

Dartmouth College

Bionet Training Schedule - Novice

August 7, 1986

9:00-10:15	Introduction - Overview of Bionet System: Login, System Commands, Mail, File structure, Databases
10:00-10:30	Break
	Overview of Programs:
10:30-10:45	GENED - Sequence Data Entry and Editing GEL - Sequencing Gel Management Program
10:45-11:15	SEQ - DNA Sequence Analysis Program PEP - Protein Sequence Analysis Program
11:15-11:35	SIZER/MAP - Restriction Enzyme Fragment Sizing and Mapping CLONER - Recombinant DNA Simulation System
11:35-12:00	QUEST - Database Similarity Searching IFIND - Database Similarity Searching

Novice Training Continued

12:00-1:00 Lunch

Hands on session:

1:00-2:15 GENED and GEL Programs

2:15-3:15 SEQ and PEP Programs

3:15-3:30 Break

3:30-4:15 SIZER and MAP Programs

4:15-5:00 QUEST and IFIND Programs

Dartmouth College

BIONET Training Schedule - Advanced

August 8, 1986

9:00-10:30	GEL - Searching and eliminating vector sequences; Semi-automatic vs. automatic merging
10:30-10:45	Break
10:45-12:00	CLONER - Simulation of the construction of pUC9
12:00-1:00	Lunch
Hands on session:	
1:00-2:30	PEP - Comparison of the Search and Align algorithms for protein sequence homology searching; setting chemical similarity matching for homology searches
2:30-3:15	QUEST - Searching using complex keys
3:15-3:30	Break
3:30-5:00	IFIND - Similarity searching between a trans- lated portion of DNA and a QUEST retrieved portion of the NBRF database

Stanford University

Bionet Training Schedule

August 27, 1988

The morning session will be geared to the novice user:

9:00-10:15	Introduction Overview of Bionet System: Logging on, System Commands, Mail, BBoards
10:15-10:30	Break
10:30-12:00	Directories, File Structure and Location; Xsearch and Find
12:00-1:00	Lunch
1:00-2:15	GENED - Sequence entry and use of ESEQ editor
2:15-3:30	SEQ - Restriction enzymes site searching
3:30-3:45	Break
3:45-5:00	PEP - Designing probes with PEP; Hybrid protein construction and hydropathicity analysis

Stanford University
BIONET Training Schedule cont'd
August 28, 1986

9:00-10:30	Sequence Alignment Algorithms
10:30-10:45	Break
10:45-12:00	Database Searches (QUEST/IFIND)
12:00-1:00	Lunch
1:00-2:30	GEL - Sequencing Gel Management Program
2:30-3:45	SIZER/MAP - Restriction Enzyme Fragment Sizing and Mapping
3:45-4:00	Break
4:00-5:00	CLONER - DNA Cloning Simulation

Stanford University
Bionet Training Schedule - cont'd
August 29, 1986

9:00-12:00	Advanced Topics Including: QUEST Searching using complex keys IFIND similarity searching using a QUEST retrieved portion of a database.
12:00-1:00	Lunch
1:00-3:00	Editors, Batch Jobs
3:00-5:00	File transfer; up and downloading of files and programs between PC's and Bionet

VIII. BIONET APPLICATION

BIONET™

Dear Researcher:

You are invited to apply for access to the BIONET™ National Computer Resource for molecular biology. Enclosed are a description of BIONET, an application form, and order form for BIONET documentation.

The BIONET Resource is a central computer facility serving the computational needs, for both research and communication, of the molecular biology community. The Resource is funded by a five year, cooperative agreement with the Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health. IntelliGenetics™, Inc. of Mountain View, California will provide the computer facilities, core software, and support. Responsibility for overseeing the Resource rests with a National Advisory Committee (NAC), comprised of Drs. Joshua Lederberg (Chair, Rockefeller), Saul Amarel (Rutgers), Fotis Kafatos (Harvard), Allan Maxam (Harvard Medical School), Thomas Rindfleisch (Stanford), Richard Roberts (Cold Spring Harbor), and Charles Yanofsky (Stanford).

The BIONET Resource has three goals:

- To provide computational assistance in data analysis and problem solving for molecular biologists and researchers in related fields.
- To serve as a focus for development and sharing of new software tools.
- To promote collaboration and rapid sharing of information among a national community of scientists.

Please read the enclosed User Agreement closely. By signing it, you will be agreeing to adhere to both the letter and the spirit of the guidelines described.

Each principal investigator must complete an application to be eligible to use the BIONET Resource. Access cannot be passed on from one principal investigator to another. Each scientist who qualifies for and currently has his or her own source of funding is considered a principal investigator.

Please type the information on your application form for legibility and accurate processing. Processing time will take approximately four weeks after receipt of your application.

If you are applying from a commercial or foreign organization, be sure that your application contains sufficient supporting material to allow the National Advisory Committee to make its judgements.

If your application is approved, we will send you a welcome notification, the "Introduction to BIONET" documentation, and instructions for logging on the BIONET

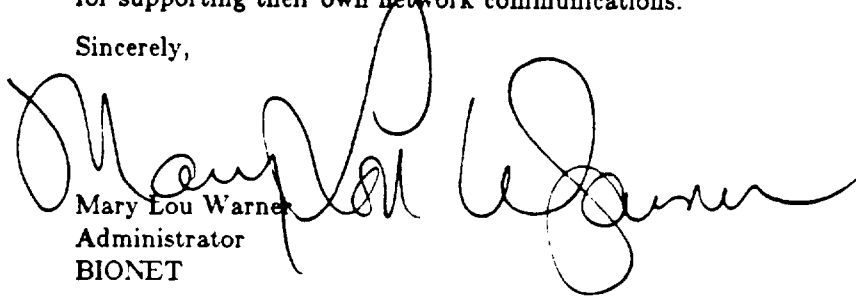
computer via the UNINET telecommunications network. We will also provide initial on-line training at your convenience.

Communications is a critical component of the BIONET Resource. On approval of your application, we will send you information on using Electronic Mail, Electronic Bulletin Boards, and File Transfer programs. These features will allow you to exchange information and ideas instantly with the BIONET staff and other users.

An annual fee of \$400 is currently being charged to all U.S. users. This fee, which covers a portion of our telecommunication charges for Uninet access, is your total cost for the BIONET Resource. Aside from the manuals, there are no other charges for this service.

Foreign investigators, including Canadians, on the BIONET system will be responsible for supporting their own network communications.

Sincerely,

A large, stylized handwritten signature in black ink, appearing to read "Mary Lou Warner". The signature is written in a cursive, flowing style with large loops and flourishes.

Mary Lou Warner
Administrator
BIONET

APPLICATION CHECKLIST

- Provided user information (page 1)
- Completed INTENDED USE OF BIONET and current grant support statement (page 3)
- Marked DRR Scientific Classifications (page 4)
- BIONET User agreement read and signed by Principal Investigator and responsible grant administrative officer (page 6)
- Filled out documentation order form (if desired)
- Copy made for your records.

Mail the completed application to:

BIONET Application
IntelliGenetics, Inc.
1975 El Camino Real West
Mountain View, CA 94040

Incomplete applications cannot be processed and will be returned. Please send all inquiries about this application to the above address. Include your name, phone number and application date in all correspondence.

Applications are processed once a month. The cut-off date is the 20th. Applications received on or after that date will be processed the following month.

Application Form for the BIONET™ Resource

See reverse for description and eligibility of user classifications

Date of Application:

Principal Investigator (full name and title):

Affiliation: Department, School and Institution:

Mailing Address (Include a Street Address for parcels shipped UPS):

Area code and phone number:

Applying for Class I, II, III or IV Use?: _____ (See Reverse for more information)

Would you like more information on the BIONET Satellite program?: _____ (See Reverse for more information)

Type of terminal or terminal emulator to be used:
(example: Tektronix 4023 or IBM-PC with VT100 emulator)

Type of communications software to be used:
(example: KERMIT or Smarterm) (note: KERMIT, a public domain communications software, is available from our lending library. Please indicate if you would like to borrow a disk for copying.)

Additional users: List up to 5 BIONET users under your direction. (Each Principal Investigator is allocated a fixed amount of space on the computer and only one user in a group can be logged in at one time.) Highlight the primary contact person for your group if not yourself. *NAC acceptance rules require individual qualifying PI's to apply separately. Additional PI's listed on this application will not be given access.*

Name	Title	Phone	Position
------	-------	-------	----------

Criteria for Eligibility

The four classes of user status are described below. Most users will be Class I users or IV users. Please call or write if you would like to be considered for Class II or III status.

CLASS I: Researchers from academic and non-profit institutions who can demonstrate that they are supported by governmental, philanthropic, or unrestricted institution funds and that their research can be assisted by the resource facilities. Exceptions will be considered on a case-by-case basis.

These users will have access to the programs in the Core, Database, and Contributed Libraries, and to the electronic mail and bulletin board facilities.

An annual fee of \$400 is charged for this access. *PI's in foreign countries, including Canada, will not pay the subscription fee but must pay their own telecommunication costs.*

CLASS II: Scientists who wish to participate in developing the BIONET Resource by providing new programs to the community. Acceptance as a Class II user is determined in part by the relevance of their programs. These programs should help achieve the goals described in the cover letter.

Class II user must meet the same eligibility requirements as the Class I users. However, they will also receive support from the BIONET staff in developing and making their contributed software accessible to the Bionet community. The Class II user will not be required to pay the subscription fee.

Please include a description, in detail, of what you intend to contribute, what support you will need from the resource and how the work will benefit the BIONET community. Also include a list of current publications in the area of intended use (for the past two years only).

CLASS III: People responsible for Department, School or Campus-wide computer facilities who wish to provide information about or access to BIONET to the community they serve. These users must provide evidence of their position and responsibilities for providing computer facilities for a local community of scientists with access to BIONET.

CLASS IV: Scientists who wish to take advantage of only the electronic communications facilities - electronic mail, bulletin boards, and file transfer programs - will be given restricted access for an annual fee of \$100. These users must meet the eligibility requirements of the Class I user.

BIONET SATELLITE PROGRAM

In addition to the above classes, BIONET, in cooperation with IntelliGenetics, is now able to offer an on-site BIONET package. Utilizing existing Digital Equipment VAX or 2060 computers, or SUN Microsystems, all of the programs, bulletin boards and electronic mail functions would be accessible at your location. Your local scientific community would benefit by having a direct and more powerful access to the resource. Accessing this service requires the purchase of a software license from IntelliGenetics. A special purchase program has been arranged to make it easy for academic institutions to join the BIONET Satellite program. If you are interested, please contact us directly or mark the appropriate response on the reverse.

Intended use of BIONET. Include a Research Title of 80 characters or less, and a Research Abstract with a minimum of 3 lines and a maximum of 350 characters. Class II and III applicants, in addition, should include additional information described in the Criteria for Eligibility on page 2. *You may attach a separate sheet if you prefer.*

Current grant support in area of intended use. Include each federal grant by Principal Investigator, title, funding institution, grant number and duration of support and a brief (three to ten line) abstract of the research. If funding is from institutional or other unrestricted funds, provide information on sources of funding sufficient for the NAC to determine if conditions for access have been met. *If this funding is scheduled to end within 12 months, state whether a renewal of the same grant/funding is pending.*

BIONETtm User Agreement

- The BIONET resource will not be used for any commercial purpose which is not specifically identified to and approved by BIONET's National Advisory Committee (NAC). Any pertinent change in sponsorship, continuity of grant support, or use made of BIONET will be reported promptly to the BIONET Resource Manager.
- The NAC will approve all access and will make the final judgment on applications that are questionable in nature, scope, or funding of research.
- Standard DEC-2060 facilities for file protection will be available to protect the integrity of your data and programs.
- Ownership of data and software developed on or contributed to the Resource will be subject to the guidelines of the Principal Investigator's institution and granting agency, to which all questions on legal issues should be directed. The BIONET Resource will retain a non-exclusive, royalty-free right to use, by approved BIONET investigators, of the data and executable versions of the software on BIONET.
- All investigators granted BIONET access must provide brief annual summaries of research results. The summaries must be included in our annual report of Resource activities to the NIH. Investigators will have sufficient advance notice to prepare the summaries.
- All publications that involve use of the Resource must acknowledge the Resource by name and NIH grant number (e.g.: *Computer resources used to carry out our studies were provided by the BIONETtm National Computer Resource for Molecular Biology, whose funding is provided by the Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health, Grant #1 U41 RR-01685.*) Investigators must send three (3) copies of these publications to the Resource Manager.
- Access to BIONET will be granted to a Principal Investigator and designated members of his or her research group. Each group will be allocated a fixed amount of disk storage space distributed by the PI and designated associates. Class II users will be granted larger amounts of disk space.
- We request that each PI limit access of his or her group to one login to BIONET at a time. Use of the Resource will be carefully monitored by the staff and the NAC.
- The BIONET Resource provides only a computer facility and associated services. It does not provide research equipment. The Resource has a small fund for fostering collaborations and will use this fund, when no other means are available, to support an effort that will advance the goals of the Resource.

I assume full responsibility for all users listed on this application form and will monitor their compliance to the conditions and restrictions for access to the BIONET Resource. I will inform the BIONET Consultant, (electronic mail address BIONET), by electronic mail, immediately about any changes in this group of users, i.e., departure of an existing user or addition of new staff qualified to use the Resource. I will inform new users of the above mentioned conditions and restrictions.

As Principal Investigator of this grant to use the BIONET Resource, I agree, by signing this application, to adhere to all conditions and restrictions for use of the BIONET Resource, as described above and such further regulations as may be issued from time to time by the NIH or the NAC.

Signature of Principal Investigator: _____

Date: _____

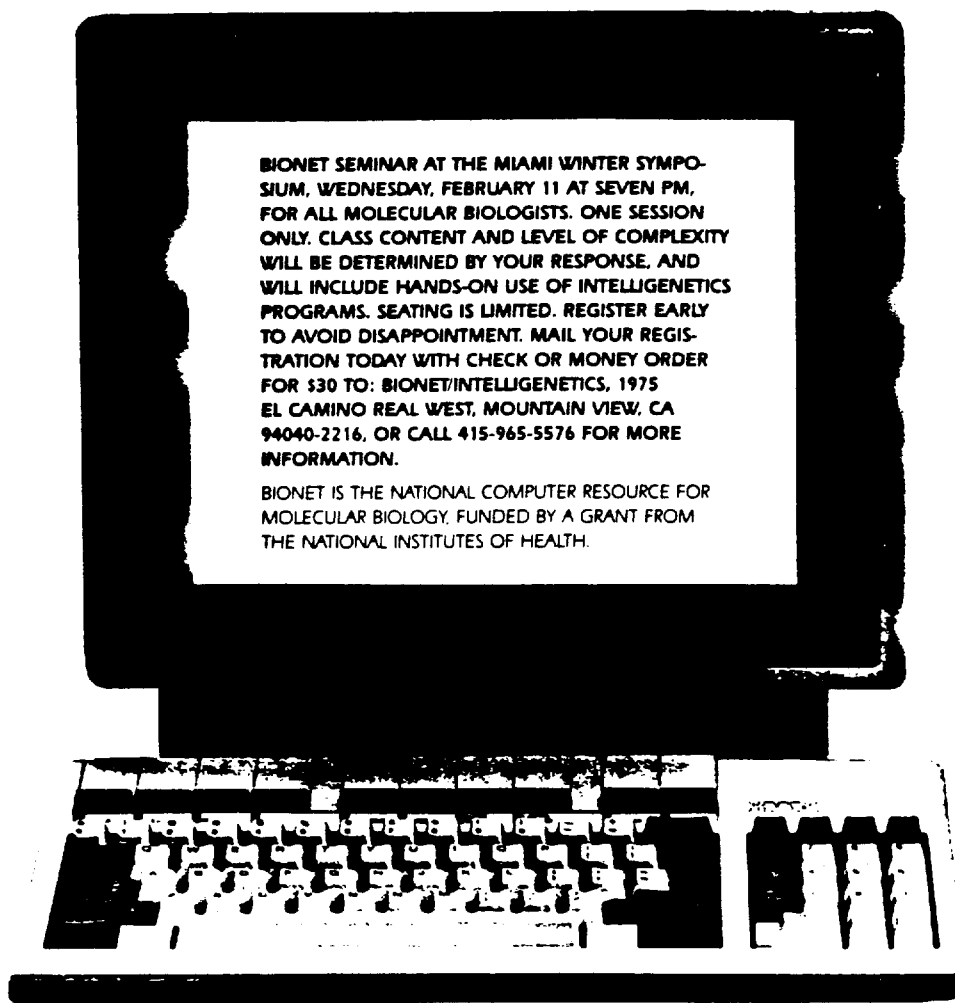
I have also furnished a copy of this application to the responsible grant administrative officer of my institution, whose name and signature are given below:

Name of official: _____

Signature: _____

IX. ADVERTISEMENT FOR BIONET TRAINING SESSION

If you are a **MOLECULAR BIOLOGIST**
you may be eligible to join **BIONET™**...



BIONET SEMINAR AT THE MIAMI WINTER SYMPOSIUM, WEDNESDAY, FEBRUARY 11 AT SEVEN PM, FOR ALL MOLECULAR BIOLOGISTS. ONE SESSION ONLY. CLASS CONTENT AND LEVEL OF COMPLEXITY WILL BE DETERMINED BY YOUR RESPONSE, AND WILL INCLUDE HANDS-ON USE OF INTELLIGENETICS PROGRAMS. SEATING IS LIMITED. REGISTER EARLY TO AVOID DISAPPOINTMENT. MAIL YOUR REGISTRATION TODAY WITH CHECK OR MONEY ORDER FOR \$30 TO: BIONET/INTELLIGENETICS, 1975 EL CAMINO REAL WEST, MOUNTAIN VIEW, CA 94040-2216, OR CALL 415-965-5576 FOR MORE INFORMATION.

BIONET IS THE NATIONAL COMPUTER RESOURCE FOR MOLECULAR BIOLOGY, FUNDED BY A GRANT FROM THE NATIONAL INSTITUTES OF HEALTH.

Clip or copy, and mail with check or money order for \$30 to: BIONET/IntelliGenetics, 1975 El Camino Real West, Mountain View, CA 94040-2216

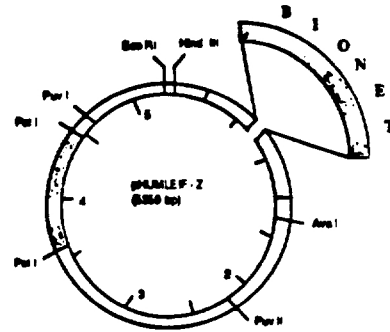
Investigator _____
Institution _____
Address _____
City _____ State _____ Zip _____

- New user Advanced user
 I am interested in hosting a session at my institution

- Please check topics of interest to you
- Managing large DNA sequencing projects
 - Restriction mapping tools
 - Simulation and design of recombinant DNA experiments
 - DNA or protein sequence database organization & search methods
 - DNA or protein sequence analysis
 - Sequence comparison methods

X. RENEWAL NEWSLETTER

**DO YOU KNOW
WHAT
YOU'RE
MISSING?**



SECOND YEAR OF BIONET IS GREAT SUCCESS!!!

**ITS TIME TO
RENEW YOUR
SUBSCRIPTION
NOW!**

BIONET has entered its third year stronger than ever. In keeping with its projected schedule, the BIONET staff has:

- added more databases
 - brought in contributed software
 - expanded the Bulletin Boards
 - upgraded existing programs
 - doubled the number of telecommunication ports
 - established a training program
- and more . . . **WITHOUT INCREASING THE SUBSCRIPTION FEE.**

We are excited about the growth and changes in BIONET and hope you are too. For more information on any of the above enhancements to the Resource, please give us a call . . . or better yet, log-in and check it out yourself!

ANNOUNCING . . . INTELLIGENETICS BECOMES JOINT VENTURE

The Amoco Corporation of Chicago has purchased a controlling interest in IntelliGenetics from IntelliCorp, Inc., of Mountain View, making IntelliGenetics a venture jointly owned by the two companies.

IntelliGenetics will continue to market its current line of molecular biology programs and will maintain its traditional emphasis on customer support.

This relationship with Amoco will provide IntelliGenetics with greater resources for the development of new software. There are plans to add several new programs to the software that runs on the SUN workstation, the VAX minicomputer, the microVAX II, and the timesharing system.

The most sophisticated new software will be Strategene, a genetic engineering workstation based on artificial intelligence technology.

Molecular biologists at the Amoco Research Center have been working with knowledge engineers at IntelliCorp for the past two years to apply IntelliCorp's KEE™, an integrated package of AI tools, to problem solving in molecular biology. The success of this work led to the formation of the joint venture with IntelliCorp.

Equipped with a mouse and windows, Strategene lets scientists rapidly simulate complex cloning experiments in a graphical environment.

The accessibility of DNA information and the ease and accuracy of simulations make it possible for scientists to experiment with a much larger number of vectors than they would ordinarily use.

Strategene uses AI techniques to organize knowledge about DNA molecules. This knowledge encompasses both descriptive information and rules for reasoning about cloning experiments. The system contains a reference library of vectors and allows researchers to enter and retrieve information from individual and laboratory libraries of constructions.

Strategene is designed to operate in conjunction with IntelliGenetics' package of analytic software. The system currently runs on a Xerox 1186.



**GET YOUR UPDATED
INTRODUCTION TO
BIONET FREE WITH
YOUR SUBSCRIPTION
RENEWAL. ACT
NOW!**



INTELLIGENETICS ANNOUNCES . . .

PC/GENE

A Personal Computer Genetic Engineering Environment

PC GENE is a comprehensive package of molecular biology software for microcomputers. It contains almost fifty different programs for analyzing peptides and nucleic acids.

You can use PCGENE as the perfect companion to BIONET, or you can run it independently. Data can be transferred efficiently through a modem connection. This allows you to perform large database searches and sequence comparisons on BIONET. At the same time you can take advantage of PC convenience and graphics capabilities to run a host of different analyses in your own laboratory.

Thus, BIONET subscribers can still get the speed and memory of a large computer when they need it.

PCGENE microcomputer software comes with the same high level of support you have come to expect from IntelliGenetics. Our scientific account representatives will offer the same degree of personal service for this new software package.

PCGENE allows biologists with little computer experience to use the programs productively in a matter of minutes. The system presents a series of choices of analyses that are expressed in terms that biologists use. It is necessary only to click the mouse or press a single key to choose a sequence to analyze, to define parameters, or to display the results in a variety of ways.

ALL UNINET DIAL-UP PHONE NUMBERS CHANGING IN SEPTEMBER

Uninet is being absorbed into GTE Teletel to form US Sprint Teletel. This means that the phone numbers to access the BIONET central resource will change.

Additionally, we regret that there will be a light change in the procedure used once you dial up. This change will occur in September. Each PI will receive a special mailing in August with all the details. Information will also be available on BIONET via the sign-on banner.

As a positive benefit of this combined network, US Sprint Teletel will have access numbers in over 50 new local dialing areas. A database of access numbers is available on Teletel to all users.

We are working to arrange a two week overlap when both the old and the new access methods will work. The Uninet dial-ups will be in service for 6 weeks following the change, but BIONET will not be accessible through them. Please consider this if you will be out of touch with BIONET for a month or more. Starting in September, access via the old UNINET dial-ups will produce an error message.

We will try to make the transition as smooth as possible. In the event of any problems reaching BIONET electronically, you can telephone the consultant at 415 324-GENE for assistance.

Some of the analyses that PC/GENE performs on peptides are:

- Computing best oligonucleotide probe
- Predicting antigenic determinants
- Searching for peptide subsequences
- Comparing sequences using the Needleman Wunsch algorithm
- Aligning two sequences
- Determining secondary structure using the Chou and Fasman or the Garnier method
- Predicting membrane associated alpha helices
- Plotting local concentrations of amino acids
- Calculating statistics of usage of di and tripeptides
- Plotting a protein's hydrophobic index

Some of the analyses that PC/GENE performs on nucleic acids are:

- Displaying tRNA in a clover leaf configuration
- Translating sequences
- Translating introns and exons using EMBL database annotations
- Searching for subsequences in nucleic acids
- Searching for coding regions using both Fickett's and Shepherd's methods
- Finding restriction sites, altering restriction enzymes lists, digesting sequences
- Creating a restriction site with a single mutation
- Comparing sequences with the Pastell dot matrix method
- Searching for hairpin loops
- Analyzing nucleic acid sequence statistics: codon usage, local base concentrations, enriched sequences

BIONET™ National Computer Resource for Molecular Biology is funded through a cooperative agreement with IntelliGenetics, Inc., by the Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health. Grant Number RR01685

IntelliGenetics, Inc. is located at 1975 El Camino Real West, Mountain View, CA 94040. Phone 415 965-5575

ELECTRONIC BULLETIN BOARDS

by Nancy Bigham

One of the three major goals of the BIONET Resource is to promote collaboration and rapid sharing of information among a national community of scientists. The BIONET bulletin boards (bboard) meet this goal by providing a facility whereby BIONET users can exchange data, laboratory techniques and ideas with others in similar fields.

During the past 6 months, the BIONET Resource has been concentrating on updating and improving the bulletin boards.

There are now 29 bulletin boards - containing articles ranging from reviews of pc-communications software to an article about Fast Fourier Transforms and related algorithms for sequence analysis (MOLECULAR-

COME ALIVE!

EVOLUTION, message #12).

Prominent members of the BIONET community have been selected to be bulletin board leaders. They will provide the bulletin boards with the most recent and pertinent information. Under this new

With the participation of the community and the work of the board leaders, the boards contain more exciting and pertinent information than ever before. Please take time to view the boards in your field of interest and to contribute information to any of the boards. For more information about reading the boards and contributing messages, see your INTRODUCTION TO BIONET manual.

CURRENT BULLETIN BOARD LEADERS

Gene-Expression	William Sofer
Genomic-Organization	Steve Harris
Libraries	Larry Kedes
Molecular-Evolution	Dan Davison
PC-Software	Doug Brutlag
Plant-Molecular-Biology	Ronald Sederoff
Politics	Michelle Cimbala
RNA-Folding	Michael Zaker

leadership, a dynamic bulletin board community is being developed by encouraging a lively interchange of information, maintaining a vital resource of community news, and archiving outdated bulletins.

VECTORBANK UPDATED By Ellen Hartzler

VectorBank is IntelliGenetics' collection of maps of common vectors designed for use in the CLONER program.

It is easier to use because we have added two new files. VECTORBANK.LST is a list of all vectors available in VectorBank, and VECTORBANK.TXT is a description of some important features of VectorBank. As in previous releases of VectorBank, we have provided multiple maps for each vector. The following list of maps for PBR322 illustrates the file naming convention.

<u>File name</u>	<u>Map Description:</u>
PBR322_6CUT.MAP	All six-cutter sites.
PBR322_COH.MAP	All cohesive cutter sites.
PBR322_COM.MAP	All prototype sites.
PBR322_FLUSH.MAP	All flush cutter sites.
PBR322_UNQ.MAP	All unique cutter sites.

Please note that we have replaced the hyphen in the all names with an underscore, i.e., PBR322-6CUT.MAP is now PBR322_6CUT.MAP.

To obtain a vector map for use in CLONER, follow these procedures:

1. Find the vector you want by looking at VECTORBANK.LST. This is a listing of all the vectors in VectorBank.
2. Enter the CLONER program and LOAD your vector from VectorBank.

If you are unfamiliar with the CLONER program, work through the tutorial on CLONER in your BIONET training manual.

NEW TRANSLATION TABLES IN SEQ

by Terry Friedemann

A new addition to SEQ provides two different ways to alter the codon tables used in SEQ for translating a DNA sequence.

• You can directly edit the codon table that contains the standard genetic code with your particular codon changes and save those changes

• Or, if you use the translation tables supplied by the program, you can translate sequences using a different genetic code without having to do any editing.

To see examples for yeast mitochondria codon changes with both the new editable codon table option and one of the new translation options, log on and send your request to BIONET using Electronic Mail.



EVOLUTIONARY ORIGIN OF HEPATITIS B VIRUS AND RETROVIRUSES

by MaryJo Lawler

Dr. William Robinson, a Professor at Stanford University and one of the first researchers to join BIONET, and Roger Miller, a postdoctoral fellow, have used BIONET in their research on the molecular structure of Hepatitis B virus.

Initially, Dr. Robinson and Dr. Miller chose to study the secondary structure of the origin of replication of known hepadna viruses. In examining stable palindromes near the origin of replication, they discovered by computer manipulation that the regions flanking these palindromes were highly conserved.

In an effort to substantiate these findings, they performed several global searches through the Genbank and EMBL databases and found that not only were these regions conserved across hepadna viruses, but they were also present in type C retroviruses. As a result of further analyses in the lab, they gathered additional evidence which led them to suggest that HBV and retroviruses have a common

evolutionary origin.

The authors have stated that without the use of computer analysis software and access to a complete and up-to-date database, the question of the genetic relatedness of the hepadna virus and retrovirus families would not have been raised.

The investigators retrieved the hepadna virus sequences from the Genbank database using the QUEST program.

They performed the initial DNA homology and palindrome analyses with the SEARCH function of the SEQ program, and employed other SEQ commands to examine base-composition and to translate sequences.

Drs. Robinson and Miller discovered that the regions were conserved in 27 viral

DNA sequences by searching over the Genbank and EMBL databases using the IFIND program. Additional searching using IFIND and the SEARCH functionality of PEP demonstrated a high degree of homology between the HBV core protein and the retroviral P30 gag nucleocapsid protein. The investigators also used the PEP program for open reading frame analysis, hydrophobicity plots and secondary structure prediction.

Dr. Miller made extensive use of the electronic mail and bulletin board facilities on BIONET to trade unpublished hepadna virus sequences with several other labs on the system.

COMPUTERS HELP ANALYZE SHOPE FIBROMA GENE by MaryJo Lawler

Grant McFadden and Chris Upton, working at the University of Alberta, have used the BIONET Resource extensively for their work on the molecular organization of the Shope fibroma gene. They have submitted for publication a paper entitled "DNA Sequence Homology between the Terminal Inverted Repeats of Shope Fibroma Virus and an Endogenous Cellular Plasmid Species."

In their paper, Dr. McFadden and Dr. Upton discuss three research findings and suggest how they correlate: the presence of an extrachromosomal autonomous DNA species, its hybridization to Shope fibroma virus (SFV), and the exchange of genetic information between host cells and cytoplasmically replicating poxviruses. The investigators used BIONET exclusively for their computer analysis.

They used the GEL program extensively for sequence entry and assembly. Once supplied with the sequence, Dr. Upton used the SEQ program to study the inverted repeat regions of the SFV DNA and to analyze the cytoplasmic DNA molecules through restriction enzyme, base composition, open reading frame, and translation analyses.

The investigators' homology comparisons were done using the SEARCH function of the SEQ program. Additional homology searches using IFIND over the Genbank and EMBL databases showed no additional homologous sequences to the inverted repeat region of SFV, but revealed similarity between

continued on page 6

PEP'S DIGEST OPTION

DIGEST is a major new addition to the protein analysis functions in PEP which is designed to help you study proteins by rapidly simulating the action of a peptide digestion with proteases or by chemicals.

The program provides a list of commonly and not so commonly used proteases and cleavage chemicals. You can add to this list, or you can create an entirely different list. When you add a new protease, DIGEST allows you to place the cleavage site before or after the recognition site. DIGEST also accommodates proteases that cleave at more than one site.

Once you are satisfied with the list of proteases, you can run DIGEST, choosing one or more proteases from the list. The resulting digestion simulation shows the location, length, and molecular weights of the fragments.

There are several additional options open to you. You can ask to see a map of the cleavage sites or see amino acid composition data for the fragments. You can treat any of the fragments as if they were independent peptides and then analyze them with any of the PEP functions. You can also ask DIGEST to simulate a peptide fingerprint by asking the program to draw a plot of the molecular weight of the fragments versus their isoelectric points.

As in all IntelliGenetics programs, you can ask for on-line help. Before you begin the option, we recommend that you read the introduction after the DIGEST: prompt.

SIMPLE SEARCHES

by Doug Brutlag and Alan Engelberg

The computer operating systems from which you use IntelliGenetics programs provide several tools on the DEC 2060 for rapidly searching unformatted databases and text files such as sequence data files.

On the DEC 2060 the fastest and simplest tool is the FIND program which is good for looking for one or a few patterns in a single file. The FIND program has the SCOPE concept from QUEST in that it allows you to look for a pattern in a line, a paragraph, or a page. It permits a limited amount of ambiguity but only allows you to examine a single file at a time. Using FIND is analogous to looking in a phone book for a person's name.

The simplest and most common application of FIND is to type FIND WITHIN <line> <pattern> IN <filename>, (see example below) leaving out all the other qualifiers.

continued on page 7

Predicting Experimental Results

When you perform a restriction digestion of a newly cloned sequence and electrophorese the resulting fragments, CLONER can save a great deal of time by quickly and accurately predicting the possible digestion patterns. In the following brief example we show how you can determine the orientation of your clone. If your vector contains more than one potential insertion site, you can use the same procedure to determine into which site you've cloned the insert.

CLONER: load pbr322.com.map We will insert our fragment into pBR322

Reading file PBR322_COM.MAP ...

1. PBR322-COM (4363 N) C ; DEFINITION PLASMID PBR322 (E.COLI CLONING VECTOR)

CLONER: new

NEW allows us to enter the restriction information about the insert. If we had sequence information, we could create a restriction map in SEQ and then load that map into CLONER.

Name for new map: FragZ

Length of new map: 1480

Topology: linear

Enter as many new lines of comments as desired; End with an extra <CR>

; Contains GeneZ

; <CR>

Please enter each site name followed by its location.

Finish with a blank entry.

Site name and cut position(s): psti 1 1480

Site name and cut position(s): hmbi 245 750

Site name and cut position(s): notI 1200

Site name and cut position(s): <CR>

FragZ is map number 2.

Loading editor help text...

MapEdit: region

continued on page 6

@FIND WITHIN line actin_myosin IN nih.lst<CR>

FIND shows each line where a hit occurs.

ACAACTI ;AMOEBA (A. CASTELLANII) ACTIN GENE-1.

BOVACT1 ;BOVINE ACTIN MRNA, 5' END.

BOVACT2 ;BOVINE ACTIN MRNA, 3' END.

BOVPR1 ;Bovine prolactin (pri) mRNA.

Since FIND simply searching for a sequence of characters, it will report hits when that sequence appears in a larger sequence, i.e., it finds a hit on actin in the word prolactin.

BOVPR1P1 ;BOVINE PROLACTIN, 5' FLANK AND EXON 1.

BOVPR1P2 ;BOVINE PROLACTIN, 5' FLANK AND PARTIAL EXON 2.

CELECT1 ;CAENORHABDITIS ELEGANS (NEMATODE) ACTIN I GENE 5' END.

CELECT2 ;CAENORHABDITIS ELEGANS (NEMATODE) ACTIN II GENE 5' END.

CELECT3 ;CAENORHABDITIS ELEGANS (NEMATODE) ACTIN III GENE 5' END.

CELECTIV1 ;CAENORHABDITIS ELEGANS (NEMATODE) ACTIN IV GENE 5' END(SEG 1).

CELECTIV2 ;CAENORHABDITIS ELEGANS (NEMATODE) ACTIN IV GENE 5' END(SEG 2).

CELMYH ;CELEGANS MAJOR MYOSIN HEAVY CHAIN (UNC-54 I) GENE, 3' END.

Here FIND reports a hit on the second pattern, myosin

CELMYUNC ;CELEGANS MAJOR MYOSIN HEAVY CHAIN ISOZYME UNC-54 I GENE.

If you want to see the name of the file where the sequence is located, you simply leave out "within line" in the FIND command line. The default scope is paragraph and that makes it possible to see the file name.

@find actin_myosin in nih.lst

*ACA.NIH

> ACAACTI ;AMOEBA (A. CASTELLANII) ACTIN GENE-1.

* ACARRS8S ;A.CASTELLANII (AMOEBA) 5.8S RIBOSOMAL RNA.

* ACARR3S ;A.CASTELLANII (AMOEBA) 3S RIBOSOMAL RNA.

*BOV.NIH

> BOVACT1 ;BOVINE ACTIN MRNA, 5' END.

* BOVACT2 ;BOVINE ACTIN MRNA, 3' END.

The pointer ">" indicates the line with the matching string of letters

PREDICTING continued from page 5

Name for new region: GeneZ
 Region boundaries: 140,1320
 Fill character: (<CR>=>)_<CR>
 Polarity (<, | or >): (<CR>=>) >
 Region GeneZ is now on level 1
 MapEdit: quit
 CLONER: list
 1. PBR322-COM (4363 N) C ; DEFINITION PLASMID PBR322
 (E.COLI CLONING VECTOR)
 2. FragZ (1480 N) L ; Contains GeneZ
 CLONER: insert 2 1 psti *We simulate the insertion of FragZ
 into pBR.*

Name for new map: (<CR>=>PBR322-COM-FragZ)
 Retain comments from PBR322-COM? (Y, N, D, ?, or *) (<CR>=>Y) no
 Retain comments from FragZ? (Y, N, D, ?, or *) (<CR>=>Y) <CR>
 Enter as many new lines of comments as desired; End with an
 extra <CR>

This map is FragZ inserted into pBR at the pstI site.
 <CR>

PBR322-COM-FragZ is map number 3.

CLONER: edit 2

MapEdit: flip *Flipping the map of the insert will allow us to
 simulate a fragment inserted with the reverse
 orientation.*

Area to invert: all

MapEdit: quit

CLONER: insert 2 1 psti *We repeat the same insertion but in
 this map the orientation of the
 insert is reversed.*

Name for new map: (<CR>=>PBR322-COM-FragZ) pbzZ-flipped
 Retain comments from PBR322-COM? (Y, N, D, ?, or *) (<CR>=>Y) no
 Retain comments from FragZ? (Y, N, D, ?, or *) (<CR>=>Y) no
 Enter as many new lines of comments as desired; End with an
 extra <CR>

FragZ inserted in pBR in the opposite direction to map PBR322-
 COM-FRAGZ.

<CR>

pbzZ-flipped is map number 4.

*To determine the orientation of the insert, we simulate a
 digestion with the enzyme chosen to analyze the clones and see
 which digestion matches the experimental results. We could
 have run simulated digestions with a number of enzymes to see
 which would give us the most distinct results.*

CLONER: digest 3 bamHI

Enzyme	Site	Length	Enzyme	Site
BamHI	(376)	3482	BamHI	(3858)
BamHI	(4363)	1856	BamHI	(376)
BamHI	(3858)	505	BamHI	(4363)

CLONER: digest 4 bamHI

Enzyme	Site	Length	Enzyme	Site
BamHI	(376)	3968	BamHI	(4344)
BamHI	(4849)	1370	BamHI	(376)
BamHI	(4344)		BamHI	(4849)

*After we have electrophoresed the restriction digest fragments
 we need only compare the gel pattern to the two sets of patterns
 above to determine the orientation of the insert.*

SHOPE continued from page 4

the extracellular DNA and a family of cellular protease inhibitors.

In addition to having access to analytical programs, Dr. Upton is pleased with the opportunity to use electronic mail to communicate with other scientists. Like many other BIONET scientists, Dr. Upton had little computer experience prior to BIONET. He has since become very active in the bulletin board communities. Dr. Upton has traded codon usage tables with several other BIONET scientists and has become one of the community's Macintosh authorities. He is currently working with an investigator in New York, whom he met through interactions on BIONET, and they are setting up what they call a "personal network" for their collective analysis needs. He foresees using BIONET even more extensively than in the past, especially because the McFadden lab has sequenced 15 to 20KB since he began work in the group.

DID YOU KNOW

When you are editing a meld, you can save your edits in three different ways.
 •"SAVE gels" is the program default. Editing changes you have made are retained for the current session only and have not been written in the .pro file. If you lose your job either because the computer crashes or because there are telecommunications problems, then you will lose those edits.

•"SAVE files" saves them permanently. These edits cannot be lost.

•"Set autosave on" will automatically save your editing changes in the .pro file if you type "Set autosave on" after the "Medit:" prompt after the "Medit:" prompt.

SEARCHES cont. from page 5

(type FIND<cr> to see a description of these qualifiers). For example, NIH.LST is a Genbank file that contains a one line entry for each Genbank sequence. On this line appears the sequence name and the first line of comments. If you were searching for a word or two that you expected to appear in the definition line, you would search the file NIH.LST as shown on page 4.

FIND can also determine

whether VectorBank contains a particular vector. You can search the file vectorbank.lst, a list of all the vectorbank maps. To see if pUC is present, type FIND WITHIN line puc IN vectorbank.lst.

Many people are using QUEST to search for simple unambiguous keys in sequences or in comments. A much faster and simpler program called XSEARCH (see example below) will allow you to search databases for keys with no ambiguous letters.

To run the program, type

XSEARCH after the "@" prompt. XSEARCH is not as convenient as QUEST in that you cannot COLLECT hits nor can you control the output. However, it searches databases 10 to 50 times faster than QUEST and is useful for an initial screen if you don't need ambiguous bases. Once XSEARCH has reported the names of the files that contain exact hits you can use QUEST to search just these files and then COLLECT or print out the

SEARCHES cont. page 8

@XSEARCH

SUBSTRING search routine (compiled 11-Jul-80) ? for help

Files to search: @nih-primatc.fls<CR>

Files to search: (continued) : <CR>

Target 1) myosin<CR>

Target 2) actin<CR>

You can search for more than one pattern.

Target 3) <CR>

Equivalences: 1// <CR>

current expression: 1 V 2

V = or, i.e. 1 or 2 is noted as a hit.

Expression:

Create .PL files? NO/<CR>

Output goes to: * TTY: // <CR>

This sends the output to your terminal.

Type DEL or RUBOUT to abort any particular file search.

Searching <SEQUENCES>APE.NIH.8510

Searching <SEQUENCES>GCR.NIH.8510

Searching <SEQUENCES>HUM.NIH.8510

Searching <SEQUENCES>HUMA.NIH.8510

Searching <SEQUENCES>HUMA1.NIH.8510

Searching <SEQUENCES>HUMAC.NIH.8510

When XSEARCH finds a match it displays the file in which the match is located and the line in which the hit occurs.

(<SEQUENCES>HUMAC.NIH.8510 1.1) {actin}

; DEFINITION HUMAN BETA-ACTIN RELATED PSEUDOGENE H-BETA-AC-PSI-1 5'END.

(<SEQUENCES>HUMAC.NIH.8510 1.4) {actin}

; KEYWORDS ACTIN; PROCESSED GENE.

(<SEQUENCES>HUMAC.NIH.8510 1.12) {actin}

; TITLE STRUCTURE OF TWO HUMAN BETA-ACTIN-RELATED PROCESSED GENES ONE

OF

(<SEQUENCES>HUMAC.NIH.8510 1.19) {actin}

; SITE 420 1540 HOMOLOGOUS TO ACTIN READING FRAME

(<SEQUENCES>HUMAC.NIH.8510 2.1) {actin}

; DEFINITION HUMAN BETA-ACTIN RELATED PSEUDOGENE H-BETA-AC-PSI-1 3'END.

(<SEQUENCES>HUMAC.NIH.8510 2.4) {actin}

; KEYWORDS ACTIN; PROCESSED GENE.

Searching <SEQUENCES>HUMMY

Searching <SEQUENCES>HUMTB.NIH.8510

Searching <SEQUENCES>HUMTR.NIH.8510

(<SEQUENCES>HUMTR.NIH.8510 1.1) {myosin} Here is a hit with another target pattern.

; DEFINITION HUMAN NON-MUSCLE (FIBROBLAST) TROPOMYOSIN GENE.

Lines recognized = 190

String Matches Unrecognized Matches

1) "myosin" 3 0

2) "actin" 205 0

Letter case ignored ("Ab" = "aB").

Files with no matches: <SEQUENCES>APE.NIH.8510, <SEQUENCES>GCR.NIH.8510, ... <S

SEQUENCES>MNKR.NIH.8510.

68 files searched, 63 without matches.

*L

..DONE.. continue to start over

FINDING NEAR-RECOGNITION SITES WITH QUEST

by Jaya Carl

QUEST's flexibility makes it possible to search for a great variety of patterns in sequences. For example, QUEST can be used to design a key to locate sequences of bases that are one base away from being a restriction enzyme site and that, if changed, would not alter the translation of the sequence. The purpose of this search is to locate a place to introduce a new recognition site to easily identify positive clones.

In the keys described below we have developed patterns that search for a set of *naar*-EcoRI sites, but the same procedure can be used to find any *naar*-restriction site that does not alter the translation.

Since we don't want to alter the translation we must determine the frame in which we are making the change. Thus the first part of the key is:

ATU]G & (...){1,}

This key represents the MET start codon immediately followed by one or more triplets.

In order not to alter the translation of the sequence when we alter the single base that introduces the recognition site, we must take advantage of the degeneracy of the genetic code. The recognition site for EcoRI is GAATTC. We can make a change in the third base of a codon. The reading frame determines which base is the degenerate one. The first key is:

ATG & (...){1,} & GAGTTC.

In this key, the reading frame is such that the first G is the first base of a codon. The third base was changed from A to G because both of these codons code for Glu.

The next key is:

ATG & (...){1,} & GAATTT.

In this key the reading frame is the same as the one above except that we are making a change in the second codon, changing the codon from TTC to TTT, since both code for Phe.

However, there is no reason to assume this particular reading frame with regard to the recognition site. Instead of there being an even set of triplets between the start codon and the recognition site, there could be one or two additional bases. For one additional base the key is:

ATG & (...){1,} & () & GAATCC.

In this case the reading frame is shifted by one, so we want to look for gaATCc instead of gaATTc, since both ATC and ATT code for Ile.

The final key is:

ATG & (...){1,} & (.) & GAATC.

This key assumes the frame shift is 2.

For an example of the way to use this key, simply log on and send your request to BIONET, using the Electronic Mail.

IntelliGenetics also maintains a database of key patterns that you can use in QUEST to help identify various structural and consensus regions in nucleic acid and protein sequences. The files are located in the <IG> directory.

We have collected the following files. If you have written a useful key, we would be delighted to include it in the KeyBank library.

AA.KEY	Identifies codons for antigenic sites.
AACOMP.KEY	Identifies codons of complementary strand for antigenic sites
AMINO.KEY	Equates one-letter amino acid code with three-letter code.
GENE.KEY	Identifies open reading frames.
KEY1.KEY	Shows keys from Quest Help Topic KEY1-EXAMPLE.
KEY2.KEY	Shows keys from Quest Help Topic KEY2-EXAMPLE.
KEY3.KEY	Shows keys from Quest Help Topic KEY3-EXAMPLE.
KEY4.KEY	Shows keys from Quest Help Topic KEY4-EXAMPLE.
NAD.KEY	Identifies dinucleotide-binding region for peptides.
OPIOID.KEY	Identifies potential DNA encoding endogenous opioid activity.
PROMOTER.KEY	Shows suggested consensus sequences for prokaryotic promoters.
REST.KEY	Identifies prototype restriction enzyme recognition sequences.
SIGNAL.KEY	Identifies consensus for leader peptide cleavage site.
ZDNA.KEY	Shows potential Z-DNA purine-pyrimidine pattern.

SEARCHES cont. from page 7

CONTEXT around the hits. XSEARCH first prompts you for "Files to Search:" and you may respond with filenames containing wildcards or indirect filenames (HUM*. * or @NIH-PRIMATES.FLS). Then it prompts you for "Targets" which are just character strings to search for. If you specify more than one target XSEARCH then prompts you for a Boolean relationship between the targets and the default is to search for target 1 OR target 2 OR target 3 OR ...

XSEARCH next asks you for equivalences and the default (obtained by hitting carriage return) is to equate upper and lower case letters. If you wish to have XSEARCH search through sequence information rather than comments then you should type the letter A (with NO carriage return!) at the "Equivalences:" prompt, and when it asks you for an "equivalence file," type SEQUENCE.XSE. This file not only equates upper and lower case, it equates T's and U's and causes XSEARCH to ignore carriage returns, line feeds, tabs and other punctuation in sequences. It is equivalent to SEQUENCE SCOPE in QUEST. Otherwise XSEARCH works exclusively in LINE SCOPE. Try XSEARCH. Type a ? (with NO carriage return) at each prompt to find out much more about XSEARCH's capabilities and limitations.

XI. BULLETIN BOARD LEADER AD

HOW TO SAVE \$400 AND HELP BRING YOUR FIELD INTO THE COMPUTER AGE

The BIONET-NEWS bulletin board has messages posted which describe the variety of uses for the bulletin board system and file transfer facilities. These uses range from having a continuous on-line scientific meeting in your research area to sending manuscripts to colleagues in distant labs. The list could undoubtedly be extended by creative people. (See also HELP MEETINGS.)

To encourage expanded use of the communications facilities, particularly the bulletin board network, we are offering a

FREE ONE YEAR BIONET SUBSCRIPTION

to users who are willing to organize and lead a bulletin board. Bulletin board leaders should be actively engaged in research in the selected area of interest.

Leading a bulletin board would involve contributing items of interest to the board, encouraging other people in the research field to participate (leaders should have plenty of contacts!), monitoring incoming messages, archiving dated material, and finally submitting a brief year-end report on the bulletin board activity to BIONET. Renewal of the position would be subject to a yearly review by BIONET. We estimate that the work involved would occupy only a few hours each month, but some responsibilities could be delegated to other lab members.

Prospective leaders should submit a proposal via electronic mail to BIONET. The proposal should include a description of the suggested bulletin board along with an estimate of the number and potential activity of participants. The activity estimate could be gathered by e-mail contacts prior to submitting the proposal. The final selection of bulletin board topics and leaders will be made in conjunction with BIONET and its National Advisory Committee. Please contact your BIONET consultant at 415-324-4363 if you have any questions.

A list of current bulletin board topics and names of leaders can be obtained by typing HELP BB-LIST after the prompt. Some of the current boards need leaders. However, new topics are especially encouraged.

As more molecular biologists and biochemists become computer-literate, participation in the bulletin board system should accelerate. As activity increases, the leadership positions will grow in influence. This is your opportunity to get involved with a new communications medium at its inception!

Somebody will eventually lead your research field into the computer age. Why not make it you?