LBL-37757
UC-406

Lawrence Berkeley Laboratory
UNIVERSITY OF CALIFORNIA

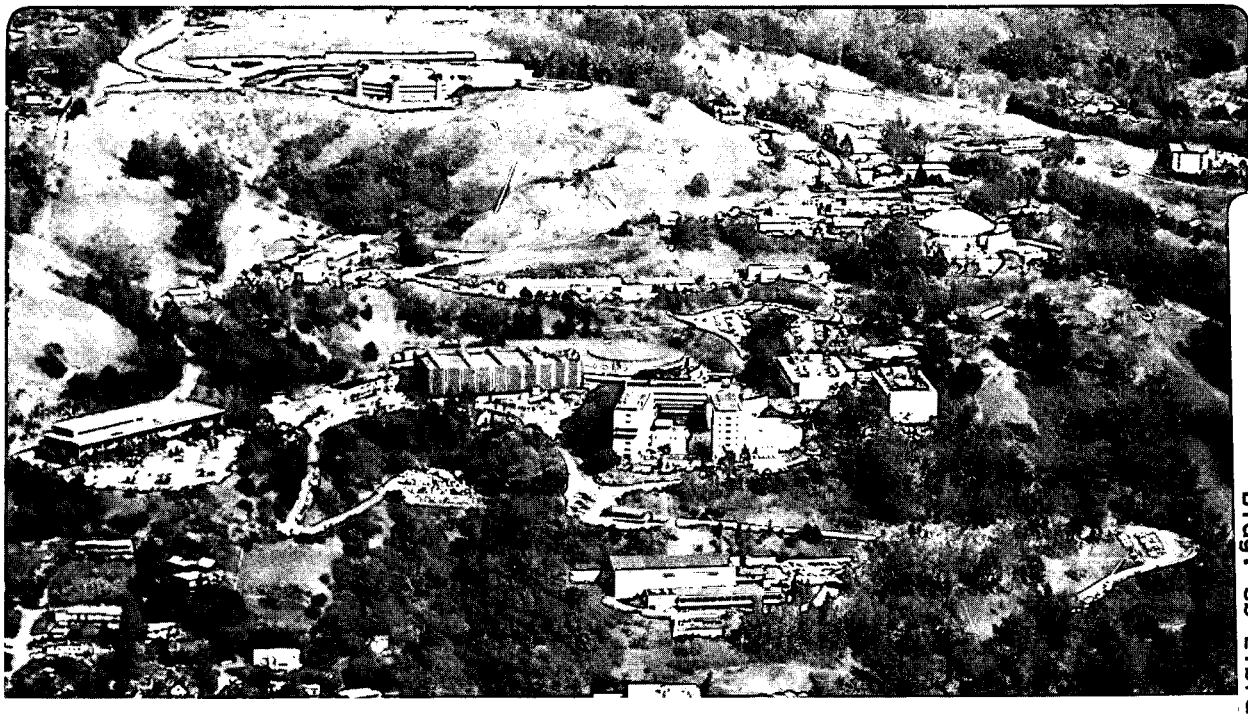
Engineering Division

Invited paper to be published as a chapter in *Automated Technologies for Genome Characterization*, John Wiley and Sons, New York, NY, in process

Automation Strategies, A Modular Approach

M.J. Pollard

September 1995



Prepared for the U.S. Department of Energy under Contract Number DE-AC03-76SF00098

DISCLAIMER

LBL-37757
UC-406

Automation Strategies, A Modular Approach

Martin J. Pollard

Human Genome Group
Engineering Division
Lawrence Berkeley National Laboratory
University of California
Berkeley, California 94720

September 1995

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

AUTOMATION STRATEGIES, A MODULAR APPROACH

Martin J. Pollard
Lawrence Berkeley National Laboratory, University of California
Engineering Division, Human Genome Group
1 Cyclotron Road, M/S 74-157, Berkeley, CA 94720

Introduction

The goals of the Human Genome Program (HGP) are to develop the genetic and physical maps and determine the DNA sequence of the human genome and the genomes of several model organisms.[1] It is an international research project involving public and private laboratories primarily in the United States and Europe. This project arose from the discovery of recombinant DNA techniques during the early 1970's and the development of the Maxim-Gilbert and Sanger methods of sequencing DNA. Of critical importance was the development of automated sequencing machines. Simultaneously, computer technology was reaching into the every laboratory in the form of automated instrumentation, personal computers, and workstations. Recombinant DNA techniques provided the biological tools necessary for a conceptually simple means of determining detailed DNA sequence information. However, the task of actually sequencing the human genome is enormous even considering the use of automated sequencing machines. It was recognized early on that this project would require the application of instrumentation technology and computing at an unprecedented scale in the field of biology.

The HGP has changed not only the way we think about biology but how we do biology in the laboratory. A primary difficulty in the task of genomic sequencing originates from the fact that the size of the DNA that can be analyzed with today's technology is extremely small relative to the size of the genome itself. In admittedly oversimplified terms, genomic research entails breaking down the genome into small pieces that can be analyzed, determining the sequence of each piece, and then collating and reassembling the information to construct the finished sequence map of the original genome. The enormous number of samples that must be physically handled and then whose information content must be processed has stimulated collaboration with engineers and computer scientists to develop new ways to do biological research.

Strategy Implementation

The large scale nature of genomic research requires strategic decisions as to how to apply engineering and computer technology to acquire accurate mapping and sequencing information in a cost effective manner. There exists several variants of sequencing protocols which differ in their emphasis on the use of hardware automation relative to computer analysis. For example, in the area of DNA sequencing the competing strategies have been random versus directed sequencing. Both strategies depend on the standard automated sequencing machines. Hardware automation is used to select and prepare samples for sequencing. Computer analysis is used to analyzed the individual sample sequence information and assemble it into contiguous finished genomic sequence.

Random strategies are designed to minimize the complexity of laboratory procedures and automation and to emphasize computer analysis and assembly of large numbers of randomly distributed sample sequences. The automation is designed to prepare large numbers of samples for sequencing in large numbers of sequencing machines. This sequence data results in a relatively large and complex computer assembly problem. In contrast, directed strategies put relatively greater emphasis on the use of automation to implement more complex laboratory procedures. Additional data is generated that must be tracked throughout the process. Some of this data is mapping information that is used to select clones to be sequenced and reject clones from regions that have already been sequenced. This results in the processing of fewer samples, the use of

fewer expensive sequencing machines, and results in a significantly simpler sequence data assembly job.

The choice of a mapping or sequencing strategy will have a profound influence on the personnel and resource needs of a Genome Center. There are undoubtedly many reasons for choosing a particular strategy. This decision must in part be determined by the availability of resources in computer science and engineering at a particular institution. There has been very little tradition of collaboration between biology and the fields of mechanical and electrical engineering. This is evident in the existence of relatively few genome centers that have strong engineering programs in these areas. Access to engineering resources will provide an additional dimension of flexibility in the development of optimum genome strategies. In the area of genomic sequencing, there has been an evolution toward strategies comprising elements of both directed and random methods. This type of dynamic strategy development requires sufficient flexibility to implement new automation hardware to modified random strategies and alteration of existing highly automated directed strategies.

The remainder of this discussion will describe some of the principles of engineering as it is applied to genome research. This will hopefully provide a systematic means for evaluating current and future large scale automation strategies as well as basis for evaluating individual instrument designs. I will focus on hardware automation for processing samples to subsequently be sequenced in automated sequencing machines. I will not discuss computer automation or informatics related to sequence assembly.

Global Automation Strategies

The expectation early in the HGP was that it would be necessary to invent revolutionary new instruments to prepare and sequence samples at a rate appropriate to complete the sequencing of the human genome at a cost of $0.50 per base over a period of 15 years. While a number innovative research programs are underway which are directed toward this goal, none have contributed to the success of the existing multi-megabase per year sequencing centers. These centers have relied on the application of conventional engineering technology to the implementation of effective production instrumentation. Revolutionary technology advancement remains an important goal to help complete the sequencing of the human genome and to make available the tools of rapid sequencing to smaller laboratories in universities, industry, and in the clinical health field.

Existing technology has been used to develop production automation that has been applied to the genetic and physical mapping of model organisms and human, and in the discovery of expressed sequence tag sites. Similarly, existing technology has helped to make possible the large scale sequencing of the genomes of model organisms (*S. cerevisiae*, *C. elegans*, *Drosophila*). Human sequencing is still in the early stages of scale-up and will build on the technology and experience of the sequencing efforts for model organisms. The sequencing of the human genome will present new technological challenges. Strategy development for human genome sequencing should reflect the still dynamic nature of the HGP.

The paradigm often invoked for genome automation is one of the "genome factory". Factory automation has been in use and has been very successful in the manufacturing field for much of this century. In recent years, with advances in automation, robotics, and computer technology, the modern factory automation environment is relatively adaptable to changes in product lines. Scientific research on the other hand has not been an environment suitable for factory-like automation except in the fields of environmental sampling and analytical chemistry. The sample handling and analysis requirements for the HGP present another area amenable to laboratory automation. It should be noted however that the application of conventional engineering factory automation to the science laboratory environment is distinct from the approach used in industry. In

the HGP the goal is to do good science, the product is information, and the quality standards are the normal standards of good scientific research.

A Bottom-Up Modular Approach

There are two contrasting approaches to laboratory automation which can be characterized as being either top-down or bottom-up. A bottom-up approach is one in which technology is implemented in relatively small discrete modules to address bottlenecks in the current state of the project at any given time. The emphasis with this approach is to get technology into the laboratory as soon as possible in the areas that need it the most to realize an increase in productivity. A top down approach would be to examine the entire experimental process and design a highly efficient and highly integrated hardware and software solution to all aspects of the problem to be solved. This is a long term view that depends on a well defined and stable "problem" to be engineered. Top-down systems are relatively inflexible to changes in experimental protocols. Systems of this design generally come on-line all at once after a considerable period of design, fabrication, implementation, and testing. The Lawrence Berkeley National Laboratory (LBNL) Human Genome Center (HGC) decided to adopt a modular bottom-up approach to the planning and application of automation technology to Drosophila and human sequencing.

The bottom-up approach practiced at LBNL recognized that in the early stages of the HGP the experimental environment both within the lab and on a worldwide basis would be quite fluid. The biological technology that led to the conceptually simple notions of how to sequence the genome was also undergoing rapid change. The period from the mid-1980's to the present has been a period of experimentation involving a variety of large scale mapping and sequencing strategies. Because of its ability to adapt to a changing experimental environment, a bottom-up modular approach is often preferred.

This early period of experimentation with biological protocols was also an opportunity to develop the working relationships between biologists and engineers that would be required for the implementation of more massive automation systems in the future. Early introduction of well-defined small scale engineering technology into the laboratory led to manageable projects to develop these working relationships. Finally, and perhaps most realistically, a modular approach is an initially less expensive means to introduce engineering technology into the laboratory during the funding ramp-up of the project.

These advantages of a bottom-up development and design philosophy must be weighed against the disadvantages of this approach. The primary disadvantage is the difficulty of building a well-integrated collection of automation modules throughout all aspects of the experiment over a long period of time. There will always be pressure to utilize new and better technology as it becomes available. Modules designed and built during the early stages of the project may constrain or influence later technology development. Alternatively modules developed early in the project will have to be rebuilt to keep up with current technology. As more modules are developed, they will begin to connect together and the interfacing of modules will become a serious concern. It is important in the adoption of a bottom up design philosophy to understand the difficulties with this approach. The disadvantages can be partly mitigated by a continuous dynamic analysis of the project throughout implementation of the automation while maintaining the distinctive features and advantages of a bottom-up approach which are: early introduction of individual automation modules, and continuous flexibility to adapt to changing protocols and new technology.

A modular approach is applicable to a wide variety of institutional environments. It is often the default approach for institutions unable to make the large initial investment of a top-down system. A number of genome centers, including the LBNL HGC, have focused on particular biological strategies to genomic sequencing. Focusing upon a particular sequencing strategy may appear to lend itself to a top-down automation design. However, the availability and productivity of

automation technology has driven the evolution of new biological strategies and has contributed to a changing laboratory environment. A modular approach is also advantageous in laboratories that may have a number of biology groups pursuing different experimental programs in genomic research. These types of laboratories push automation development into areas of common interest among biology research groups with different goals. Finally, a modular approach is convenient for laboratories that choose to go to outside contractors for engineering expertise and equipment. It requires laboratories to define discrete instruments for purchase.

Laboratory Automation and Robotics

Decisions regarding global automation strategies as well as designs for particular instruments should be based on a realistic expectation of what benefits can be achieved by automation. The principles learned and applied to factory automation can also be applied to genome laboratory automation. An understanding of the types of processes that are automatable and a reasonable expectation of the performance that can be achieved is essential for designing a high quality, efficient, and cost effective strategy for mapping and/or sequencing.

The following is a list of benefits attributed to factory automation that are frequently cited throughout the business and the engineering world. It is worth considering their implications to genome research. This list is not ordered with respect to importance. It is should also be realized that any global or particular automation design does not have to meet all of these criteria. It only has to realize an increase in productivity. In general, productivity in genome research can be defined in terms of quantity, quality, and the cost of the finished data. In the case of the HGP, productivity is the measurement of high quality unique sequence at the lowest cost in the shortest possible time period.

- **Speed/High-Throughput**

  Automated instruments can often perform a task much faster than manual methods to achieve higher throughput. Throughput is the number of samples processed per unit time. One area where this is clearly the case is in the use of imaging technology. The imaging of fluorescent samples in polyacrilamide and agarose gels, coupled with the initial computer analysis of these images, is considerably faster than manual methods. Similarly, film exposures to radioactive probes can be digitally scanned and analyzed.

  Automated instruments can also perform mechanical tasks faster. The machine may actually move faster than a person performing the same operation. The important elements to consider in order to achieve an increase in the speed of a mechanical operation are: the speed of the drive elements, the weight of the samples (total load), and the distance to be traveled. Mechanical speed will increase for light loads, moved over short distances, very fast. Average overall speeds can be increased with machines that operate at a steady pace for long periods of time.

  Throughput can be increased by operating on many samples at once or doing multiple tasks at one time - multiplexing. An example of multiplexing is the thermal cycling of samples in an air oven. While the cycling time of an air oven is relatively slow, the use of a dozen 864 well microtiter plates results in a very high sample throughput.

- **Repetition/Boredom**

  Tasks that are highly repetitive and boring are good candidates for automation. Repetitive tasks are often a poor use of skilled personnel and often lead to production errors.

- **Cost Effectiveness**

  Automated machines can reduce material and personnel costs. Training for complex tasks can be reduced and the people operating machines may be able to do more tasks or process more samples. Machines can also improve cost by reducing the amount of consumables such as reagents and plasticware. In some cases a custom instrument may be less expensive than a modified commercial instrument. The actual cost of any instrument must be amortized over the life of the instrument.

- **Better Quality Results/Can't Do It Any Other Way**

  Automated machinery can often do tasks that cannot practically or efficiently be done by manual methods. The manipulation of objects on a small scale is an example. Accurately moving a probe (pipet tip or a colony picking needle) to a small target (a well in a 384 well plate or a bacteria colony) is best done by an automated instrument. Another example is the placement of a gridding tool for very high density gridding of samples on filters.

- **Complexity**

  Automated machines can often do complex tasks provided the task can be broken down to simple discrete operations. These operations can be chained together into a single complex task performed by a machine. A machine can perform the overall task with more consistency than can be done manually. Predictable results from consistent machine operation can be used to plan and optimize subsequent steps in the laboratory. Some tasks are complex enough to require a significant learning period for new employees to perform them adequately. A high turnover in personnel for this task will result in constant training and low productivity. It may be simpler and faster to teach operators to use a machine. DNA sample preparation for sequencing is an example of a complex operation that can be automated.

- **Standardization**

  Automation at some level requires materials and format specifications in order to process the samples. Standardization on, for example, the 96 well microtiter plate format provides a known reference format upon which everything else should conform. Standardization can enforce a common input and output for each module within an integrated system. It can also make it possible for instruments to be transferred or commercialized for wide distribution.

- **Safety**

  Safety considerations are a major reason for the introduction of automation in industry. Automated machines and robots can be used to perform tasks in environments that are hazardous to humans. The handling of toxic chemicals and radioactive materials are good examples of where automation can provide essential or desired health and safety protection. However experience has shown that the use of radioactive and other chemicals commonly used in biology laboratories, can be done safely with proper training. The use of radioactive and toxic chemicals is generally a waste disposal problem rather than a safety problem in the genome laboratory. Often, rather than solving safety problems, automated machines themselves can be dangerous and the safety issues become one of protecting laboratory personnel from the machines.

Fixed vs. Flexible (Robotic) Automation

The hardware configuration of an automated system can be classified as either fixed or flexible. Both top-down and bottom-up automation strategies can use both fixed and/or flexible automation.

The term "fixed automation" is used to describe specialized machines that have been designed to perform a particular task or set of tasks very efficiently to achieve very high throughput. Fixed automation cannot easily be reconfigured to perform a substantially different task. This is an appropriate automation strategy for an unchanging process that demands high throughput. The entire cost of the machine must be justified by the throughput of the process over the period of time during which the fixed process remains useful.

A flexible or robotic instrument is defined as: A reprogrammable, multifunctional manipulator designed to move material, parts, tools, or specialized devices through various programmed motions for the performance of a variety of tasks (Robot Institute of America, 1979). The cost of the robot is justified by the longer lifetime achieved by adaptability to changing processes. Robots generally take the form of a coordinate system frame and an arm, and a end effector (i.e. gripper) that moves within the work envelope. A stated advantage of flexible automation verses fixed automation is the speed of implementation of new and different protocols. Robots can be purchased "off the shelf". The remaining work is then limited to design and fabrication of peripheral components and programming of the final system. Since few parts are custom manufactured there is less mechanical debugging. Finally, the "reaction time" to implement a new protocol on a robot platform is shorter than the time to design and build a custom automated machine.

Flexible automation tends to be slow because the single arm and gripper performs tasks one at a time. These robots are used to best advantage when servicing multiple custom applications modules located within the work envelope thereby providing a multiplex capability to the whole system.

There are various classification schemes to describe robots but the categories of robots generally seen in the biology laboratory are the following:

- An articulated arm operating in generally a Cartesian (XYZ) coordinate system. The Hewlett Packard ORCA and the CRS T265 are examples of this type of robot that are commonly found in genome laboratories. Both of these robots move on a linear rail that can be as long as 2 or more meters and therefore a very large work envelope. They have very dexterous jointed arms and grippers that can tilt and rotate giving 6 degrees of freedom.
- A cylindrical coordinate system robot can access tools and materials by moving a gripper radially and circularly around a vertical center axis as well as vertically on the center axis. The robot produced by Zymark and the jointed robot produced by Adept and CRS operate in this type of coordinate system. These robots access sidestations that are located radially around the center of the robot.
- Cantilever and Gantry robots operate in a Cartesian coordinate system but they are built from linear rails oriented along the XYZ axes. These robots generally do not have articulated arms though wrists with additional degrees of freedom can be attached to the gripper mechanism. The Beckman Biomek 1000 and 2000 Workstations and the Hamilton Microlab Workstation are designed with this format.

The advantages of having a capable automation group (mechanical technicians, engineers, and programmers) to implement flexible automation cannot be understated. Commercial robots are more often than not deficient in either applications hardware and/or software. The off-the-shelf

advantages of commercial robots are overstated relative to the ease in which a custom gantry robot system can be assembled by a competent mechanical engineer. Finally, the reconfigurability feature is seldom utilized in a production environment. Most robot systems are configured for one application and are seldom used for anything else. It is not uncommon to see two identical robot systems each configured to a particular application. This "actual" use of robot platforms should be weighed against the high throughput advantages of a fixed automation platform.

Large vs. Small Machines

Important considerations in modular instrument design are space, access, and reliability. Automated instruments and robots come in a wide range of sizes depending on their functions and applications. Instruments can be as small as a microtiter plate filler (9"w x 12"d x 10"h) to a Beckman Biomek Workstation (3'w x 2'd x 30"h) to machines that will fill a large room. The size of the instrument will influence how it is used on a daily basis.

Large machines can be characterized as having large work envelopes to provide access to many microtiter plates, filters, or specialized automation module side stations. These instruments occupy a significant amount of laboratory space and are often located some distance from the general laboratory area. They must be designed to operate unattended for significant periods of time so that there will be less frequent operator interaction for an instrument that is located a long distance from the main laboratory. Large instruments are generally more complex and expensive to meet the criteria of handling large numbers of samples and to achieve reliable operation. They will take a long time (1-2 years) to design, fabricate and test; therefore the resulting throughput and productivity goals should be set very high to justify the investment of time and money. Machines of this type are generally only built to implement very stable experimental procedures. The largest machines are often found in top-down strategies and in mature bottom-up strategies.

Small machines can occupy a small amount of lab space and therefore it is more practicable to locate them in proximity to the main laboratory activity. This means that operators can interact with them more frequently if necessary. Often they will perform limited tasks and are therefore less complex to design, fabricate, and they can be implemented fairly quickly (1-12 months). This also means that they are easily duplicated. There may be an advantage to having multiple small instruments versus a single large instrument. A single large instrument that is down for maintenance or repairs for any significant amount of time can seriously effect productivity.

Robot Capabilities

An essential element in designing automation modules is to understand the type of things that robots and automated machinery can do for biologists in the laboratory. Much of the automation developed to date is devoted to materials handling and processing, although a few instruments, typically imaging systems, generate information to be used in later stages of analysis. This reflects what normally happens at the laboratory bench. Any task that is proposed for automation must be reduced to its elemental robot compatible components. The automation solution may look very different from the manual procedure but will achieve the same result. An appreciation of these capabilities is essential for developing biological protocols that are automatable.

The following is list of the type of practicable robot capabilities that can be used as building-blocks in order to automate benchtop protocols:

- **Materials Handling**

  Robots can move objects from one place to another place. This is often done with grippers configured with custom designed fingers to accommodate the item being moved and the

location into which the item is moved. Materials can be grasped with suction cups, pushed and pulled, moved with conveyor belts or shuttle mechanisms. Sophisticated robots have the capability of changing grippers and tools. The typical actions are to move microtiter plates from one location to another and to move pipetting or dispensing tools to microtiter plates or reservoirs. Colony picking machine needles move bacteria from a colony to a microtiter plate by adhesion of the cells to a picking needle.

The major performance specifications for materials handling and motion control using a robot are:

repeatability - the ability to return to the same point
accuracy - the ability to go a specified distance from a known reference point
velocity - how fast the robot can move from one location to another
acceleration - the speed with which a load can be brought from rest to a particular velocity
load - the amount of mass that can be lifted, pushed, or pulled
dexterity - the ability to trace complex paths
tool interchangability - the ability to attach and operate a variety of tools

- **Liquid Handling**

Aspirate/Dispense - Liquid delivery is usually performed using robotically compatible pipettors (to aspirate and dispense using fixed and removable tips) and dispensers. These tools can either attach to the robot arm and gripper or reside as a fixed module within the robot work envelope.

Pipetting tools can be used to mix liquids and, in combination with movement, transfer liquids from one location to another. Pipettors can be found in two configurations. Syringe pump pipettors have a remote motorized syringe station connected by tubing to a dispensing nozzle. The second configuration is similar to manual pipettors where the entire plunger mechanism is in a compact enclosure attached to the gripper. This configuration of the pipetting tools used in the Beckman Biomek robots. Pipettors and dispensers are usually configured with 1, 4, 8, or 12 liquid delivery channels. Pipettors can operate in a volume range from 1000 µl down to 0.5 µl though this entire range is generally not available in a single tool.

A dispenser delivers liquid from a large reservoir. This is done using pumps or by pressuring the reservoir and releasing fixed volumes by opening a valve for a calibrated period of time. An interesting dispenser research area is the use of inkjet dispenser mechanisms to dispense submicron volumes of liquid. Liquid delivery at volumes less than 2 ml can be difficult due to the surface tension of the liquid. Careful selection and design of pipet tips and dispenser nozzles is required.

Adhesion—Liquid transfer can also be achieved by simple adhesion of liquid to a pin as is the case in 96/384 pin replicating tools used to copy bacteria libraries by inoculating microtiter plate wells from a source microtiter plate.

Pumping—Machines can actively pump liquids from one reservoir to another or to be used as a thermal transfer medium for heating and cooling.

- **Heating and Cooling**

Heating and cooling sidestations can be found in various formats. Temperature controlled heating and/or cooling blocks transfer heat by conduction through a metal block in close contact with a thin-walled microtiter plate. Heat is also transferred by convection using circulating

water baths or air ovens. Heating by quartz lamps or hot air heating elements is used to sterilize picking needles in colony picking machines.

- **Image Acquisition**

  Image acquisition is accomplished with a digital camera or scanning device. It is used to record fluorescent images, such as ethidium stained DNA in an agarose gels, or to locate objects, as is done in a colony picker or pick and place robot. The typical colony picker camera is a room temperature CCD black and white camera with a 480 x 640 pixel detector. Cooled digital cameras (for high sensitivity to weak fluorescent signals) and high resolution cameras (larger pixel array detectors) are very expensive and are less commonly used. Color cameras are also available but these typically reduce the pixel resolution to 1/3 of a black & white camera. Additional hardware requirements are a camera lens, a digital frame grabber card, and an image analysis software package.

- **Data Analysis**

  The computers that control a robot can process data as well. An agarose gel imaging station can acquire the image, calculate calibration curves and calculate the length of the DNA samples recorded in the gel. Computers can also be used to track bar code data as well and interface with LIMS or other information/materials tracking software.

- **Sensing**

  Robots and automated machines can sense a variety of physical parameters. Camera vision images can be processed to guide a robot to an object (and verify its presence and correct placement). Other sensors can measure temperature, pressure, humidity, proximity (i.e., is an object located near the sensor), color, liquid level, force, and a variety of chemical parameters.

Computers and Software

Typically both flexible and fixed automation modules are controlled by stand alone personal computers, usually either an IBM style PC or a Macintosh computer. Most laboratory scale automation does not require powerful computers. The computationally intensive features of motion control and robot kinematics are built into custom input/output boards, amplifiers, and other stand alone instrumentation modules. Much of the computer power of a 486 PC is used to execute graphics applications such as a graphical Windows user interface or digital image display and analysis software. Robot systems designed for industrial use may require a more powerful custom computer and electronic control equipment system.

The actual operation of automated systems is through simple software interfaces intended to be used by nonspecialized personnel. A good user interface is one that is intuitive and simple to use. The software interface should be well laid out visually on the screen, should not produce unexpected and confusing messages or machine responses, should remember previous user parameters, and should operate with a minimum of user input. Ideally the user interfaces for all modules should have a similar "look and feel" to minimize training on each instrument. Custom software interfaces should be designed to accommodate the users. It is extremely important to realize that a user can declare a functional instrument unacceptable simply based on experience with a poorly designed software interface.

Small commercial liquid handling robots such as those produced by Beckman, Hamilton, and Tecan are design to be "programmed" by novice users. Users can typically learn to program these systems in a single day. Implementing a protocol usually involves entering simple hardware

parameters and the chaining together of a library of preprogrammed robot actions into a complete protocol. Most robot systems, both commercial or custom built, required expert dedicated personnel to program protocols and maintain the software. The types of software skills necessary to program automation modules range from learning the complex and powerful custom macro languages of a commercial articulated robot, programming in high level languages such as Visual Basic, low level languages such as C or assembly code, and finally programming in custom low level command languages that operate specific input/output boards, motor driver boards, etc.

Software platforms should be chosen with the following criteria in mind: The ability to develop a well-designed user interface, to be fully capable of controlling the equipment, and to be easily maintained by software personnel. It is very important that more than one person be familiar with any software platform that is implemented. At LBNL, we have chosen Visual Basic as the standard software package used to write motion control applications because it provides a powerful Windows based graphical user interface, is simple for anyone to program and maintain, and can be linked to other Windows compatible software (i.e., the C programming language).

Buy or Build?

In the course of instrumentation module development and implementation, the question of whether to buy or to build an instrument will invariably be encountered. If a commercial instrument meets the stated goals, then it is the best to purchase the instrument. This has the obvious advantages of a machine that has been extensively tested, which may have a historical track record of laboratory use, and has the technical support and expertise of the manufacturer.

If no commercial instrument meets the stated goals, then it still may be possible to purchase a commercial instrument and modify its hardware and/or software capabilities. The manufacturer may or may not cooperate in a modification of their instrument. Many manufacturers maintain research groups that are often willing to work with customers on special projects usually with the expectation that they can develop something new to add to their product line. They may recommend a third party engineering consultant or direct you to customers who may have solved similar problems.

Building a large scale instrument should be considered a major undertaking requiring proper resources in personnel and laboratory and machine shop space. Requesting a custom instrument from an engineering group is not the same as purchasing an instrument from a commercial manufacturer. It will require not only a financial commitment but an intellectual and time commitment from the biology management to be an integral participant in the project. This means close interaction between engineers and the biology management during the design phase and with biology users during the installation and beta testing phase.

A custom instrumentation project will generally require one or more engineers (mechanical, electrical engineering, and computer science) as well as mechanical and electrical technicians and machine shop personnel. It will also require facilities and space to fabricate parts and assemble the instruments separate from the final installation point. Depending on the size and complexity of the project it can take from 2-12 months to design and fabricate an instrument and 2-6 months to debug the instrument and integrate it into production.

Many laboratories do not have the proper engineering personnel and facilities for instrumentation development and must seek assistance from outside engineering firms. If the instrument is to be built by an outside engineering firm, then the commercial aspects of the project should be discussed early in the project. One option is to contract for a specific one-time special project to be built by an engineering firm for a particular laboratory. It is also possible to collaborate with a commercial company to develop an instrument with the intention of commercializing the instrument for sale in a wider market.

The build or buy question touches on the issue of technology transfer. Technology transfer is only going to be successful for instrument designs that are easily adaptable to new environments. Both custom and commercial instruments will have to be designed for a range of performance, size, and cost for a given customer base. However, the demands of genome research is for efficient high throughput (specialized) instrumentation. It is not surprising that there are many unique instruments in genome centers around the world that do not have wide applicability in other genome laboratories or in the commercial sector. It also illustrates that strong engineering capabilities are necessary to take full advantage of commercial instruments.

Automation Modules

A modular approach in the application of engineering solutions to automation at the LBNL HGC and elsewhere can best be illustrated by a list and description of some of the custom instruments that have been produced. These are instruments that exist in the laboratory and are in use today. This list contains instruments that range from modifications of commercial instruments to entirely custom fabrications. The following list, description, and notes on instrumentation are devoted primarily but not exclusively to the instruments pursued by the LBNL HGC. In some cases similar instruments have simultaneously been developed elsewhere. The descriptions also list some of the primary automation benefits achieved by each machine. This is an illustrative list and is not intended to be a survey of all of the instruments developed worldwide.

Instrumentation Modules

- **Imaging**

  There are a number of custom and commercial imaging instruments dedicated to recording radioactive or fluorescent sample signals or scanning photographic films. Imaging of hybridization filters is common in many laboratories. The imaging instrumentation at LBNL (Figure 1) is dedicated to imaging ethidium bromide stained agarose gels. It consists of UV illumination of the gels, a cooled CCD camera for image capture, and commercial and custom image analysis software. There is also an essential imaging component to colony picking machines.

  Automation benefits—Digital image capture and analysis are much faster than manual methods, it is cost effective, it performs repetitive work, and image instrumentation can perform complex computational analysis of the data.

- **Colony Picking**

  This LBNL custom colony picker consists of a digital imaging station used to determine the location of the colonies in 10x10 cmn agar plates and a separate colony picking machine (Figure 2). The machine consists of 12 picking needles mounted on a carousel above two XY tables. The XY tables move the source colony plate and the destination microtiter plate underneath the picking needles. The needles pass through a sterilizing station with each cycle. The digital imaging of the plates, the colony picking, the placing of the colonies into the deep well microtiter plate, and the needle sterilization all take place simultaneously. The machine requires that the user continuously image and feed plates to the colony picker. The entire system picks colonies at a rate of about 1,200 colonies per hour. The typical use of the colony picking machine is for low volume continuous colony picking which consists of multiple (20-30) experiments of 96 colonies every couple of days. Another mode of use, more prevalent at other laboratories, is to pick large libraries (25,000 - 100,000 colonies) on an infrequent basis.

Automation benefits—The machine performs extremely repetitious work and it picks colonies faster, on average, than manual methods. Picking colonies from 10x10 cm agar plates into deep well microtiter plates is unique to this machine and is particularly well suited to the colony picking experiments done at the LBNL HGC.

Other implementations—Colony picking has been implemented in a large number of laboratories including the Sanger Centre, the Imperial Cancer Research Fund, Stanford University (plaque picking), etc. Commercial instruments have been developed by Hybaid Ltd., Genetix Ltd., and BioRobotics in England.

- **Liquid Handling Robots**

Commercial liquid handling robots produced by Beckman, Hamilton, and Tecan are common in genome laboratories. They are accurate, easy to program, and have relatively high throughput if configured with 8 channel pipetting tools. Automated pipetting at LBNL is implemented on the Beckman Biomek 1000 and the Hewlett Packard ORCA robot. The Biomek 1000 has proven to be an accurate and reliable low volume (1-20 ml) multichannel pipetting machine. Virtually all of our programming on the Biomek 1000 is done in Microsoft QuickBasic coupled with a custom low level software package (BiomekQB). With the ability to completely control the Biomek 1000 through our custom software, we were able to develop custom hardware modules to implement PCR reaction setup, sequencing reaction setup, sequencing reaction pooling, and agarose gel loading applications.

The Hewlett Packard ORCA is used for pooling and library replication applications that require a large number of source and destination microtiter plates and therefor a large work envelope. We developed a disposable tip 12 channel pipetting head that is used extensively for row, column, and plate pooling protocols. A separate fixed tip 16 channel manifold is used for plate filling applications.

Automation benefits—Both the Biomek and the ORCA perform repetitive work, implement complex pooling protocols, and are cost effective as unattended machines that allow personnel to perform other tasks.

- **Thermal Cycling**

LBNL developed a custom water based thermal cycler. The instrument cycles water at different temperatures underneath a thin-walled microtiter plate to achieve rapid temperature transitions that produce very high quality PCR products. The four independently programmable microtiter plate stations were designed to be loaded with the Hewlett Packard ORCA robot. The LBNL Thermal Cycler, the HP ORCA, and the Biomek 1000 have been combined into a larger integrated PCR processing system.

Automation benefits —The microtiter plate PCR systems performs faster than commercial machines. It is a cost effective machine relative to commercial instruments. It produces high quality PCR products which means there is less noise in the PCR data. It is robotically compatible to an HP ORCA robot platform.

Other implementations—Water based thermal cycling has been implemented primarily as robotically controlled baskets of sealed microtiter plates immersed into temperature controlled reservoirs. While the individual temperature cycles are fairly long, the overall throughput is very high in some of these machines that can run many 384-well microtiter plate samples. Circulating temperature controlled air ovens can also thermally cycle samples in a batch mode. Numerous commercial thermal cycling instruments are on the market based on resistance

heating and water or Peltier cooling. The commercial machines typically cycle one microtiter plate at a time.

- **DNA Synthesis**

LBNL developed a custom 12-channel DNA synthesizer that can produce 20 nm of 21 mer DNA at about $0.10/per base in 2.5 hours. The instrument was designed to have dedicated liquid delivery lines for each reagent to minimize reagent loss that occurs in commercial instruments when reagent manifolds must be flushed between each step. The instrument is manually loaded with the correct CPG beads in each channel, all of the dispensing steps and chemistry are automated, and the final deprotection and capping steps are done manually. The instrument has been designed for scale up in multiples of 12 channels but the LBNL HGC has not needed additional capacity in a single machine.

Automation benefits—This instrument is extremely cost effective. It has significantly higher throughput than commercial instruments.

Other implementations—Stanford University developed a 96-channel DNA synthesizer. There are a number of commercial DNA synthesizer instruments.

- **Microtiter Plate Pooling/Microtiter Plate Filling/Library Replication**

These applications have been implemented on a Hewlett Packard ORCA articulated robot platform (Figure 4). A substantial amount of peripheral hardware was developed around the robot including: 96/384 pin replicating tools, microtiter plate locating and stacking plasticware, a 12-channel syringe pump pipettor, and a multipin tool ultrasonic bath and heater sterilizer. Applications programming was done using the software package provided with the ORCA robot.

Automation benefits—Performs repetitious replication, dispensing, and pipetting tasks, accurately completes large and complex pooling schemes, a cost effective stand alone (hands-off) instrument that frees personnel for other duties.

- **DNA Preparation**

LBNL is currently constructing a custom machine to prepare DNA for sequencing. We will be implementing a modified boil preparation protocol. The instrument will consist of a gantry robot equipped with a pneumatic gripper, 5 eight-channel dispensers, and 2 Beckman Biomek eight-channel 200 ml pipettors. Centrifugation will be performed in an Eppendorf 5416 centrifuge located in the work envelope and modified for remote robot control. The system is designed to prepare 184 samples in 2-3 hours of unattended operation.

Automation benefits—The instrument will achieve consistent sample preparation relative to labor intensive manual preps. It will run unattended allowing personnel to perform other duties. It will have moderate throughput. DNA preparation using a boil prep p*rotocol is very cost effective.*

Other implementations—There are no instruments, commercial or custom, that perform high throughput DNA preparation. Some instruments have implemented centrifugation, others depend solely on filtration, and some incorporate magnetic bead technology as part of the protocol. DNA preparation is highly sensitive to the biology and all of the manual benchtop kits and automated procedures have required significant protocol development at each laboratory in order to achieve reliable sample preparation.

- **Gridding**

  Gridding machines spot samples at very high density on filters. Gridding applications have been implemented on a HP ORCA at the Lawrence Livermore National Laboratory. Genetix Ltd. has implemented gridding applications on a relatively large gantry robot and the Sanger Centre has implemented this application on a somewhat smaller gantry robot.

  Automation benefits—These instruments can grid much faster than manual methods, they can grid at much higher density than is practicable by manual methods, it is cost effective and it performs a repetitive task.

Conclusions

Much but not all, of the instrumentation developed around the world in various genome centers could be described as modular. In many cases this has been by default because most of the laboratories starting out in the new field of genomics could not afford or risk a top-down designed automation system. In some cases a bottom-up modular automation system was created by design as was the case at the LBNL. A key element to this approach is adaptability to change within a new field within Biology.

Currently, with the rise of several multi-megabase per year sequencing centers, it is conceivable that new genome centers will begin to choose from existing experimental strategies and adopt the hardware and software solutions developed elsewhere. This would be the beginning of the adoption of a rational top-down strategy. This is not to say that a new center would not bring new automation ideas into the field but that there would be great pressure to enter the field quickly at a high technical level relative to existing centers. Recently, an instrument system that could be described as a top down implementation has been installed at the Whitehead Institute as part of the genomic mouse mapping project. The Genomatron is a large scale, specialized, high throughput, and highly integrated system of machinery to perform sample preparation, thermal cycling, and sample gridding on filters. It was developed at a mature phase in the experimental mapping protocols and justified by a cost analysis of a predictable set of laboratory protocols, goals and stable funding. Similarly, as sequencing strategies and technology mature there will be a tendency for new sequencing centers to adopt automation strategies based on current technology. The choice to pursue either a top-down or bottom-up approach will be base on many considerations but some of them should be a rational understanding of what automation can do and what resources are necessary to be successful. Some of these concepts have been described here.
Large scale genomic research is a relatively new field in biology that will require a multidisciplinary approach to accomplish the stated goals of the HGP. While current technology should be capable of sequencing the human genome, the development of more advanced revolutionary technology will be required to bring genomic research capabilities to university and industry laboratories pursuing research in mapping and sequencing of other organisms and into the clinical health field. Engineers and biologists have not been traditional partners in research. This is beginning to change in genome research. Engineering and computer science experience and resources should be equally available to biologists to develop optimal strategies. This can only happen if each field can develop at least a working knowledge of the other fields. It is to be hoped that some of the ideas and concepts outlined here will provide a window to decades of engineering experience that can be applicable to genome research.

Acknowledgments

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, Human Genome Program, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Disclaimer

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

References

1. Collins, F., and Galas, D., "New Five-Year Plan for the U.S. Human Genome Project", Science, **262**, 43, (1993).

Figure Captions

Figure 1 - Agarose gel imaging station with cooled CCD camera and data analysis computer.

Figure 2 - Bacterial colony picking machine. This machine picks bacterial colonies at a rate of 1200 colonies per hour. The digital imaging station is not shown.

Figure 3 - 12-channel DNA synthesizer. The machine synthesizes 12 different 20-mers in 2.5 hours for approximately $0.10 per base.

Figure 4 - The Hewlett Packard ORCA robot is used for pooling, plate filling, and library replication applications.

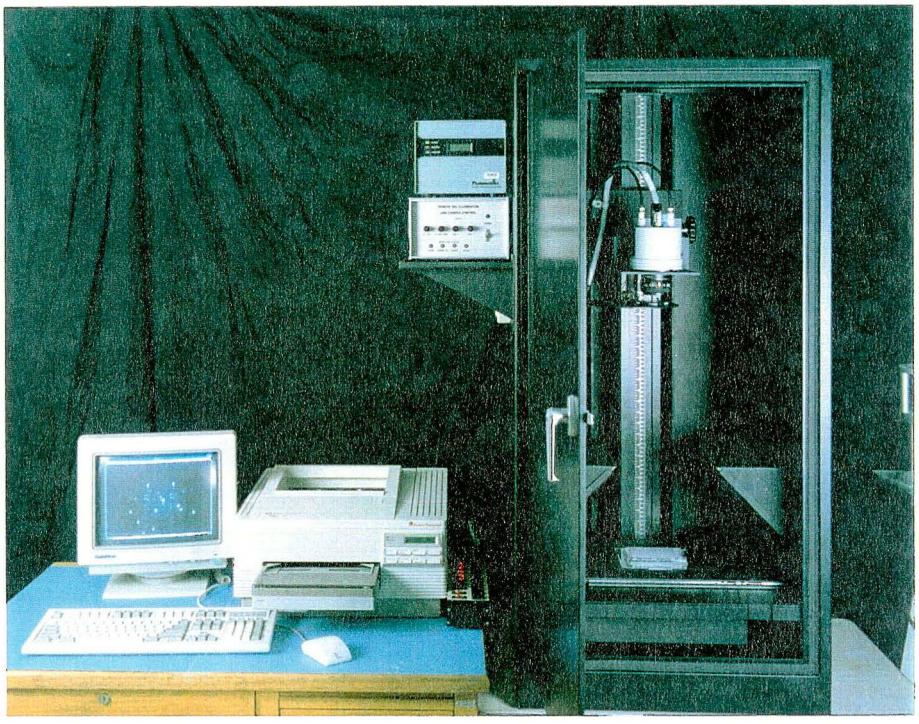


Figure 1

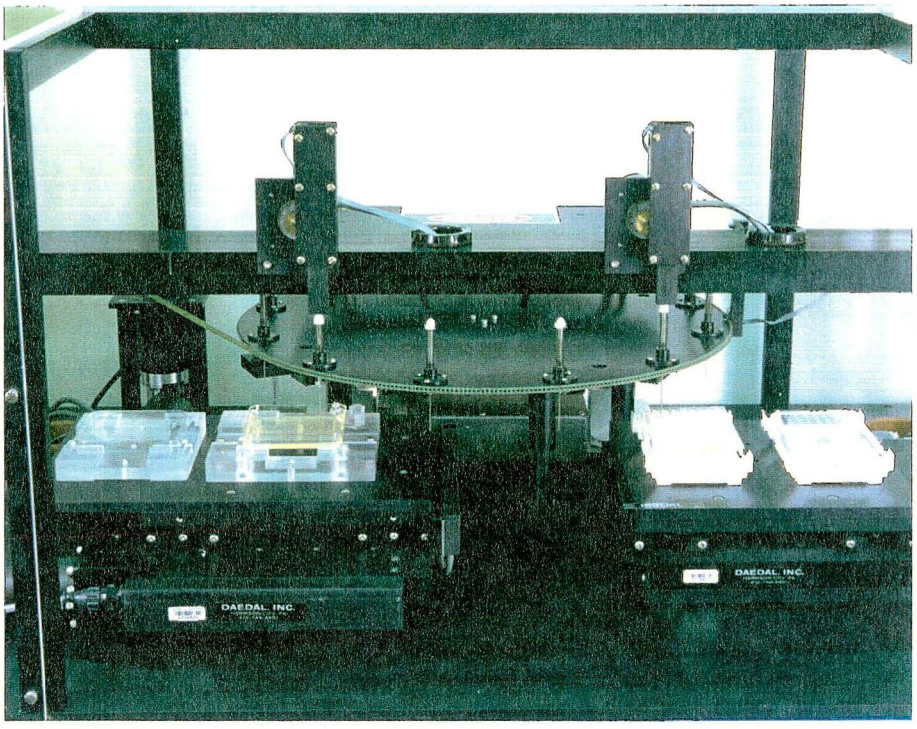


Figure 2

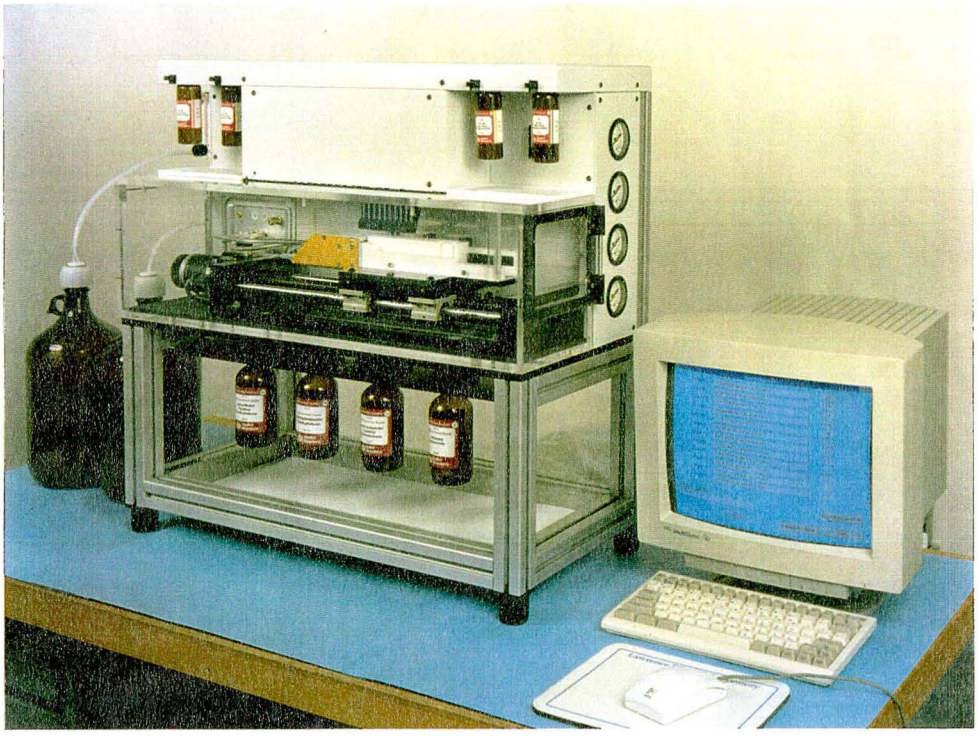


Figure 3

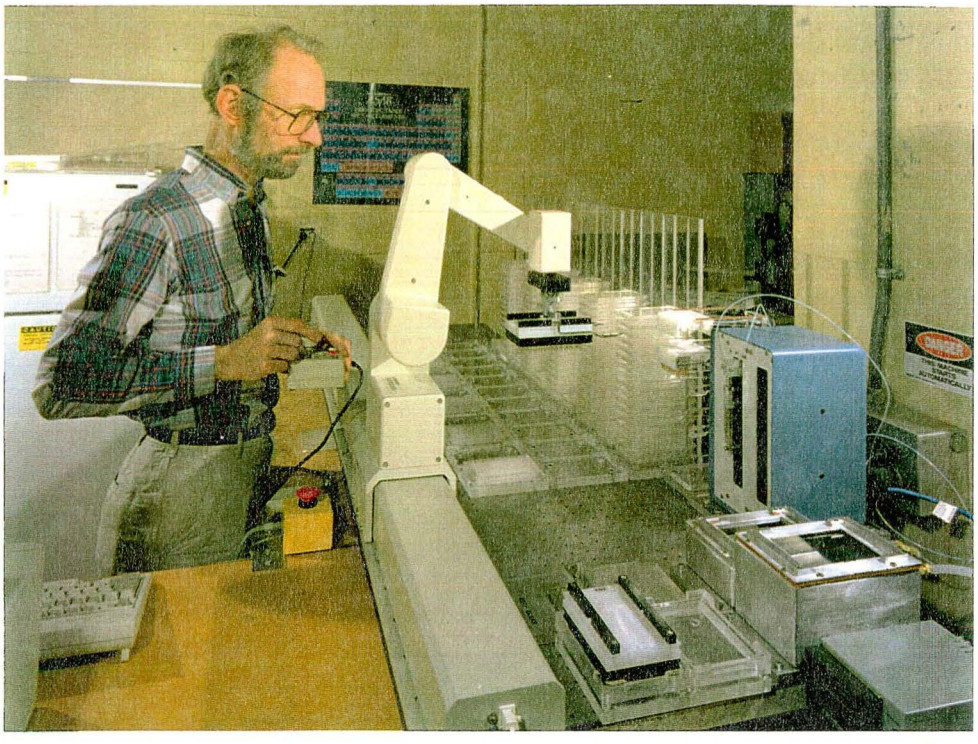


Figure 4

LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
TECHNICAL INFORMATION DEPARTMENT
BERKELEY, CALIFORNIA 94720