EMSL) High Field Magnetic Resonance Facility (*4*) houses numerous state-of-the-art spectrometers (including 750, 800, and soon 900 MHz) that, as part of the EMSL national user facility, are available for use by external researchers (independently or as part of a collaboration with EMSL researchers) on a competitive proposal basis. The EMSL Computing and Information Sciences (CIS) directorate has been researching and developing collaborative tools as part of a broad Department of Energy (DOE) Collaboratory development effort for several years. The VNMRF project, a collaboration between CIS and the EMSL

Macromolecular Structure and Dynamics directorate, was seen as a way to help EMSL users, test Collaboratory technologies in scientific research, and learn about the social and research process changes required to work effectively with remote colleagues.

BACKGROUND

Implementing remote control of a scientific instrument is often assumed to be sufficient for enabling distributed research teams to effectively perform experiments on the instrument. In practice, many other capabilities are needed. For remote collaborations to be as fruitful as collaborations with colleagues down the hall, one must also have real-time electronic access to both collaborators and noninstrument shared resources. Remote researchers (scientists not at the facility site) need to be able to request time on instruments, schedule their experiments, and work with technicians or collaborating researchers at the instrument site to receive training and learn local procedures. Then, during the actual instrument time, the remote researchers need to guide local sample preparation and consult on experiment setup and data acquisition for one or more experiments. Once data exist, remote researchers need access to the data files and local analysis tools. They may again wish to consult with their colleagues at the instrument site during analysis or during the preparation of documents based on the work. Throughout this process, a central repository of background information, plans, ideas, progress, reports, and decisions is needed to allow the team members to coordinate their actions.

As modern research problems become increasingly complex in scope and scale (sequencing and understanding the human genome is a good example), overall project teams are becoming larger and more interdisciplinary. These teams and individual work groups (researchers doing a specific experiment) are becoming more geographically and administratively dispersed than in the past. Collaboratories and virtual facilities appear to be a natural fit to facilitate the work of such groups. Collaboratories are under development in many fields, from space physics (5) and fusion (6) to combustion (7) and materials microcharacterization (8), and they are changing the way research is being done in their target communities.

The capabilities available in collaboratories and virtual facilities have increased tremendously in the 10 years since the word "collaboratory" was coined (9). While early efforts were often limited to text exchange and streaming data, modern collaboratories provide videoconferencing, shared whiteboards, shared computer displays, electronic notebooks, discussion groups, and more. Project websites, e-mail lists, and a variety of means to make it easy to transition between tools help meld the whole into a virtual collaboration space. In a *scientific* collaboratory, scientific resources—instruments, data, laboratory notebooks, analysis software, and the scientific literature—are integrated into this shared virtual space.

APPROACH

The development of our VNMRF was undertaken as a true collaboration between NMR and computer science researchers. The principal local and remote spectroscopists agreed to conduct their research without travel, acknowledging that additional effort would be required to learn how to use collaborative tools, discover how to make effective use of them, and work with software developers to iteratively improve the VNMRF over the life of the project. The computing team installed stable versions of their evolving collaborative tools, helped set up secure communications between the sites, and agreed to provide "whatever was necessary" in terms of training, support, and the development of additional collaborative capabilities. All agreed to measure success in terms of the efficacy of supporting the NMR research and to actively participate in an iterative plan–develop–deploy–evaluate–repeat cycle.

Several criteria were used in selecting the NMR research project that would be the focus of the VNMRF development effort. First, the NMR project had to be representative of those expected at the EMSL facility. We also required that the project be "real," i.e., the successful conclusion of the project would include peer-reviewed publications of the NMR research results. The NMR team (authors Keating and Ellis at the EMSL and Pelton and Wemmer at Lawrence Berkeley National Laboratory (LBNL)) decided upon a series of experiments to determine the structure of the DNA-binding domain of the heat shock transcription factor (HSF). The structure of a 92-residue fragment of the DNA-binding domain of HSF from yeast has been studied previously by both NMR (*10–12*) and X-ray crystallography (*13*) and HSF from *Drosophila melanogaster* has been studied by NMR (*14, 15*). In the proposed study, the domain from yeast would be extended by 24 residues at the C-terminus (molecular weight 13.7 kDa). This region serves as a linker between the DNA-binding domain and a coiled-coil trimerization domain. Information about the structure of the linker gained via NMR would provide insight into how the three DNA-binding domains are oriented with respect to the DNA. Further, the LBNL group believed, based on their experience studying the 92-residue protein at 600 MHz (their highest field instrument), that studying the larger construct using the EMSL 750-MHz instrument would more than compensate for the added complexity of the larger protein, leaving the collaborative tools and our ability to use them effectively as the major risk factor.

To select the hardware and software to be used for collaboration, we used a process analogous to that for choosing the NMR project. Since the VNMRF project was a test case for the operation of a user facility, the hardware and software solution chosen was required to be broadly available to both local and remote researchers. Hardware and software for use at the remote site(s) had to be inexpensive, easy to install, and simple to learn. One early cost-related decision was that we could not

require remote users to purchase a new computer, which led to a requirement that our solution run on common scientific computing platforms (Mac, Windows, Unix). We were more willing to accept costs and complexity local to the EMSL where maintenance and support would be available from computer operations staff. Compatibility with the existing EMSL computing infrastructure was also considered. The overall system was required to be "bleeding edge," using technologies that could be expected to become widely available and supported over the next 3–5 years. We also required the system to be modifiable and extensible so that we could act on feedback and incrementally improve the system during the project.

The EMSL Collaboratory team has been designing collaboration tools based on these criteria for several years, in anticipation of projects such as the VNMRF. The EMSL has taken a very pragmatic approach to software development, partnering with universities and national computing centers on development, integrating commercial and open-source components into the system to avoid duplicate efforts, and providing automated Web-based distribution, installation, and support mechanisms. All of this contributes to our ability to rapidly develop or integrate new functionality requested by users. EMSL's collaborative software developments have been and are funded primarily through coordinated multinational laboratory DOE projects. One of DOE's goals in these projects is to quickly move state-of-the-art technologies from research software projects into usable software systems suitable for at least pilot usage in scientific collaboratories. EMSL Collaboratory software is in use in research and education projects at multiple sites around the globe. Accordingly, we chose to leverage this effort as the basis for the VNMRF system.

Hardware and network setup. The primary computers used for collaboration were the researchers' existing desktop machines. These were a 400-MHz Windows95 PC with 128 Mbytes of memory (PNNL) and a Sun Ultra1 with 128 Mbytes of memory (LBNL). Several other machines with similar capabilities were used at various times to allow collaboration from other offices and laboratories. PNNL and LBNL are currently connected via a T3 network link (45 Mbits/s), part of DOE's Energy Science Network. We estimate that a T1 (1.2–1.4 Mbits/s) network connection is sufficient to effectively use all of the tools for one collaborating group (two to three participants) although this still does not provide full-motion video. For the remote user securely logged into an EMSL spectrometer, the response was usually real-time as far as entering spectrometer commands and looking at resulting graphics. Even a modem provides sufficient bandwidth for using electronic notebooks and some screen sharing, although it is too slow for use of the audio/video conferencing tools.

As part of the VNMRF project, we installed cameras and echo cancelers on each machine. Echo cancelers are small hardware devices that attach to the audio input and output of the computer and serve as both microphone and speaker, allowing both parties to speak at the same time, as with a telephone. (Without one of these devices, only one party can speak at a time, as with walkie-talkies.) Headsets provide a similar reduction of feedback, but they are cumbersome and limit the conversation to a single participant per site. In our NMR labs there is generally too much background noise from the air conditioning systems to be able to effectively use anything but a headset, prompting us to collaborate primarily from our offices. We consider echo cancelers essential pieces of equipment to allow natural conversation via computer.

We have experienced intermittent delays in audio, video, and screen sharing transmissions (fractions of a second to 1 s) due to bursts of other traffic on our network. While this could be somewhat annoying for screen sharing and video, it could render audio completely unintelligible, forcing the use of a telephone. (Some of our audio and video problems have recently been traced to technical problems in routing audio and video traffic across the PNNL firewall. These have now been resolved and audio interruptions have become much rarer.) Faster networks and the ability to prioritize traffic (to guarantee a requested quality of service) should provide a technical solution to these problems over time.

Software. The EMSL Collaboratory software has been described in more detail elsewhere (*16–18*) and information about the current release and the software itself is freely available from our website (*19*). We will provide a brief description here, focusing on aspects important in the VNMRF. Specific information on how these tools were used to support the NMR project appears in a following section.

CORE2000. The suite of real-time software tools developed at EMSL is called the COllaborative Research Environment, or CORE2000. CORE2000 is an extension of the National Center for Supercomputing Applications Java-based Habanero (*20*) environment. CORE2000 adds shared computer screens, remote cameras, and third-party audio and video conferencing to Habanero's whiteboard, chat box, and other tools. The CORE2000 client allows users to start or join sessions by supplying the session name, the server hostname (or IP number), and optional port number. (Users can access a continuously running CORE2000 server maintained at EMSL or start their own locally.) Currently, the collaborators must agree upon the session name and the server to use ahead of time by e-mail or phone. For example, a session might be named "HSF Analysis," running on escher.emsl.pnl.gov, port 2000. When a user starts or joins a session, they see a palette of icons representing the available tools (Fig. 1). Anyone in the session can then click at any time on whatever tools are needed. CORE2000 starts each tool simultaneously on whatever mixture of PC, Mac, and Unix systems the remote collaborators are using. A future version of CORE2000 will allow collaborators to start, monitor, and join sessions via a Web page.

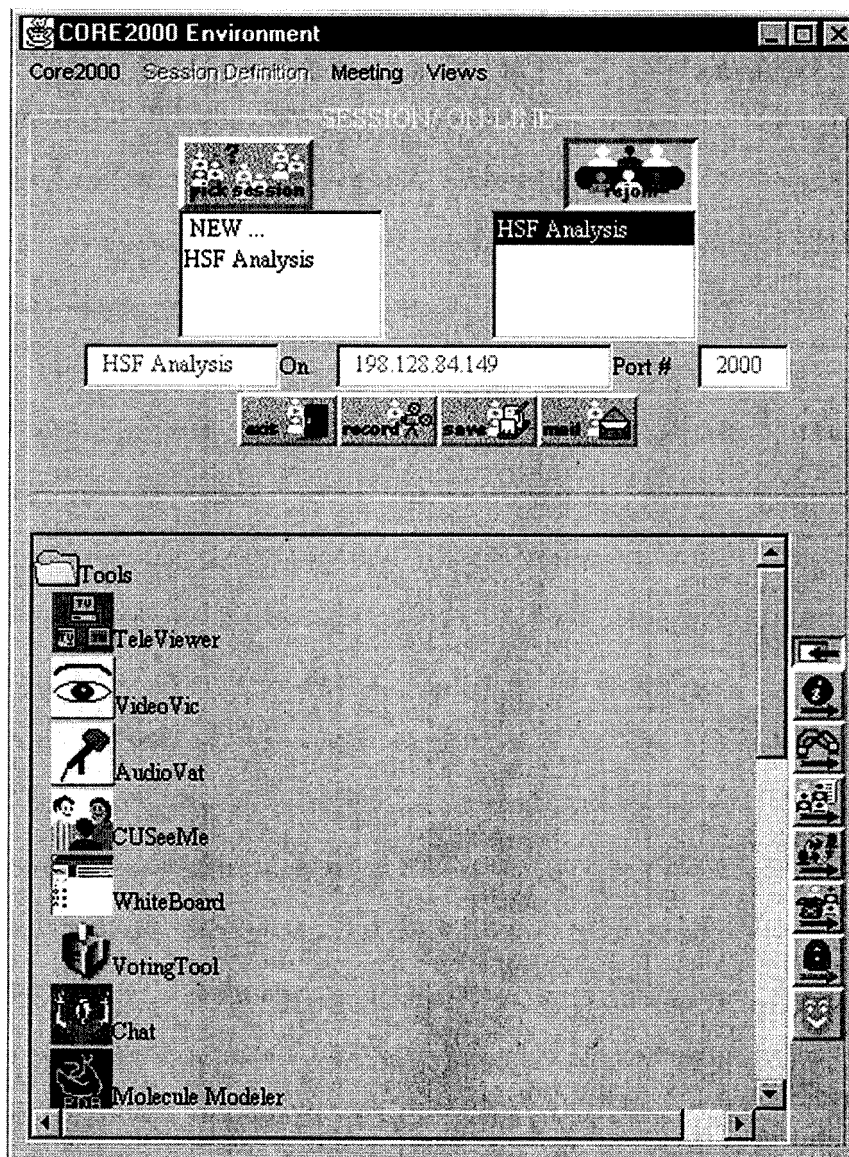CORE2000 third-party audio and video capabilities are self-

FIG. 1. The CORE2000 client interface.

explanatory—they allow participants to converse and to see each other. CORE2000 can launch the publicly available Mbone (*21*) audio and video tools, the option used in our VNMRF project, or CUSeeMe (*22*) (limited to non-Unix participants) audio and video. The CORE2000 chatbox tool is used to exchange short text messages. The whiteboard tool allows users to create sketches and diagrams together using a variety of pen colors. Users can drag-and-drop geometric shapes (lines, rectangles, ellipses, etc.), type text, or draw freehand on the whiteboard. They can also import GIF or JPEG images, such as NMR spectra, pulse sequence diagrams, or molecular models onto the whiteboard and mark them up as the discussion proceeds.

The TeleViewer is CORE2000's dynamic screen sharing tool. Developed several years ago at the EMSL, the Tele-Viewer allows users to transmit a live view of any rectangle or window on their screen to all session participants. Collaborators simply click-and-drag a rectangle over the area they wish to share and transmission to the group begins. Any collaborator can dynamically change the screen being shared by doing a local click-and-drag on their own screen to share the view of an application from their desktop screen. Figure 2 shows a Tele-Viewer window running as part of a CORE2000 client on a Windows desktop, with a Varian console display (the X-windows display from the Unix-based spectrometer computer) being transmitted.
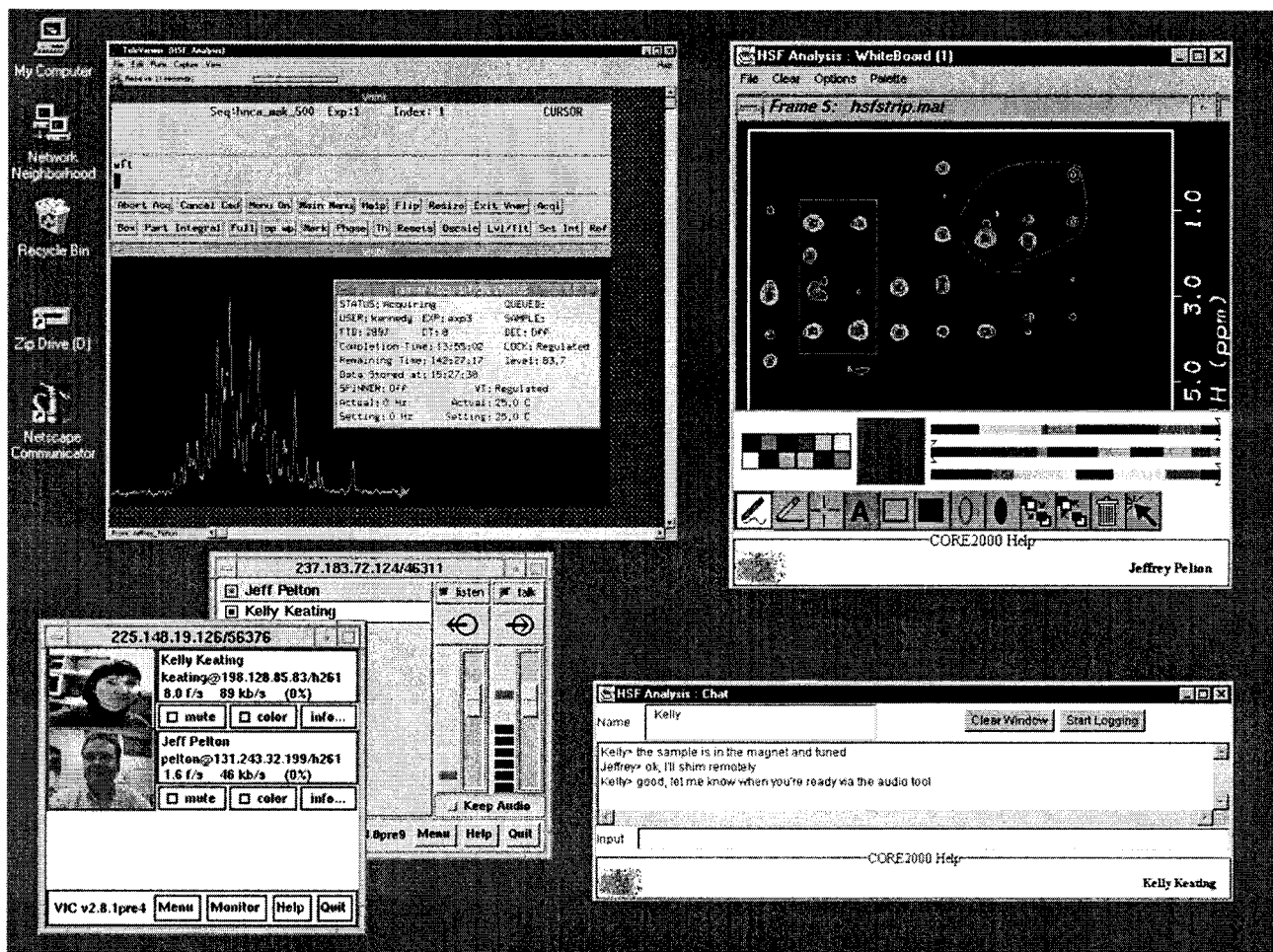
FIG. 2. A PC computer desktop layout of CORE2000 real-time tools. (Top left) The TeleViewer, broadcasting a view of the Varian NMR spectrometer console. (Top right) Whiteboard with a NOESY strip plot pasted onto it and annotated. (Bottom right) Chatbox. (Bottom left) Audio and video tools.

Two other real-time tools available from CORE2000 are worth noting here. They allow viewing and rotation of molecular models: the Molecular Modeler tool to view pdb-formatted molecular structures and the 3D XYZ tool to view molecules in an xyz format.

CORE2000 also has a simple programming interface in common with Habanero that allows new tools to be added as needed. Various groups have used this interface (20) to develop, for example, collaboratively controlled geographical information system viewers and image analysis software for the Visible Human project. At PNNL, a data acquisition system for a mass spectrometer was developed with the programming interface. During the VNMRF project, we used this interface to develop a camera controller tool that provides collaborative remote pan–tilt–zoom control for cameras (we used a Canon VC-C1 Communication Camera) positioned in the EMSL NMR lab.

Electronic laboratory notebook (ELN). An ELN is a rough analog of a paper laboratory notebook, designed to allow distributed teams to record and share a wide range of notes, sketches, graphs, and other information. The Web-based EMSL ELN was developed as part of a DOE collaboratory project with researchers at LBNL and Oak Ridge National Laboratory (23). The EMSL ELN (Fig. 3) presents an initial log-in screen requiring the user's name and password and then displays a main window containing a table of contents with a user-defined hierarchy of chapters, pages, and notes. The content of the currently selected page appears in a separate browser window. All entries are key word searchable. Notes on a page are created using a variety of "entry editors" which are launched from the main window. The notebook currently includes editors to create text (plain, HTML, or rich text), equations (LaTeX), and whiteboard sketches (using the CORE2000 whiteboard), to capture screen images, and to upload arbitrary files. Once a note is created, a click on the "submit" button publishes it to the notebook page and makes it available to other authorized users of the notebook. A simple programming interface allows new editors to be created and added to existing
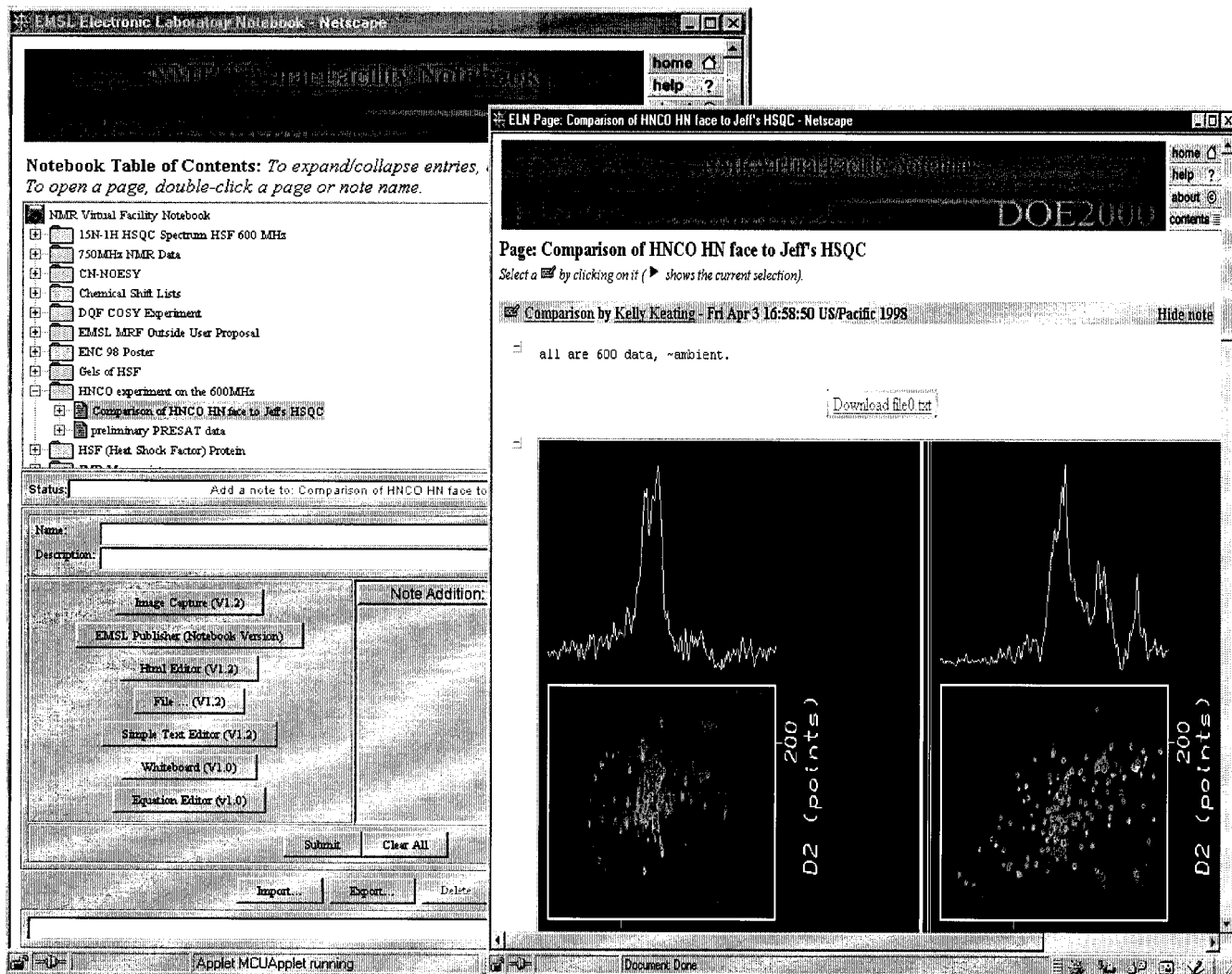
FIG. 3. The NMR Spectroscopists' Electronic Notebook. The table of contents (left) shows a list of chapters, each of which can be expanded to show its pages; "Comparison of HNCO HN face . . ." is a page in the chapter "HNCO experiment on the 600 MHz." On the right is the page, which comes up in a separate Web browser window.

notebooks. Entries are shown as part of a page, tagged with the author's name and the date and time of entry. Each "note" can be rendered by the browser (e.g., text, images), by using external applications (e.g., Microsoft Word), or by using Java applets (e.g., equations, molecular structures). The display of entries is fully customizable. While the current ELN restricts access to group members, it is not fully secure against hackers. The next version of the notebook will include certificate-based user authentication, encrypted data transmission, and digital signatures to address security and to begin to address the issues related to using a notebook as a legally defensible document.

Secure instrument control and data access. At the beginning of the VNMRF project, the EMSL already had mechanisms in place to allow remote users to access EMSL computer resources and data. Since the NMR spectrometer console software is based on Unix and X-Windows, these mechanisms

were also sufficient to allow remote users to control the spectrometer. (The spectrometer manufacturer often takes advantage of this to install and troubleshoot spectrometers over the Internet.) However, we felt the existing mechanisms were insufficient for a virtual facility in two respects. First, we felt it would be necessary to allow collaborative versus simply remote access to the spectrometer so that local experts and others could observe the spectrometer console in real-time to advise and/or learn from the remote operator. Our simplest real-time collaboration solution is to have the remote operator capture and share the spectrometer console using the Tele-Viewer, allowing the rest of the group to observe, but not control, the spectrometer. An advantage of this method is that only the remote operator needs to be running X-Windows. Other collaborators can then use CORE2000 and do not need to have an X-window session going. The second issue was

security. While X-Windows makes it possible to run programs over the Internet, it provides no protection against "session hijacking" and other attacks that could allow hackers to take control. We felt more protection would be needed, especially as we began advertising the continuous availability of EMSL's expensive, high-profile spectrometers. We collaborated with EMSL's Computing and Network Services group to set up and use secure shell (ssh) (24), a publicly available tool that provides authenticated, encrypted telnet functionality and provides similar protection for X-windows. Since the successful incorporation of ssh into the VNMRF project, the ssh protocol has become part of the overall EMSL computing infrastructure, implemented on a "gateway" machine rather than on the spectrometers, primarily for ease of administration. We are currently investigating ways to provide secure group control of the spectrometer, again in a platform-independent manner.

RESULTS AND DISCUSSION

Evolution of the VNMRF during the Heat Shock Factor Protein VNMRF Project

Authors Pelton and Keating started acquiring NMR data using the initial VNMRF capabilities in September 1997. All sample preparation was done at LBNL, and the samples were mailed to EMSL. Since then, a full series of 2D and 3D heteronuclear NMR experiments on HSF, as needed for structure determination, have been completed this way. Data analysis is proceeding, with both real-time and electronic-notebook-based exchanges occurring frequently. As part of our general process, we used time between individual experiments to review how well the collaborative technologies performed, identify tasks that were difficult to do via the Internet, brainstorm ways to make these tasks easier, and install and test new capabilities. We now describe how the collaborative tools were used to support different tasks related to the experiment and highlight some of the improvements we made in response to the NMR team's experiences. For the purpose of discussion, we divide these tasks into three categories: preexperiment, experiment (data acquisition), and postexperiment.

During the preexperiment stage, collaborators need to get used to using the Collaboratory tools, exchange preliminary information, and design and plan the project. Pelton and Keating used CORE2000's audio and video conferencing for introductions (Pelton and Keating had never met in person), discussions of sample preparation, etc. The telephone was often used instead of the Internet audio, because of its higher quality. Traditionally collaborators at the same institution can sit together and lay materials out on a desktop and together look at a computer screen to view a pulse sequence or a molecular model. The TeleViewer filled this role in the VNMRF and made discussions of these types of works in progress possible, without requiring special preparation for a video conference. The TeleViewer also made it possible to debug problems

during the configuration of other software and for Keating to guide Pelton as he learned to use the Collaboratory tools and became familiar with the specifics of the Varian spectrometer console software used at the EMSL. The ELN provided a shared space for literature references, molecular structures from experiments on the smaller HSF construct, and plans and decisions reached in real-time sessions. One of the first customizations of the notebook involved linking in a Java applet viewer for protein structures. After a brief search and some initial tests, we integrated the freely available WebMol Java applet (25). WebMol displays pdb-formatted molecular structures in a 3D, rotatable format. It also allows users to display interatom distances and angles, enough information to allow quick analysis and comparisons without having to launch a stand-alone analysis package. We worked briefly with the WebMol developer to make cosmetic modifications, e.g., making it possible to run WebMol within an existing Web page instead of having it pop up its own window, and upgraded when a new version became available.

During the experiment stage, real-time interactions dominate. The researchers log in to the spectrometer console, work with each other in order to set up the experiment, and bring in facility staff for assistance with sample insertion, tuning, and discussion of any problems that arise. Thus, logging into the EMSL network and the NMR console software via ssh and starting CORE2000 become part of the daily routine. Other than the physical operations involved in sample handling, insertion of the sample into the spectrometer, and probe tuning, all aspects of the NMR experiments can be done collaboratively over the Internet. Pelton and Keating completed several experiments in this manner, with Pelton in direct control of the spectrometer software, working with Keating via CORE2000. It is worth noting that some tasks, such as shimming the NMR magnet, can easily be accomplished by one person, and, at various times, would be done without video conferencing or screen sharing, locally by Keating or remotely by Pelton. Direct communication became more important during configuration of the experimental parameters, discussion of adjustments, evaluation of trial spectra, and finally deciding when to start the experiment. In some experiments, CORE2000 was used to bring a third person into the conversation to address a particular issue.

For long-term monitoring of the experiment (up to several days), either Pelton would stay logged into the spectrometer and leave its display in the TeleViewer which could be viewed by Keating for the duration, or Pelton and Keating would log in independently as necessary. Both mechanisms also allowed Keating to monitor experiments from home or while on travel. As noted above, we added a live laboratory camera to CORE2000 over the course of data collection. During data acquisition, this allowed the researchers to get a good view of the spectrometer's status modules. With a few button clicks Keating or Pelton, in their respective offices, could move from observing general lab activity to viewing the digital tempera-

ture readout from the probe and confirming that the pulsed field gradient lights were flashing as expected.

In the postexperimental stage, researchers need to archive the data, process and analyze it, create databases, exchange progress reports, prepare for further experiments, and eventually prepare written reports. The VNMRF makes use of the existing EMSL data archive. The archive provides drag-and-drop access to several hundred gigabytes of disk storage backed by a multiterabyte robotic tape library. Users on the EMSL network, local or remote, can easily move files or directories from a local EMSL host (including the spectrometer) to and from the archive. Remote users of the EMSL VNMRF must then use ftp as a second step if they wish to move the files back to their home institution. (Although the technical means exist to eliminate the need for the ftp step by allowing remote computers to seamlessly share EMSL's distributed file system, we did not feel the costs and complexity were justified.)

Over the course of the project, an NMR Spectroscopists' "version" of the ELN was created to make it easier to record information related to NMR. The changes ranged from "bug fixes" necessary to allow the ELN server and underlying Web server to recognize files with a ".fid" extension as a valid, non-HTML file type, to the integration of the WebMol viewer, to the development of automated means of creating ELN entries directly from the NMR console command line. In the latter case, an "ELN Wizard" was developed that can be called from within other programs to automate transfer of parameter sets and screen snapshots immediately to a user-specified chapter and page, without requiring the user to open a browser and log in to the ELN. In addition, we developed a Java applet for the display of the parameter file (the Varian procpar file in our case) that shows the parameters not as a long text list, but in an interactive window format that displays only the lines of text associated with the selected parameter. Selections are made either by scrolling down a list or by typing the first letter or two of the parameter name. Later in the project, we added a small scanner to Keating's desktop PC that allowed her to conveniently scan gels of the purified protein samples and other paper documents into the ELN.

Once an experiment was complete, Pelton and Keating moved the data from the spectrometer to the EMSL archive and then made local copies for themselves when they were ready to begin analysis. Data analysis included both joint, real-time viewing and discussion of data and independent analysis using local software, followed by later real-time comparisons. During the VNMRF project, Pelton and Keating each used the analysis package most common at their site and verified each other's results by comparing their respective crosspeak assignment databases. While this mode of collaboration presents an interesting contrast to that used in data acquisition (both researchers using the same software), the choice was primarily personal preference, and we did not attempt to quantify which mode was more effective. Using CORE2000 and the Tele-

viewer, Pelton and Keating could interact closely, watching as menu items were chosen, cursors were moved, and peaks were selected. The whiteboard tool proved convenient for the purpose of viewing and annotating 2D slices of a processed 3D spectrum.

A variety of means were used to exchange peak assignments ranging from sending faxes of handwritten notes to e-mail and entry of Excel spreadsheets into the electronic notebook. Although it is possible to enter an Excel file into the current version of the notebook, only users with Excel can view the contents. However, if the file is first converted to "comma separated value" format, the notebook will display an HTML table of the spreadsheet contents and allow the remote user to download the file itself, providing an immediately readable, cross-platform solution. We are currently using the Wizard described above to create "Save To Notebook" macros within Microsoft Word and Excel to simplify transfer of these types of information to the ELN. Finally, as projects near completion, text and figures for publication need to be developed. Again, real-time interactions prove to be useful in defining plans and schedules, negotiating differences, and reaching decisions, while e-mail and the ELN support the exchange of drafts and provide a shared storage space for final versions.

Future Improvements for the VNMRF

As one might expect, during the course of this project we identified many additional possible improvements that we did not have time or resources to fully implement. In many of these cases, partial solutions exist. In others, we expect to be able to provide the requested functionality in the next version of the EMSL Collaboratory tools. Online NMR reservation systems and online spectrometer logbooks, both using the same usernames and passwords as the ELN and other tools, which allow administrators full access while limiting what members of an individual group can see, were one suggestion. We are currently experimenting with several tools to provide some functionality in these areas. Another suggested improvement was to provide databases (of chemical shifts, for example) that can be shared. Also, it may be useful to record a CORE2000 session for later playback. Currently, CORE2000 has a limited ability to record sessions. Unfortunately, several of the CORE2000 tools, including audio, video, and the TeleViewer, cannot be recorded as part of the session (however the audio and video tools can be recorded separately). Such records would be useful to a colleague who was unable to join the real session. It would also be useful for collaborators to be able to see existing sessions and to "dial" colleagues using CORE2000 rather than having to prearrange sessions and to worry about session names and host and port numbers. The next version of CORE2000 will include a Web-based session directory that will list all CORE2000 sessions and allow people to join a session simply by clicking a "join" button.

There are several useful extensions planned for the ELN.

Automatic e-mail notification will allow collaborators to be notified when their colleagues add information to their shared notebooks. Strong encryption (128-bit secure socket layer encoding) will protect data in transit to and from the notebook server, while digital signatures and timestamps will help provide legal defensibility for notebook entries. Work is also in progress to support interactive visualization of individual slices from a processed 3D NMR spectrum on an ELN page, without requiring the prohibitive time delay of having the full dataset ($>$100 Mbytes) sent to the local browser.

Observations after Completing a Series of NMR Experiments Using the VNMRF

Overall, we found "working together apart" to collect and analyze NMR data between remote sites with the VNMRF to be effective. There are certainly differences from working with local colleagues and local spectrometers. Many of these differences such as the lack of nonverbal feedback from body language and the intrusion of technology are well known in the computer-supported cooperative work community and have been discussed elsewhere (*18*). The overall effect of working remotely on the 750-MHz spectrometer via the VNMRF has been likened by one of us to "driving a Cadillac"; it's a luxury vehicle, but there is no "road feel."

Planning projects, running NMR experiments, and analyzing results with remote colleagues require a few extra moments at startup for remote log in and launching collaborative tools. However, with some willingness to adapt, the time savings and improvements in the quality of work to be had through on-demand access to remote resources and frequent feedback from remote colleagues more than offset the initial delay. This conclusion has now been reached by several independent groups using the VNMRF. Given the example of Pelton's and Keating's interactions, many of the outside users who have used the EMSL NMR facility since 1998 opted to conduct some or all of their experiments remotely. Some traveled to EMSL for training and to run their first experiments and then later consulted with an EMSL NMR researcher via CORE2000 or ran NMR experiments remotely while others never made a physical visit. All have reported positive experiences. In this section we give specific observations, primarily from the original VNMRF team, of the pros and cons of working remotely.

Installing and configuring hardware and software to successfully complete a first session is arguably the most frustrating part of working remotely. While the complexity of the installation of the Collaboratory software itself was a factor in the beginning, automated installation scripts and other improvements have made this less of a factor. Similarly, finding cameras that work across the wide range of machines found at user sites has also become easier over time. However, several factors continue to contribute to the frustrations of the initial setup. At the core of the problem is the fact that proper operation requires that cameras, browsers, computers, and all networks in between the participants be properly configured and in working order. We've discovered over the course of the project a variety of subtle misconfigurations, at both computer and network levels, that may allow other applications, even e-mail and Web browsers, to function while preventing the operation or degrading the performance of other collaborative tools. The fact that the machines and the networks in between may all have different administrators adds to the confusion. The significant differences in Java and browser features between versions and across operating systems introduce additional complications. Nevertheless, most of the complexity is faced by the virtual facility; an individual will likely encounter only a few problems, although they may learn more about the Internet than they first expected. Fortunately, the overall operation is reliable once set up. With a few preexperiment practice sessions with an EMSL researcher, users quickly became familiar with the ssh remote log in process, CORE2000, and the ELN. Planning to do setup and training well in advance of an experiment, analogous to making travel reservations, certainly helped avoid last minute surprises.

Although Pelton did not visit the EMSL prior to beginning experiments, subsequent groups have reported that the initial visit to the facility was useful (particularly if it was a new spectrometer type to them). The ability to both make introductions and observe spectrometers and collaborative tools in action before having to start getting accounts and installing software helped provide initial momentum. Video appears to play a similar role within collaborative sessions. Usually, after the initial startup of the video tool and an exchange of greetings, the video window was iconified and not reopened during the remainder of the session except to introduce additional colleagues (people walking into one of the researcher's offices or someone joining the CORE2000 session).

Throughout the project, input from the NMR researchers led to improvements. The NMR researchers were asked to analyze the way they conducted experiments and identify the tasks that compose the experiment lifecycle. They were also asked to help identify on an ongoing basis which of these tasks would benefit most from close collaboration and what information needed to be shared to accomplish these tasks (data files, parameter sets, literature references). While this did lead to software improvements, in many instances this introspection and simply using the electronic collaboration tools gave the researchers ideas for new ways to work together. For example, over time, researchers began using the chatbox for two specific purposes: to notify remote colleagues when they stepped away from their desk and to transmit unfamiliar terms and numbers. The latter is especially useful during NMR experiment setup when the researchers are discussing spectrometer commands, such as lengthy pulse sequence names. It is faster and more accurate to type information such as these sequence names, other program commands, or URLs into the chatbox than to spell them out loud and request that the other party repeat the information as confirmation. Copy and paste then makes it easy

to transfer the information from the chatbox to where it will be used.

Verbal dialog also changed slightly. The relative lack of feedback from remote participants and their lack of "presence" prompted many participants to ask status questions, analogous to a lecturer asking "Can you hear me?" and "Can you see this slide in the back?" With more experience, users started to realize that more feedback is useful, and they automatically began saying "Do you see that in the TeleViewer?" or typing messages in the chatbox to confirm that they saw updated information, to explain long pauses, or to inform their colleagues of off-camera events.

The TeleViewer, at first used only for live screen sharing of the NMR console, quickly became a central tool in most sessions. It proved extremely useful for live discussions of data processing and analysis, particularly since each researcher was using different NMR processing software. It also began to eliminate the need to e-mail, ftp, or fax documents and plotted spectra in advance of phone calls. We found having a real-time discussion via the audio tool while displaying processed spectra using the TeleViewer worked as efficiently as seeking help from a colleague down the hall. In fact, we became so used to the TeleViewer that we've experienced a new frustration during phone calls as we've tried to explain a situation verbally to colleagues who are not familiar with the TeleViewer, knowing how easy it would be to simply show them rather than explain.

The ELN has also prompted changes. It's rare anymore for us to print a multiple-page hard copy of the parameters from an NMR experiment since it is much more convenient to upload them into the ELN and have them efficiently displayed in the Java applet. The ELN is also the primary archive for short notes, structures, and pictures. We still tend to maintain hard copies of vital data, such as 3D matrix slices, and, while it might be convenient to be able to link to an online version of journal articles on an ELN page, we find it more comfortable to read a paper copy of an article rather than an on-screen version.

Perhaps the most exciting changes were those in the overall distribution and scheduling of tasks within distributed groups. From our initial discussions about how NMR researchers traditionally worked with local versus remote colleagues it became clear that local interactions are less formal and more frequent. Researchers expect to see colleagues in the hallway, discover problems, brainstorm solutions, and change plans as needed on a daily or weekly basis. With remote colleagues, who may visit the facility only during the actual experiments, work is divided into larger, more independent chunks. Sample preparation is done at one site without much consultation with remote collaborators, as is analysis. We believe that researchers collaborating via the Internet, although they originally expect to work as they traditionally do with remote colleagues, evolve their collaboration to be more like local ones over time.

Two specific examples of this are worth noting. In early experiments, Pelton and Keating did their analyses in parallel, without much interaction, over the course of several months. In later experiments, they took advantage of the availability of CORE2000 and the TeleViewer to verify that each had performed the initial data processing (deconvolution, apodization, Fourier transform, phasing, etc.) properly. Similarly, Pelton and Keating began to confer more on tentative crosspeak assignments and other pre- and/or postexperiment tasks that were initially tackled independently. These informal "comparing notes" sessions helped them to quickly catch each other's oversights and avoid blind alleys. While the bulk of the work was still done individually, as is the case for local researchers, this ability to meet as needed resulted in more frequent communication, a reduction in "lost work," and a pattern of activity much closer to the norm for colocated groups.

The other shift toward a local style of work is the VNMRF researchers' ability to make opportunistic use of spectrometers. When spectrometer time suddenly becomes available because the scheduled researcher on a given day is unable to use the time, replacement users no longer need to be limited to in-house staff. Since samples can be stored at EMSL, and in general a remote user's sample can be shipped overnight, VNMRF users can prepare for an experiment on the spectrometer in about the same amount of time as an in-house staff researcher can. Pelton and Keating have so far been able to complete several additional experiments in this manner, helping them address questions that arose during analysis. This capability benefits the facility's users and helps facility management optimize utilization of the spectrometers.

THE FUTURE OF VIRTUAL NMR FACILITIES

With the success of the first project done at the EMSL VNMRF, we have shown that virtual facilities, already becoming successful in other domains, are applicable to NMR. Despite their current drawbacks, the utility of electronic collaboration technologies today and the necessity of these technologies for future ultrahigh field spectrometer facilities is clear. Given the escalating costs of high field (900+ MHz) spectrometers, it is unlikely that many existing NMR sites will be able to independently afford to establish a stand-alone facility for only in-house researchers. A more likely scenario, as recently envisioned by the Committee for High Field NMR, is the creation of a set of complementary ultrahigh field user facility sites, or "sectors," that would become part of a "National Magnetic Resonance Collaboratorium" (1). That committee's report also urged that these sectors coordinate policies and have interoperable software to provide the NMR community with shared resources that transcend facility and discipline boundaries. For example, a researcher could draw upon several sites for the most appropriate expertise, instruments, and other resources. We believe it is particularly timely for communities using instrument-intensive techniques such as high field NMR to incorporate the virtual facility concept into existing centers and to consider it in the construction of new facilities—plan-

ning for state-of-the-art Internet connections, audio/video- and computer-equipped conference rooms, fewer guest accommodations, and so on. We anticipate that such virtual facilities will quickly become a standard way for researchers to access expensive state-of-the-art equipment and to work with remote colleagues more efficiently.

CONCLUSIONS

The EMSL VNMRF was developed as a true collaboration between the Macromolecular Structure and Dynamics directorate and the Computing and Information Sciences directorate within EMSL. It provides a useful combination of secure remote operation of EMSL NMR spectrometers with Internet-based video conferencing, remote-controlled laboratory cameras, real-time computer display sharing, a Web-based electronic notebook, and other capabilities. The EMSL VNMRF is proving to be as or more effective than travel in allowing external researchers, working independently or as part of a collaboration with EMSL researchers, to complete experiments using the numerous state-of-the-art spectrometers housed in EMSL's High Field Magnetic Resonance Facility. These researchers are finding the VNMRF to be a convenient means to enhance their access to EMSL resources before, during, and after use of the spectrometers as evidenced by their comments to EMSL's NMR operators and the growing popularity of the VNMRF (new users, repeat users, and users who have chosen to use the VNMRF after an initial trip). Given the rapid development of collaborative technologies and users' changing expectations as they become more sophisticated in their use of the VNMRF over time, we will be evolving and improving the VNMRF for many years to come. Working to hide the complexities of remote interaction from VNMRF users and responding to their requests for integrated scheduling and logging, shared data summaries (e.g., crosspeak databases), and other tools should make later versions of the VNMRF even more useful.

We note that while the majority of projects at the EMSL NMR Facility are expected to involve an external NMR researcher working with local research staff and NMR operators, as discussed in this paper, education-related projects and collaborations between theoreticians and experimentalists are also anticipated. Previous work identifying different types of collaborations (peer-to-peer, mentor–student, interdisciplinary, producer–consumer) and the differences in their communications needs (*18*) suggests that a VNMRF optimized for peer-to-peer interactions may not be optimal for supporting a theoretician working with an NMR collaborator or a student learning a new NMR technique. Thus, in-depth consideration of the requirements for supporting these types of collaborations will undoubtedly suggest additional directions for VNMRF enhancements.

We believe that our cooperative process for the development of the VNMRF—deployment of an extensible base of collab-

oration tools to the researchers, up-front analysis, and iterative development, deployment, and feedback—has been critical to our success. The choice of extensible base platforms, such as EMSL's CORE2000 and ELN software, has made it possible to quickly and incrementally incorporate new technologies and to be responsive to users needs. Our systematic approach, involving collaboration between NMR researchers, computer scientists, and researchers studying the dynamics of distributed groups, has also been very effective. We believe that the rapid advances being made in all three fields and the dynamic, nonlinear coupling between user expectations, scientific processes, and collaboration technologies make such broad expertise essential in the development and management of virtual facilities.

Collaboratories have the potential to remove the walls around departments, organizations, and universities and could lead to the creation of metalaboratories with capabilities in both expertise and equipment that far exceed those available in any one laboratory. Virtual NMR facilities have the potential to bring high field NMR as a research tool to a broader audience of faculty, students, industrial, and government researchers. Such collaboratories may also grow to include complementary fields such as mass spectroscopy and X-ray crystallography, allowing researchers to apply complementary techniques to address complex research problems. For all of these reasons, we feel that virtual NMR facilities should be seen not only as a cost-saving measure, but as additional means, along with higher field magnets, better probes, and new pulse sequences, to expand NMR's role as an integral and effective part of the nation's research infrastructure.

ACKNOWLEDGMENTS

REFERENCES

1. "National Magnetic Resonance Collaboratorium. A Report by the Committee for High Field NMR: A New Millenium Resource," (1998), http://www.magnet.fsu.edu/NMRcoll/.

2. K. Pervushin, R. Riek, G. Wider, and K. Wuthrich, Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biomolecules in solution, *Proc. Natl. Acad. Sci. USA* **94,** 12366–12371 (1997).

3. These include NMRFAM in Wisconsin, NHMFL in Florida, the Resource for Solid-State NMR of Proteins in Pennsylvania, the Center for Magnetic Resonance at MIT in Massachusetts, and recently our Environmental Molecular Sciences Laboratory User Facility at Pacific Northwest National Laboratory (PNNL) in Washington, among others listed at http://micro.ifas.ufl.edu/facilities_s.html.

4. EMSL High Field Magnetic Resonance Facility, http://www.emsl. pnl.gov:2080/docs/msd/mrf_guide98/index.html.

5. G. M. Olson, D. E. Atkins, R. Clauer, T. A. Finholt, F. Jahanian, T. L. Killeen, A. Prakash, and T. Weymouth, The Upper Atmospheric Research Collaboratory (UARC), *Interactions* **5,** 48–55 (1998).

6. The Remote Experimental Environment, http://www.fusionscience. org/collab/REE/.

7. Diesel Combustion Collaboratory, http://www-collab.ca.sandia. gov/Diesel/ui/.

8. Telepresence Microscopy and Advanced Analytical Electron Microscope Collaboratory, http://tpm.amc.anl.gov/.

9. Committee on a National Collaboratory, "National Collaboratories: Applying Information Technologies for Scientific Research," National Academy Press, Washington, DC (1993).

10. F. F. Damberger, J. G. Pelton, C. J. Harrison, H. C. M. Nelson, and D. E. Wemmer, Solution structure of the DNA-binding domain of the heat shock transcription factor determined by multidimensional heteronuclear magnetic resonance spectroscopy, *Prot. Sci.* **3,** 1806–1821 (1994).

11. F. F. Damberger, J. G. Pelton, C. Liu, H. Cho, C. J. Harrison, H. C. M. Nelson, and D. E. Wemmer, Refined solution structure and dynamics of the DNA-binding domain of the heat shock factor from *Kluyveromyces lactis, J. Mol. Biol.* **254,** 704–719 (1995).

12. H. S. Cho, C. W. Liu, F. F. Damberger, J. G. Pelton, H. C. M. Nelson, and D. E. Wemmer, Yeast heat shock transcription factor N-terminal activation domains are unstructured as probed by heteronuclear NMR spectroscopy, *Prot. Sci.* **5,** 262–269 (1996).

13. C. J. Harrison, A. A. Bohm, and H. C. M. Nelson, Crystal structure of the DNA binding domain of the heat shock transcription factor, *Science* **263,** 224–227 (1994).

14. G. W. Vuister, S.-J. Kim, A. Orsoz, J. Marquardt, C. Wu, and A. Bax, Solution structure of the DNA-binding domain of *Drosophila* heat shock transcription factor, *Nat. Struct. Biol.* **1,** 605–613 (1994).

15. G. W. Vuister, S.-J. Kim, C. Wu, and A. Bax, NMR evidence for similarities between the DNA-binding regions of *Drosophila melanogaster* heat shock factor and the helix–turn–helix and HNF-3/ forkhead families of transcription factors, *Biochemistry* **33,** 10–16 (1994).

16. R. T. Kouzes, J. D. Myers, and W. A. Wulf, Collaboratories: Doing science on the Internet, *IEEE Comput.* **29,** 40–46 (1996).

17. J. D. Myers, N. Chonacky, T. H. J. Dunning, and E. R. Leber, Collaboratories: Bringing national laboratories into the undergraduate classroom and laboratory via the Internet, *Council Undergrad. Res. Q.* **17,** (1997).

18. A. Schur, K. A. Keating, D. A. Payne, T. Valdez, K. R. Yates, and J. D. Myers, Collaborative suites for experiment-oriented scientific research, *Interactions* **3,** 40–47 (1998).

19. EMSL Collaboratory, http://www.emsl.pnl.gov:2080/docs/collab/.

20. NCSA Habanero, http://havefun.ncsa.uiuc.edu/habanero/.

21. S. McCanne, Scalable multimedia communication using IP multicast and lightweight sessions, *IEEE Internet Comput.* **3,** 33–45 (1999).

22. CuSeeMe videoconferencing for Macintoshes and PCs, http://www. wpine.com.

23. DOE 2000, http://www.mcs.anl.gov/DOE2000/.

24. Secure shell software (ssh), http://www.cs.hut.fi/ssh/.

25. D. Walther, WebMol—a Java based PDB viewer, *Trends Biochem. Sci.* **22,** 274–275 (1997).