

BIONET™
National Computer
Resource
for
Molecular Biology

Grant No. RR01685-02
Annual Progress Report
March 1, 1985 - February 27, 1986

The BIONET Resource is funded through a cooperative Agreement with Intellicore, Inc., by the Biomedical Research Technology Program, Division of Research Resources, National Institutes of Health

BIONET Resource
c/o Intellicore
1675 E. Coast Road West
Mountain View, CA 94039
415-947-0500
415-947-0501

Table of Contents

Part I. Title Page	1
Part II. Description of Research Progress	2
II.A. Scientific Subprojects	3
II.A.1. Collaborative Research and Service	3
II.A.2. Core Research and Development	5
II.A.3. Training	7
II.B. Books, Papers and Abstracts	9
II.C. Resource Summary Table	13
Part III. Narrative Description	15
III.A. Summary of Research Progress	15
III.A.1. Service	16
III.A.1.a. Scientific Consulting - Class I, III and IV Support	17
III.A.1.b. Scientific Case Studies Using BIONET	19
III.A.1.c. KERMIT Lending Library	23
III.A.1.d. Subscription Fee and Its Effects	23
III.A.1.e. Reapplication Procedure	24
III.A.2. Collaborative Research	26
III.A.3. Core Research	29
III.A.3.a. Hardware Text Searching Machines	29
III.A.3.b. BIONET Satellite Program	31
III.A.4. BIONET Training Program	33
III.A.4.a. A Brief Review	33
III.A.4.b. Some Lessons Learned	34
III.A.4.c. A New Strategy	34
III.A.5. Resource Facilities	36
III.A.5.a. Computer Hardware and Telecommunication Networks	36
III.A.5.b. Summary Statistics on Machine Use	38
III.A.5.c. Computer Software - Core Library	50
III.A.5.d. Computer Software - System Library	50
III.A.5.e. Computer Software - Contributed Library	52
III.A.5.f. Database Library	52
III.B. Highlights	53
III.C. Administrative Changes	53
III.C.1. Facilities	53
III.C.2. Personnel	53
III.D. Resource Advisory Committee and Allocation of Resources	54
III.E. Dissemination of Information on Resource's Capabilities	55
III.E.1. Community Interactions and Awareness	55
III.E.2. Electronic Communications	56
III.E.2.a. Bulletin Boards	56
III.E.2.b. Bibliographies	59
III.F. Suggestions and Comments	61
I. Letter to BIONET Scientists	63
II. Justification of \$400 Fee	67
III. Letter on Class IV Access	69
IV. Reapplication Form for BIONET Access	70
V. Program for the Rutgers/Waksman Workshop	74

VI. Descriptions of the BIONET Satellite Program	76
VII. Text of Advertisement to Appear in Nucleic Acids Research	79
VIII. The BIONET Brochure Mailed to NIH Grantees	80

List of Figures

Figure III-1:	Pieslice Allocation of the DEC-2060 Computer	40
Figure III-2:	Actual Use of the DEC-2060 for the Month of October, 1985	41
Figure III-3:	BIONET's Percentage Use of the DEC-2060, 12/84 - 11/85	42
Figure III-4:	BIONET's Prime Time Use of the DEC-2060, 12/84 - 11/85	45
Figure III-5:	BIONET's Non-Prime Time Use of the DEC-2060, 12/84 - 11/85	47
Figure III-6:	BIONET's Total Use of the DEC-2060, 12/84 - 11/85	49
Figure III-7:	Total Telenet and UNINET Network Use, 12/84 - 11/85	51

List of Tables

Table II-1:	BIONET User Community	2
Table III-1:	Summary of Distribution of Questions.	17
Table III-2:	Breakdown of Program Questions by Function	18
Table III-3:	Breakdown of Program Use, by Category	25
Table III-4:	BIONET Prime Time CPU Minutes	44
Table III-5:	BIONET Prime Time Connect Hours	44
Table III-6:	BIONET Non-Prime Time CPU Minutes	46
Table III-7:	BIONET Non-Prime Time Connect Hours	46
Table III-8:	BIONET Total CPU Minutes	48
Table III-9:	BIONET Total Connect Hours	48

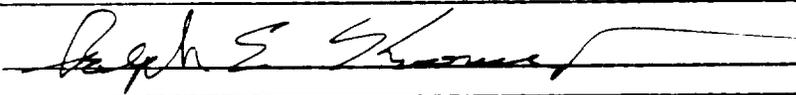
DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
National Institutes of Health

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

1. PHS AWARD NUMBER:

1	U	4	1	R	R	0	1	6	8	5	-	0	2
---	---	---	---	---	---	---	---	---	---	---	---	---	---
2. TITLE OF AWARD BIONET, National Computer Resource for Molecular Biology
3. NAME OF RECIPIENT INSTITUTION: IntelliGenetics, 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	5
---	---	---	---	---	---	---	---
- 5b. TO (Month, Day, Year):

0	2	-	2	8	-	8	6
---	---	---	---	---	---	---	---
6. PRINCIPAL INVESTIGATOR:
- 6a. NAME: Ralph E. Kromer
- 6b. TITLE: President, IntelliCorp
- 6c. SIGNATURE: 
7. DATE SIGNED (Month, Day, Year): _____
8. TELEPHONE (Include Area Code):

4	1	5	-	9	6	5	-	5	7	5	8
---	---	---	---	---	---	---	---	---	---	---	---

Part II. Description of Research Progress

This section of the Report provides statistical information on use of the BIONETtm Resource. The period covered is 12/84 - 11/85. Its individual sections have been prepared under guidelines discussed with BRTP staff. In the one year since BIONET formally began accepting and approving applications for access, the BIONET community has expanded to 533 Principal Investigators (PI's), representing about 1800 total investigators, whose classification is as summarized in Table II-1. As shown in the Table, of the 537 accepted Class I PI's, 27 chose to drop their membership when the subscription fee for telecommunication access was announced. To review, Class I PI's represent the service component of the Resource, Class II the collaborative component, and Class III is reserved for those persons who are responsible for local computing facilities, and who have agreed to act to support the local community of scientists accessing BIONET.

Table II-1: BIONET User Community

Class I PI's	537
Declined	(27)

Total Class I	510
Class II PI's	12
Class III PI's	3
Nat. Advis. Comm.	8

Net Total	533

II.A. Scientific Subprojects

II.A.1. Collaborative Research and Service

We report Collaborative Research and Service first, because it represents the bulk of the Resource's activities in this grant year. We have not reported the research abstracts in the interests of saving space. These abstracts are maintained on our PC database of BIONET users, and are available to any interested party.

We report the "Usage Factor" as both central processor units (cpu time), in minutes, and connect time, in hours, for each PI. These values are the sums of all usage by the PI and his or her group members ("Sub-I's"). We report data only on those PI groups that have accessed the Resource. Of the 522 Class I and II PI's, 361, representing about 1010 individual investigators, have logged on to the Resource. The data for Sub-I's is maintained on our PC database and is available to any interested party.

We do not report Resource staff hours nor BRTP funds allocated for individual PI's because it is impossible to allocate these rationally to such a large user community. Summary information for all PI's is given in Part II, Section C., the Resource Summary Table.

II.A.2. Core Research and Development

We report summary information for our two current Core Research projects. The B RTP funds allocated are calculated on the basis of the sum of individual salaries for the time spent, plus a percentage of the total Resource budget devoted to providing supporting services.

The Resource's primary commodity is cpu cycles made available to all investigators, including BIONET staff. Thus, the percentage used in the above calculation is the cpu time used by the indicated investigators compared to the total cpu time used on the system. The calculation for the actual cost is this percentage times the basis of costs of providing this commodity to local staff; i.e., the budget categories of Supplies, Computer Expenses, Software Licenses and Services and Documentation. This basis is, using our anticipate expenditures in these categories through February 28, 1986, \$265,500.

DSR SCIENTIFIC SUBPROJECT FORM

PAGE II, SECTION A

INSTITUTION: IntelliGenetics

REPORT PERIOD: March 1, 1985 to February 28, 1986

AWARD NUMBER: 1 U 4 1 R 0 1 6 8 5 - 0 2

FILL out a separate Subproject Form for Core, Collaborative or Training. Check one of the following:

CORE RESEARCH & DEVELOPMENT
 COLLABORATIVE RESEARCH & SERVICE
 TRAINING

Descriptive Title (80 characters)

Abstract

BIONET
Satellite Program

Distribution of the computational and communication facilities of BIONET to other, geographically remote sites (Satellite Resources); linking these Satellites for rapid exchange of electronic mail and bulletins, data files and programs

Sequential Search Machine

Exploring the new technologies of hardware, sequential search machines for applications to biological sequence retrieval and intercomparison

	(2) Science Axis I	(3) a. Investigator(s) Name (Last Name, First Name & Middle Initial) b. Department c. Non-host Inst.	Resource Technology %	(4) USAGE FACTOR		BRTTP Full Allocation
				CPU MIN USED	Resources Staff Hours	
9	40, 42	a. Smith, Dennis H. Roode, R. David b. IntelliGenetics	DEC 2060 "	22 284	94 148	\$ 11,3
9	42, 68 70	a. Smith, Dennis H. Brutlag, D. b. IntelliGenetics	DEC 2060 "	45 276	188 40	\$ 16,20
CUMULATIVE TOTALS:	2			627	470	\$ 27,54

II.A.3. Training

We report summary information for our Training program. The sites at which BIONET provided some level of training are named here, and are discussed in more detail in Part III, the Narrative Description. The BRTP funds allocated include the salaries of the personnel involved, the estimated costs of the training program itself for this year (\$5,500), and a percentage allocation of cost for cpu use, as discussed in the previous section.

DBR SCIENTIFIC SUBPROJECT FORM

INSTITUTION: IntelliGenetics

REPORT PERIOD: March 1, 1985 to February 28, 1985

AWARD NUMBER:

1	U	4	I	R	R	0	1	6	8	5	-	0	2
---	---	---	---	---	---	---	---	---	---	---	---	---	---

FILL out a separate Subproject Form for Core, Collaborative or Training. Check one of the following:

CORE RESEARCH & DEVELOPMENT
 COLLABORATIVE RESEARCH & SERVICE
 TRAINING

Descriptive Title (60 characters) Abstract	(2) Science Axis		(3) a. Investigator(s) Name (Last Name, First Name & Middle Initial) b. Department c. Non-host Inst.	Resource Technology a/	(4) Usage Factor		(5) BRTP Fund Allots
	I	II			CPU MIN USED	Staff Hours	
BIONET Training Program (FASEB, Rutgers/Waksman, Nature update in Molecular Biology, Biotech '85, Miami Mid-Winter Symposia)	9	40, 42 68	a. Mansfield, Elaine Lawler, MaryJo Azhir, Ari Kedes, Laurence Brutlag, Douglas Smith, Dennis H. b. IntelliGenetics	DEC-2060 " " " " "	447 68 2279 71 276 22	151 255 203 40 40 47	8
CUMULATIVE TOTALS:	1				3163	736	\$61,348

a/ Identify Resource Technologies Used. b/ Give Hours Resource Technologies Used. See Instructions, page 11.

II.B. Books, Papers and Abstracts

We report the publications by members of the BIONET scientific community on a version of the special form as generated by our local database management system. We report only the category of Collaborative Research and Service. There have been no publications in the Core Research program. The published materials used to support our Training program (the *Introduction to BIONET*, the *BIONET Reference Manual* and the *BIONET Training Manual*) have been described before and thus are not reported separately.

Part II, Section B

INSTITUTION: Intelligenetics

Award Number 1U41RR01685-02

REPORT PERIOD: March 1, 1986 to February 28, 1986

COLLABORATIVE RESEARCH

- * Vogt, M., Haggblom, C., Swift, S., Haas, M., Envelope gene and long terminal repeat determine the different biological properties of Rauscher, Friend, and Moloney Mink Cell Focus-inducing viruses, J. Virol. 55, 184-192, 1985
- * Upton, C, McFadden, G., DNA Sequence Homology between the Terminal Inverted Repeats of Shope Fibroma Virus and an Endogenous Cellular Plasmid Species, Molecular and Cellular Biology, Jan. 1986
- * Andersen, R.D., Birren, B.W., Taplitz, S.J., Herschman, H.R., The Rat Metallothionein-1 Structural Gene and Three Pseudogenes, One of Which Contains 5'-Regulatory Sequences, Molecular and Cellular Biology, 1985
- * Andersen, R.D., Taplitz, S.J., Briston, G., Herschman, H.R., Rat Metallothionein Multigene Family, in Proceedings of the Second International Meeting on Metallothionein and Other Low Molecular Weight Metal-binding Proteins, Birkhauser Verlag, Boston, Aug. 21, 1985
- Katzen, A.L., Kornber, T.B., Bishop, J.M., Isolation of the proto-oncogene c-myb from *Drosophila melanogaster*, Cell 41: 449-456, June 1985
- Simon, M., Drees, B., Kornberg, T., Bishop, J.M., The Nucleotide sequence and tissues specific expression of *Drosophila* c-src, Cell, Oct 1985
- * Lohe, Allan R., Brutlag, Douglas L., Multiplicity of Satellite DNA Sequences in *Drosophila Melanogaster*, PNAS, 1985
- * Calhoun, David, Bishop, David T., Bernstein, Harold S., et. al., Fabry Disease: Isolation of a cDNA clone encoding human alpha-galactosidase A, PNAS, 1985
- * Cooke, N.E., David, E.V., Serum Vitamin D Binding Protein is a Third Member of the Albumin and alpha-fetoprotein gene family, J. Clin. Investig., 1985
- * Glaichenhaus, N., Leopold, P., Cuzin, F., et. al., Changes in the expression of cellular genes inn cells immortalized or transformed by polyoma virus, Cancer Cells, Cold Springs Harbor Laboratory, 1985
- Biggs, J., Searles, L.L., Greenleaf, A.L., Structure of the Eukaryotic Transcription Apparatus: Features of the Gene for the Largest Subunit of *Drosophila* RNA Polymerase II, Cell 42: 611-621, Sept. 1985
- * Hamori, Eugene, Novel DNA Sequence Representations, Nature 314: 585, 1985
- * Nakauchi, H., Nolan, G.P., Herzenberg, L.A., et. al., Molecular cloning of Lyt-2, a membrane glycoprotein marking a subset of mouse T lymphocytes: Molecular homology to its human counterpart, Leu-2/T8, and to immunoglobulin variable regions, PNAS USA 82: 5126-5130, 1985
- * Hogness, D.S., et. al., Regulation and Products of the Ubx Domain of the Bithorax Complex, Cold Springs Harbor Symposia Vol L, 1985
- * Allison, L.A., Moyle, M., Shales, M., Ingels, C.J., Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases, Cell 42: 599-610, 1985
- * James, D., Leffak, I.M., Polarity of Replication Through the Avian Alpha-Golbin Locus, Mol. Cell. Biol., 1985
- * Singer, P.A., Oshima, R.G., Molecular Cloning and Characterization of the Endo B Cytokertin Expressed in the Preimplantation Mouse Embryos, J. Biol. Chem., 1986
- * Knott, T.J., Rall, S.C., Scotts, J., et.al., Human Apolipoprotein B. Structure of Carboxyl-Terminus

COLLABORATIVE RESEARCH

Domains, Sites of Gene Expression, and Chromosomal Location, Science, 1985

- * Regier, J.C., Pacholski, P., PNAS USA 82: 6035-6039, 1985
- * Robinson, H.L., Miles, B.D., Avian leukosis virus-induced osteopetrosis is associated with the persistent synthesis of viral DNA, Virology 141: 130-143, 1985
- * Miles, B.D., Robinson, H.L., High frequency transduction of c-erbB in avian leukosis virus-induced erythroblastosis, J. Virology 54: 295-305, 1985
- * Shank, P.R., Schatz, L.M., Robinson, H.L., et. al., Sequences in the gag-pol-5' env region of avian leukosis viruses confer the ability to induce osteopetrosis, Virology 145: 94-104, 1985
- * Robinson, H.L., Jensen, L., Coffin, J.M., Sequences outside of the LTR determine the lymphomogenic potential of Rous associated virus-1, J. Virology 55: 752-759, 1985
- * Robinson, H.L., Avian leukosis viruses as vectors for the development of vaccines, Proceedings of 34th Annual National Breeders Roundtable: 64-92, 1985
- * Robinson, H.L., Gagnon, G.C., Patterns of proviral insertion and deletion in avian leukosis virus induced lymphomas, J. Virology, 1985
- * Miller, R.H., Robinson, W.S., Common Evolutionary Origin of Hepatitis B Virus and Retroviruses, PNAS, 1985
- * Schumacher, M., Camp, S., Taylor, P., et. al., Primary Structure of Torpedo californica Acetylcholinesterase Deduced from dDNA Sequence, Nature, 1985
- * Timmerman, K.P., Tu, D., Complete sequences of IS3, N A R 13: 2127-2139, 1985
- * Kemper, B., Molecular Biology of Parathyroid Hormone, Critical Reviews in BioChem, 1986
- * Maratea, D., Young, K., Young, R., Deletion and Fusion Analysis of the Phi OX174 E Lysis Gene., Gene, 1985
- * Machida, C.A., Bestwick, R.K., Kabat, D., A Weakly Pathogenic Mutant of Rauscher Spleen Focus-Forming Virus Has Lost the Carboxyl-Terminal Membrane Anchor of Its Envelope Glycoprotein, J. Virol. 53: 990-993, 1985
- * Machida, C.A., Bestwick, R.K., Boswell, B.A., Kabat, D., Role of a Membrane Glycoprotein in Friend Virus-Induced Erythroleukemia: Studies of Mutant and Revertant Viruses., Virology 144: 158-172, 1985
- * Bestwick, R.K., Hankins, W.D., Kabat, D., Roles of Helper and Defective Retroviral Genomes in Murine Erythroleukemia: Studies of Spleen Focus-Forming Virus in the Absence of Helper., J. Virol. 56, 1985
- Li, J.-P., Bestwick, R.K., Machida, C.A., Kabat, D., Role of a Membrane Glycoprotein in Friend Viral Erythroleukemia: Nucleotide Sequences of Non-Leukemogenic Mutant and Spontaneous Revertant Viruses., J. Virol. 57, 1986
- * Gustafson, T.A., Markham, B.E., Morkin, E. , Analysis of Thyroid Hormone Effects on Myosin Heavy Chain Gene Expression in Cardiac and Soleus Muscles Using a Novel Dot-Blot mRNA Assay., BBRC Vol. 130, No.3 1161:1167, 1985
- * Morkin, E., Sheer, D., Gustafson, T.A., et. al., Regulation of Cardiac Myosin Isoenzymes by Thyroid Hormone, UCLA Symposia on Mol., Vol.20, 1985
- Black, D.L., Chabot, B., Steitz, J.A., U2 as well as U1 Small Nuclear Ribonucleoproteins are Involved in Pre-Messenger RNA Splicing, Cell 42, 1985
- * Smith, D. H., BIONET: National Computer Resource for Molecular Biology, Abstracts, Federation of

Part II, Section B

Award Number 1U41RR01685-02

INSTITUTION: Intelligenetics

REPORT PERIOD: March 1, 1985 to February 28, 1986

COLLABORATIVE RESEARCH

American Societies for Experimental Biology, Apr 22-25, 1985, Anaheim, CA

* Smith, D.H., BIONET: National Computer Resource for Molecular Biology, Abstracts, International Congress on Computers and Biotechnology, Jan 30-31, 1986, Baltimore, MD

* Smith, D.H., Brutlag, D., Friedland, P., Kedes, L., BIONET: National Computer Resource for Molecular Biology, Nucleic Acids Research, 1986

II.C. Resource Summary Table

The Resource Summary Table includes the totals from the previous sections under Training and Core Research and Development. The totals for the Collaborative Research and Service categories were arrived at as follows. The Usage Factor, again in cpu minutes, represents the grand total of all use summarized on the previous forms, plus staff use not allocated to the above two categories. The B RTP funds allocated include the remainder of the direct costs estimated for this year (see Section III of the "Application for Continuation Grant"), i.e., those funds not allocated to the above two categories.

The category of Administration/Miscellaneous includes only the Usage Factor of BIONET's share of the cpu time (minutes) for computer facility staff and DEC-2060 system overhead accounts (see Table III-8). No funds for this cpu time or staff time are allocated; such funds are considered part of the support of the user community and are distributed on the basis of cpu time to the categories above, as described previously. The Funds Allocated include only the items of capital equipment purchased during the year.

The category of Down Time includes the sum of scheduled and unscheduled maintenance on the DEC-2060 computer. In the period 12/84 - 11/85, there was a total of 139 hours (8340 cpu minutes) of down time:

- 93 hours (5580 cpu minutes) were scheduled down time, including 69 hours of downtime for the move of the computer over the weekend of August 17-18, 1985; the remainder was scheduled maintenance.
- 46 hours (2760 cpu minutes) of down time were due to unscheduled maintenance.

The down time reported on the Summary Table is BIONET's 50% of the total, or 4,170 cpu minutes. Note that the unscheduled maintenance of 2760 cpu minutes is 0.5% of the total cpu time available for the year. Thus, the DEC-2060 system has been available more than 99% of the time, 24 hours a day, seven days a week. No funds have been allocated to this category.

PART II, SECTION C RESOURCE SUMMARY TABLE REPORT PERIOD March 1, 1985 to February 28, 198

AWARD NUMBER	RESOURCES					(3) Number Investigators	(4) USAGE FACTOR Resource a/ Technology	(5) BRTP Funds Allocated \$	(6) Resource Fees \$ Collected	(7) Other Funds \$
	(1) Number Subproject	(2) Number Publications	(3) Number Investigators	(4) CPU MIN USED	(5) Resource Staff Hrs					
1041R01685	2	2	3	627	470	\$ 27,543	- 0 -	- 0 -		
CORE RESEARCH & DEVELOPMENT										
	361	40	1010	138,095	6507	540,760	- 0 -	\$97,600		
COLLABORATIVE RESEARCH & SERVICE										
		0	6	3,163	736	61,348	- 0 -	- 0 -		
TRAINING										
				17,418		5,000				
ADMINISTRATION/MISCELLANEOUS										
				4,170						
DOWN TIME										
GRAND TOTALS	363	42	1029	43,473	7713	\$634,551	- 0 -	\$97,600		

See Instructions. Page iv.

Part III. Narrative Description

III.A. Summary of Research Progress

This has been a difficult, yet exciting and successful second year for the BIONET Resource.

Problems and Solutions

The year was made difficult by the unanticipated reduction of this year's grant award at the very end of the first year. Substantial dislocations in our level of staff and planned activities were the result. The Resource has substantial fixed costs in providing computer facilities and telecommunication access to the scientific community. Thus, our final, approved budget had severely reduced funds in categories of personnel, documentation and services, training and other support of Core and Collaborative research activities. In addition, our projections showed that we would exceed our budgeted funds for network communications about eight months into the grant year. These facts led us to take the following steps to preserve the availability of the Resource to the community.

Documentation. We could no longer afford to sell BIONET reference and training manuals to BIONET scientists for the cost of production, because of staff time involved in processing and filling orders. Therefore, we turned the processing of orders for these materials over to IntelliGenetics. We continue to provide, free of charge, the "Introduction to BIONET" for all new PI's.

Computer Networks. We searched for and found an alternative, lower cost vendor for telecommunications network support, and made the transition from Telenet to UNINET in the first part of July, 1985. This switch has allowed us to reduce our costs/access port to the point where, coupled with subscription fees (see next paragraph), we could increase the number of lines available to BIONET scientists outside our immediate area. Further information on this transition and the current network configuration can be found in the letter to BIONET scientists in Appendix I, and in the subsequent Subsection III.A.5, Resource Facilities.

Subscription Fees. Our original proposal had anticipated some level of cost recovery or access charges beginning in the third year of our grant. However, the Year 2 budget cut forced us to impose an access charge, with very short notice for the BIONET community. After negotiations with the NIH Division of Research Grants, we were able to impose a subscription fee, recognizing the monies collected as telecommunications access charges that were not to be regarded as grant-related income (as grant-related income, the monies collected would be used to reduce the total grant award). The letter announcing the NIH view of the access charges is attached to Section III of our "Application for Continuation Grant". A discussion of the fee was sent to all BIONET users (see Appendix I). New BIONET applicants receive a revised version of that text with their application forms. The fee was set at \$400 per PI per year, for reasons summarized in Appendix II.

Subsequently, we announced a new category of membership in the BIONET Resource, Class IV, for those who wish to use only the communications aspects of the Resource, electronic mail and electronic bulletin boards. The fee for this method of access is \$100 per PI per year. The letter announcing Class IV access, sent to all persons who had not contacted us about payment of subscription fees, is given in Appendix III. A version of this letter was also posted on the BIONET-NEWS bulletin board on BIONET. Information on Class IV access is now part of our standard application package.

A Successful Year

Despite our fiscal and administrative difficulties, we have been able to accomplish a great deal this year. These accomplishments are described in detail in the subsequent discussion, but here, in brief, are some notable successes:

- We have approved 560 PI's for access to the Resource, representing about 1800 individual scientists.
- Over 50 publications have appeared or are in press, where the facilities of the BIONET Resource played a key role in obtaining results. This is particularly notable in that more than half of the community has had access for only a few months.
- The computer facility is under new management; the facility was moved to our new offices over one weekend, and was available again to the community with less than three days total downtime. Facilities have been improved substantially.
- The imposition of subscription fees has been successful. Not only have they generated enough support to solve our fiscal problems, the impact on the Resource and its community has been relatively minor. With very few exceptions, those choosing not to pay were not significant users of the Resource.
- We have been able to make significant progress in our Core and Collaborative research programs, despite the budget cuts. Two major Core projects have been begun, our BIONET Satellite program, and investigation of hardware solutions to the problem of biological sequence database search, retrieval and analysis.
- Our reapplication forms for another year of access for PI's have led to many words of praise for the Resource and its crucial role in the PIs' research. A few PI's had significant complaints, and we are working with those PI's on their problems.

III.A.1. Service

The Service component of the BIONET Resource includes primarily Class I investigators who wish to use the BIONET facilities to support their research. It also includes Class III investigators, a classification reserved for persons responsible for Department, School, or Campus-wide computer facilities who wish to provide information about or access to BIONET to the local community they serve. There are currently three investigators who have been granted this status, K. Manly at Roswell Park, P. Vitek at the Imperial Cancer Research Foundation, and J. Claverie at Institut Pasteur.

More recently, we have established a new class of BIONET participants, Class IV. This class is reserved for those investigators who wish to take advantage of the communication components of BIONET, but do not wish to run the analysis programs. This policy should address desires for this level of access expressed at our March, 1985 NAC meeting. There will be a telecommunications-related subscription fee for this level of access as well, and we have set this fee at \$100. We have already informed all PI's who chose not to pay the full subscription fee about this policy (see letter in Appendix III). Interestingly, none have chosen to accept this option. We also have sent out a bulletin on the system to inform all current subscribers. We now include this option as part of our standard information and application packet sent out to potential BIONET scientists. Only one new applicant has so far requested this status.

III.A.1.a. Scientific Consulting - Class I, III and IV Support

An important function of the BIONET consultants is to provide support to BIONET users. This support takes the form of answering questions and making available information about the BIONET Resource. Questions come by phone, by electronic mail, and by terminal links from investigators to staff. A survey of the past year's phone, mail messages, and terminal links from BIONET Resource scientists gives a good overall indication of the usage of the different aspects of the Resource.

A total of 2043 inquiries were tallied from November 1, 1984 to November 13, 1985. The data were divided into five categories by subject. These categories are programs (Core and Contributed Libraries, electronic communication), TOPS20 system, personal computers and personal computer software, telecommunications, and BIONET administration. The distribution among these categories of questions is shown in Table III-1).

Table III-1: Summary of Distribution of Questions.

Category	Number of Inquiries	Percent of Total Inquiries
Programs	745	36
TOPS20 System	423	21
PC and PC software	313	16
Telecommunications	287	14
BIONET Administration	275	13
TOTALS	2043	100

The first category in Table III-1, Programs, has been further subdivided into scientific categories of questions, as summarized in Table III-2). The first column indicates the category. The second column shows the total number of questions received in each category. The third column shows the number of

questions received as a percent of the total questions on programs. The last column shows the number of questions in each category as a percent of the total of 2043 questions tallied above.

Table III-2: Breakdown of Program Questions by Function

Scientific Category	Number of Program Inquiries	Percent of Program Inquiries	Percent of Total Inquiries
Sequence and Gel Data Entry & Manipulation	104	14	5
Consensus Sequence Determination	58	8	3
Protein and DNA sequence Manipulation	199	27	10
Databases, Database Searches and Sequence Alignment	259	35	12
Experiment Planning	19	2	1
Communication with the BIONET Community	74	10	4
Other	32	4	1
TOTALS	745	100	36

As shown in Table III-2, there were 745 requests for information concerning programs offered by the BIONET Resource; these questions comprised 36 percent of the total requests for information. Questions about database access and sequence comparison represented more than 1/3 of the total program inquiries. Questions about biological sequence manipulation, for example, restriction mapping, alignment, translation, codon frequencies, represented 27% of the total. Sequence entry, determination of consensus sequences, and electronic communication make up the bulk of the rest of the questions.

Returning to Table III-1, questions about the TOPS20 system represented 21% of the total. These questions concerned accessing programs, manipulating files and directories, using the text editors, and controlling output on the terminal.

A wide variety of personal computers (PC's) and PC software is in use by BIONET subscribers. The most frequently asked questions concerned IBM PCs and Apples and their related software. The majority of information requested concerned the transfer of files to and from the user's computer and the BIONET computer. There were a total of 313 questions, comprising 16 percent of the total, on this topic.

The telecommunication category includes phone and mail correspondence concerning the quality of communications between remote users and BIONET, and the procedures involved in connecting to the BIONET computer. These questions made up 14 percent of the total with 287 inquiries.

The final category of inquiries concerns BIONET Administration. This category does not include all the correspondence as the majority of questions were directed to the BIONET administrator, and not to the BIONET consultants. There were 275 requests for information. These inquiries covered requests for and complaints about the manuals, account or application status, adding or deleting users, and possibilities for training sessions. This category comprised 13 percent of the total.

Our data indicate that the majority of users have needed assistance on use of the Core Library software to solve specific scientific problems. As the community has matured in its use of the software, the questions have become deeper, indicating a real awareness of the basics of using computers to solve such problems. The large number of questions about the TOPS-20 system, telecommunications, and PC access to the Resource reveal that the community is distributing its computational tasks. PC's are used to collect sequence and other data, and to up load this information to BIONET. Results obtained from analysis on BIONET are then down loaded to the PC's, where they are used to guide new experiments and eventually used in preparation of scientific publications.

III.A.1.b. Scientific Case Studies Using BIONET

In just over a year of user access, BIONET has grown to service almost 1500 researchers in the field of molecular biology. In one way, BIONET's contribution to research can be inferred by the number of subscribers. However, another way to demonstrate BIONET's current and potential efficacy is through the examination of the work of a few researchers who have submitted research findings for publication and had used BIONET in the process of discovering these findings.

Using information obtained through our reapplication procedures (see below) about BIONET use and resulting publications, we have selected three examples which we present as brief case studies: "Common Evolutionary Origin of Hepatitis B Virus and Retroviruses," PNAS, in press (1986), by R. Miller and W.S. Robinson (Stanford University); "DNA Sequence Homology between the Terminal Inverted Repeats of Shope Fibroma Virus and an Endogenous Cellular Plasmid Species," Mol. Cell Bio., in press (1986), by C. Upton and G. McFadden (University of Alberta); and "Molecular Cloning and Characterization of the Endo B Cytokeratin Expressed in Preimplantation Mouse Embryos," J. Biol. Chem., in press (1986), by P. Singer, K. Trevor, and R. Oshima (La Jolla Cancer Research Foundation).

"Common Evolutionary Origin of Hepatitis B Virus and Retroviruses"

Dr. William Robinson, a Professor at Stanford University, was among the first researchers to join

BIONET. Since the acceptance of his application on October 31, 1984, he and Roger Miller, a post doctoral fellow, have used over 64.5 CPU hours in 540 connect hours on 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, Drs. Robinson and Miller 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, Dr. Robinson and Dr. Miller 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.

Although BIONET is not the only source of DNA and protein analysis programs, it was the primary resource for Dr. Robinson and Dr. Miller. 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 function 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, hydropathicity plots and secondary structure prediction. Dr. Miller attributed their use of BIONET as a resource for their DNA and protein analysis programs to the system's comprehensive suite of software and its inexpensive and easy access.

When asked which components of the resource he thought needed attention, Dr. Miller felt the documentation was a problem for computer naive users. We feel this shortcoming will be addressed with the implementation of more training programs and the addition of an example manual for new users (see Subsection III.A.4). Dr. Miller was, however, particularly positive about the electronic mail and bulletin board facilities on BIONET. As a result of interactions on BIONET, the Robinson lab traded unpublished hepadna virus sequences with several other labs on the system. In addition, Roger Miller actively participated in the community bulletin boards. For the benefit of other BIONET users, Dr. Miller reviewed and documented how to use Michael Zuker's program for predicting RNA secondary structures, BIOFLD. He posted his review on both the BIONET-NEWS and the CONTRIBUTED-

SOFTWARE bulletin boards, and it has proven to be one of the most widely read and used bulletins (see Subsection III.A.2). It is interesting to note that Roger Miller was a novice computer user before joining BIONET. He had previously used a personal computer for only a month.

"Molecular Cloning and Characterization of the Endo B Cytokeratin Expressed in Preimplantation Mouse Embryos"

Dr. Robert Oshima of the La Jolla Cancer Research Foundation in La Jolla California joined BIONET on December 31, 1984. Since that time, he and a Post Doctoral Fellow in his lab, Phillip Singer, have used over 9 CPU hours in 190 connect hours on BIONET in their research of Endo B filament proteins.

Much of the research for this paper had been done prior to the lab's BIONET access, but Dr. Singer used BIONET extensively for the confirmation of their research results. Having already constructed a cDNA clone (Endo B) from parietal endodermal mRNA, Drs. Oshima and Singer characterized their cDNA clone through analyses of amino acid composition of Endo B, open reading frame positioning, and homology comparisons against other known filament proteins and type I keratins.

Dr. Oshima and Dr. Singer used the QUEST and IFIND programs for retrieving known type I keratin sequences and aligning those with the Endo B protein. These comparisons, along with additional searches using the PEP program, confirmed the investigators' findings that the Endo B protein had a 54-68% homology with type I keratins. Dr. Singer also used PEP to further confirm their characterization through hydropathicity and secondary structure analyses. A final search over the Genbank and EMBL databases revealed approximately three to five genes in the mouse genome that showed homology to the Endo B cDNA.

Unlike the Robinson lab, Dr. Oshima had additional sources for computer analysis programs: Roger Staden's nucleic acid analysis programs on the U.C.S.D. computer and the PCS and Schwindigger programs for the IBM personal computer. Because he had prior experience with analysis software, Dr. Singer used the programs in a more sophisticated way, as reflected by his questions to the BIONET Scientific Consultant. He felt that the load on BIONET was unfavorably high during midday hours, but has since resorted to doing his searches in batch to alleviate that problem. However, after doing much of the sequence assembly for sequences reported in the above paper by hand, the Oshima lab now uses BIONET extensively, specifically the GEL program. Because Dr. Singer feels it is useful and convenient to use one set of compatible software when initiating a new research project, he and other members of the lab are using BIONET for the cloning and characterization of an analogous human keratin. They expect this paper to be completed before the end of the year.

"DNA Sequence Homology between the Terminal Inverted Repeats of Shope Fibroma Virus and an Endogenous Cellular Plasmid Species"

Grant McFadden, an Assistant Professor at the University of Alberta, was accepted to BIONET on November 30, 1984. This group, particularly Chris Upton, a post doctoral fellow, has been the heaviest user of BIONET, using the system for over 152 CPU hours in 725 connect hours. Their area of research is the molecular organization of the Shope fibroma gene.

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, and the exchange of genetic information between host cells and cytoplasmically replicating poxviruses. The investigators used BIONET exclusively for their computer analysis.

The investigators 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 the extracellular DNA and a family of cellular protease inhibitors.

The McFadden lab is a small lab which, Dr. Upton said, could not afford adequate computer resources without BIONET. His major complaint was the inability to search the entire database, sequences and comments, in one run, a problem that arose from the recent large increases to the databases. Obtaining a solution to this problem is among IntelliGenetics' top priorities. In addition to having access to analytical programs, Dr. Upton is pleased with the opportunity to communicate with other scientists. Like Roger Miller from William Robinson's lab, Dr. Upton had little computer experience prior to BIONET. He has also 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. Dr. Upton 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.

Our primary service goal at BIONET is to provide a full range of computational tools and a high level of user support. BIONET was also designed to enhance communication among the research community. The results of the above investigators' use of BIONET demonstrates that all Resource components are important in effective use of BIONET for scientific research. All three lab groups stated that they plan to use BIONET even more actively than they had in the past, which was also one of the most frequent comments from the BIONET reapplication forms. In our opinion, we are meeting our service goals and BIONET is becoming established as a useful and effective molecular biology computing resource.

III.A.1.c. KERMIT Lending Library

The BIONET Resource has been providing access to Kermit software and documentation through a lending library. We have provided this service because of the tremendous number of BIONET scientists with personal computers who wish to access the Resource and require terminal emulation and file transfer. KERMIT provides both.

Kermit is a protocol for transferring sequential files between computers of all sizes over ordinary asynchronous telecommunication lines. Developed at Columbia University, Kermit is non-proprietary, documented, and in wide use. It is presently available for more than 100 different machines and operating systems and additional versions are under development. Because of its wide acceptance, ease of use and minimal cost, we promote the use of Kermit by making it readily available to our users for copying.

We have responded to approximately 70 requests for Kermit software since February, 1985. The lending library provides diskettes and documentation for duplication for the Apple II, IBM-PC and its lookalikes, Macintosh, and the TRS-80. The user can also download the Kermit files directly from the DEC 2060 for other systems. As the popularity of Kermit increases in the BIONET community, we expect a heavier demand for this software. Some users may prefer to purchase Kermit diskettes and documentation (at a minimal cost) from outside vendors rather than duplicate and return our copies. Columbia University is now selling diskettes for some systems at \$10.00 a copy and there are other companies selling diskettes for other systems. As more information becomes available, we will pass it along to our users.

III.A.1.d. Subscription Fee and Its Effects

The rationale and policies for a subscription fee for telecommunication access to BIONET were introduced at the beginning of Section III.A. Class II and III users are exempt, as are foreign users who pay their own telecommunication costs.

There was no way to determine in advance the effects of this fee on the community. Although the community was polled by electronic mail, few replied and only one or two had serious concerns. From the beginning, we established the policy that no one would be prevented from accessing the Resource because of inability to pay. However, we have also requested that those who cannot pay give us a target date (e.g., for pending grant renewals) when payment might be possible.

We have looked at statistics on the community and its use of BIONET to determine the effects of the fee. The pattern of use of the system as measured by numbers of cpu minutes and connect hours consumed is discussed in detail in Paragraph III.A.5.b. Briefly, we saw no significant decrease in use of the system during the initial period when the fee was imposed that we could ascribe to the fee itself. Summer vacations are a better explanation for a slight decrease in use, because the use has again begun to climb dramatically.

More revealing are the effects on the numbers of PI's accessing the Resource. As summarized in Table II-1, of the 537 Class I PI's who have been accepted, only 27 refused, in writing, to pay the fee. However, nearly 200 PI's have never contacted us at all, so their accounts have been frozen awaiting further word. This was actually not surprising, because only about 65% of the PI's accepted ever used the system even before the fee was imposed. Those who declined and those who have not contacted us have been, almost without exception, very minor users of the system if they logged in at all. Almost 300 PI's have paid the fee or have made arrangements with us that the fee will be paid in the near future. 28 are exempt (Class II, foreign, NAC), giving us a total of 322 active PI groups as of December 1, 1985. Thus, the active community represents about 60% of the total users accepted over the last 15 months. Some percentage of the 40% who are not active users were probably deterred by the fee, but the precise number is difficult to determine. Those who remain find the system useful for their research, and many have told us that the fee is minimal compared to the services available on BIONET.

III.A.1.e. Reapplication Procedure

One of the original policies of BIONET was that each PI should be required to reapply each year. We have devised a simple procedure that collects information from each PI which we are required to report as part of our Annual Report to the NIH, especially a publication list and rationale for use of BIONET. At the same time, we ask for updates on information supplied previously to determine eligibility for access. We also ask for changes in status of funding, changes in the PI's user group, and so forth. At the request of Dr. Lederberg of our National Advisory Committee, Each PI is again required to supply his or her signature and that of a responsible administrative official,

We have followed this procedure for the preparation of this Report. The reapplication form is shown in Appendix IV. The publication list in Section II.B was compiled from the responses. Along with collection of factual information, we also asked for information on how BIONET was used and for comments on the Resource.

Use of the Resource, by Scientific Category. The data collected from the reapplication forms on how BIONET was used is shown in Table III-3. Although not all scientists indicated their patterns of use, and some had not had access long enough for them to comment, the patterns of reported use are almost exactly the same as the pattern of queries to the Scientific Consultants (compare to Table III-2).

Comments on the BIONET Resource. 34% of the reapplication forms had some type of comment about the Resource. The majority of comments were favorable; many users felt that the programs and support were excellent. The most frequently cited complaints were about the program documentation and the telecommunication networks. The complaints about the telecommunications centered around slow response time, unexpected disconnections, and difficulty in accessing the BIONET computer. The

Table III-3: Breakdown of Program Use, by Category

Scientific Category	Program Usage	Percent Of Total
Sequence and Gel Data Entry and Manipulation	35	10
Consensus Sequence Determination	26	7
Protein and DNA sequence Manipulation	121	33
Database Searches/Sequence Alignment	123	34
Experiment Planning	9	2
Communication with the BIONET community	32	9
Other	17	5
TOTALS	363	100

reapplicants also included many useful suggestions on programs or features that they would like to see included in the BIONET Resource. The comments from the reapplicants indicate that the users are actively participating in BIONET and are willing to help us better serve the BIONET community.

The specific complaints and suggestions are already being incorporated into our plans for next year. For example, complaints about the manuals (primarily that they are not designed for the naive user), have already prompted new directions in our Training program (see Subsection III.A.4). We have solved most of the network problems, but this is a continuing challenge. Limitations to access have been alleviated somewhat, as we have gone from eight UNINET network ports to twelve.

The suggestions about new programs give us some specific guidance on areas to which we are directing our Core and Collaborative Research efforts. Our Core project on hardware serial search machines is prompted by complaints about the cpu time required to search the growing databases (see Core Research, Subsection III.A.3). Requests have been made for programs for multiple sequence alignment and methods for investigating evolutionary relationships among biological sequences. The former is already being addressed (see Collaborative Research, Subsection III.A.2). The latter has resulted in us approaching scientists with programs for this application as potential collaborators on BIONET.

III.A.2. Collaborative Research

Despite our budgetary restrictions, we have been able to make some progress in our Collaborative Research program. The most important step we have taken with respect to our program was to modify slightly our criteria for membership in Class II, our category for collaborative researchers. When BIONET was first established, the intent of Class II was for persons doing substantial development work on BIONET. We have decided this requirement is too strict for the following reasons:

- There are many programs of potential value to the BIONET community that have already been developed on other systems. Developers of these programs are willing to contribute them and do the work necessary for them to run on the 2060. Giving such contributors, selected for their community spirit, demand for their software and its complementary nature to other software available, Class II status seems to us eminently reasonable.
- Given our limited staff, it makes more sense to support a larger number of contributors who face only program compatibility problems, rather than a smaller number of developers each of whom might need substantial staff support. In any case, we have received several reasonable proposals from contributors and very few from developers.

For these reasons, we are entertaining, and have begun reviewing and accepting, proposals for Class II access from those who wish to contribute software to BIONET.

A second important step has been taken through the establishment of joint accounts at other molecular biology computing resources. We can now communicate through electronic mail with MBCRR, GenBank, and the PIR (see T. Smith, W. Goad and W. Barker in the list below). When our ARPANET connection (see Subsection III.A.3) has been established, communication and file transfer will be much easier, and we look toward direct exchange of programs and data over the network.

The following is a summary of our Class II community as of December, 1985. As of this date, this community has used about 8050 cpu minutes of computer time, and 1450 connect hours to BIONET. These figures represent about 6% and 5%, respectively, of the total BIONET Class I-III use of the system. These figures are indicative of the facts that several collaborators have just been accepted and have not yet contributed their software, and the others have primarily contributed their software or data and have required little development time.

M. Kanehisa/NIH. Dr. Kanehisa has contributed his IDEAS (Integrated Database and Extended Analysis System for nucleic acids and proteins) software. This suite of nine programs is a partial implementation of his VAX/VMS version of IDEAS, and contains eight programs for homology searches, three of which allow rapid database search, and an RNA secondary structure folding program. He installed these programs on the DEC-2060 with our help, recompiled all the software, and pointed all programs to our standard set of database directories and files. He has posted a message describing the availability of the programs on the BIONET-NEWS bulletin board; the bulletin has subsequently been

moved to the CONTRIBUTED-SOFTWARE bulletin board. This bulletin has been followed by a review by BIONET staff members Drs. Azhir, Brutlag and Kedes, posted to the same bulletin board. This review pointed out some strengths and limitations of the programs, and contained some helpful hints on their use. Our records show that one or another of his individual programs have been accessed about 170 times since they were made available in April, 1985. This represents significant use by the community.

M. Zuker/NRC, Canada. Dr. Zuker contributed his program BIOFLD for RNA secondary structure folding to BIONET in March, 1985. He announced its availability through a message to the BIONET-NEWS bulletin board; this bulletin has subsequently been moved to CONTRIBUTED-SOFTWARE. R. Miller of W. Robinson's group at Stanford volunteered to review this program for the community. He has since posted a series of bulletins giving information on use of BIOFLD, suggestions on use of parameters and observations about the behavior of the program. BIOFLD has been accessed over 170 times by the BIONET community, representing significant use. Recently, Dr. Zuker has announced a PC version of BIOFLD to the BIONET community, and provided some preliminary documentation and instructions on obtaining a copy of the program.

D. Brutlag/Stanford. Dr. Brutlag and his student, B. Siegel, have begun a project to extend a program called MULTAN, for MULTiple nucleotide sequence ANalysis. This program was developed originally by B. Bains at Stanford, and will be available to the BIONET community. New developments will include translation into a more portable language, exploration of additional heuristics to improve upon initial selection of consensus sequence, improvement in the program's accuracy and application to analysis of polypeptide sequences.

H. Ginsburg/Minnesota. Dr. Ginsburg, in the laboratory of R. Dale, has begun a study of computer-based approaches to maintaining large collections of strains. This study will be carried out in the LISP programming language because of the utility of list representations to manipulations of strains and their genetic markers.

W. Pearson/Virginia. At his request, Dr. Pearson has licensed to IntelliGenetics, on a non-exclusive basis, the recent Lipman/Pearson "DFASTP" program for rapid protein homology searches. He has worked closely with us on producing a DEC-2060 version in the KCC C language compiler (see Paragraph III.A.5.d). The program is now running on the 2060 and will be released to BIONET after additional testing is performed. He has modified the program to read directly the original format of the Protein Database, maintained on the 2060 in the <NBRF> directory. We will continue working closely with Dr. Pearson as we extend DFASTP and produce the necessary documentation, but support for this version will be supplied by IntelliGenetics and BIONET.

D. Mount/Arizona. Dr. Mount originally proposed to make his PC software package available to the

BIONET community through down-loading of software to PC's. However, the slowness of file transfer programs such as Modem and KERMIT at 1200 baud make the time required prohibitively long. Recently, the Molecular Biology Computer Research Resource (MBCRR) at Dana Farber has begun floppy disk export of Mount's software. A bulletin to that effect was posted on BIONET-NEWS and has subsequently been moved to the CONTRIBUTED-SOFTWARE bulletin board. Thus, we do not expect to continue with our earlier plans to distribute the software directly from BIONET, but will direct requests to the MBCRR. We have written a letter of support to Dr. Mount for his application for a molecular biology computing resource. We feel that close collaboration among resources is essential to avoid duplication of effort.

T. Smith/Dana Farber. Dr. Smith, Director of the Molecular Biology Computer Research Resource, has been granted Class II access to BIONET by courtesy. This was done to facilitate cooperation and collaboration between BIONET and the MBCRR. Dr. Smith uses the bulletin board system on BIONET to announce availability of new software and data on the MBCRR system. For example, the Workshop on Problems in Genetic Sequence Analysis, scheduled for August, 1986, was announced to the BIONET community this way. Recently, the MBCRR has contributed to BIONET a version of the NBRF protein database restructured into functional categories. For example, all DNA-binding proteins, all immunoglobulins, and all cytochromes are grouped in individual files, and the files are in the standard format for use in the Core Library programs. We are currently testing the database prior to release of it to the BIONET community.

C. DeLisi/NIH. Dr. DeLisi has proposed contributing software for prediction of higher-order protein structures. Currently, the programs he feels are of most importance are still under development on his DEC-VAX facility.

G. Rose/Pennsylvania State. Dr. Rose has recently been accepted as a Class II collaborator and will be contributing software for protein secondary structure prediction.

C. Lawrence/NYS Dept. Health. Dr. Lawrence has recently been accepted and will be contributing software for statistical analysis of molecular biological data. He requires access to a library of statistical routines on BIONET, and the IMSL package of subroutines for statistics has been ordered for him and for other persons requiring access to these tools.

G. Stormo/Colorado. Dr. Stormo has recently been accepted and will be contributing software for quantitative sequence evaluation, analysis of binding sites, and sequence "landscapes" to display patterns of strings shared by two or more sequences. The last application represents another approach to solving the multiple sequence alignment problem.

W. Barker/NBRF,PIR. Dr. Barker is Director of the Protein Identification Resource (PIR), and has been given Class II status by courtesy. She represents our liaison with the PIR community.

W. Goad/Los Alamos. Dr. Goad heads the Los Alamos efforts related to the collection of nucleic acid sequence data for GenBank. He has been given Class II status by courtesy to foster communications with the GenBank Resource. He also collects sequences from BIONET submitted to him by electronic mail. He will contribute to BIONET programs for form-driven entry of sequences so that community members can submit their data in the correct format directly to GenBank.

R. Roberts/Cold Spring Harbor. Dr. Roberts is a member of our National Advisory Committee, so is grouped on the system in that category. However, he has spent a substantial amount of time working with BIONET on automated methods for updating his restriction enzyme database. Recently, he was able to transfer to us the latest version of this database in a format directly compatible with the Core Library software. Work remains to be done on automatic sending of messages about updates, and automatic logging of changes and testing of the new file, and we will assist him in completing these tasks. The goal is simple. We want BIONET scientists to have access to the latest data on restriction enzymes, rather than having to wait many months for its appearance on-line. Separately, Dr. Roberts is supplying a file of commercially-available enzymes, and we have already organized that into a form such that a user can programmatically select just those enzymes available from a selected supplier.

III.A.3. Core Research

Because of budgetary restrictions and the almost complete devotion of BIONET personnel and resources during the previous year to developing and consolidating the service, training, and collaborative components of the resource, Core Research has been limited to detailed planning of two major research goals for the next year of BIONET operations:

- **Hardware Text Searching Machines.** We are investigating specialized text searching hardware to optimize biological database searching;
- **BIONET Satellite Program.** We are investigating both hardware and software methods for the networking of BIONET with other regional, national, and international biologically-related computational resources.

III.A.3.a. Hardware Text Searching Machines

A common operation on BIONET involves the searching of one of the major nucleic acid or protein sequence databases for specific patterns of nucleotides or proteins. The Core Library of software has two programs that access these databases. The first is IFIND, which searches the database for sequence homologies using a specific query sequence. The second is QUEST, which is a sequence database search and retrieval program. QUEST uses a finite state machine that allows complex, often ambiguous

patterns to be found in a database. Searches using either program against a large database such as the rapidly growing GenBank may require execution times ranging from cpu minutes to hours. Indeed, the growing use of batch jobs during nights and weekends (see Paragraph III.A.5.b) is a measure of the time required. Such searches represent a major use of cpu time on BIONET. Anything that can be done to reduce this time is not only scientifically interesting, it is essential in freeing up time for other scientists to perform their computations.

Recent hardware developments have led us to believe that we can vastly decrease the search time for complex patterns in QUEST. Such hardware may also increase the speed of the first phases of IFIND searches, and this application will be pursued after QUEST. One device, the Fast Data Finder (FDF) produced by TRW, Inc., can pass an entire database as one long character string through pattern matching hardware at a rate between 7 and 9 million characters per second. The databases are stored on a Fujitsu 2350 hard disk (474 Megabyte unformatted) driven by a Concept 21 disk controller which allows the formation of a very rapid data stream by interleaving data from several disk reading heads on the Fujitsu simultaneously. This multiplies the fundamental disk streaming rate from 1 to 1.5 megabytes per second per head, up to 7 to 9 megabytes per second, the limit of the FDF hardware. Transient rates above 10 megabytes per second are buffered in cache memory. The implications of these speeds are profound. For example, the GenBank database of nucleic acid sequences is now 12-14 Mbytes, including all comments. The FDF is capable of searching this database in 1.5 - 2 seconds.

The pattern to be found is stored in a series of cells in the FDF, one character per cell, and the data stream is passed through this series of cells. As the stream is passed from cell to cell through the FDF it reports a hit on the target when the pattern in each cell matches. The minimum number of cells (we are proposing initially 1,000 cells with the ability to upgrade to 10,000 cells in one year) would allow a maximum target size of 1000 characters. Much of the standard QUEST search key syntax (strings, ranges, fixed and variable length don't cares, Boolean relations etc.) is already built into the FDF hardware so that a straightforward translation of QUEST keys to FDF syntax is possible. We are proposing that TRW provide us with translations from our current pattern matching language into their syntax and also provide us with access to their Programmer Interface Language for interacting with the FDF. This will allow us to emulate QUEST in the easiest fashion.

The FDF has several advantages over the current QUEST program. First, the cells in the pattern matching hardware can be subdivided so as to search for several patterns simultaneously (maximum 248 patterns and each pattern utilizes a minimum of 24 cells although the patterns themselves may be smaller than this). Secondly, the FDF also allows up to seven mismatches within a defined character string within the pattern. These abilities to search for many patterns simultaneously and to permit mismatches in strings will allow the future development of DNA sequence alignment algorithms including rapid

searches for homologies that including indefinite insertion/deletion gaps in addition to mismatches. For this later important application (after one year of use) we will need additional pattern matching cells, preferably near the hardware limit of 10,000. A further important property of the FDF is the ability to report regions of high density of specific sequence patterns. This has long been a major aim for QUEST development.

We intend to use our standard QUEST program on the BIONET DEC 2060 as the interface for users to prepare their search keys and to specify the database to be searched. We would also like the physical interface between the FDF (currently integrated with a SUN workstation) and the DEC 2060 to be flexible enough so that the FDF could be driven by identical software running on a VAX or on a SUN, the two other major machines that run QUEST. The simplest solution to this would be for the FDF to receive patterns and return results via a SUN based Ethernet connection.

The eventual goal is for the QUEST program to recognize when a proposed search will take more than a few moments of elapsed time and then ship the request over the Ethernet to the FDF hardware. The results of the search will be passed back to the QUEST program, so that the scientist using QUEST need make no special provision for long searches.

III.A.3.b. BIONET Satellite Program

We have begun the BIONET Satellite program in earnest. This program has the goal of distributing the BIONET Resource among computers throughout the academic community, while at the same time establishing better communication links among BIONET, its Satellites and other computing resources in molecular biology. Descriptions of the program with a more detailed statement of goals and objectives can be found in Appendix VI. As can be seen from the Appendix, the actual software license is a business arrangement between the Satellite institution and IntelliGenetics. BIONET's responsibility is to forge the communication links to ensure that scientists can communicate easily with one another.

We have previously described the initial, collaborative arrangements established with other resources in Subsection III.A.2. This is the first step toward the goal of linking the Resources. We currently have a Satellite established at the Salk Institute, and will soon establish two others, one at the US Department of Agriculture, the other at Fort Dietrick (US Army RIID).

We are following two approaches to communication with other facilities, ARPANET and a phone line based network that we are simply calling the BIONET Network for the moment.

ARPANET. BIONET has arranged Internet access to the ARPANET through a DARPA-funded project with IntelliCorp. In exchange for our assistance with the mechanics of the connection to ARPANET, BIONET will be able to make use of this connection for communications, especially electronic mail.

DARPA approved the IntelliCorp connection in October, 1985. We expect our connection to be operational in April, 1986, following the necessary lead times for the leased line provided by DARPA according to government procedures. Network services available through ARPANET include file transfer, mail, virtual terminal service, and others.

Since there are mail gateways from the ARPANET to other communications networks, this connection will do much to expand BIONET's reach. Most notably, mail interchange will be possible both with BITNET which includes EARN in Europe and with the NSF-originated CSNET. BITNET/EARN was undertaken collaboratively by a number of Universities with some help from IBM.

Additionally, since the ARPANET uses the TCP/IP internetwork protocols, a great many other networks with gateways to ARPANET will be fully accessible as well. These include the MILNET and local area networks at many major universities and research centers around the US and even in some foreign countries.

BIONET's central DEC-2060 resource will need to be connected to an IntelliCorp local area network which will in turn be a part of the Internet which includes ARPANET. This will mean that we will have to license software for the TCP/IP protocols for use on the 2060, and obtain an Ethernet interface to the local area network. At the same time, the IntelliCorp DARPA contract is purchasing the necessary gateway which will connect the IntelliCorp network to the leased line provided to an ARPANET network node processor, or IMP. The bandwidth of the ARPANET connection will be 56 kilobits per second, which will of course be shared by BIONET with the IntelliCorp DARPA users.

BIONET Network. As in the case of BITNET and CSNET, the ARPANET will form only a part of the communications backbone for the BIONET Network. The anticipated BIONET Network sites, or satellites, will vary in size and funding and an economical communications option is needed. We are currently examining options for hardware and software to provide this service.

We anticipate that asynchronous dial-up modems will be used to provide the economical link. As CSNET-RELAY does in CSNET, the BIONET central DEC-2060 resource will serve as the relay host for communication between BIONET's network sites. Most of the BIONET satellites are expected to be some model of the DEC VAX computer. BIONET has no-cost access to a MicroVAX II at IntelliGenetics and will develop the mechanism for mail exchange on this computer. We may wish to add a cache buffer memory to the DEC-20 front-end processor in order to increase the throughput possible for such communication.

III.A.4. BIONET Training Program

III.A.4.a. A Brief Review

The training program for BIONET has been severely restricted this year due to the budget cuts mentioned previously. However, we have been able to perform some trainings and demonstrate the use of BIONET at several national and regional meetings of molecular biologists. The presence of BIONET at meetings is not training in a formal sense, but there were many opportunities to answer specific questions and demonstrate use of BIONET for specific problems. These meetings also provide opportunities to inform potential BIONET applicants about the Resource.

The following summarizes our previous activities and those planned prior to the end of the current grant year.

- **FASEB Meeting.** The Federation of American Societies for Experimental Biology meeting was held April 22-25, 1985, in Anaheim. At this meeting we made a formal presentation about the BIONET Resource, in a *Workshop on International Genetic Sequence Resources*. In addition, we participated in a booth, jointly sponsored by IntelliGenetics and BIONET.
- **Rutgers/Waksman Institute Workshop.** A workshop, entitled *INTRODUCTION TO BIONET: A National Computer Resource for Molecular Biology* was held under the auspices of the Waksman Institute of Microbiology, at the Piscataway campus of Rutgers University, June 17-19, 1985. There were two parts to the Workshop, a one-day lecture program on June 17, attended by 79 persons, followed by two additional days for 23 people, all of whom attended the first day. The program for this Workshop is shown in Appendix V. The two-day session allowed all attendees access to terminals connected to a DEC-2060 machine at Rutgers running the Core Library software and emulating the BIONET bulletin board and electronic mail systems. The reports from all attendees on their reactions to the training were extremely positive. All left feeling they know much more about the use of computers in molecular biology in general, and the use of BIONET in particular. The most frequent negative comment was that there was too much material covered in the one-day session.
- **NATURE Meeting.** The *NATURE* meeting entitled *Update in Molecular Biology* was held October 7-9, 1985 in San Francisco. IntelliGenetics and BIONET jointly sponsored a booth at the show.
- **BIOTECH '85.** The BIOTECH '85 International Conference and Exhibition was held October 21-23, 1985 at the Washington Convention Center, Washington, DC. BIONET and IntelliGenetics jointly sponsored a booth at the exhibition.
- **International Congress on Computers in Biotechnology.** This congress will be held January 30-31, 1986 at the Baltimore Convention Center. A talk will be presented on the BIONET Resource in a session titled "Systems and Resources". BIONET information will be available at the IntelliGenetics booth set up in conjunction with other, overlapping conferences sponsored this same week at the Convention Center.
- **Miami Mid-Winter Symposia.** BIONET will sponsor a booth at the Mid-Winter Symposia in Miami, February 3-7, 1986. We are arranging for two training sessions at the meeting, organized around new training materials discussed below.

III.A.4.b. Some Lessons Learned

The trainings at Stanford late in the first year of our grant, the training at Rutgers/Waksman and our experience in assisting the scientific community at trade shows and in our extensive scientific consulting all lead to the same conclusion. People, especially those unfamiliar with computers, get very little out of lectures on use of software. Without the ability to use a system under careful guidance, the amount of information transferred is only slightly above zero. There must be terminals and/or PC's, at least one per two trainees, access to the BIONET software and communication facilities if not the actual computer itself, and carefully chosen examples to illustrate use of both system and application software. Despite our efforts to write documentation for the new user, it is clear that available documentation and training manuals are useful only after a person has mastered some basic techniques.

III.A.4.c. A New Strategy

We are going to develop a new training program, built around examples of application of our software to problems described in the language of molecular biology. This will differ substantially from our current materials, which are focused on specific programs and what they will do, rather than on a specific problem and how to solve it. Our experience has shown us that the following kinds of topics would cover the questions asked most often (these examples are part of a bulletin that was sent to potential participants at Miami):

- BIONET: FACILITIES AND COMMUNICATIONS.
 - What programs and features are available to BIONET users: descriptions of what each is typically used for and how you can access them
 - How to master UNINET
 - How to find important information of the bulletin boards
 - How to keep your directory within allocation
 - How to send electronic mail--including how to find out who else is on BIONET
 - How to make your backspace key work
- ENTERING AND EDITING DNA AND PROTEIN SEQUENCES
 - Using the screen-oriented editors (ESEQ); deciding what type of "terminal" you are for GENED; how to move the cursor in the editor
 - How and when to use ambiguity codes
 - Entering proteins by three-letter codes
 - Creating subsequences out of known sequences
 - Selecting and saving a sequence from the database for your own use

- GENERATING RESTRICTION MAPS—FINDING RESTRICTION ENZYME CUT SITES
 - Listing all or a subset of restriction enzyme cut sites of your sequence
 - Generating restriction maps from fragment size or mobility data
 - Generating restriction maps of a given sequence
 - Creating and using an individualized restriction enzyme list
- CONSTRUCTING VECTORS
 - Locating and using existing maps of common vectors
 - Cleavage and recombination of fragments
 - Generating a cloning vector restriction map
 - Excising fragments to customize recombinant plasmids
 - Testing directional cloning and insertional inactivation in cloning vectors
- ASSEMBLING SEQUENCES TO GENERATE A CONSENSUS SEQUENCE
 - Entering gel sequence information
 - Automatically merging together data from multiple gels
 - Editing consensus sequence—how to propagate changes through to constituent gels
 - Error checking and sequence comparison
 - Handling of both dideoxy and chemical sequencing data
- SEARCHING and ALIGNMENTS
 - How to find out if your sequence is in the database
 - Comparison of your sequence vs. the entire database
 - Comparison of your sequence vs. taxonomic or some functionally similar partitions of the database
 - Explanation of indirect files
 - How to search for sequences with key words or literature references
 - What alignment methods are available, and which to use when
- OTHER COMMON ANALYSES
 - Searching for optimal regions to design probes
 - Reverse translation

- Hydropathicity plots (and what each method's graphs mean)
- Secondary structure prediction
- Calculating amino acid composition
- Translation
- Searching for dyad symmetries
- Locating internal repeats
- Calculating base composition
- FILE TRANSFER
 - How to get your PC to act like a terminal
 - How to get data to and from BIONET

III.A.5. Resource Facilities

There have been several changes in the management of and personnel assigned to the BIONET computer facilities. These changes are summarized in Section III.C, Administrative Changes. The present section is devoted to a description of the current facilities and summary statistics on use of the Resource. The statistics cover the twelve months since our last Annual Report, 12/84 - 11/85.

III.A.5.a. Computer Hardware and Telecommunication Networks

Hardware. The BIONET Central Resource Machine is a Digital Equipment Corporation 2060 computer. The configuration was augmented this year to include an additional RP07 disk drive. Rather than simply providing additional disk space, this drive allows us a fallback in the event of the failure of one of the primary RP07 drives. (This happened during the month of October, 1985, and the existence of the additional RP07 did in fact greatly reduce the necessary downtime.) The primary drives are combined into a single disk structure and must both be functional in order for the system to run. In addition, the third RP07 is used as an additional storage place for files which are not essential in a short-term fall-back operation.

The hardware configuration is as follows:

KL10-E Model R Processor:

- 2 MF20/MG20 Memory controllers
- 2 MW MG20 Memory
- .75 MW MF20 Memory
- MCA20 Cache Buffer Memory
- 2 RH20 Massbus Channels

Console and Front End Processor:

PDP-11/40 CPU, 32 KW 16 bit memory
 RX02 Dual floppy disk drives
 8 DH11 Terminal interfaces 8 * 16 TTY lines each = 128 lines
 RH11 Massbus Channel
 LP20 Line printer interface

DN20 Front End Processor:

PDP-11/34 CPU, 128 KW 16 bit memory
 DMR11 Network interface

Peripherals:

3 RPO7 disk drives 111MW each
 RPO6 disk drive 39MW
 372 MW Total disk storage
 TU78 1600/6250-BPI tape drive
 LP26 600 LPM Line printer
 Imagen Imprint-8/300 Laser Printer

Disk space (data storage)

Public structure (PS:) disk space use on the 2060 is dynamic. The following snapshot is representative of typical usage, and is taken from December 1985.

Total disk space	433,000	(pages--222 million words)
Overhead/Common	<148,000>	(Core, System and System Support Libraries)
Swapping Space	< 25,000>	
File system Overhead	< 70,000>	(Directories and index pages)

	190,000	
BIONET Allocation	95,000	(Half of the available space)
Bionet Usage 12/85	< 53,000>	

Unused space	42,000	(Available for BIONET growth)

Note that file system overhead varies greatly depending on the size of the files involved. Since BIONET users have many small files, BIONET growth may increase file system overhead, altering the above distribution.

Terminal Lines

Because the usage of a particular terminal line varies greatly, and because many BIONET users share a single line in succession, there was in the past an imbalance in the allocation to BIONET of terminal lines. However, with the departure of the IntelliCorp KSD users from the system (see Section III.C), additional terminal lines were freed for BIONET. These are not regularly needed by BIONET at this time,

but may be used intermittently or for growth. Current system terminal line distribution is as follows:

Total lines	128	
Overhead	< 10>	(Shared devices, BCRG staff)

	118	
Allocated BIONET	59	(Half of the available lines)
BIONET Users	< 18>	(Public Data Network, Local Dial-Ups)
BIONET Staff	< 6>	

Unused lines	35	(Available for BIONET growth, temporary use for trainings, replacement of a bad line before it is repaired)

Public Data Network Connection. BIONET is accessed principally over the UNINET Public Data Network. An X.25 PAD (packet assembler/disassembler) is located on-site. This is known as the Host PAD, or HPAD. It provides individual terminal ports which are cross-connected to those on the DEC-20. The Uninet trunk line operates at 9600 baud synchronously, and the PAD converts this into up to 16 asynchronous ports whose speed is typically 1200 baud. A handshaking protocol is employed to smooth over bursts of data during the multiplexing.

UNINET we originally chosen as a replacement for Telenet because of its better response time and its lower cost. The lower cost was achieved through a very favorable fixed price per port arrangement that we negotiated with UNINET. Currently 12 UNINET host ports are used by BIONET, and usage is monitored carefully in the event more are needed. The ports are accessed in sequence, with those higher in the sequence not being used while any lower port is free. The number of connect hours per month drops off after the first 6 ports. The usage on these first 6 ports therefore represents many more sessions than does the usage of ports 7 through 12. Our monitoring of the port use also has revealed that it would be cheaper for BIONET to lease the higher-numbered ports on a use, or traffic, basis. We currently are leasing 8 ports fixed, 4 on traffic, and will change this distribution as required for the lowest possible cost.

We have been examining the replacement of the UNINET-supplied leased HPAD with a BIONET owned HPAD. The consideration is the savings of lease charges while maintaining adequate reliability. We plan to make this replacement before the end of the current grant year.

III.A.5.b. Summary Statistics on Machine Use

The cpu cycles of the DEC-2060 computer are allocated to the user community, including BIONET, by the system's class scheduler. This scheduler is given the percentage of the machine to allocate to each class of users. Any cycles not consumed by a given class (*windfall*) are available to the rest of the user

community. This method was chosen so that cpu cycles not consumed by one segment of the community could be used by other segments if needed, i.e., no cpu cycles are wasted if someone needs them.

The current percentage allocations ("pieslices") are shown in Figure III-1. As summarized in the figure, BIONET Class I (and III and IV) are allocated 30% of the machine, and Class II and staff 10%. The 20% overhead (system overhead, batch and computer staff and operations) is allocated one-half to BIONET, for a total of 50%. These allocations remain the same as last year. However, there are substantial changes to the other classes of users for reasons discussed in Section III.C. Note that the BATCH class is assigned 1% of the system during prime time. In off prime time, the percentage allocation is increased substantially in response to demands by the BIONET community.

The actual use of the machine by the BIONET community is now substantially greater than 50% of the total cpu cycles actually used. As an example, the percentage use of the machine for the month of October, 1985 is shown in Figure III-2. It is clear that BIONET is receiving more than its fair share of the cpu cycles. Note that BIONET scientists' use of BATCH is charged to the individual accounts by the accounting program. Thus, extensive use of BATCH shows up in this pie chart as BIONET Class I (or II) use, rather than in the category BATCH Jobs.

The data for BIONET percentage of system use are plotted in histogram form in Figure III-3. This figure demonstrates that BIONET has consumed more than 50% of the total cpu cycles **used** (data on % of **available** are given below) on the 2060 since February, 1985, and is now consistently consuming 65 - 75% of the total cpu cycles used on the system.

In the following series of tables and figures, we provide further details on the actual use of the system by the BIONET community. Looking first at use of the system in prime time (8 AM - 8 PM, M-F, PST), data for cpu time and connect hours for the indicated segments of the community are given in Tables III-4 and III-5 by month, and totals. The cpu data in Table III-4 is also plotted in histogram form in Figure III-4. (The figures for the facilities group staff and overhead for November, 1985 are artificially low because the statistics were computed before Thanksgiving weekend, before the end of the month operator totals were added in.)

There are several important facts that can be determined from these data. Looking first at cpu time, and given that there are about 12,000 cpu minutes (total cpu minus 20% for overhead) **available** prime time in the average month for the entire system, BIONET (Users plus Staff) has been consuming well over 50% of available cycles. The category of BIONET Users (Classes I-III) compete for 30% of the machine. The class has consumed more than 30% of available cycles since March, 1985, and have thus been able to take advantage of considerable windfall.

Figure III-1: Pieslice Allocation of the DEC-2060 Computer

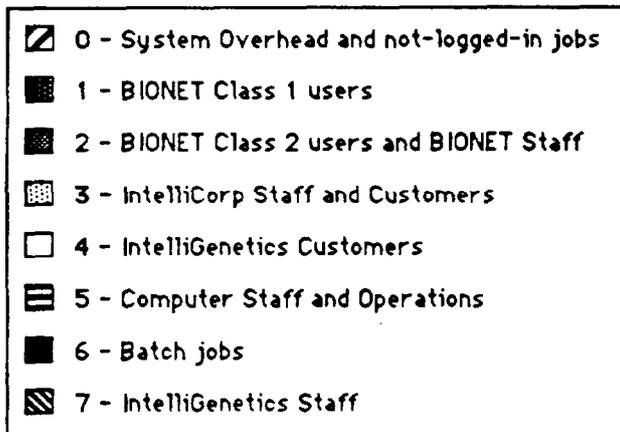
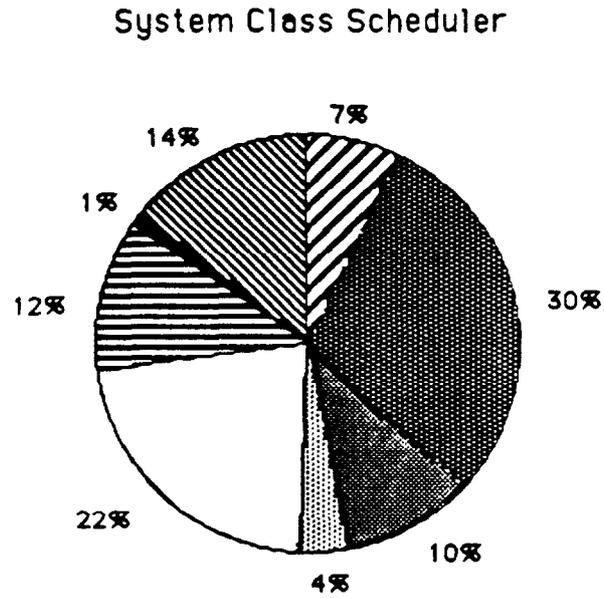


Figure III-2: Actual Use of the DEC-2060 for the Month of October, 1985

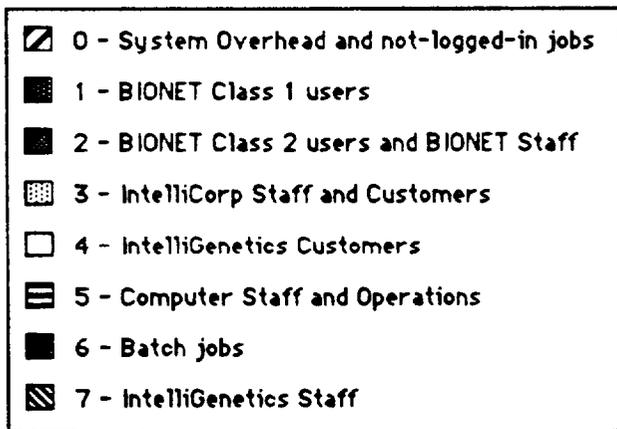
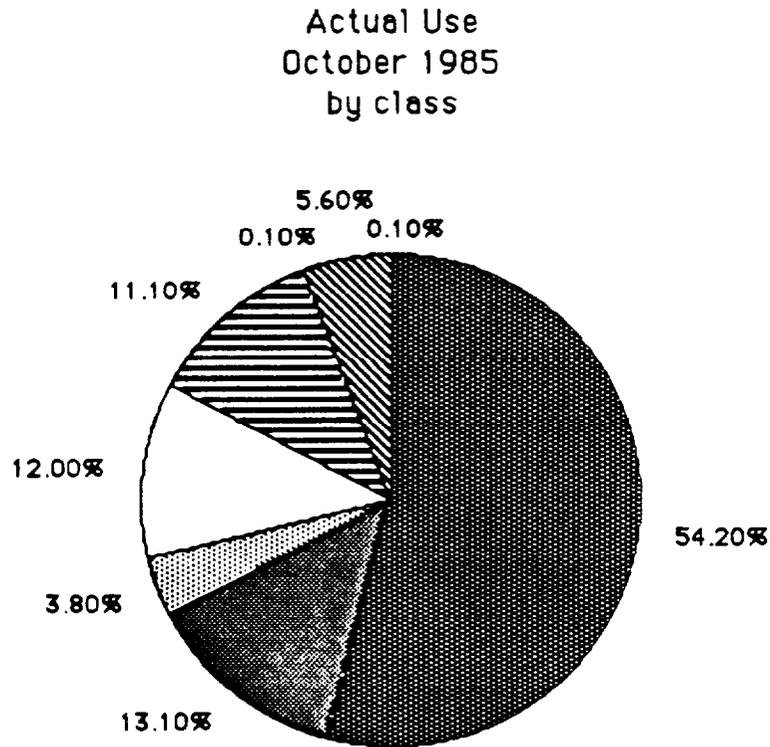
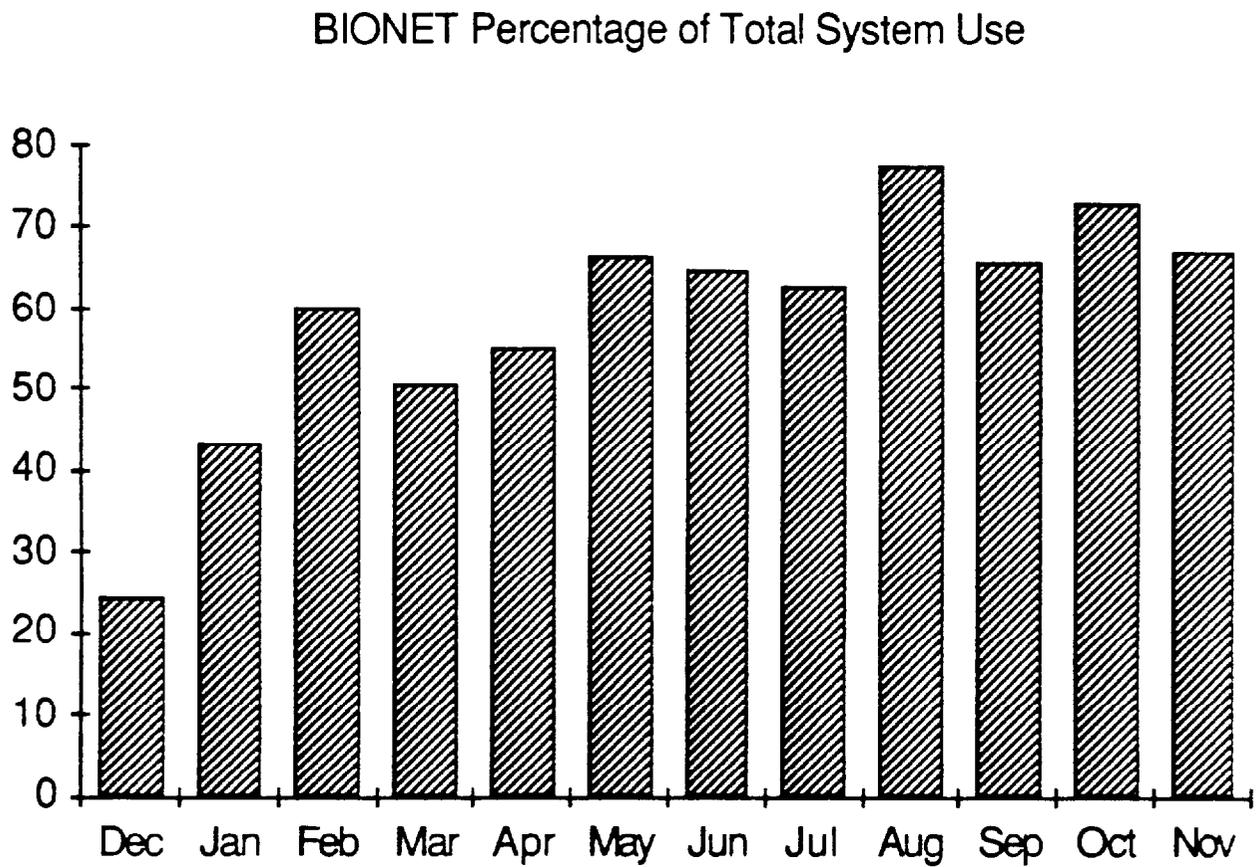


Figure III-3: BIONET's Percentage Use of the DEC-2060, 12/84 - 11/85

The total number of connect hours, prime time (Table III-5), for the category BIONET Users has remained in the range 1800 to 2200 since May, and the relationship between connect hours and cpu minutes remains relatively constant over those months.

The data for non-prime time (weekends and 8 PM - 8 AM M-F) are shown in Tables III-6 and III-7, and the data on cpu time are plotted in histogram form in Figure III-5. Particularly notable in these data are the dramatic increases in cpu time over the past year, especially in the last three months, due almost entirely to BIONET use. These increases are due primarily to the extensive use of overnight batch runs to perform time-consuming analyses involving database searches, using the IFIND homology and the QUEST database search and retrieval programs. Thus, the community has gravitated naturally toward off-hours use of these programs for such analyses.

Given that there are about 22,000 cpu minutes (total minus 20% overhead) available each month in non-prime time, BIONET (Users plus Staff) has recently been consuming more than 50% of the amount available. Given low use of the system by other classes in non-prime time, BIONET consumes most of the cpu cycles actually used during these times.

The data for total use of the Resource by BIONET are presented in Tables III-8 and III-9 and the total cpu time is summarized in Figure III-6. BIONET Users and Staff, since May of 1985, have consumed 40% or more of all the cpu cycles available on the system (total minus 20% overhead).

One important conclusion from all these data is that the Resource is rapidly approaching saturation. Certainly, during prime time, the system load is becoming a barrier to rapid computation. At this point, limitations on the number of access ports keep the load average under control by limiting the number of concurrent users. However, as we add additional telecommunication ports, we will quickly become limited by available cpu time.

Another important conclusion we have reached from these data is in regards to the effects of the subscription fees on use of the Resource. The total use by BIONET scientists (not including staff) increased steadily from November, 1984 through May, 1985. In the summer months of June through August, use leveled off, beginning before the subscription fee was announced, which we attribute to summer vacations more than any effect of subscription fees. Beginning in September, 1985, use increased steadily again to a level substantially above the months prior to initiation of the fee.

Summary data for use of our telecommunications network are presented in Figure III-7 by month for the past 12 months' use of the Telenet (until mid-July, 1985) and UNINET (beginning early July, 1985) networks. Three factors distort this Figure. First, the value for July is artificially high because we were running the two networks simultaneously and performing extensive tests on UNINET. Second, we noticed

Table III-4: BIONET Prime Time CPU Minutes

	BIONET Users (except staff)	BIONET staff	BCRG & System Overhead	Total BIONET Use
December	769	397	385	1551
January	2598	1054	579	4231
February	3368	1091	644	5103
March	4236	571	473	5280
April	5169	861	529	6559
May	6791	776	515	8082
June	5004	905	530	6439
July	5575	1094	564	7233
August	5132	1248	508	6888
September	4854	798	509	6161
October	6476	1330	455	8261
November	6135	473	88	6696
TOTAL	56107	10598	5779	72484

Table III-5: BIONET Prime Time Connect Hours

	BIONET Users (except staff)	BIONET staff	BCRG & System Overhead	Total BIONET Use
December	328	519	1218	2065
January	761	1164	1368	3293
February	1137	829	1340	3306
March	1206	638	347	2191
April	1452	764	1353	3569
May	2177	737	1473	4387
June	1908	577	1567	3690
July	2291	916	1661	4643
August	1846	700	1374	3767
September	1777	606	1585	3810
October	2101	763	1688	4537
November	2187	689	156	3032
TOTAL	19171	8902	15130	42290

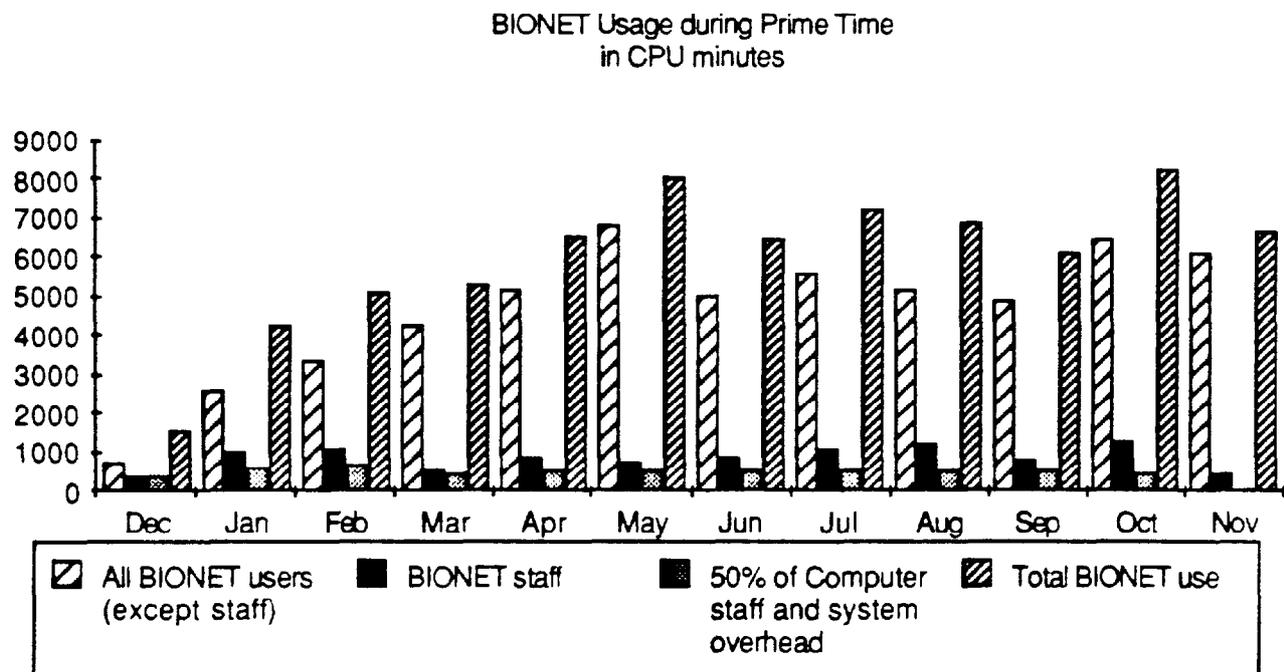
Figure III-4: BIONET's Prime Time Use of the DEC-2050, 12/84 - 11/85

Table III-6: BIONET Non-Prime Time CPU Minutes

	BIONET Users (except staff)	BIONET staff	BCRG & System Overhead	Total BIONET Use
December	366	91	225	682
January	1673	128	826	2627
February	3848	357	159	4364
March	4169	26	404	4599
April	3386	356	1370	5112
May	6777	206	1300	8283
June	6567	1129	1415	9111
July	6956	850	1613	9419
August	5396	1244	1238	7878
September	7056	1192	876	9124
October	9553	1407	1103	12063
November	12326	111	86	12523
TOTAL	68073	7097	10615	85785

Table III-7: BIONET Non-Prime Time Connect Hours

	BIONET Users (except staff)	BIONET staff	BCRG & System Overhead	Total BIONET Use
December	117	159	1751	2027
January	420	145	1749	2314
February	562	208	1680	2450
March	697	121	142	960
April	601	149	1859	2609
May	949	221	2002	3172
June	1246	194	2210	3650
July	1197	230	2519	3946
August	887	192	1843	2922
September	1109	202	2590	3901
October	1213	190	2343	3746
November	1746	173	117	2036
TOTAL	10744	2184	20805	33733

Figure III-5: BIONET's Non-Prime Time Use of the DEC-2060, 12/84 - 11, 85

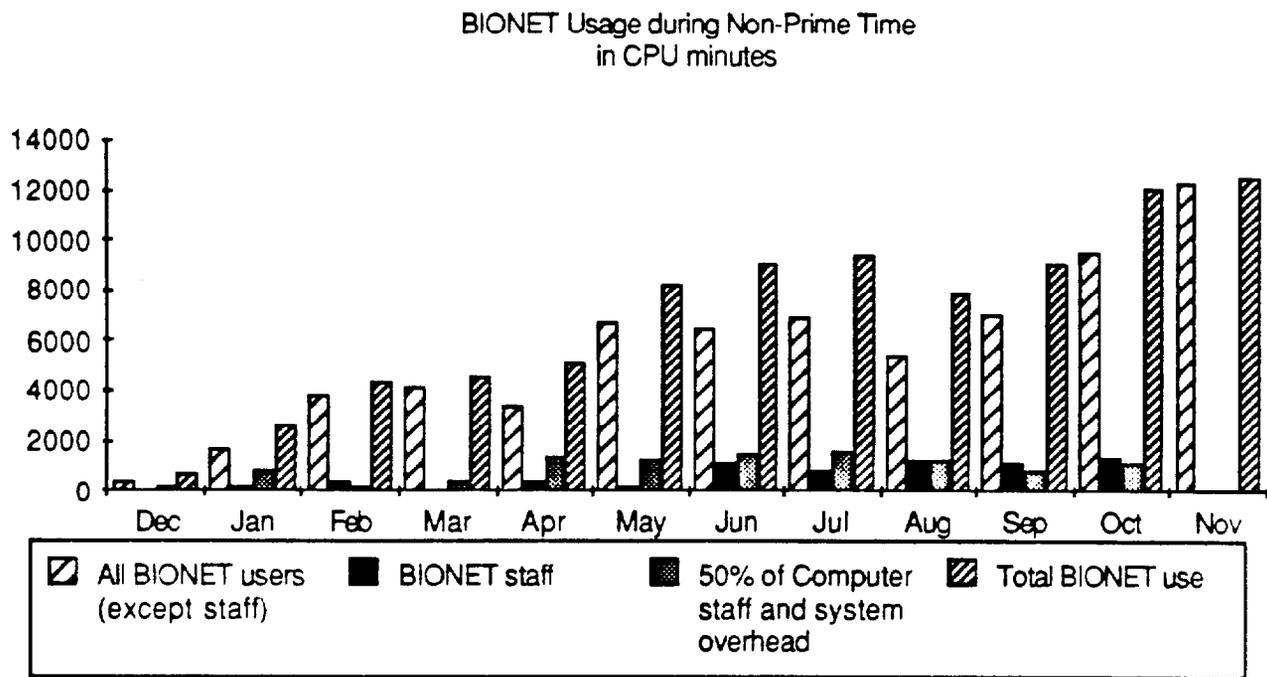


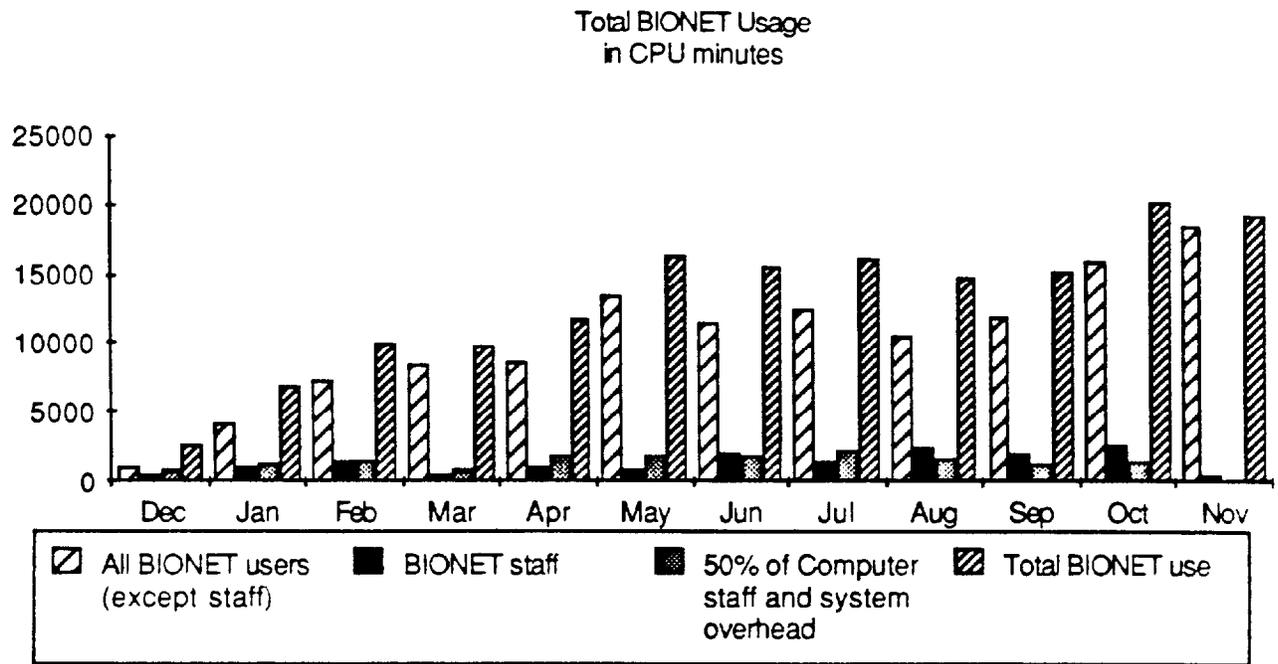
Table III-8: BIONET Total CPU Minutes

	BIONET Users (except staff)	BIONET staff	BCRG & System Overhead	Total BIONET Use
December	1136	489	1015	2640
January	4271	1182	1407	6860
February	7216	1449	1414	10079
March	8405	597	877	9879
April	8556	1217	1899	11672
May	13568	982	1816	16366
June	11571	2035	1946	15552
July	12531	1945	2178	16654
August	10528	2492	1747	14767
September	11911	1990	1386	15287
October	16029	2737	1559	20325
November	18462	585	174	19221
TOTAL	124184	17700	17418	159302

Table III-9: BIONET Total Connect Hours

	BIONET Users (except staff)	BIONET staff	BCRG & System Overhead	Total BIONET Use
December	445	678	2969	4092
January	1181	1309	3117	5607
February	1699	1037	3020	5756
March	1903	759	489	3151
April	2053	913	3212	6178
May	3126	958	3475	7559
June	3154	771	3777	7340
July	3488	1146	4180	8589
August	2733	892	3217	6689
September	2886	808	4175	7711
October	3314	953	4031	8283
November	3933	862	273	5068
TOTAL	29915	11086	35935	76023

Figure III-8: BIONET's Total Use of the DEC-2060, 12/84 - 11/85



that many users were leaving their terminals after completing their work without logging off BIONET, thereby tying up the network port and preventing other users from accessing that port. Therefore, we implemented an "idle zapper" which monitors the cpu use for each BIONET job, sends a warning message after 10 minutes of cpu idle time, and detaches the job after 5 more minutes of idle time, as a good compromise based on comments on the idea from the user community. Thus, an idle job can tie up a port for no longer than 15 minutes. The job is still available to the user, who can reattach to it and continue from where he or she left off. The zapper has been very effective in freeing up network ports. Third, the data for October, 1985 are artificially low because of UNINET network problems, which have since been resolved.

III.A.5.c. Computer Software - Core Library

Through our license agreement with IntelliGenetics, we have provided all Core Library software releases to the community. There have been two major releases so far this grant year, and another will occur at the end of January, 1986.

One important addition to the Core Library was requested by Dr. Yanofsky of our National Advisory Committee, the addition of the DIGITIZER program to the suite of software. Up until recently, access to the software to use a sonic digitizer for entry of gel data (restriction digests, sequencing ladders) has not been possible for BIONET scientists. We have made arrangements to modify the software license agreement with IntelliGenetics, and digitizer access is now possible. A bulletin to that effect has gone out to the community, and a small number of laboratories have purchased the necessary hardware to use DIGITIZER.

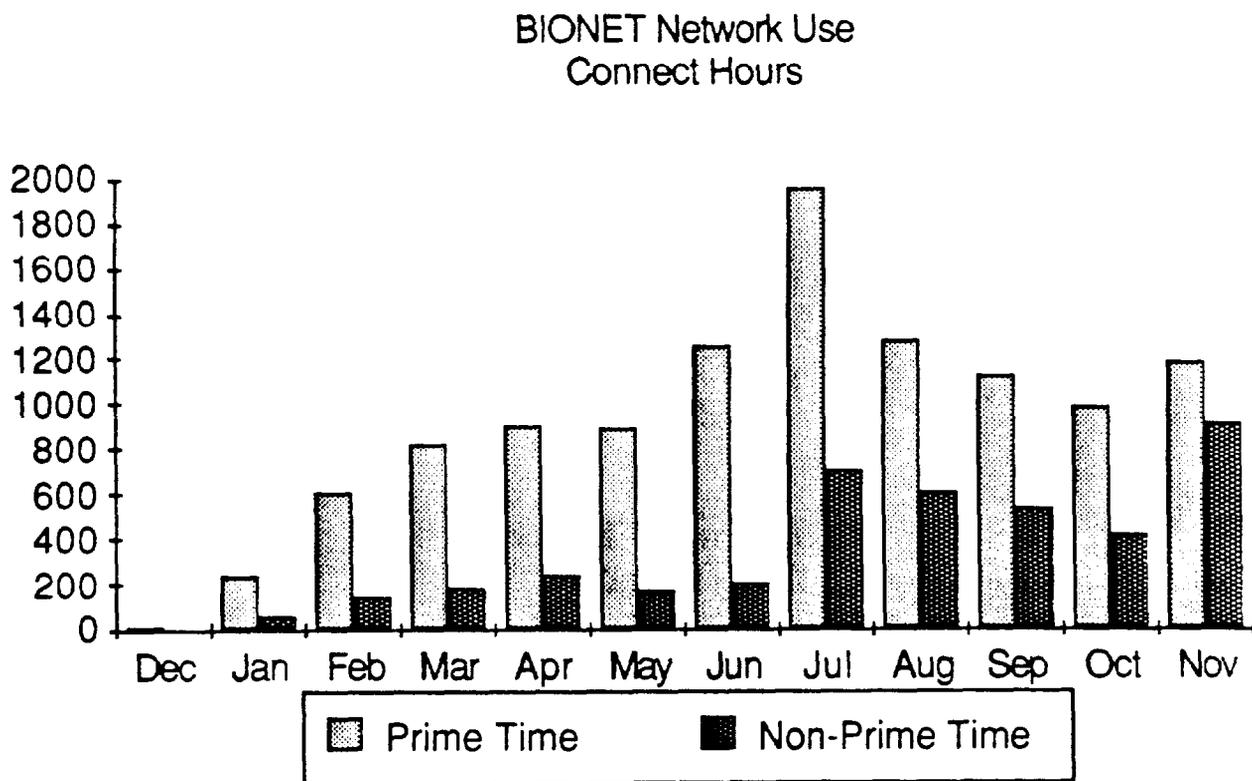
III.A.5.d. Computer Software - System Library

During the course of the year, the following additions have been made to the system support library described in last year's report.

Communication.

FINGER--Displays an information message or "plan" optionally provided by a user for other users advising them of travel itinerary or other contact information, and also displays the date the user in question was last on the BIONET system.

WHOIS--Directory lookup program for BIONET investigators. During the course of the year this utility was upgraded to have more generalized search capability and to permit searches of mixed case text. The WHOIS database of BIONET users was extended to include research titles for each PI, to enable other PI's to identify investigators with similar research interests.

Figure III-7: Total Telenet and UNINET Network Use, 12/84 - 11/85

File Search.

FIND, DFIND--These utilities were added. They permit a concise display of matching text in a file, displaying respectively either a whole paragraph or a single line containing the match. Complicated search keys may be specified and the output can be directed to a file.

System utilities.

Several of these were added: TTYINI (sets terminal parameters more accurately by default), CALC (numeric calculations), RTTY (terminal display of a file in reverse order), Dired and CLEAN (display and hard copy directory management), MPW (suggest good random passwords to users), LASTN (displays last N lines of a file).

Programming Languages.

C--We installed the KCC compiler for the C programming language from Stanford University together with a new version of the FAIL assembler.

III.A.5.e. Computer Software - Contributed Library

We have set up a special directory on BIONET, the <CONTRIBUTED> directory, as a repository for contributed software and databases from the outside community. This directory is protected so that staff and BIONET investigators have access, but other users of the DEC-2060, for example customers of IntelliGenetics, do not. This was done so that persons who wish their software to be accessible only for not-for-profit research have a mechanism to do so.

The software and databases that have been contributed are summarized in Subsection III.A.2, above.

III.A.5.f. Database Library

We maintain all major collections of biological sequence data on BIONET, including the GenBank and European Molecular Biology Laboratory (EMBL) nucleic acid sequences, and the Protein Identification Resource database of protein sequences. We also maintain the Cold Spring Harbor database of restriction enzymes (contributed by Dr. Roberts, see Subsection III.A.2). We provide VectorBanktm, from IntelliGenetics, for use in programs designed to model cloning experiments. These databases are updated immediately on receipt from the suppliers.

Where necessary for use in Core Library programs, we reformat these databases; the original databases are also maintained in separate directories for programs designed to access those formats.

III.B. Highlights

The following are highlights of the BIONET Resource's second year of operation.

- 560 PI's have been granted access to the BIONET Resource, substantially exceeding our early estimates of the size and interest of the community. These scientists have already published over 50 scientific papers in which the Resource played an important role in obtaining the results of their investigations.
- Collaborative research projects have brought to the Resource both new computer software to complement what was available already, and new databases, bibliographies, and computer-readable textbooks. These augmentations of the Resource are widely used by BIONET scientists.
- Our investigations of special computer hardware for rapid searches of nucleic acid and protein sequence databases have led to identification of and writing of specifications for a machine to be used in conjunction with existing facilities. Such a device would dramatically improve such searches, thereby increasing the amount of computer time available to other investigators.
- Our BIONET Satellite program for establishing a loosely-linked network of computers has gotten off to a promising start. The communications and networking facilities of BIONET will be used to maintain electronic mail and bulletin boards accessible by all Satellite resources.

III.C. Administrative Changes

III.C.1. Facilities

In July, 1985, IntelliGenetics acquired control of the DEC-2060 and related facilities from IntelliCorp. Thus, all facilities used by BIONET are now totally under the management of IntelliGenetics. 50% of the machine is still devoted to BIONET as part of the Cooperative Agreement under which we are funded. As discussed previously in Paragraph III.A.5.b, BIONET actually consumes substantially more than 50% of the computer (cpu) time.

As part of these changes, the Knowledge Systems Division (KSD) of IntelliCorp separately purchased several DEC-VAX systems and over the period of September and October, 1985, reduced its share of the 2060 from 20% to 4%. The pie chart describing the current percentage allocations of the system to each category was presented previously in Figure III-1.

III.C.2. Personnel

As part of the reorganization of the computer facility, several changes in personnel have taken place. R. David Roode joined IntelliGenetics in the role of Biotechnology Computer Facilities Manager. Mary Yardley became Operations Supervisor and Lauri Kanerva continues as Senior Computer Operator. Andrea Gorman and John Shelton assumed other roles in IntelliCorp. Other major changes in personnel

were brought about by budget cuts as discussed at the beginning of this report. We are currently engaged in a search for a BIONET Scientist to assume the role of Ms. Ari Azhir, who has recently left the company.

On November 15, Marcia Allen was added as a second Scientific Consultant to the BIONET staff. This was made possible by the improved financial condition of the Resource. This addition was chosen specifically to improve our user consultation services and to enable us to revitalize our training program.

III.D. Resource Advisory Committee and Allocation of Resources

Our methods of allocating resources (staff and computer time) are relatively simple. The DEC-2060 computer uses its windfall scheduler to allocate cpu time to the various categories of users and overhead, as described in detail in Subsection III.A.5. These cpu cycles are distributed on a first-come, first-served basis. Because of the limited number of communication ports into the 2060 (currently 18, see Paragraph III.A.5.a), we have asked the community not to have more than one person per PI group using BIONET at the same time during prime time. The community has done an excellent job in complying with this policy.

In October, 1985, we doubled the disk space available to each PI's group. This has eased substantially the problems several groups encountered in managing large sequencing projects or individual databases of sequences.

We have devoted most of our staff time to the Collaborative Research and Service components of the Resource. In the next year we will devote additional staff time to foster collaborations and to our Core Research activities. This will be possible because the community is becoming more sophisticated in its use of BIONET and we have already augmented our staff in support of the Service component.

On August 21, 1985, we held a meeting with "local" members of the National Advisory Committee (see below) Tom Rindfleisch, Joshua Lederberg and Charles Yanofsky, to take advantage of Joshua Lederberg's visit to the Bay Area. One question which was posed at that meeting was how fairly the Resource, in terms of computer cycles, was being distributed to the community. One answer to this has been obtained by summary statistics on the top 20 users of BIONET over the past twelve months. Our data indicate that 38% of the total cpu cycles were delivered to the top 20 users. This statistic must be interpreted in the light of two important facts. One, the top 20 users on a month-by-month basis show substantial differences, reflecting the nature of computing in the area (significant use of the computer is often followed by additional laboratory studies suggested in part by the computational results). Second, the needs of different PI's differ substantially. Those who require frequent access to the databases will use substantially more cpu time than other PI's who may be doing restriction mapping or assembly of

consensus sequences. The former group will use a disproportionate share of the Resource, and this fact is what has prompted us to seek an alternative solution by applying new hardware to serial search, as described previously under our Core Research program (Subsection III.A.3).

The last meeting of our full National Advisory Committee was held March 23, 1985. The next scheduled meeting is February 24, 1986. The current membership of the Committee is as follows:

- Professor Joshua Lederberg, MD, PhD. (Chair), President, The Rockefeller University.
- Dr. Saul Amarel, PhD., Director, Information Processing Techniques Office, Defense Advanced Research Projects Agency, Department of Defense.
- Professor Alan Maxam, PhD., Dana Farber Cancer Institute, Harvard Medical School, Harvard University.
- Dr. Richard J. Roberts, PhD., Senior Staff Investigator, Molecular Biology, Cold Spring Harbor Laboratory
- Thomas Rindfleisch, MS, Director, Knowledge Systems Laboratory, Department of Computer Science, Stanford University.
- Professor Charles Yanofsky, PhD., Department of Biological Sciences, Stanford University.
- Professor Fotis Kafatos, PhD., Department of Cellular and Developmental Biology, Harvard University.

III.E. Dissemination of Information on Resource's Capabilities

We discuss two areas related to dissemination of information about the Resource that we have pursued this grant year. The first is interactions with the scientific community through participation at conferences, advertising the availability of BIONET, and mailing information about the Resource to NIH grantees. The second is use of the electronic mail and bulletin board facilities of the Resource itself to keep the BIONET community aware of changes and improvements.

III.E.1. Community Interactions and Awareness

We have used three methods this year to inform the community about BIONET and to solicit applications for access to the Resource. The first method has been participation at major conferences, where we have presented papers and/or have had booths at exhibitions. These efforts were summarized previously under Training, Subsection III.A.4. At these conferences, we have distributed the standard application packets to scientists, after demonstrating to them the capabilities of the Resource.

The second method is advertising. Due to our limited budget, we have placed only one advertisement this year, in the special computer issue of *Nucleic Acids Research*, which appears in January, 1986. The text of the ad is provided in Appendix VII.

The third method is a mass mailing to NIH grantees whose research areas were characterizable as related to molecular biology, biological chemistry, and so forth. This list was provided to us by Dr. Charles Coulter, BRTP/DRR/NIH, who obtained the list and associated mailing labels by searching the NIH database of information on research areas of its grantees. The keywords chosen for the search were purposefully general to capture as many potential investigators as possible. Thus, there were some investigators chosen who are working in tangential areas, for example, classical genetics.

There were approximately 4030 mailing labels from this list. A brochure describing BIONET was sent to each of these investigators on November 10, 1985. The brochure itself is shown in Appendix VIII. The brochure is two-sided, and folded in half along perforations. Persons wishing an application packet need only fill out the requested information, tear off the return half and send it back to us. So far we have received 315 requests back for application forms. Many of these returns also have requested information on the BIONET Satellite program.

III.E.2. Electronic Communications

The electronic communication facilities of BIONET provide another important way to disseminate information about the Resource. In addition, electronic mail and bulletin boards provide a mechanism for scientific and technical interchanges among members of the community. Information on the types of electronic mail communications with BIONET was summarized previously in discussion of the Service component of the Resource (see Subsection III.A.1). We have also established a second mechanism for sharing information electronically, on-line bibliographies. Bulletin boards and bibliographies are discussed in the next sections.

III.E.2.a. Bulletin Boards

The electronic bulletin boards are an important component of the BIONET Resource. They provide BIONET users with a facility for the exchange of data, laboratory techniques and ideas with others of like mind. For example, a laboratory just beginning a study of the conservation of DNA sequences might experience some frustrating technical problems. A message to the MOLECULAR-EVOLUTION bulletin board, describing the problem, will probably reach and be read by more than 1000 BIONET users, some of whom will have experienced similar problems and can offer solutions. Obviously, the users represent a wealth of knowledge. Communication is the key to accessing and disseminating that knowledge.

BIONET's bulletin board system consists currently of 20 bulletin boards of varied topics. The topics were selected from user requests and from a survey of the most frequently asked questions. We have designed the system so that each BIONET user automatically receives messages that are of concern to all users, but can decide independently which other bulletin boards he or she would like to subscribe to. Subscribing to a bulletin board is an automated procedure which results in the automatic presentation of new messages

upon logging onto the system. However, all users have access to all bulletin boards, whether or not they are subscribers, through the electronic mail facility.

The bulletin board topics are as follows:

BIBLIOGRAPHIES	Instructions for using BIONET's online bibliographies that have been contributed by members of the BIONET community.
BIONET-NEWS	Information relevant to all BIONET users. All BIONET users automatically subscribe.
CONTRIBUTED-SOFTWARE	Reviews/instructions for using the software that has been contributed to BIONET by BIONET users.
EMPLOYMENT	Postings of job opportunities by BIONET users.
IMMUNOLOGY	Information/inquiries relating to immunology.
LIBRARIES	Requests/postings of availability for/of cDNA libraries or codon usage tables.
MOLECULAR-BIOLOGY-LAB-METHODS	Information/inquiries relating to laboratory techniques.
MOLECULAR-EVOLUTION	Information/inquiries relating to the study of evolutionary relationships of genes or proteins.
ONCOGENES	Information/inquiries relating to oncogenes.
PC-COMMUNICATIONS	Instructions/inquiries relating to using a PC to communicate and to transfer files to and from the BIONET computer.
PC-SOFTWARE	Reviews/inquiries of software packages for any type of personal computer.
PLANT-MOLECULAR-BIOLOGY	Information/inquiries relating to the study of plant genetics.
POLITICS	Concerns/opinions which may or may not be related to research in molecular biology.
PROGRAM-APPLICATIONS	Instructions/suggestions for using the BIONET programs for special applications or research projects.
PROTEIN-ENGINEERING	Information/inquiries relating to protein engineering.
RESTRICTION-ENZYMES	Information/inquiries relating to restriction

enzymes.

STARTUP	For new users, a quick introduction to BIONET including the most frequently asked questions from new users and their solutions.
TOPS20-HINTS	Instructions/suggestions for managing your directories, copying files, running batch jobs or any other system facilities/commands.
VECTORS	Information/inquiries relating to vectors and vector construction.
YEAST-GENETICS	Information/inquiries relating to yeast genetics.

There are about 170 different messages on the bulletin boards, 18 are suggestions for lab techniques, 17 which involve the trading of libraries, and 21 reviews of software used for molecular biology research. Surprisingly enough, the most popular bulletin boards are the PC topics, and the least used bulletin board is ONCOGENES. The number of messages reflects an under-utilization of this resource, which we will address with our more aggressive plan described below. But interviews with BIONET users and feedback from the reapplication forms indicates that the number of messages on the bulletin boards does not entirely reflect the level of interaction among members of the BIONET community. In many instances, the bulletin boards have served as a catalyst for the dissemination of information across the community.

With limited staff, it was impossible to actively promote the bulletin board communities. However, with the implementation of subscription fees and the hiring of additional personnel, we have begun to select active members of the BIONET community to serve as bulletin board leaders. These scientists will submit and solicit articles, reviews and information from the community and post them on a bulletin board. They will receive special access privileges, and will monitor and update the messages. The bulletin board leaders will also post monthly updates on BIONET-NEWS describing the new messages sent to their bulletin board. This way, every member of the community will have the opportunity to remain informed about collaborations or new developments in molecular biology without having to read messages that may not be in his or her field. It will also eliminate the problem of having messages on multiple bulletin boards. We feel that with the implementation of our more aggressive plan for bulletin board community leaders, the bulletin boards on BIONET will prove to be an even more important component for the dissemination of information among the molecular biology research community.

III.E.2.b. Bibliographies

In response to a suggestion from the National Advisory Committee, Bionet set up a procedure whereby Bionet users could contribute their personal bibliography files for the use of others. The various format-independent text searching programs available on Bionet makes this feasible. The program FIND allows users to search for particular text words or other character patterns in single files and to print out specified amounts of text around the patterns. The program XSEARCH allows users to search many files for patterns, and the program QUEST permits the combined flexibility of both FIND and XSEARCH. The most important aspect of these tools is that they are general, context-independent searching methods so that bibliographic data in almost any format could be searched.

Users were informed of these text searching tools by a BBOARD message and two sample bibliographic files were announced by the Bionet co-investigators. They included the following files:

```

PS:<BRUTLAG.BIB>
CHROMOSOME.BIB.25      ; References on chromosome structure
COMPUTER.BIB.30       ; Computer algorithms and methods
DNA.BIB.5             ; DNA structure and topology
DROSOPHILA.BIB.16    ; Drosophila molecular genetics
GENETICS.BIB.17      ; Molecular and classical genetics
METHODS.BIB.29       ; General laboratory methods
RESTRICTION.BIB.16   ; Rich Roberts restriction enzyme refs
TOPOISOMERASE.BIB.48 ; DNA topoisomerase references

PS:<KEDES>
MUSCLE.BIB.1         ; References on muscle proteins and genes

```

These files were not greatly referenced nor was there a lot of feedback from the community. One other user, (Tom Broker) volunteered a complete bibliography of work performed on papovaviruses and we are helping him mount this extensive bibliographic database.

The primary problem with this approach is that personal databases are just that, very personal, limited in scope and generally not kept up to date. When scientists want to access bibliographic information online, they usually go to more complete collections such as MEDLINE or DIALOG.

What would be much more useful to the BIONET community would be to provide access to full text copies of well known reference works, such as the bibliographies of Drosophila by Herskowitz, Genetic Maps by O'Brien, Genetic variations of Drosophila melanogaster etc. Most of these works are not prepared in computer readable form.

Genetic Variations of Drosophila melanogaster

Fortunately, Dr. Dan Lindsley did prepare this extensively used reference work in a computer readable

form. It is fondly known as the Redbook and it has been the bible of *Drosophila* genetics and a primary research resource for 17 years. Recently Dr. Lindsley has undertaken to produce a new edition of this work. Moreover, both he and his publisher have agreed to make both editions of the Redbook available on BIONET for online access. The advantages of having the book available online is that one can effectively cross index the entire book for any word that appears in the book. For example, using the text searching tools mentioned above, one can find all mutations that affect bristles as well as all suppressors of bristle mutants. This kind of cross indexing is not possible in any other way. Methods for finding all known mutants at any genetic map position or in any region of the polytene chromosome are also possible. While there are Appendices in the back of the Redbook to allow the location of the genes by location, online access allows one to have descriptions of all genes that are similar in 1) location, 2) function or 3) which interact with each other.

We have made the Redbook available in both chapter form, identical with the chapters as they appear in the book itself, and in two large sections of the books including all the point mutations (in the file <REDBOOK>MUTANTS) and all the chromosome rearrangements (in the file <REDBOOK>OTHERS). These latter files aide in finding mutants using the FIND program. We have just received a tape containing initial chapters of the new version of the Redbook and we intend to make this available in the same way, with updates as the work is completed. Currently the new version is about 60% complete.

Cloned Segments of the Drosophila Genome

In addition to *Genetic Variations of Drosophila melanogaster*, Dr. John Merriam (UCLA) has compiled a list of all the molecular markers on the *Drosophila* chromosomes. This compilation includes all cloned segments of *Drosophila* DNA that have been mapped to specific genetic positions as well as all rearrangements whose break points have been cloned. These cloned segments are extremely useful to molecular biologists who wish to isolate specific genes from *Drosophila* using the walking procedure. Moreover, this compilation will eventually develop into a complete molecular map of the *Drosophila* genome to complement the genetic one. We are working with Dr. Merriam to provide this useful resource on Bionet.

Once both the new edition of the *Genetic Variations of Drosophila melanogaster* and the *Cloned Segments of the Drosophila Genome* are available, this will be written up for publication as letters to the editors of various molecular and genetic journals. A description of the databases will be submitted to *Drosophila* Information Services, where most geneticists currently go for this kind of information.

Once these works are made available and properly announced we will evaluate their usefulness to the community as judged by the number of read accesses and how many individuals access these databases.

We will also determine how many new users apply for BIONET use primarily to access these genetic databases. Currently WHOIS reports that there are 14 laboratories who mention the word *Drosophila* in their research title. There are probably many other laboratories that are concerned with *Drosophila* as a research organism but who do not have *Drosophila* in their research title (i.e. HOGNESS).

We can also consider recruiting other genetic databases which currently exist or will shortly in a computer readable form. Some examples would be the *E. coli* genetic map (BLATTNER) and the human genetic maps (RUDDLE and WHITE). These extensive genetic and restriction maps are a natural complement to the DNA and protein sequence information that BIONET now provides and we hope that these higher order rearrangements of genetic function will be as useful in the future as are the sequence databases at the present.

III.F. Suggestions and Comments

We have two suggestions that would dramatically improve the productivity of the BIONET Resource, and would increase its availability and utility to the scientific community.

The first suggestion we have relates to the relationship between the Resource and the NIH staff. We have, in general, received a great deal of support from staff at the Office of Grants Management and B RTP itself. However, we simply must have more warning regarding decisions at the federal level that affect our budget. The dislocations we experienced in the first months of this grant year were substantial, and much time was wasted on administrative, as opposed to scientific, problems. The NIH is in a delicate position, in that it cannot afford to alarm its awardees about potential budget cuts that might not take place. But the down side risks of this approach, in our opinion, are more dangerous. We know that there may be additional NIH budget problems in fiscal 1986. We request that we be kept informed, even if information is tentative and subject to change, about any decisions that could affect our Year 3, and subsequent awards.

The second suggestion relates to the fact that there is increasing number of computer resources for molecular biology and related areas funded through DRR, alone. These resources make up a significant fraction of the B RTP/DRR budget, yet each has been funded through the traditional grant proposal and review process. What has resulted is a number of resources with complementary goals and facilities that have no means of communication with one another to easily exchange their stock in trade, electronic mail, new software and updated databases. Our program for BIONET Satellites is designed to take some steps toward network interconnections, but we do not have the budget nor the mandate to solve this problem alone. Obviously, a proposal can be generated for the necessary funds to build such a network, but it is not clear whether there would be sufficient new science to be successful. The networking problem has already been solved by others, so any proposal would be primarily technological. We suggest that DRR

Staff and Council, who must already be cognizant of the situation, investigate and propose some future programmatic goals. We would be pleased to participate in this effort.

I. Letter to BIONET Scientists

BIONET™ National Computer Resource
for Molecular Biology
c/o IntelliGenetics
124 University Avenue
Palo Alto, CA 94301
14 June 1985

***** IMPORTANT ANNOUNCEMENTS ABOUT BIONET *****

SUBSCRIPTION FEE

SWITCH to UNINET

BULLETIN BOARDS

BATCH JOBS

Dear BIONET Scientist:

I am writing this letter to bring you up to date on a number of changes that will be occurring in the near future on the BIONET Resource. By most measures, our first year of operation has been very successful. We currently have more than 450 Principal Investigators, representing over 1400 individual scientists, who have received approval for access to BIONET. In late March, our National Advisory Committee met with us to review progress and set directions for the next year (we are on a March 1 to February 27 funding cycle). Some of the changes that will be occurring are an outgrowth of that meeting, especially with regards to funding problems, communications and the use of bulletin boards and electronic mail.

As most of you know from a bulletin on BIONET (you can review the bulletin by issuing the MM command BBOARD BIONET-NEWS and reading the bulletin entitled "Budget cuts for BIONET") budget cuts at the NIH have led to substantial reductions in our budget for this grant year. For the past three months, with the help of consultants and our Advisory Committee, we have been studying solutions to the problems posed by these cuts. To reiterate, we have had to reduce staff from 8.5 full time equivalents to about 5.5, and we project that our telecommunications budget will be consumed late this fall unless some prompt action is taken. The following sections on **Subscription Fee** and **Switch to UNINET** represent our best solutions to these problems, solutions that will enable us to continue to provide you with high quality service and to build the Collaborative and Core Research aspects of BIONET. Most importantly, the changes mean that you will not have to pay the full cost of your telecommunications. This was crucial to us in order that those laboratories that are suffering

from funding problems still be able to access BIONET.

Subscription Fee

There are several possible solutions to the problems of staff reductions and a substantial projected deficit in our telecommunication budget. We have initially focused our attention on the latter problem, because without telecommunications access, there would be no Resource! Many solutions to this problem would have the effect of making access prohibitively expensive for those laboratories with limited resources; such laboratories are often those that need access to BIONET the most. Fortunately, we have found what seems to be a good compromise. With the concurrence of our National Advisory Committee, we are instituting a nominal subscription fee to BIONET, which, when combined with available funds, will make up the deficit. The fee will also allow us to expand our telecommunication services as the BIONET community grows, by increasing the number of network lines to the computer.

We regret to have to impose a fee on such short notice. Our long range plans indicated that some type of access charge would have to be leveled eventually, but we had not anticipated it happening this quickly. We received notice of our budget cuts on February 26, three days before the beginning of our second year. Since that time we have tried unsuccessfully to have the cuts rescinded, but the NIH, Congress and the OMB are still at odds over the 1985 NIH budget and no rapid solution appears to be forthcoming. We are also exploring other funding sources, but have no assurances at this time that any additional money will be available.

To cover our costs, the fee we must request is \$400 per year per Class I Principal Investigator (PI). We hope that virtually every PI on BIONET will be able to afford this nominal expense. This fee should be received by us on or before August 1, 1985. Checks should be made payable to IntelliGenetics, marked clearly as BIONET Subscription Fee, and attached to the enclosed Subscription Fee form. Class I PI's in Alaska, Hawaii, Canada and Mexico must pay the subscription fee because their telecommunications costs will be billed to BIONET by UNINET. *PI's in other foreign countries, who will be billed separately by their network vendors, will NOT have to pay a subscription fee.*

Switch to UNINET

We will soon be moving over to a new vendor for telecommunications services, UNINET. The primary reason for this change is that UNINET has offered very attractive pricing for use of BIONET. They will be charging us a fixed rate per month per communications line, independent of use. This charging scheme means that we can do much more precise planning for our current and future budgets. In addition, other organizations that have recently switched from Telenet, our current vendor, to UNINET report better network response and reliability, plus more rapid action on problems with the network. We will still pay careful attention to minimizing your need for long periods of connection to BIONET. In particular, we are working on simplifying BATCH submissions (see below) so that you can logout from BIONET, leaving the line available for another person to use. We have also implemented a procedure that will detach your job after a period of time of non-use (currently two hours) should you accidentally forget to logout your job.

This switch will take place on or about July 1; after that time Telenet will no longer be available for access to BIONET. You will receive a mailing from UNINET with complete instructions on use of the network prior to that time. You will be reminded by bulletins on the BIONET-NEWS bulletin board when you login about precise schedules. Use of UNINET is even less complicated than Telenet. The primary difference will be in the telephone number that is dialed plus slightly different recognition characters to "wake up" the network. Depending on location, and because UNINET's distribution of local nodes is different from Telenet's, you may have to call a longer distance for access, or you may

be lucky, and find a UNINET node nearer to you!

Bulletin Boards

The primary use for bulletin boards is the sending of messages to a group of people interested in specific topics. It is often much simpler to send a bulletin than to send electronic mail because for the latter you must maintain an accurate mailing list. The BIONET system keeps track of who wants to read which bulletin boards, thereby saving you considerable effort.

Our current system is based on the bulletin board topics you selected as part of your application to BIONET. These topics are automatically added to your LOGIN.CMD file as part of setup of your account on BIONET. Everyone has BIONET-NEWS in this file. You are notified automatically when you login that you have bulletins to read, and are given their titles and a choice to read them or to ignore them.

We have had some successful use of the bulletin boards, but also several disappointments. On the positive side, there have been several valuable exchanges among members of the community, and several of you have expressed an interest in being a community leader, albeit on topics more specific than the current, generic ones. On the negative side, many people are ignoring important bulletins by skipping over the reading of them on login to BIONET. Also, the current topics are too limited to inspire use of the bulletin boards.

For these reasons, we are going to change our approach somewhat, to include the following steps:

- We will use a new list of topics, selected from your suggestions and current patterns of use of the bulletin boards. Periodically, we will add new topics to the list based on your suggestions. This will make the set of topics more useful to you and your colleagues.
- We will revise your existing LOGIN.CMD file to contain only the BIONET-NEWS bulletin board, and make it simple for you to add other topics at your discretion.
- We would like BIONET-NEWS to be read by *EVERYONE*, all the time. We will continue to encourage you to post bulletins that are of wide community interest on BIONET-NEWS. To prevent the occasional user of BIONET from being overwhelmed by bulletins on this bulletin board, we will periodically move outdated bulletins to topical bulletin boards, or delete them, so that only a few important and timely bulletins remain to be read.

Attached to this letter you will find a revised list of bulletin boards that represents a start to meeting some of our goals, together with instructions on how to add one of the new topics to your LOGIN.CMD file so that you are notified about interesting subjects when you login.

BATCH Jobs

Many members of the community have asked how they can submit lengthy computations to the system and logout and go away, to return at some later time to review the results. That capability is available through the BATCH system on the BIONET computer. Attached to this letter are some simple instructions on how to SUBMIT such computations to run in BATCH overnight. We hope they are useful to you.

In conclusion, I want to thank you for your participation in what has been a successful experiment in making powerful computational facilities available to a very large number of scientists. Your enthusiastic contributions to this success are appreciated, and we at BIONET hope they will continue unabated despite the budgetary problems discussed previously.

Sincerely,

Dennis H. Smith
Resource Manager, BIONET

II. Justification of \$400 Fee

The following message was sent by electronic mail to Dr. Rich Roberts, a member of BIONET's National Advisory Committee, in response to questions from him on the justification for setting the fee at \$400.

7-Aug-85 12:17:13-PDT,3352;000000000001
 Mail-From: SMITH created at 7-Aug-85 12:16:59
 Date: Wed 7 Aug 85 12:16:58-PDT
 From: Dennis Smith <SMITH>
 Subject: Re: charges
 To: ROBERTS
 cc: AMAREL, KAFATOS, LEDERBERG, MAXAM,
 RINDFLEISCH, YANOFSKY, SMITH
 In-Reply-To: Message from "Richard J. Roberts <ROBERTS>" of
 Wed 7 Aug 85 06:34:44-PDT

Rich,

The figure was determined by taking a hard look at our budget to see what was needed and then to estimate, based on the profile of accepted PI vs. those who actually used the system, how many would sign up.

In brief - we need about \$50,000 to meet telecommunications costs, i.e., to avoid going into the red. We desperately need another consultant and a scientist, the latter for core research, and that would cost 50-60K depending on obvious factors. Given a target of roughly \$100K, we'd need 250 PI's signed up at \$400 each. Our estimates out of a group of 450 PI's was that 175 - 200 was a good guess. The first 125 would pay our projected deficit and the remainder MIGHT allow us to hire another user consultant, our highest priority.

The message to the community soliciting comments yielded only a few comments, almost all of which were essentially, "Siggghh, but it's worth it" and two complaints.

Since the fee was announced, we have received nearly the 125 target of checks or requests for invoices, and 8 refusals. None of the eight was a significant previous user of the system.

I have made it clear that we would not turn anyone away for lack of funds. There are two people in this situation, and I will not cut them off if they can't raise the full amount. One was the type of person you described, who was mainly interested in communication.

We give people at least 30 days after subscription is due, and a much longer period if they let us know about their problems, so they have at least 60 days and more if they need it.

So, those are the current facts. An issue remaining from your message is treatment of those who wish only access to communication facilities, and that should certainly receive some discussion. There are several mechanical problems with this, but, more importantly, I have been discouraged by the community's response to the communication facilities. We have excellent use of electronic mail, but only sporadic use of bulletin boards. The reality of the last 8 months of reasonable use of the system is that lots of people are banging away at it to solve their problems, and use mail and links, and advising, and phone calls to generally very good purposes - they have some really interesting problems that involve sophisticated use of the programs, i.e., little chaff or nonsense. But it seems to be a cottage industry out there, and despite our efforts to simplify further the bulletin boards and our letter and messages to demonstrate their use and make it easy to add and promote new topics, only that segment of the community with previous "electronic" experience is making any use of those facilities.

Comments and suggestions welcomed.

Dennis

PS - I am trying to get in touch with Lederberg, who is now in CA to schedule at least a local meeting of NAC members and the next full meeting. Do you have any constraints over the next few months? -----

III. Letter on Class IV Access

BIONET™

Important Notice on

Reduced Subscription Fee for Communication

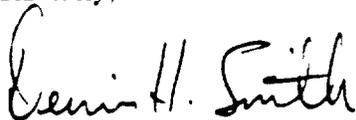
Dear BIONET Scientist:

I am writing to you because our records indicate that you either have not logged in to BIONET or that you have used the system but have not chosen to pay the subscription fee we had to implement late this summer. Your account still exists on the system and will remain there temporarily. It is not currently accessible by you pending further information about your intentions to remain a participating member of the BIONET community.

We know that the current funding situation is as serious for many of you as it has been for BIONET. Many scientists have made special arrangements with us to maintain access to the system pending availability of new grant funding, institutional support and so forth. We want to be sure that you understand that we will not prevent access to BIONET simply due to lack of funding. If you have special circumstances and need access to BIONET to support your research, I encourage you to contact Mary Warner or myself ((415) 965-5576,77).

One of the goals of BIONET is to foster active communication and collaboration among molecular biologists. Therefore, we would like everyone to be able to access at least the electronic mail and bulletin board facilities, even if access to analysis programs is not desired. Therefore, we have instituted a new class of BIONET membership, Class IV, which offers access to electronic mail, electronic bulletin boards and file transfer programs. These facilities will allow you to interact and exchange information with your colleagues on BIONET. Because such interaction does make use of our telecommunications facilities, we must also charge a subscription fee for access, but at the reduced rate of \$100 per Principal Investigator per year. We have included a Subscription Fee Form which lists this option. We very much hope that you will choose at least this method for participating in the BIONET community.

Sincerely,



Dennis H. Smith

Resource Manager, BIONET

IV. Reapplication Form for BIONET Access

BIONET™

26 September 1985

Dear BIONET Principal Investigator:

The National Institute of Health requires, as part of our Annual Report, that we review the status of all BIONET subscribers each year. Thus, we need information from you on any changes, from your original application, with respect to institutional affiliation, address, funding status, sub-investigators, etc. Most importantly, we need a list of all publications in which BIONET played a role. We also ask that at this time you reaffirm your original agreement for access to BIONET.

Use the enclosed forms to:

- Update your title, affiliation, mailing address and/or phone number
- Update the list of your sub-investigators
- Note change in status of funding
- Provide a list of current publications resulting, in part, from the use of the BIONET Resource (Remember to cite the BIONET Grant # 1 U41 RR-01685-02 in all such publications.)
- Provide a brief description on how BIONET was used in your research.

We would also like to give you this opportunity to comment on the BIONET resource - what role it is playing in your research and any suggestions/requests for improvement.

Because we must prepare our Annual Report in December, we need you to return this re-application to us no later than November 1, 1985. Thank you for your cooperation.

Sincerely,



Mary Lou Warner
BIONET Administrator

BIONET RESOURCE**Reapplication****Fiscal 1986**

Principal Investigator (full name and title):

Affiliation: Department, School and Institution (Changes Only):

Mailing Address (Changes Only):

Area code and phone number(Changes Only):

BIONET Agreement

As Principal Investigator of this grant to use the BIONET Resources, I agree to adhere to all conditions and restrictions for use of the BIONET Resource, as described in the document "The BIONETtm Resource, Description and Applications Form" and such further regulations as may be issued from time to time by the NIH or the NAC.

The BIONET Resource will not be used for any commercial purpose which is not specifically identified to and approved by the NAC. Any pertinent change in sponsorship, continuity of grant support, or use made of BIONET will be reported promptly to the BIONET Resource Manager.

I have also furnished a copy of this re-application to the responsible officer of my institution, whose signature appears below.

I also assume full responsibility for all users listed on this applications 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 existing user or addition of new staff qualified to use the resource. I will inform new users of the above mentioned conditions and restrictions.

Date: _____

Name of official: _____

Signature of Principle Investigator

Signature of Official

P.I.Name _____

BIONET re-application page 2**Current Sub-investigators**

If your group of sub-investigators has changed since your last correspondence, please note the changes below:

<u>NAME</u>	<u>Title</u>	<u>Phone number</u>	<u>Change</u>
-------------	--------------	---------------------	---------------

Funding Status

Please note any change in status of your funding, including Institution, Grant Number, title of grant, and duration of grant.

Current Publications

Please list current publications resulting, in part, from the use of the BIONET Resource (use standard bibliographic format). Remember to cite the BIONET Grant # 1 U41 RR-01685-02 in all such publications. A sample citation would be: *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-01685-02.*

BIONET Re-application Page 3**Use of BIONET**

Briefly describe how BIONET has been used in conjunction with your research:

COMMENTS

We invite your comments, suggestions and requests about the BIONET Resource. Which programs are the most useful to you - the least? Should the bulletin boards be broader in scope - more specific? Would you like more interaction with other users? What else would you like to see included in the BIONET Resource, for example, other computer programs. Would you like more information about the BIONET Satellite Resources?

V. Program for the Rutgers/Waksman Workshop

If you receive more than one copy of this brochure, please pass extra copies on to a friend or co-worker.

One-Day or Three-Day Workshop
 Introduction to *BIONET*[™]:
 A National Computer Resource for Molecular Biology

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS
 Waksman Institute of Microbiology
 P O Box 759
 Piscataway, New Jersey 08854-0759

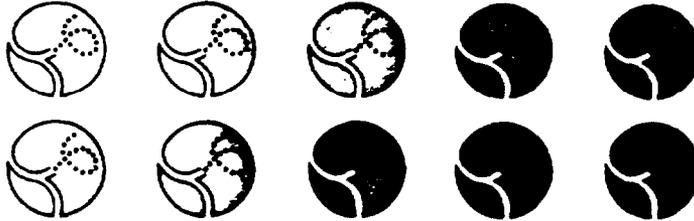
INTRODUCTION TO *BIONET*[™]: A National Computer Resource for Molecular Biology

One-Day Workshop—June 17, 1985

Three-Day Workshop—June 17-19, 1985

The Waksman Institute
 of Microbiology

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS
 Campus of New Brunswick



ABOUT THE WORKSHOP LEADERS

Douglas Brutlag, Ph.D., Associate Professor of Biochemistry at Stanford University

Dr. Brutlag, coinvestigator of *BIONET*, is an authority on nucleic acid enzymology. His work has focused on the evolution of nucleic acid sequences and the assembly of chromosomes. Dr. Brutlag has also had a long-term interest in computer systems and programming and was a coinvestigator of the MOLGEN project at Stanford.

Laurence Kedes, M.D., Professor of Medicine at Stanford University School of Medicine

Dr. Kedes, coinvestigator of *BIONET*, was also a coinvestigator of the MOLGEN project and has been responsible for overseeing the development of several software projects in molecular biology. As a molecular and cellular geneticist, Dr. Kedes spends the majority of his research efforts investigating the nature of gene organization and regulation in animal cells.

Elaine Mansfield, Ph.D., Training Manager and Consulting Scientist for *BIONET*, IntelliGenetics, Inc.

Dr. Mansfield has worked for IntelliGenetics, Inc., for two years in Marketing and Customer Training and Support. She has conducted numerous on-site trainings and written user manuals for several of the IntelliGenetics programs. Dr. Mansfield currently oversees the training programs and collaborates on user support services for *BIONET*. She holds a Ph.D. in human biochemical genetics from the University of California at Berkeley.

Introduction to **BIONET**[™]: A National Computer Resource for Molecular Biology

One-Day Workshop—June 17, 1985

Three-Day Workshop—June 17-19, 1985

The Waksman Institute of Microbiology

PURPOSE OF THIS WORKSHOP

BIONET is a national computer resource sponsored by the N.I.H. and established to provide academic scientists with an interactive timesharing computer, up-to-date sequence databases and analysis programs, and powerful communication tools for rapidly exchanging information with colleagues. This one- or three-day workshop will provide participants with an understanding of the programs, database organization, and advanced computer resources available on **BIONET**.

On the first day, the workshop leaders will use video projection systems to show actual on-line interaction of protein and DNA sequence analysis programs, electronic bulletin boards, and electronic mail facilities. A more in-depth presentation will be conducted as a two-day hands-on session for **BIONET** principal investigators or their representatives. Applications presented will include how to use the computer for efficiently managing DNA sequencing projects, sequence comparison methods, and optimal probe design. In addition, the procedure of transferring data from a personal computer to the **BIONET** computer will be demonstrated.

PROGRAM CONTENT

- An overview of **BIONET** and the computer system
 - Resource organization and goals
 - Core program library, sequence databases, and development library
 - Introduction to electronic mail and bulletin boards — communication tools for collaboration
 - System commands and directory organization
- Restriction mapping tools
 - Restriction fragment length calculation
 - Construction of restriction maps from enzyme digests
 - Strategies for generating large restriction maps
- Simulation and design of recombinant DNA experiments
- Sequence entry, verification and editing *The Genetic Editor*
- Managing large DNA sequencing projects
 - Methods for eliminating vector sequences from your gels
 - Customization to chemical- or dideoxy-sequencing methods
 - Strategies for assembling multiple gels and generating reliable consensus sequences
- Sequence database organization and searching methods
 - Comparing N.I.H. GenBank, EMBL, and NBRF databases
 - Rapid sequence alignment and similarity searches
 - Sequence retrieval using exact or ambiguous patterns
- Sequence analysis programs
 - Nucleic acid sequence analysis, comparison and manipulation
 - Amino acid sequence analysis, comparison and manipulation
- Sequence comparison methods
 - The algorithms, speed, precision and limitations
- Distributed processing: moving data to and from the **BIONET** computer

Schedule:

8:30 A.M.	Registration and coffee
9:00-noon	Morning session
1:30-4:30 P.M.	Afternoon session

LOCATION

Waksman Institute of Microbiology
Rutgers, The State University of New Jersey
Hoes Lane and Frelinghuysen Road
Piscataway, New Jersey 08854-0759

WHO SHOULD ATTEND

This session should be of interest to all **BIONET** recipients. In addition, principal investigators involved in molecular biology research would benefit from learning about the resource. IntelliGenetics commercial customers, and other molecular biologists are welcome to attend the one-day session. Attendance at the two-day hands-on session is designed primarily for **BIONET** recipients and may be limited to one representative per laboratory group.

WORKSHOP LEADERS

Douglas Brutlag, Ph.D., Department of Biochemistry, Stanford University

Larry Kedes, M.D., Department of Medicine, Stanford University Medical Center

Elaine Mansfield, Ph.D., **BIONET** IntelliGenetics, Inc., Palo Alto, California

REGISTRATION AND FEE

The registration form and fee must be received in advance of the program date. This fee includes course admission, class materials, and beverage breaks.

For additional information or to reserve a space, contact Selma Gitterman, Director, Continuing Professional Education Program at 201/932-4258 between 1:00-4:00 P.M., E.S.T.

CONTINUING EDUCATION UNITS

CEUs are awarded to participants in this program. The CEU gives formal recognition to persons continuing their education and keeping up-to-date in their chosen field or profession.

ATTENDANCE

Attendance is limited for both workshops. The three-day session will be limited to 30 participants. Please register as early as possible.

INTRODUCTION TO **BIONET**[™]: A National Computer Resource for Molecular Biology

June 17, 1985 Course Fee: One-Day Workshop: \$100
June 17-19, 1985 Three-Day Workshop: \$450

Check workshop desired One Day Three Day

Dr., Mr., Ms _____

Position/Title _____

Company/Institution _____

Address _____

City _____ State _____ ZIP _____

Telephone _____

Please make check payable to **Rutgers, The State University of New Jersey**. Return form with payment to the Continuing Professional Education Program, Waksman Institute of Microbiology, P.O. Box 759, Piscataway, New Jersey 08854-0759.

PLEASE POST

How did you learn about this course?

Direct Mail Other _____

From Supervisor _____

From Posted Material _____

Other workshop topics I would be interested in _____

Please keep me on the mailing list

Yes No

Also add the following persons to the mailing list

Name _____

Company/Institution _____

Address _____

City _____ State _____ ZIP _____

VI. Descriptions of the BIONET Satellite Program

I am writing because you have expressed interest in learning more about our new program for BIONET Satellites. As you undoubtedly realize, the community requests for access to BIONET are already pushing the resource to its limits during the middle of the day, and applications for access continue unabated. The response from the community of molecular biologists was expected, given our past experiences with the community's need for access to computer software. What was not expected was that the Resource would approach saturation after only six months of operation!

The issue was the topic for discussion at the March 23, 1985 meeting of BIONET's National Advisory Committee (NAC). At that meeting, we proposed a program for BIONET Satellites that would essentially distribute the BIONET Resource to a number of additional sites. In other words, rather than trying to enlarge the existing central timesharing computer, we are trying to take advantage of a large installed base of computers accessible to molecular biologists. This approach to expanding BIONET reflects the changing emphasis in use of computers throughout the scientific community, from centralized to more distributed systems.

The NAC was supportive of this approach because it offers the possibility for a rapidly-growing network of computers and cooperating scientists. At the same time, the NAC and the BIONET staff agreed that distribution of the Resource carries with it the danger of isolation of Satellites from the rest of the community. Therefore, an integral part of our proposal is to extend the collaboration and communication possible on BIONET to the Satellites. In this way, access to electronic mail and bulleting boards, and file transfers of programs and data, will maintain communication within the community.

In a separate letter from Dr. Michael Kelly, General Manager of IntelliGenetics, the BIONET Satellite program is described in more detail, including information on the requirements for full participation as a Satellite. I am hoping that you will become an integral part of BIONET's efforts to make available to the community the latest in analysis programs and sequence data.

Sincerely yours,

Dennis H. Smith, Ph.D.
Resource Manager, BIONET

IntelliGenetics

Dear BIONET Scientist:

The BIONETTM computing resource, a cooperative agreement between IntelliGenetics and the National Institutes of Health is now in its second year of operation. The success of this program is now beginning to outdistance the available resources with over 1200 investigators currently using the system. In order that we may serve the molecular biology community with more of the capabilities of BIONETTM, we have instituted a new and exciting program, **Bionet Satellites**.

Bionet Satellites are designed to provide all of the functionality of BIONET at the local level. Utilizing existing Digital Equipment VAX or 2060 computers on your site all of the programs, bulletin boards and electronic mail functions of BIONETTM will be available.

Bionet Satellites existing service capability to the user is available now. It will reflect the three major goals of BIONET:

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

Messages.

Collaborative research and other community interactions depend on electronic mail and bulletin boards. Exchange of messages across a distributed network is required to maintain this important aspect of BIONET.

File Transfer.

Rapid sharing of software and data among the Satellites and the central 2060 is required to ensure that investigators have access to the latest programs and DNA and protein sequence data. Up- and down-loading of files among computers on the network will make this possible.

In designing our plan for BIONET Satellites, we are paying special attention to communication. We have identified communication as an important Core Research project for BIONET. We have begun implementing both short and long range plans to accomplish this goal:

- **Short Term.** We will take advantage of existing communication software and our telecommunication network, augment it for our purposes, and use it for low bandwidth transfer of messages and files among the computers comprising a distributed BIONET. This software will be made available to Satellites to allow relatively transparent communication with the 2060 and other Satellites.
- **Long Term.** Longer term, the major barrier to high bandwidth, automatic transfer of messages and files is the current lack of access of the BIONET community to existing networks (ARPANET, CSNET) that support such activities. In other words, the technology

exists, but the community and many potential sites for Satellite do not have ready access to it. It is our goal to make that technology available to a distributed BIONET. We are currently exploring funding resources and existing networks that will allow us to meet that goal.

We also plan to link BIONET with other National Resources, including GenBankTM, the Protein Identification Resource (PIR), the newly-established Molecular Biology Computer Research Resource at Dana Farber, and other national and international resources. For example, it would be possible to use the communications facilities described above to obtain new programs from Dana Farber, use them in concert with the latest GenBank data and complementary analysis software on BIONET or one of its Satellites, and forward resulting sequence data directly to GenBank or the PIR. Your participation as a BIONET Satellite in this enterprise will allow you to take immediate advantage of these facilities.

A Bionet Satellite can be installed on your site in a short period of time. The BIONET staff will provide training to assist your colleagues in using the core group of genetic engineering programs.

Accessing this service only requires the purchase of a software license from IntelliGenetics. A special purchase program has been arranged to make it easy for academic institutions such as yours to join the Bionet Satellite program. The cost of the license for this service is:

DEC MicroVAX II (DH-630Q4) - \$20,000 per year for three years
 DEC VAX 11/750 or larger - \$20,000 per year for three years
 DEC 2060 - \$24,000 per year for three years.

DEC 2060 the cost per user is \$400 per year. This price includes software at least six times a year. At the end of the three year period you may purchase a maintenance agreement which is

Assuming 50 users on the VAX or 60 users on a DEC 2060, the cost per user is \$400 per year. This price includes the update of databases and upgrading the software every three years. At the end of the three year period the software will be maintained for you at a cost of currently \$6,000 per year.

If you would like to join this extremely useful computing resource, please write to me.

If you would like to join this extremely useful computing resource, please write to me.

M.D.

Michael J. Kelly, Ph.D.
 General Manager
 IntelliGenetics, Inc.
 1975 El Camino Real
 Mountain View, CA

J West
 94040-2216

We will give your application immediate attention and provide advice in setting up your Bionet Satellite. Thank you for your interest in this program. We look forward to providing you the ultimate in computing and communication for molecular biology.

We will give your application immediate attention and provide advice in setting up your Bionet Satellite. Thank you for your interest in this program. We look forward to providing you the ultimate in computing and communication for molecular biology.

Sincerely yours,



Michael J. Kelly, Ph.D.
 General Manager
 IntelliGenetics, Inc.

October 1985

VII. Text of Advertisement to Appear in Nucleic Acids Research

IntelliGenetics INVITES YOU TO JOIN BIONET™

AN N.I.H. COMPUTER RESOURCE FOR MOLECULAR BIOLOGY

OVER 1500 SCIENTISTS ALREADY BENEFIT FROM:

- ACCESS TO THE LATEST DATABASES OF SEQUENCES
- ACCESS TO COMPUTER PROGRAMS
 - TO ENTER AND ANALYZE SEQUENCES
 - TO COMPARE SEQUENCES TO DATABASES
 - TO HELP PLAN CLONING EXPERIMENTS
 - TO PROVIDE TOOLS FOR PROGRAM DEVELOPMENT
- ACCESS TO HUNDREDS OF YOUR COLLEAGUES AND THEIR RESEARCH RESULTS USING ELECTRONIC MAIL AND ELECTRONIC BULLETIN BOARDS

CONTACT US TODAY...

BIONET™

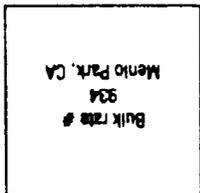
e/o IntelliGenetics, Inc.

An IntelliCorp Company
1975 El Camino Real West
Mountain View, California 94040-2216
Telephone (415) 965-5575

VIII. The BIONET Brochure Mailed to NIH Grantees

AN N.I.H. COMPUTER RESOURCE FOR MOLECULAR BIOLOGY

JOIN US...



An IntelliCorp Company
1975 El Camino Real West
Mountain View, California 94040-2216
Telephone (415) 965-5575

BIONET™
c/o IntelliGenetics, Inc.

Dear Scientist:

Your name is included in a list of National Institutes of Health grantees given us by the NIH grant management staff. This indicates that your research may be in MOLECULAR BIOLOGY or a related field involving ANALYSIS OF PROTEIN AND NUCLEIC ACID SEQUENCE DATA:

IF you would benefit from:

- Access to the latest databases of sequences
- Access to computer programs
 - to enter and analyze sequences
 - to compare them to databases
 - to help plan cloning experiments
 - to provide tools for program development
- Access to hundreds of your colleagues and their research results via electronic mail and electronic bulletin board facilities

return this card for an application to the BIONET Resource.

BIONET™ is a national computer resource sponsored by the National Institutes of Health and established to provide scientists at non-profit institutions with an interactive timesharing computer, up-to-date sequence databases and analysis programs, and powerful communication tools for rapidly exchanging information with colleagues:

- The Core Library, consisting of nine programs that manipulate and analyze nucleic acid and protein sequence data, plus additional programs contributed by the BIONET community.
- The Database Library, containing existing databases of nucleic acid and protein sequences, including GenBank™, the European Molecular Biology Laboratory (EMBL) database, the National Biomedical Research Foundation (NBRF) library of protein sequences, VectorBank™, and Cold Spring Harbor Restriction Enzyme database.
- The System and Programming Support Library, providing tools for program development, including: programming languages (Fortran, C, Pascal, BASIC, MAINSAIL, and Interlisp) and system utility programs (MLAB, EMACS, TVEDIT and SCRIBE™); facilities for electronic mail and electronic bulletin boards; KERMIT and MODEM for file transfer; the UNINET™ telecommunications network.

The **BIONET** resource will admit researchers from academic and non-profit institutions who can demonstrate that they are supported by governmental, philanthropic, or unrestricted institutional funds and that their research can be assisted by Resource facilities. The **BIONET** staff will consider applications funded from proprietary or restricted sources, and make recommendations to its National Advisory Committee which will make final decisions on all access to the Resource. There is an annual subscription fee of \$400 to cover telecommunication costs.

If you would like more information about the BIONET Resource and an application form, please fill in and mail the attached card.

BIONET has also instituted a program of Satellite Resources, whereby investigators can run BIONET software on local DEC 20™ or VAX™ or SUN™ computers. Check the box on the return card to obtain more information.

Dr., Mr., Ms. _____

Position/Title _____

Company/Institution _____

Address _____

City _____ State _____ ZIP _____

Telephone _____

(Please print clearly—This will be used as your mailing label)

**PLACE
STAMP
HERE**

BIONET™

c/o IntelliGenetics, Inc.
An IntelliCorp Company
1975 El Camino Real West
Mountain View, California 94040-2216

Check this box to obtain more information on BIONET Satellite Resources