DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health

Division of Research Resources
Biotechnology Resources Program
Annual Progress Report
PART I, TITLE PAGE

1. PHS GRANT NUMBER: | 1 | U | 4 | 1 | R | R | 0 | 1 | 6 | 8 | 5 | — | 0 | 2 |

2. TITLE OF GRANT: BIONET: National Computer Resource for Molecular Biology

3. NAME OF RECIPIENT INSTITUTION: IntelliGenetics, Inc., A Subsidiary of IntelliCorp

4. HEALTH PROFESSIONAL SCHOOL (If applicable): N/A

5. REPORTING PERIOD:

5a. FROM (Month, Day, Year): | 0 | 3 | — | 0 | 1 | — | 8 | 4 |

5b. TO (Month, Day, Year): | 0 | 2 | — | 2 | 8 | — | 8 | 5 |

6. PRINCIPAL INVESTIGATOR:

6a. NAME: Ralph E. Kromer

6b. TITLE: President, IntelliCorp

6c. SIGNATURE:

7. DATE SIGNED (Month, Day, Year): 12/21/84

8. TELEPHONE (Include Area Code): | 4 | 1 | 5 | — | 8 | 5 | 3 | — | 5 | 5 | 4 | 1 |

1

Table of Contents

List of Figures

2. Part II. Description of Program Activities

2.1 Scientific Subprojects

We have discussed the preparation of this part of our report with Ms. Barbara Perrone of the Division of Research Resources, NIH. BIONET is distinguished from most other Research Resources by its very large user community; we must therefore treat service, training and dissemination of information differently from traditional approaches. This first report is made more difficult because we have only just begun providing access to the Resource to the first group of approved investigators. Thus, we must allocate most of the personnel and associated costs to administrative categories. In the future, we will be able to distribute such costs uniformly among the large numbers of users because we will provide very similar levels of service and training to all members of the BIONET community. For this annual report, we have agreed with Ms. Perrone to report our statistics as described below. For future reports, we have discussed with Ms. Perrone the possibilities of providing this information directly to the NIH in computer-readable form. We have adopted the following methods for reporting the three major categories of scientific subprojects:

- Core Research and Development - As discussed in Chapter 3, we have operated under a mandate from our National Advisory Committee to deemphasize this aspect of BIONET until the Service and Collaborative components were firmly established. Therefore, because we have performed no core research in this reporting period, we include no forms for this section.

- Collaborative Research and Service - We are currently reviewing the first set of proposals for access to BIONET for collaborative research (our methods for solicitation of applications from potential collaborators are discussed in Chapter 3, Section 3.5). Therefore, all usage reported in this category is for service use of BIONET. We report only that usage accrued from November 26, the first day of actual use of BIONET, through December 16, the last date that statistics were summarized before submission of this report. The actual numbers, names, and affiliations of approved principal investigators as of December 16 are given in Chapter 3, Section 3.1. In reporting this information, we have generated reports from our IBM PC-based database of information on applications and accepted investigators. Although these reports are not written directly on the forms provided, we have supplied all the necessary information in the order requested, one sheet per Principal Investigator. We report staff hours and BRTP funds allocated in summary form in the Resource Summary Table (see page 3 for a description of allocations).

- Training - We have begun to assemble an extensive training program designed to provide basic training to large numbers of investigators. Our first training will be held December 20, 1984. Because this training (and all future trainings) is not aimed at specific investigators, we cannot provide a meaningful subproject form for this effort. Rather, we have described our training program in narrative form in Chapter 3, Section 3.6, and provide a summary of staff hours devoted to the training programs in the Resource Summary Table.

2.2 Books, Papers and Abstracts

There are no books, papers or abstracts to report because BIONET has only just made the Resource available to its user community.

2.3 Resource Summary Table

For this reporting period, we have prepared the Resource Summary Table in the following way.

- Under numbers of subprojects, we have included *only* BIONET PI's that have used the machine, and totaled them under Collaborative Research and Service.

- There are no publications reported by staff or external PI's because the Resource has just begun operation.

- Under number of investigators, we have included the total number of users that have used BIONET so far, i.e., all PI's and their user group members (56), plus all staff members allocated to this category (8). We report this under Collaborative Research and Service. The two investigators listed under Training and the three under Administration are staff members allocated to those categories.

- Resource Technology

    o Usage Factor -- all staff and external PI's use the DEC2060 computer.

    o How Used -- We report total time used, as CPU and connect time (both in hours). for staff allocated to administration and training, and total time used, CPU and connect. for the remainder of staff *and* all PI user groups, under Collaborative Research and Service.

    o Resource Staff Hours -- We total resource staff hours in each of the indicated categories, separating out personnel time devoted to administration and training, and place the remainder in Collaborative Research and Service.

    Under funds allocated we include:

    o Administration/miscellaneous; direct and indirect costs for personnel plus capital equipment costs.

    o Training; direct and indirect personnel costs.

    o Collaborative Research and Service; direct and indirect personnel costs plus all other expenses not included in the above two categories.

The Resource fees collected are from charging for our major documentation, the *BIONET Reference Manual*.

3. Part III. Narrative Description

3.1 Summary of Research Progress

The BIONET Resource has three major scientific goals:

- Provide a symbolic computing facility to the entire national community of molecular biologists and researchers in related fields.

- Serve as a forum for the rapid exchange of scientific information among that community.

- Serve as a focus for development and sharing of new software tools.

These are ambitious goals to achieve with a limited staff. However, our work and planning are now paying off and we have established a firm base for the next several years of the BIONET Resource.

The first nine months of BIONET Resource funding have been devoted to the substantial tasks involved in configuring, acquiring, and testing the hardware, software, and ancillary components of the Resource. This effort has not been without difficulties, but these have been met and solved. This section will describe the progress that has been made in various areas in bringing the BIONET Resource to fruition. The culmination of this first phase of effort was on November 26, 1984, when the resource became operational with an initial user group of 63 approved Principal Investigators (PI's) encompassing 199 total user accounts. Detailed statistics on the total numbers of PI's and user accounts are given in Subsection 3.1.8. That we have received and responded to about 1100 requests for applications shows the level of community interest. We have provided a copy of *The BIONET Resource: Description and Application Form* as Appendix I of this report.

3.1.1 Hardware

A Digital Equipment Corporation 2060 computer was chosen as the best currently available time-sharing system for a resource of this size. The 2060 has a long history of successful performance at hundreds of major academic centers. In addition, the collection of associated software tools for facilitating the communications and collaboration goals of the BIONET Resource have reached a level of maturity unique to the 2060 system.

The BIONET Resource machine has been fully configured and tested. We believe the current disk space (see below) will adequately serve the needs of the first 400 BIONET Class I Principal Investigators (approximately 1000 user accounts). The adequacy of the processor itself remains to be determined. Obviously, when many concurrent users are running CPU intensive jobs, the system response begins to degrade. For the time being, however, the number of concurrent BIONET jobs is limited by the number of Telenet network ports available (see below, Subsection 3.1.4).

The hardware configuration is as follows (note that exactly half of the entire system is devoted to BIONET use):

Mainframe:

DEC-System 2060 FE. FCC Certified 36-bit time-sharing mainframe

   2.75 Megawords MOS memory
   128 Asynchronous terminal lines

Mass Storage:

  (2) DEC RP07 Winchester disk with 516MB each storage
  (1) DEC RP06 Removable disk with 167MB of storage
  (1) DEC TU78 Tape drive. Normal and high (1600, 6250 BPI)
       density, 125 inches per sec writing speed.

Printing Devices:

    DEC LP26 Band printer. 600 lines per minute.

    Imagen 8/300 Laser printer. 300 dot per inch, 8 pages
      per minute.

Communications Devices (Bionet specific):

  (6) VADIC 3451 "Triple" modems, 300 or 1200 baud.
  (1) GTE Telenet TP3010 Network Processor

Disk space on the 2060 is allocated as follows:

```
Total disk pages is 432,752   (two RP07's at 216,376 pages each,
                               and one RP06 at 76,000 pages)
Overhead/Common     (212,000) (Core, System and System Support Libraries)
Swapping space      ( 15,000)
                    ---------
                    281,752


BIONET ALLOCATED    140,876  (1/2 OF THE AVAILABLE AMOUNT)
```

     The 2060 is partitioned for Class Scheduling with windfall allocated as follows:

Overhead	10%
Bionet-Class 1	30%
Bionet-Class 2	10%
IntelliCorp (staff)	20%
Commercial	20%
Overhead	9%
Batch	1%

The Class Scheduler is so far doing an excellent job in partitioning the 2060, ensuring that all classes get their fair share, with unused cycles in one class available for other classes.

3.1.2 Software

BIONET software falls into two major classes that specifically designed for problem-solving in molecular biology, the Core Library and the Contributed Library; and the System Support Library, general purpose software for programming, text processing, electronic communications, etc. In the initial phases of the Resource, the first class of software consists only of the Core Library, the IntelliGenetics' programs that were described in detail in the original BIONET proposal. Since the date of the proposal acceptance, we have added programs and have made substantial improvements to existing programs. A major new release was made available to the community shortly after BIONET became accessible to the community. IntelliGenetics has agreed to provide all of those improvements to the BIONET Resource without additional charge. Summarized below are the nine programs and a description of major enhancements made since the original proposal.

3.1.2.1 The Core Library

During the first week of December, IntelliGenetics released the latest update, Version 4.0, of its molecular biology software. Version 4.0 contains many minor, transparent changes to improve function and correct known deficiencies in all of the core programs, and three programs, SEQ, PEP, and GEL, were extensively modified. These programs are all now available to the BIONET community.

Some of the improvements increase the ease with which users can perform more varied tasks. For example, in the SITE option of SEQ, the user can build a customized subset of restriction enzymes and save the subset as a private database. Also in the SITE option, the user can either specify the intersection of two subsets or the addition of two subsets of restriction enzyme classes. For example, he or she can specify a list containing all flush-cutters and all 4-cutters, or a list of all flush-cutters which are also 4-cutters. Both of these enhancements makes it easier for the user to create restriction enzyme databases for their personal needs.

SEQ uses a new, faster algorithm for finding the occurrences of restriction sites. Now, large sequences can be analyzed more quickly. The display of large sequences can be graphically displayed, compressing what would be many lines of output into a single line, if desired.

We are meeting new analytical needs with the implementation of the SITE option. Restriction map output can display one or three frame translations, allowing easier and faster design of site specific mutagenetic experiments.

By giving the PEP program the capacity to handle the amino acid ambiguities, B (ASX) and Z (GLX), we have added user flexibility in amino acid analysis. We have increased experimental flexibility with the WINDOW option, which calculates any amino acid properties as arithmetic or geometric averages in windows along sequences, and with the match option, which allows the user to set match tables in

homology searches. For example, the user can compare sequences for similar distribution of hydrophobic, hydrophilic and charged residues.

For increased clarity in both PEP and SEQ, we have rewritten on-line help. The programs now display the selected options as part of the prompt wherever multiple options are possible.

The new version of GEL is radically new and has been redesigned to manage sequencing project that use either the Sanger dideoxy sequencing technique or the Maxam and Gilbert chemical technique. The old version could only properly handle chemical sequencing.

Through a number of user settable features, users can now tune GEL to reflect their sequencing methodology. For example, when set to dideoxy sequencing, the matching table's ambiguity code includes additional symbols useful for dideoxy sequencing. GEL search for restriction sites, such as insertion sites, and also can cut automatically at those sites to remove unwanted ligations. GEL can search for the occurrences of vector sequence in the gels and allows the user to edit them out. The automatic merging mode allows the user to quickly assemble good merges. Improved heuristics for finding melds results in the best possible matches being found faster.

GEL projects are no longer limited by the number of gels or the redundancy of overlap. The new meld editor gives the user great flexibility and control in realigning individual gels within a meld and can propagate changes throughout the overlaps at the users discretion. A data entry checking facility speeds data entry and looks for errors during entry.

- NINE INTELLIGENETICS BIOTECHNOLOGY PROGRAMS: CLONER, GENED, GEL, IFIND, MAP, PEP, QUEST, SEQ, SIZER

  o GENED--Edits Sequences
  GENED contains two nucleic acid and protein sequence editors: EDIT, for use with all terminals including printing terminals, and ESEQ, for use with video terminals. The editors accept protein sequences in one- or three-letter codes and nucleic acid sequences containing symbols for ambiguous bases.

  o MAP and SIZER--Produce Restriction Enzyme Maps
  MAP generates and displays all of the restriction enzyme maps that fit a set of restriction enzyme fragment size data. SIZER calculates restriction fragments lengths from fragment mobility data/ You can enter SIZER from MAP or directly from the operating system.

  o GEL--Assembles DNA Sequencing Data
  GEL assembles your DNA sequencing data into completed consensus sequences. It accepts and analyzes data from all popular sequencing methods, including the Maxam-Gilbert chemical method, the Sanger random fragment method, and the primer extension method.

  o SEQ--Analyzes Nucleic Acid Sequences

SEQ performs a wide range of analyses on nucleic acid sequences in your own data files (created in GENED or GEL) or on data from a sequence database.

o PEP--Analyzes Polypeptide Sequences
PEP performs a variety of useful analyses on protein sequences from your own data files (created in GENED) or from a sequence database.

o CLONER--Models Recombinant DNA Experiments
CLONER models complex recombinant DNA experiments and provides database management for your plasmid constructions. CLONER can help predict the feasibility of a recombinant DNA experiment.

o IFIND--Searches Databases for Similar Sequences
IFIND rapidly searches large databases for specified DNA or protein sequences, using the Wilbur and Lipman method, and then aligns the related sequences with the query.

o QUEST--Searches Databases and Retrieves Sequences
QUEST efficiently searches a database for any pattern of characters or keywords in headings, comments, references, or sequences. The power of QUEST lies in its ability to find ambiguous patterns.

3.1.2.2 The Contributed Library

This library currently is empty. Collaborators will contribute software to this library and, with the approval of the National Advisory Committee and the assistance of BIONET staff, move mature programs into the Core Library.

3.1.2.3 The System Support Library

The second class of BIONET software, the System Support Library, encompasses all of the programs necessary for general resource operation. The current list is modified somewhat from that shown in the initial BIONET proposal to reflect availability, improvements, costs, and the like. Summarized below are the major software tools currently available on BIONET.

- **Communication**

  o MM
  An electronic mail system that allows users to send, read, and manage their messages and read and write messages on public bulletin boards.

  o BBOARD
  Allows users to read public "bulletin boards" for common distribution of messages.

  o TALK
  Lets users communicate with each other via their terminals.

  o ADVISE
  Lets an expert simulate typing on the keyboard of another user.

- **Text Processing Programs**

o TEXT EDITING

- EMACS
  Powerful screen-oriented editor that can be customized.

- TVEDIT
  Simpler screen-oriented editor.

- TECO
  Character-oriented editor.

- SOS
  Simple line-oriented editor.

o PUBLICATION

- SCRIBE
  A powerful document formatter which produces high quality output on a wide range of printers.

- RUNOFF
  A simple document formatter which produces fair quality output on a limited range of printers.

- SPELL
  A spelling checker/corrector that can be used in conjunction with EMACS and MM.

o FILE SEARCH

- XSEARCH
  An extremely powerful and efficient search program; can be used to find occurrences of text strings within files.

o FILE COMPARISON

- SRCCOM
  Produces a list of differences between two text files.

- **File Transfer Programs**

o KERMIT
  Allows transfer of files over telephone lines to or from other computers that use KERMIT. KERMIT is supported on a wide range of computers including personal computers.

o MODEM
  Allows transfer of files over telephone lines to or from other computers that use MODEM. MODEM is supported on a wide range of computers including personal computers.

- **Standard Programming Languages**

o C - currently on order

o FORTRAN
o MAINSAIL
o BASIC
o PASCAL

3.1.3 Databases

Access to the rapidly growing primary data of the field of molecular biology is vital to the successful operation of the BIONET resource. That data consists of nucleic acid and protein sequences, in addition to information about restriction enzymes, cloning vectors, and so on. We have established formal working arrangements with the major nucleic acid sequence libraries, GENBANK and EMBL, to provide immediate access for BIONET users, and with the National Biomedical Research Foundation to provide access to their well-known protein sequence library. In addition, BIONET, in cooperation with Dr. Richard Roberts of Cold Spring Harbor Laboratory, provides an up-to-date database of restriction enzyme sequences and cleavage sites. Finally, IntelliGenetics makes available to BIONET a database containing proven and potential cloning vectors called Vectorbank. These databases will be tested for accuracy and updated frequently. New databases, as they become available and useful, will be provided to the BIONET community.

Below we provide some details on contents and size of each of the major currently available BIONET databases:

- The EMBL database contains 1481 sequence entries, comprising 1,654,863 nucleotides abstracted from nearly 1000 references. This database is updated approximately annually.

- GenBank, the Genetic Sequence Data Bank, contains 4393 loci, 3,689,752 bases from 5756 reported sequences. (Dec. statistics: 4526 loci, 3,813,396 bases, from 6019 reported sequences.) The GenBank database is updated monthly.

- The NBRF, Protein Sequence Data Bank, contains 2784 sequences and 557,759 residues. The NBRF database is updated quarterly.

- VectorBank contains 76 maps of 18 different frequently used cloning vectors. This database is updated twice a year.

- The Restriction Enzyme Library contains a complete list of restriction enzymes, derived from the annual Cold Springs Harbor compilation published in *Nucleic Acids Research*.

The BIONET community can access the restriction enzyme database directly or from the SEQ, PEP, and GEL programs. This database contains 400 isoschizomers and 109 prototype restriction enzymes. The restriction enzyme database is updated annually.

BIONET provides the nucleic acid (EMBL and GenBank) and protein (NBRF) databases in two forms, formatted and original. The formatted forms are compatible with the Core Library programs. The original forms appear as received from the database sources.

3.1.4 Network Communications

The single area that caused the most difficulty during the first nine months of BIONET operation was the choice and operation of the communications network. For reasons of cost and effectiveness of communications, we initially chose the CompuServe national network. Soon after the BIONET link with CompuServe was installed, it was apparent that an error had been made. Response time, which had been tested thoroughly when we were choosing a network and which at that time was eminently satisfactory, had degraded so seriously that it was as poor as 15 seconds between keystroke and echo in some locations. After extensive investigation, we determined that CompuServe load had expanded enormously over the last several months and they would not be able to remedy the problem for many months. Since we had no desire to subject BIONET users to such unsatisfactory performance, we canceled (at no cost to the BIONET Resource) our contract with CompuServe. After reexamining the major alternative networks, Tymnet and Telenet the Resource staff decided that Telenet offered the best cost/service proposal, and a Telenet link was established. This link was formally accepted as operational on November 8, 1984, and as indicated above, installation of Resource hardware was completed prior to acceptance and initiation of the first sets of PI accounts on BIONET.

Since that date, Telenet service has been stable and satisfactory. We have tested response time from all major nodes and are happy with the results. We have investigated configuration of network parameters to assure reliable, error-free communication as well as the ability to upload and download files to microcomputers. The network is now configured such that all operations should proceed transparently, i.e., no reconfiguration by the user is required to accomplish routine, interactive computing or file transfers.

Nine ports of the eighteen port Telenet communications processor are available for BIONET. Continual monitoring of the ease and reliability of networked communications will be a major task of the BIONET Resource staff. As longtime users of such facilities, we fully understand the necessity of keeping response-time and information transfer errors to an absolute minimum.

As discussed in the budget sections of this report, the network costs for use of Telenet will be substantially higher than anticipated. Thus, the network will be a limiting factor to the number of concurrent users we can support, both in terms of numbers of ports and their cost. We anticipate that the only way we will be able to expand this service will be to require users to pay for the costs of Telenet communications, leaving BIONET to pay for the rental of the hardware itself. This is an issue that will be addressed by our staff and the National Advisory Committee in the near future.

3.1.5 BIONET Documentation

Because of the enormous size of the user community and the very limited size of the BIONET Resource staff, excellent documentation will have to provide much of the user assistance that ideally would be available on a personalized basis. For that reason, the BIONET staff has spent hundreds of man-hours on a comprehensive introductory guide, the *Introduction to BIONET*, to the Resource that provides step-by-step instructions on resource utilization to a computer-naive molecular biologist. This guide is 66 pages long and is distributed at no charge to all new BIONET users. We have included a copy with this Annual Report. In addition, a comprehensive manual, the *BIONET Reference Manual*, for the programs currently in the BIONET Core Library is being made available to all BIONET users at cost. Finally, we provide the *BIONET Training Manual* when BIONET users participate in trainings. The training manual contains introductory material on the Resource's organization, the use of electronic mail and bulletin boards, and program examples which illustrate common problems in molecular biology. For other details of the training manual, refer to section 3.6.1.

3.1.6 Electronic Bulletin Boards

Since the rapid exchange of scientific information is one of the major goals of the BIONET Resource, we have established electronic bulletin boards to facilitate that exchange. Electronic bulletin boards almost instantaneously allow for the free flow of information on any numbers of topics.

Bulletin boards are easily set up by any member of the BIONET staff for the BIONET community as a whole. In addition, we encourage BIONET users to set up personal bulletin boards that would be of interest to smaller groups of users. These can be set up by any member of the BIONET community. Personal bulletin boards will reside on individuals' directories and be protected in such a manner as to give access to anyone in BIONET. Currently, no personal bulletin boards have been installed. The creation of such bulletin boards will be announced by mail or by posting to the BIONET-NEWS bulletin board.

3.1.6.1 Electronic Communities

All BIONET users are required to indicate in their applications a preference for at least one Electronic Community (bulletin board). The list of bulletin board preferences is added to the user's LOGIN.CMD file and read by the BBOARD command at each login. New messages and unseen messages automatically appear on the user's terminal each time they login. BIONET users may add or delete bulletin boards from their interest list by appropriately editing their LOGIN.CMD files. Our consultants are available for assistance in this process if necessary.

We are in the process of locating individuals interested in maintaining the bulletin boards of the Electronic Communities. Until suitable community leaders are identified, the BIONET staff will maintain the bulletin boards.

- Electronic Communities

    o DNA-RECOMBINATION
    o DNA-REPAIR
    o DNA-REPLICATION
    o DNA-SEQUENCING
    o GENE-EXPRESSION
    o GENOMIC-ORGANIZATION
    o IMMUNOLOGY
    o METABOLIC-REGULATION
    o PROTEIN-ENGINEERING
    o VECTOR-CONSTRUCTION

3.1.6.2 Service Bulletin Boards

Other bulletin boards, called Service Bulletin Boards, have been installed to assist new BIONET users in using BBOARD and MM (the STARTUP bulletin board), announce BIONET news items (the BIONET-NEWS bulletin board), and provide a summary of staff and their responsibilities to the resource (the BIONET-STAFF bulletin board). We will be installing other bulletin boards in response to the needs of the BIONET community. Already, the first of these, the PC-SOFTWARE bulletin board, describing useful tips on various topics related to personal computers, has been established.

- Service Bulletin Boards

    o STARTUP
    o BIONET-NEWS
    o PC-SOFTWARE
    o BIONET-STAFF

3.1.7 Application and Review Process

Applications are solicited (see Section 3.5 and Appendix I) from scientists identified as Principal Investigators on their existing funding. These applications are reviewed internally by BIONET staff, which refers any difficult applications (for instance, for-profit affiliations, lack of qualifications to be a PI) to the Executive Committee and the National Advisory Committee. To date, only a few applications remain deferred pending receipt of additional material, and only one application has been rejected (a request for Class I access from an employee of a commercial company). These applications request that the PI indicate Class I or Class II status.

As part of the application form, PI's can request that other members of their research groups also be named as BIONET users. When an application is approved, the PI and the named additional users each get an account on BIONET. The total amount of disk space allotted for each PI is fixed and must be shared among the PI and all named users.

3.1.8 Class I Users - The BIONET Service Component

We have differentiated between two classes of BIONET users. The first class constitutes the service component of the Resource. Investigators approved for access in this category use the Resource and its software to help solve their current research problems in molecular biology. These users receive limited disk space, and do not have access to staff to aid in program development. Otherwise, they have access to the full complement of software, databases, and consulting and training services.

The following table summarizes the history to date of accepted applications. The "Date Submitted" column refers to the date on which accounts were ordered from our computer resources group, based on approved applications. As the Table indicates, a total of 125 PI's have been approved for access to BIONET, comprising a total of 405 separate user accounts. To date over 1100 applications have been distributed in response to injuries requesting applications.

Date Submitted	PI Total	SUBI Total	Total Accts
10/31/84 \ 11/1/84 /	63	136	199
11/19/84	14	32	46
11/30/84	32	73	105
12/7/84	12	30	42
12/10/84	3	8	11
Grand Totals	124	279	403

3.1.9 CLASS II Users - The BIONET Collaborative Research Component

We have begun the process of establishing a group of Class II (collaborative) users on the BIONET Resource. These users will receive significantly greater computational resources (both CPU allocation and disk space) than Class I users and will be expected to contribute substantially to the Resource goal of serving as a center for the development of advanced software for molecular biology. We have sent invitations for research proposals to all of the active participants in this area known to the BIONET staff and co-investigators. The solicitation letter can be found in Section 3.5, and the list of its recipients in Appendix IV. We expect from 5 to 15 Class II collaborators to be selected from those sending in proposals. As agreed to at a National Advisory Committee meeting, the BIONET associate-investigators (Brutlag, Friedland, and Kedes) will serve as an initial peer review group for such proposals with the NAC itself serving as final review authority. For the time being, as soon as we receive an acceptable

application from a potential Class II collaborator, we grant Class I access until the review process for Class II can take place.

In the next year the BIONET staff and co-investigators will have the important task of nurturing these collaborative linkages. It is our intent to insure the dynamic growth of the BIONET community by providing a continually improving base of software to that community. A large percentage of that software will come from our Class II users.

3.1.10 Other Resource Activities

Apart from a brief description of the recruitment of collaborative researchers, this section of the Annual Report has mainly discussed activities involved in establishing the service component of the BIONET Resource. Later sections of this report will detail the substantial preparatory work that has gone into providing for the training and dissemination components of the Resource. Finally, planning for the core research component of the Resource has just begun; resource staff and the NAC had agreed that all staff time should be spent on service component and related activities until the current status of BIONET activity had been reached.

3.2 Highlights

Since the bulk of the last nine months has been spent on bringing the BIONET Resource to an operational status, there are no research highlights to report. We have described the milestones in preparing the hardware, software, and other components of the Resource in the previous section of this report. However, the enormous initial user community response to the resource--125 Class I Principal Investigators, 405 total user accounts, and over 1100 requests for applications as of December 17, 1984-- provides ample evidence of the need within the molecular biology community for the BIONET Resource.

3.3 Administrative Changes

Several major management manpower goals were achieved in BIONET in 1984 and several key positions underwent staffing changes.

Dr. Ralph E. Kromer became the Principal Investigator of BIONET commensurate with his assumption of the role of President and Chief Executive Officer of IntelliCorp. Dr. Kromer, who has a Ph.D. in Statistics from Stanford, had been Director of the Computer Science Laboratory at Texas Instrument, Inc., before joining IntelliCorp.

Completion of Staff Assignments

Dennis Smith, Ph.D. (50%), was hired in June to serve as Resource Manager. Dr. Smith has extensive experience with biomedical computer support facilities, including his tenure as Co-Investigator of the Dendral project on SUMEX, and membership on the Biotechnology Resources Review Committee of the Division of Research Resources of the National Institutes of Health (1978-81). Dr. Smith had headed the Research Applications Development computer group at Lederle Laboratories before joining IntelliCorp, and has had extensive experience in providing computer support for biomedical scientists. Dr. Smith replaces Dr. Tom Kehler in this position.

John L. Shelton, M.S., (50%) serves as Director of IntelliCorp's Computing Resources Group, managing day to day operation of BIONET computing resources as well as other computers, and planning growth. Mr. Shelton had previously been with the Knowledge Systems Division of IntelliCorp.

Andrea Gorman (50%) was appointed as Operations Manager, Computing Resources Group. Ms. Gorman had previously helped to provide computational services for molecular biologists in her role as Scientific Account Representative for IntelliGenetics. She is familiar with the problems of new users of the DEC2060 computer, is responsible for management of BIONET accounts, and supervises the maintenance of both systems' and BIONET specific software.

Mary Yardley (50%) has been appointed Senior Systems Operator in the Computing Resources Group. She is responsible, together with Ms. Gorman, for establishing directories and accounts on the BIONET computer, and for performing the many maintenance tasks required on these accounts and system utility programs.

Liz Martin has been appointed as full-time Administrator. She has direct responsibility for managing BIONET applications and maintenance of our PC-based database which is used to track applications, approved PI's, and user accounts.

Alan Tway, M.S., was appointed full-time User Consultant for BIONET. Mr. Tway has been associated with IntelliGenetics since 1981 and has had extensive experience in working with IntelliGenetics software users, in training, and document production.

Elaine Mansfield, Ph.D., Thomas Bonura, Ph.D., (50%), Ari Azhir, Ph.D., and Alan Engelberg, M.A., were all recently added to the scientific staff to cope with the enormous tasks associated with establishing user consulting, collaborative efforts, training, and documentation. They previously played part-time roles in BIONET until the computer facilities were established and operational. All have their degrees in molecular biology and related fields, and all are familiar with programs in the Core Library. Although

many responsibilities in the above areas are shared, each has a primary focus. Dr. Mansfield has primary responsibility for the training program, Dr. Bonura for the electronic communities, Dr. Azhir for user consultation and collaboration, and Mr. Engelberg for documentation.

3.4 Resource Advisory Committee and Allocation of Resources

3.4.1 National Advisory Committee

Our National Advisory Committee consists of the following members:

- Professor Joshua Lederberg (Chair), President, The Rockefeller University

- Professor Saul Amarel, Department of Computer Science, Rutgers University

- Professor Alan Maxam, Dana Farber Cancer Institute, Harvard Medical School

- Dr. Richard J. Roberts, Senior Staff Investigator, Molecular Biology, Cold Spring Harbor Laboratory

- Mr. Thomas Rindfleisch, Director, Heuristic Programming Project, Computer Science Department, Stanford University

- Professor Charles Yanofsky, Department of Biological Sciences, Stanford University

- Professor Fotis Kafatos, Department of Cellular and Developmental Biology, The Biological Laboratory, Harvard University

In 1984, this committee met with the BIONET staff on two occasions, in April, to review plans for initial organization, and in August, to review progress and establish application and approval procedures for access to BIONET. Our next meeting is scheduled for mid- to late March, 1985.

3.4.2 Allocation of Resources

We grant equal means of access to the BIONET 2060 computer for all approved applicants, whether for service (Class I) or collaboration (Class II), i.e., all have access to the same telecommunications facilities. Class I users are given 30% of the available CPU resources of the 2060. Each PI is given 200 disk pages to allocate among his or her research group of approved users, and we provide a special system utility to allow the PI to establish and change those allocations. Class II users will be given 10% of the available CPU. They will be assigned disk space commensurate with their needs for program development and sharing. Both classes have equal access to the Core Library, System Development Library, and Database Library. Both classes receive the same documentation, and will have access to the same training sessions.

The primary difference between Class I and Class II users is their access to staff time for consulting and program development. Our scientific consultants will devote additional time to responding to Class II

requests, and a systems programmer will be available to help members of this class to contribute and make available to the community as a whole, new software packages.

3.5 Dissemination of Information on Resource's Capabilities

We have used three major methods to disseminate information about availability of the BIONET Resource. These include:

- Demonstrations at National Scientific Meetings.
- Advertising in *Science* and *Cell*.
- Class II User Direct Mailing.

3.5.1 Demonstrations at National Scientific Meetings.

Prior to the BIONET Resource facility installation, the BIONET staff held demonstration sessions at the FASEB meeting in St. Louis, Missouri, and the Biological Chemistry meetings during the spring of 1984. We distributed announcements of the Resource and kept mailing list of all interested investigators. We used the mailing list for our initial mailings of application forms.

The BIOTECH '84 meeting in Washington D.C. in September 1984 took place at the same time as the initial mailing of BIONET applications. The BIONET staff attended this meeting as part of a larger IntelliGenetics group, demonstrated the BIONET facilities, and distributed applications to interested Principal Investigators. We plan further dissemination and training at upcoming meetings as described in subsection 3.6.3.

3.5.2 Advertising in *Science* and *Cell*.

We have played an active role in announcing the availability of BIONET to the community, advertising BIONET in two major publications. The following page contains a copy of the advertisement that appeared in *Science* magazine (**1984**, *225*, 1250). This advertisement generated over 600 requests for applications. The same copy was run in *Cell* (**1984**, *38*, 908).

In addition, BIONET has been mentioned and often discussed at length in articles appearing in *Science*, *Nature*, *Bio/Technology*, *Industrial Chemical News*, *Chemical Week*, *Esquire*, and *Genetic Engineering News*.

3.5.3 Class II User Direct Mailing.

The following letter was sent to potential collaborators, Class II users of BIONET. Their names and addresses are list in Appendix IV. To date, we have received six requests for Class II access. As mentioned earlier, we have granted qualified applicants Class I access pending internal and NAC review for Class II access.

Figure 3-1: Copy of Advertisement as it Appeared in *Science* and *Cell*

BIONET Resource
124 University Avenue, Suite 300
Palo Alto, CA 94301

Dear Colleague:

The Investigators and staff of the BIONET National Resource would like to invite you to apply for resource collaborator status (Class II membership). One of the major goals of the BIONET Resource is to promote the development of advanced and innovative software specifically for problem-solving in molecular biology.

For a limited number of collaborative researchers, the resource will provide significant computing resources to further the state of the art of this field. Based upon your previous experience in software development in molecular biology, we believe that both you and BIONET might profit from a research relationship and we solicit your application.

We have enclosed a description of BIONET goals and resources as will as an application form for collaborative membership. For Class II membership, your application will be treated like a research proposal; these competive applications will be reviewed in confidence by the BIONET staff and National Advisory Committee. We anticipate admitting 6 to 12 Class II members in this initial application phase.

Please feel free to contact us or Dr. Dennis Smith, Resource Manager, at (415) 328-4870 for further details about the BIONET Resource. Alternatively, you may wish to contact the BIONET Co-Investigators directly. Dr. Peter Friedland may be reached at (415) 497-3728, Dr. Douglas Brutlag at (415) 497-6593 and Dr. Laurence Kedes at (415) 493-5000 X 5318. We are happy to answer any questions you may have and look forward to working with you.

Sincerely,

Elaine Mansfield, Ph.D.

Thomas Bonura, Ph.D.
Consulting Scientists, BIONET

3.6 Training Program for BIONET

Although IntelliGenetics scientists have previous experience training commercial customers on the use of the programs in the BIONET Core Library, we have never before had to train so many new users at the same time. Our traditional software training program consists of a two-day intensive hands-on training. Since our limited staff and the large numbers make it impossible to deliver this level of training, we are adopting three approaches to explore the best combination of methods for training great many scientists with a small of staff.

- Lectures and Demonstrations - Videotaping
- Hands-on Training at IntelliGenetics
- Trainings at National Meetings

3.6.1 Lecture and Demonstrations.

Our first approach is to hold one-day training sessions for a significant number of users, where there are facilities available to support our needs. The text of the following letter, sent out to the initial set of approved applicants, summarizes our strategy and requirements for facilities. Because it is clear that not all investigators on BIONET will be able to take advantage of the more traditional hands-on approaches summarized in the following sections, we will be videotaping this first session. If the result is of sufficient quality to be distributed widely, we will establish a video tape lending library in lieu of actual trainings.

Dear BIONET User,

Welcome to the BIONET Resource
A National Computer Resource for Molecular Biology

Training is an important aspect of using the BIONET computer effectively. It will enable you to learn more efficiently to use the varied tools on the computer. The training will also be a good opportunity to meet other investigators and potential collaborators who also use the BIONET Resource.

As one of the first investigators on BIONET, we extend a special invitation to a resource training on December 20, 1984 at Stanford University Turing Auditorium. You will find a map and schedule of the day attached.

To provide individual video monitors, attendance must be limited to 80 persons accepted on a first-come, first-served basis. Although we anticipate being able to accommodate up to 4 people from each group, priority will be given to representatives from different laboratory groups.

To reserve your place, please return the attached reply card by December 10, 1984 . Indicate the name and phone number of each person you wish to send. Lunch and a comprehensive BIONET Training Manual will be provided at the training session. Before attending the training session, all participants MUST:

1. Read the enclosed *Introduction to BIONET*.
2. Log in to the computer at least once.
3. Be willing to teach others in the lab what they learn at the training session.

If your schedule prohibits attending this initial training, we will be conducting future sessions. In particular, the Waksman Institute of Continuing Education at Rutgers University is sponsoring a 1- or 3-day workshop for BIONET on June 17-19, 1985. This special opportunity will allow in-depth, hands-on training for up to 30 attendees. Applications for this workshop will be mailed when they become available. Other trainings will be held in conjunction with national scientific meetings.

Sincerely,

Elaine Mansfield
Training Manager
Consulting Scientist, BIONET

To give you an idea of the scope of such one day trainings, the following is the schedule for the December 20, 1984 session.

Computer Resource for Molecular Biology
BIONET Training Outline

Thursday December 20, 1984

Turing Auditorium, 111 Polya Hall
Stanford University

MORNING SESSION

9:00-9:30   Dr. Dennis Smith, BIONET Resource Manager
                          Welcome
               An overview of BIONET Resource

9:30-10:30   Getting Started
               Dr. Thomas Bonura, Consulting Scientist BIONET

               Using Electronic Mail and Bulletin Boards
               System Commands and Directory Organization
               Sequence Database Organization

10:30-10:45   BREAK

10:45-12:00   Sequence Entry and Assembly
               Alan Tway, BIONET User Consultant

               GENED - The Genetic Editor
               GEL - DNA Sequencing Project Manager

12:00-1:00   LUNCH

AFTERNOON SESSION

1:00-2:15   Database and Searching Methods
               Dr. Elaine Mansfield, BIONET Training Manager

               QUEST - Database Search and Retrieval
               IFIND - Sequence Similarity and Alignment Program

2:15-3:30   Cloning and Restriction Mapping
                          Tom Bonura

          SIZER - Restriction Fragment Length Calculation
          MAP - Construction of Restriction Maps from Enzyme Digests
          CLONER - Simulation and Design of Recombinant DNA Experiments

3:30-4:30   Sequence Analysis Programs
                       Elaine Mansfield

          SEQ - Nucleic Acid Analysis, Comparison and Manipulation

PEP - Peptide Analysis, Comparison and Manipulation

4:40-5:30   Question and Answer Sessions
                   Special Interest Groups

To provide broad exposure to a large audience, we will hold this session at the Turing Auditorium. a computer training facility at Stanford University. This facility provides individual video monitors for all attendees to see all computer interaction close at hand. Whereas this mode is not as good as hands-on training, it will afford many more BIONET recipients to be trained quickly.

The initial BIONET training session takes place at the same time as a major new release of IntelliGenetics software. A major staff commitment was made to develop new training materials particularly for BIONET users. We have developed "Annotated Examples," illustrating how common experimental problems may be solved by the computer were developed. These examples were specifically chosen to allow the BIONET users return to their laboratory and run similar sessions with their own data. These "Annotated Example" are part of the *BIONET Training Manual*.

We will be further developing these examples so that BIONET users may run command files on-line after the training sessions to automatically reproduce the annotated examples. The BIONET staff will further develop the training examples into interactive tutorial sessions over the next grant year.

3.6.2 Hands-on Training at IntelliGenetics.

Following the one-day tutorial monitor training, we will offer BIONET investigators an opportunity to participate at a later date in the on-going IntelliGenetics two-day training program. Due to limited staff resources, these trainings have been held only monthly. We will make every effort to increase this number if demand merits it and staffing permits.

3.6.3 Trainings at National Meetings.

The intention of BIONET Resource staff, from the initial proposal to the present, is to hold training sessions in conjunction with major national professional meetings. During the first grant year the BIONET staff will hold BIONET trainings in conjunction with two national meetings.

The Recombinant DNA meetings will be in San Francisco, on February 3-7, 1985. The Turing Auditorium has been reserved for a follow-up day-long training on February 8th. The BIONET staff will also demonstrate the Core Library software and resource facilities at the Miami MidWinter Symposium on February 11-15, 1985. We plan to use the training videotapes at these and future meetings.